A more robust unscented transform

James R. Van Zandt[a]

[a]MITRE Corporation, MS-M210, 202 Burlington Road, Bedford MA 01730, USA

ABSTRACT

The unscented transformation is extended to use extra test points beyond the minimum necessary to determine the second moments of a multivariate normal distribution. The additional test points can improve the estimated mean and variance of the transformed distribution when the transforming function or its derivatives have discontinuities.

Keywords: unscented transform, extended Kalman Filter, unscented filter, discontinuous functions, uncertainty distribution

1. INTRODUCTION

The representation and estimation of uncertainty is an important part of any tracking application. For linear systems, the Kalman filter[1] (KF) maintains a consistent estimate of the first two moments of the state distribution: the mean and the variance. Twice each update, it must estimate the statistics of a random variable that has undergone a transformation—once to predict the future state of the system from the current state (using the process model) and once to predict the observation from the predicted state (using the observation model). The Extended Kalman Filter (EKF)[2] allows the Kalman filter machinery to be applied to nonlinear systems. In the EKF, the state distribution is approximated by a Gaussian random variable, and is propagated analytically through the first-order linearization (a Taylor series expansion truncated at the first order) of the nonlinear function. This can introduce substantial errors in the estimates of the mean and covariance of the transformed distribution, which can lead to sub-optimal performance and even divergence of the filter.

Julier and Uhlmann have described the *unscented transformation* (UT) which approximates a probability distribution using a small number of carefully chosen test points.[3] These test points are propagated through the true nonlinear system, and allow estimation of the posterior mean and covariance accurate to the third order for any nonlinearity. The *unscented Kalman filter* (UKF) uses the UT for both the transformations required by a Kalman filter. For linear functions, the UKF is equivalent to the KF. The computational complexity of the UKF is the same as the EKF, but it is more accurate and does not require the derivation of any Jacobians.

Julier and Uhlmann have extended the UT to stochastic functions.[4] We show how this extension may be adapted to increase the number of test points, and illustrate how the additional test points can improve the accuracy of the estimated mean and variance of the transformed distribution.

In Section 2 we briefly review the UT and apply it to a simple 1D function with a discontinuous first derivative. In Section 3 we extend the UT to use additional test points. We apply the revised algorithm to the 1D example and to a discontinuous 2D function. The results are discussed in Section 4.

2. ESTIMATION IN NONLINEAR SYSTEMS USING THE UNSCENTED TRANSFORM

In this paper, we will consider a single nonlinear transformation $\mathbf{y} = \mathbf{g}(\mathbf{x})$, where $\mathbf{x}$ is assumed to be multivariate normally distributed with mean $\bar{\mathbf{x}}$ and covariance $\mathbf{P}$, and we wish to estimate the mean and variance of $\mathbf{y}$.

The UT is superficially similar to the Monte Carlo method, but uses a small deterministically chosen set of sample points. In the simplest form,[3] one chooses for an n dimensional state vector $\mathbf{x}$ a set $\boldsymbol{\sigma}$ of $2n$ points as the columns of $\pm\mathbf{A}$ where

$$\mathbf{A}\mathbf{A}^T = n\mathbf{P}. \tag{1}$$

Further author information: E-mail: jrv@mitre.org

This set of points has zero mean and covariance $\mathbf{P}$. We then compute a set of test points $\mathcal{X} = \boldsymbol{\sigma} + \bar{\mathbf{x}}$ with the same covariance but mean $\bar{\mathbf{x}}$. Transform each point as $\mathcal{Y} = \mathbf{g}(\mathcal{X})$. Finally, compute the estimated mean and covariance of the transformed state as the mean and covariance of the points $\mathcal{Y}$:

$$\bar{\mathbf{y}} = \frac{1}{2n}\sum_{i=1}^{2n}\mathcal{Y}_i \tag{2}$$

$$\mathbf{P}_{yy} = \frac{1}{2n}\sum_{i=1}^{2n}[\mathcal{Y}_i - \mathbf{y}][\mathcal{Y}_i - \mathbf{y}]^T. \tag{3}$$

One may also use a central point with a different weight[3] and scale the pattern of test points.[5] Those variations are omitted here for simplicity.

There are an infinite number of matrix square roots satisfying Eqn. 1. They include a triangular matrix which may be found using Cholesky decomposition and a symmetric square root. All such roots are related by orthonormal transformations.[6] If $\mathbf{U}$ is orthonormal, then $\mathbf{A}_2 = \mathbf{AU}$ is another suitable root because

$$\mathbf{A}_2\mathbf{A}_2^T = (\mathbf{AU})(\mathbf{AU})^T = \mathbf{AUU}^T\mathbf{A}^T = \mathbf{AA}^T = n\mathbf{P}. \tag{4}$$

2.1. 1D Example

The EKF and UT were applied to a simple one dimensional function with a discontinuous first derivative. A particle with an initial velocity of -1 strikes a barrier at the origin at time zero and bounces elastically. Thus,

$$y = g(x, t) = |x + v_0 t|. \tag{5}$$

The original distance from the particle to the barrier is assumed to be normally distributed with unit variance. The transformed partial distribution function is the sum of two truncated normal distributions, as shown in Fig. 1 at time $t = -0.3$, just before the mean of the original distribution arrives at the barrier. The mean of the transformed

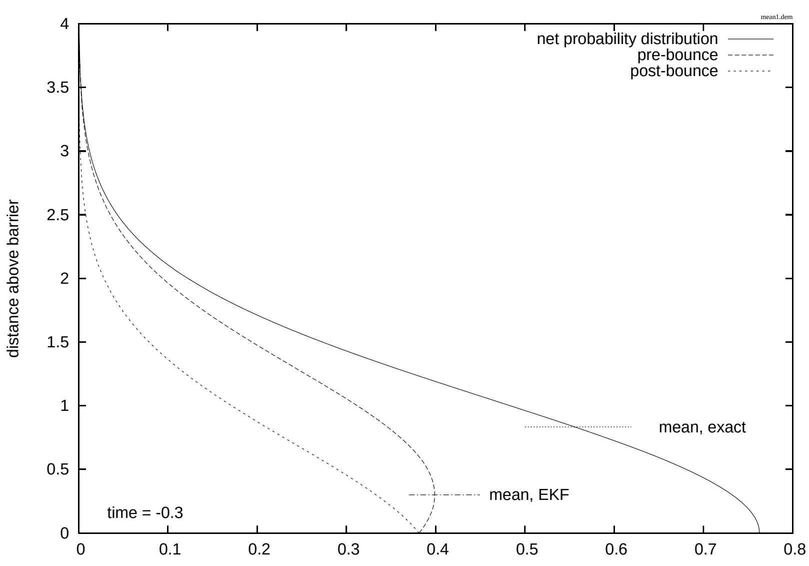

Figure 1. True partial distribution function of particle at time $t = -0.3$, for 1D example.

distribution was calculated with the EKF and the unscented transform. The software used for this was adapted from the reference implementation by Julier.[7] It uses the symmetric matrix square root. Estimated means are shown in Fig. 2, along with the exact values,

$$\bar{y} \equiv E[y] = \int_{-\infty}^{v_0 t}(v_0 t - z)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz + \int_{v_0 t}^{\infty}(z - v_0 t)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz = v_0 t \text{ erf}\left(\frac{v_0 t}{\sqrt{2}}\right) + \sqrt{\frac{2}{\pi}}e^{-\frac{1}{2}v_0 t^2}, \tag{6}$$

evaluated using MAXIMA.[8]

Since the EKF examines the function at only one point, it gives no hint that the distribution mean never actually reaches the barrier.

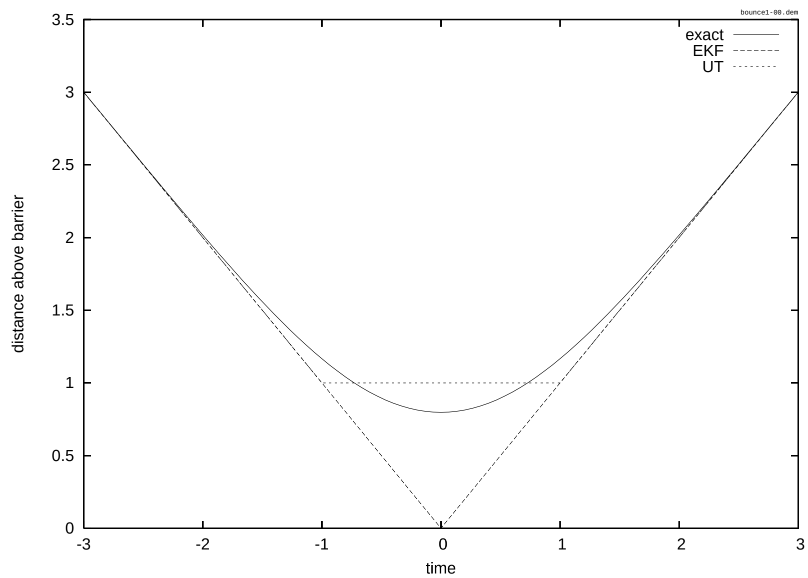

Figure 2. Estimates of distribution mean for 1D example.

The estimated root variance of the distribution is shown in Fig. 3. The EKF estimates the variance as a constant. The unscented transform uses two test points which pass one another, so the estimated variance momentarily goes to zero. The actual variance,

$$E\left[(y-\bar{y})^2\right] = 1 + v_0^2 t^2 - \left(v_0 t \,\mathrm{erf}\left(\frac{v_0 t}{\sqrt{2}}\right) + \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2} v_0^2 t^2} \right)^2, \tag{7}$$

shrinks during the collision but remains finite.

3. A MORE ROBUST UNSCENTED TRANSFORM

Tracking problems commonly use a stochastic process model that involves process noise. We can accomodate this in the Monte Carlo method by independently choosing, for trial i, $\mathbf{x}_i$ and $\mathbf{q}_i$ randomly from their respective distributions. We can proceed analogously for the unscented transform. Suppose the $\mathbf{q}_i$ are normally distributed with covariance $\mathbf{Q}$. We can form an augmented state vector

$$\mathbf{s} = \begin{bmatrix} \mathbf{x} \\ \mathbf{q} \end{bmatrix} \tag{8}$$

from the original state components and the process noise vector, with covariance*:

$$\mathbf{S} = \begin{bmatrix} \mathbf{P} & 0 \\ 0 & \mathbf{Q} \end{bmatrix}. \tag{9}$$

We may then construct test points from the square root of $\mathbf{S}$.

In general, the added test points required to account for process noise will have distinct values for the state variables. If that is not true for the chosen type of matrix square root, one can ensure it by introducing a suitable

*If the distributions of process noise and the state variables were cross correlated, then $\mathbf{S}$ could be full rather than block diagonal.

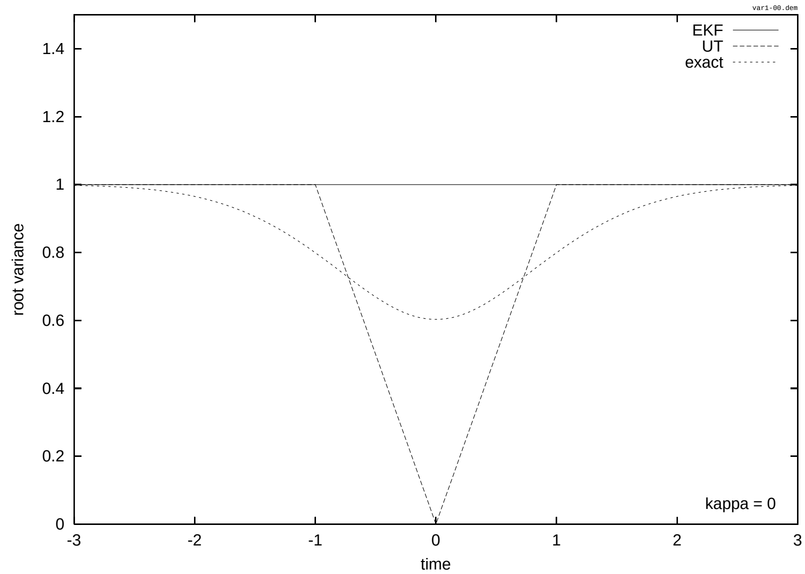

Figure 3. Estimates of distribution root variance for 1D example.

orthogonal transformation. These extra points can improve the mean and covariance estimates for the transformed distribution. We propose this as a natural way of improving the unscented filter: pretend there is some process noise even if there is none. Enlarge the state vector and covariance matrix as if accommodating several more "hidden variables", compute the test points, and use only the elements actually present in the state vector. We used unit variance for each hidden variable, but the actual value does not matter. This procedure may also be viewed as a way of conveniently constructing a non-square matrix root satisfying Eqn. 1. (Julier and Uhlmann mention the possibility of using a non-square matrix root.[3])

The possible improvements in estimation of the mean and variance are shown in Figs. 4 and 5, respectively.

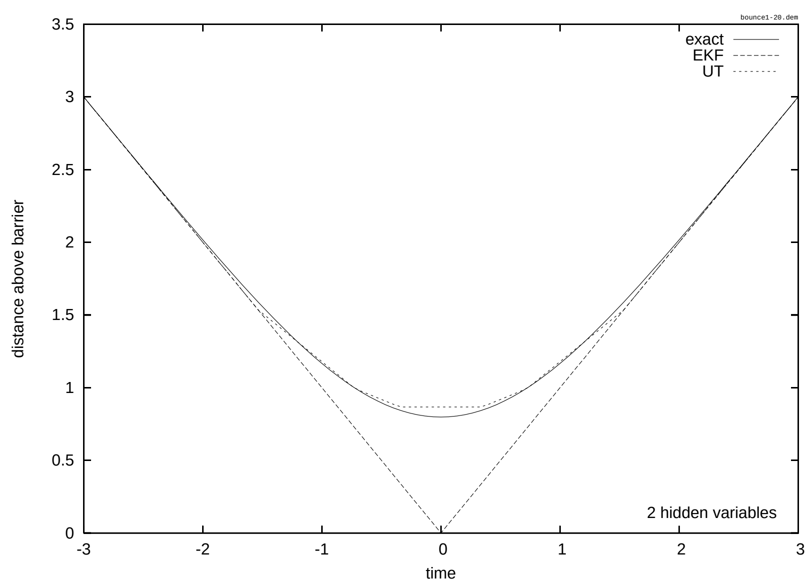

Figure 4. Estimates of distribution mean for 1D example, with two hidden variables.

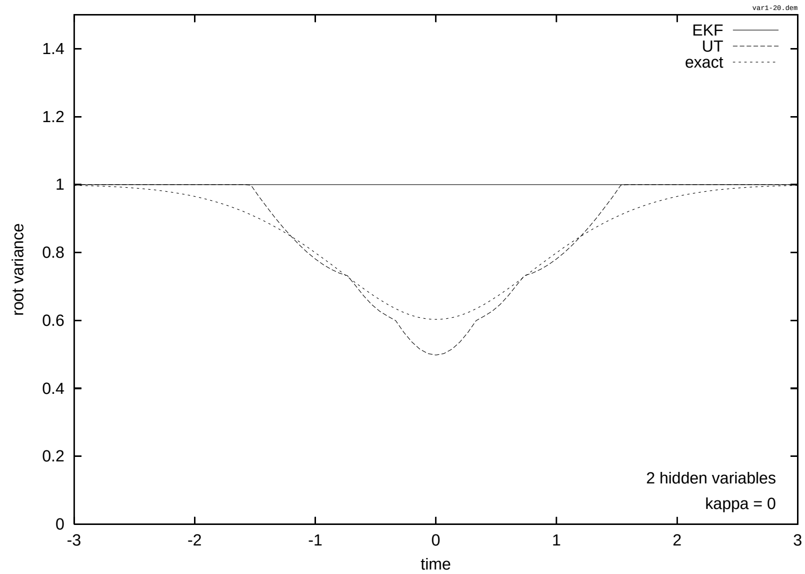

Figure 5. Estimates of distribution root variance for 1D example, with two hidden variables.

We hasten to point out that these results depend on a particular orthogonal transformation

$$\mathbf{U} = \begin{bmatrix} -0.416245 & -0.198412 & -0.887340 \\ 0.905154 & 0.002127 & -0.425078 \\ 0.086228 & -0.980116 & 0.178708 \end{bmatrix}, \tag{10}$$

which was chosen to minimize the mean square error in the estimated mean during the collision. This transformation generates the test points shown in Fig. 6. Since the covariance is assumed diagonal for this example, the test point values are a scaled version of the first row of $\mathbf{U}$:

$$\boldsymbol{\sigma} = \pm\sqrt{(n+h)\mathbf{P}}\mathbf{U} = \pm\sqrt{3\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}\mathbf{U} = \pm\begin{bmatrix} -.7209575 & -.3436597 & -1.5369180 \\ \# & \# & \# \\ \# & \# & \# \end{bmatrix}, \tag{11}$$

where $h = 2$ is the number of hidden variables, and we have suppressed the test point elements that are not actually used. Note that the cusps in Figs. 4 and 5 correspond to times when test points reach the barrier. The orthogonal transform in Eqn. 10 is in fact a little better for this problem than

$$\begin{bmatrix} \sqrt{1/35} & \sqrt{9/35} & \sqrt{5/7} \\ \sqrt{9/10} & -\sqrt{1/10} & 0 \\ \sqrt{1/14} & \sqrt{9/14} & -\sqrt{2/7} \end{bmatrix}, \tag{12}$$

which yields equally spaced test points.

Figs. 7 and 8 show results for a "random" orthogonal transformation.

We conjecture that, at least for the symmetric square root, orthogonal transforms can be found that will be effective for a variety of nonlinear functions, provided the number of real and hidden variables match.

3.1. 2D Example

The UT was also applied to a 2D example similar to one described by Julier and Uhlmann.[3] A particle with a binormally distributed initial position $\mathbf{x} = [x_1, x_2]^T$ and known velocity $[0, -v_0]$ approaches the corner of a barrier, such that it will either suffer an elastic collision or continue unimpeded depending on x_1. The transformation is

$$\mathbf{y} = \mathbf{g}(\mathbf{x}, t) = \begin{cases} [x_1, x_2 + v_0 t]^T & \text{if } x_1 < -0.25 \\ [x_1, |x_2 + v_0 t|]^T & \text{if } x_1 \geq -0.25 \end{cases}. \tag{13}$$

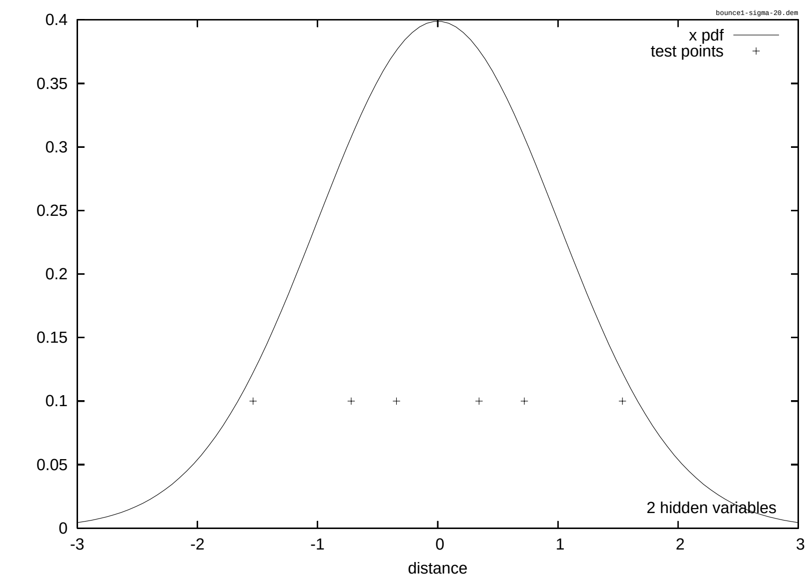

Figure 6. Test points for 1D example with two hidden variables.

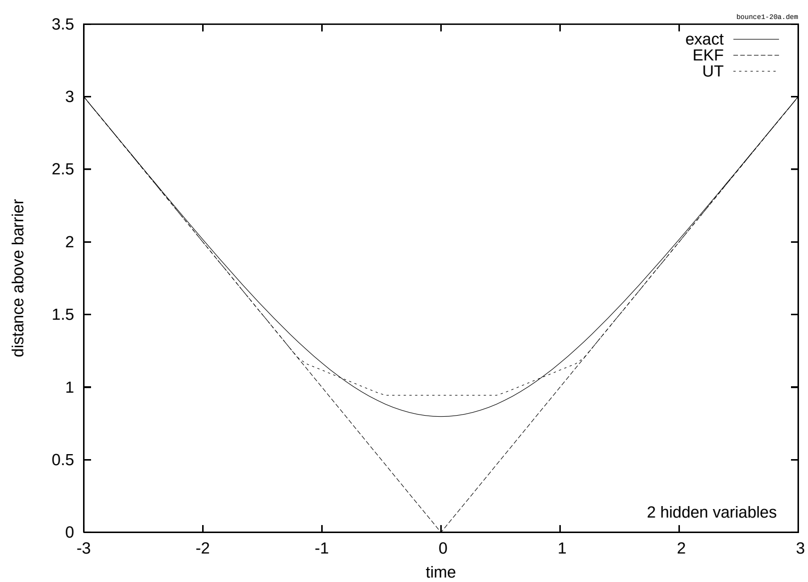

Figure 7. Estimates of distribution mean for 1D example, with two hidden variables and an unoptimized orthogonal transformation.

We wish to estimate the mean and variance of $\mathbf{y}$ as a function of time. The distribution at time $t = 6$ is shown in Fig. 9. The UT uses four test points, of which three collide and one continues. The estimated means are shown in Fig. 10.

We can improve our estimate of the mean by adding test points that sample more of the distribution. By adding two hidden variables (hence four more points) and using the orthogonal transformation

$$\mathbf{U} = \begin{bmatrix} -0.9561578 & 0.0521817 & 0.2880479 & -0.0082342 \\ -0.0845242 & -0.8885920 & -0.1070800 & 0.4379427 \\ -0.1652688 & -0.3078994 & -0.5151937 & -0.7825980 \\ -0.2265043 & 0.3359752 & -0.8000861 & 0.4423559 \end{bmatrix}, \tag{14}$$

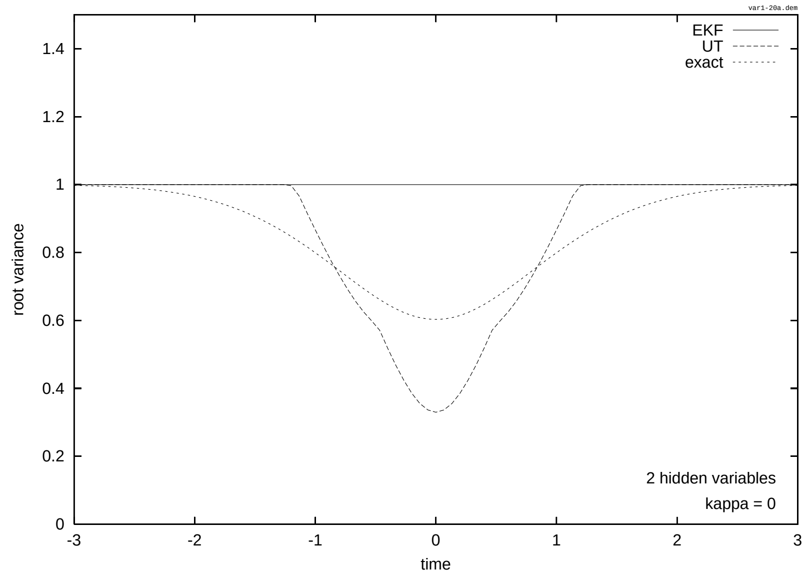

Figure 8. Estimates of distribution root variance for 1D example, with two hidden variables and an unoptimized orthogonal transformation.

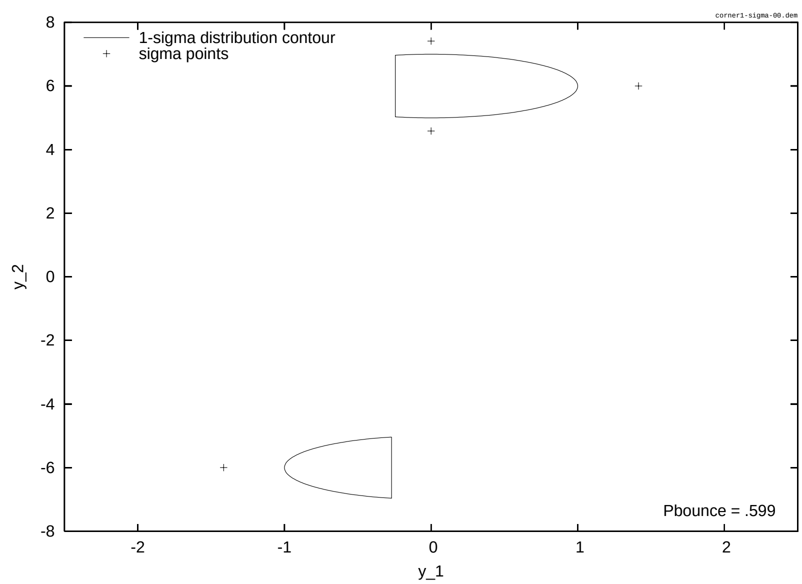

Figure 9. True transformed distribution and UT test points at time $t = 6$ for 2D example.

we obtained the test point distribution in Fig. 11 and the estimated mean in Fig. 12. Estimates of root variance with and without the hidden variables are shown in Fig. 13.

4. DISCUSSION

We have demonstrated that the mean and covariance estimated by the UT can be improved by adding more test points, and that a convenient way to add test points is to introduce "hidden variables" and a suitable orthogonal transformation. This technique may be usefully applied to functions with discontinuities. Possible applications include the tracking of targets with abrupt changes in dynamics, such as a missile at thrust termination or an aircraft entering a turn.

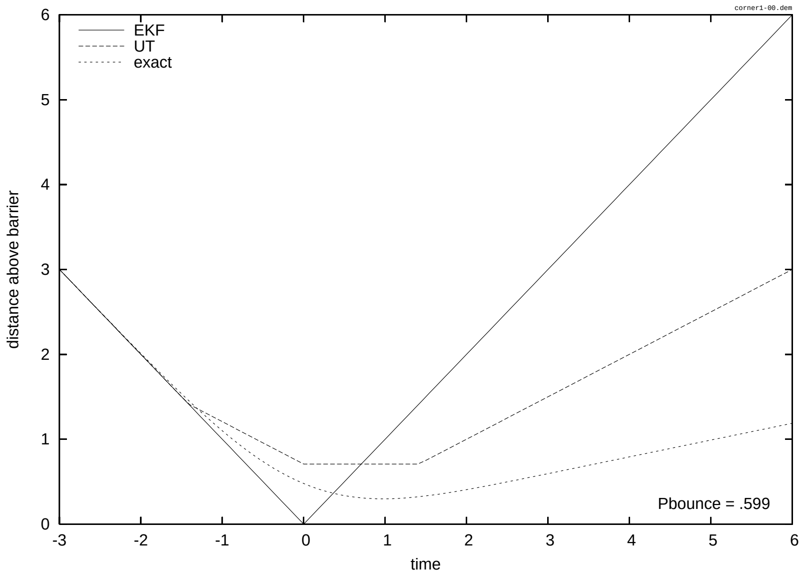

Figure 10. Estimates of distribution mean for 2D example.

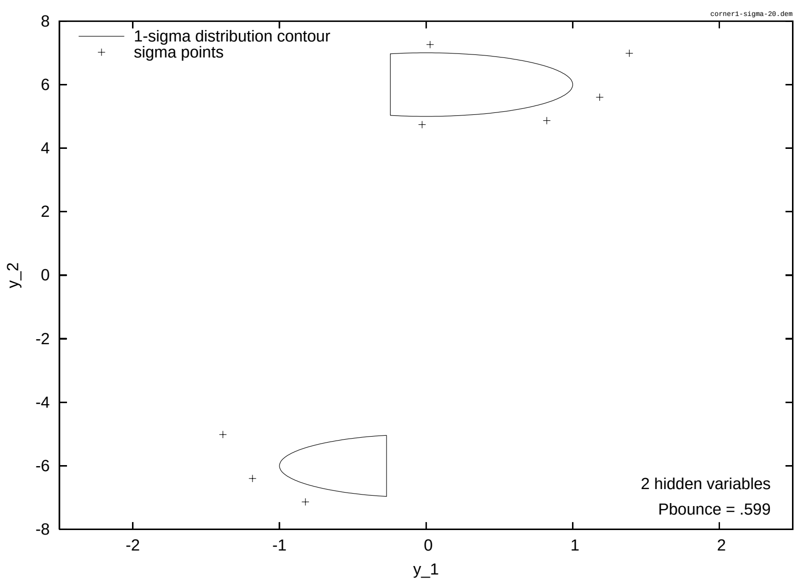

Figure 11. UT test points for 2D example with two hidden variables.

Even with the added test points proposed here, the UT is still accurate only to third order. For nonlinear functions introducing substantial fourth order errors, such as conversions between spherical and cartesian coordinates, a more suitable method is the fourth order UT developed by Julier and Uhlmann[4] requiring $O(n^2)$ test points.

Following the principle of reproducible research as advocated by Buckheit and Donoho,[9] the software which produces all figures in this paper is available from the author.

Acknowledgments

Thanks go to O. Drummond for encouraging this research and to J. Uhlmann for supplying his reference implementation of the unscented filter, to John W. Eaton for Octave,[10] and to William F. Schelter for MAXIMA.[8]

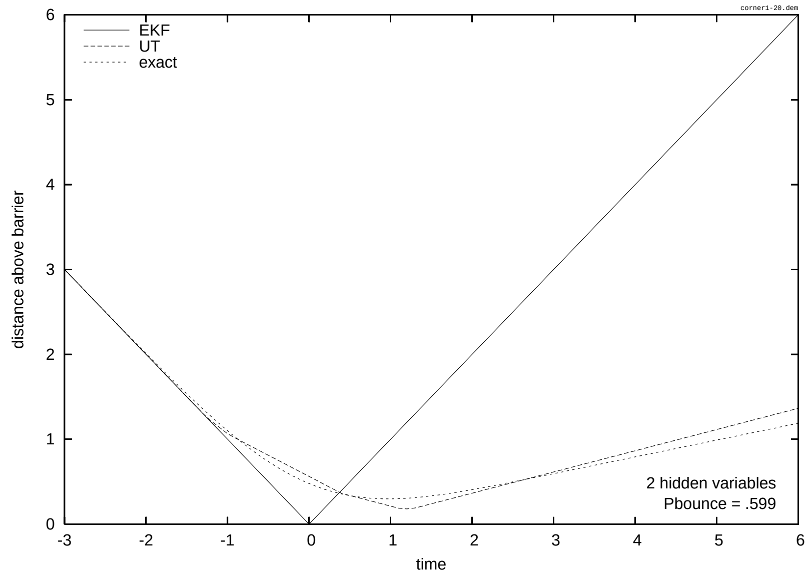

Figure 12. Estimates of distribution mean for 2D example with two hidden variables.

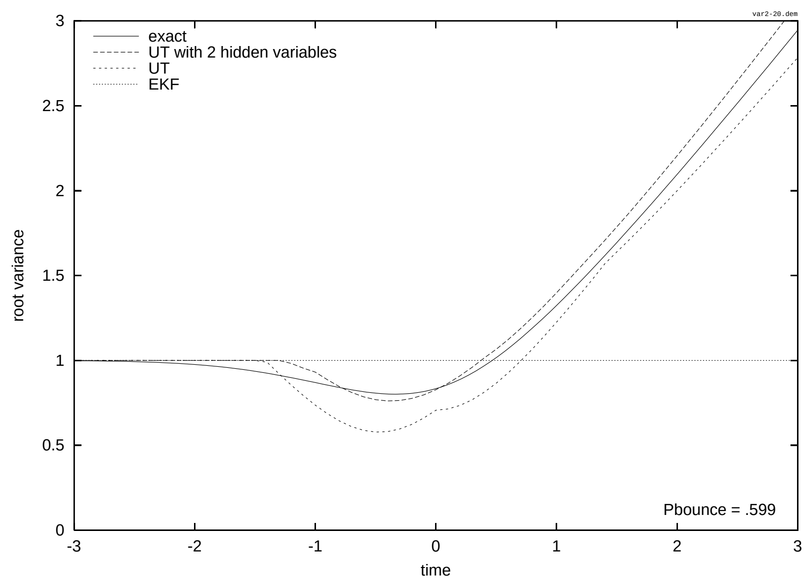

Figure 13. Estimates of distribution root variance for 2D example.

REFERENCES

1. R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME, Journal of Basic Engineering* **82**, pp. 34–45, Mar. 1960.
2. A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.
3. S. Julier and J. K. Uhlmann, "A general method for approximating nonlinear transformations of probability distributions," tech. rep., Robotics Research Group, Department of Engineering Science, University of Oxford, 1996.
4. S. J. Julier and J. K. Uhlmann, "A new extension of the kalman filter to nonlinear systems," in *The Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls, Orlando, Florida*, 1997.
5. S. J. Julier, "The scaled unscented transformation." Feb. 2000.

6. G. Strang, *Linear Algebra and its Applications*, Harcourt Brace Jovanovich, San Diego, 3rd ed., 1988.
7. S. J. Julier, 2001. http://www.robots.ox.ac.uk/ siju/work/uns.tar.gz.
8. W. F. Schelter. http://www.ma.utexas.edu/maxima.html.
9. J. B. Buckheit and D. L. Donoho, “Wavelab and reproducible research,” in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, eds., vol. 103 of *Lecture Notes in Statistics*, pp. 55–81, Springer Verlag, Nov 1995.
10. J. W. Eaton *et al.* http://www.octave.org/.