A TRANSFORMATION-BASED DERIVATION OF THE KALMAN FILTER AND AN EXTENSIVE UNSCENTED TRANSFORM

Friedrich Faubel, Dietrich Klakow

Spoken Language Systems,
Saarland University, D-66123 Saarbrücken, Germany
{friedrich.faubel, dietrich.klakow}@lsv.uni-saarland.de

ABSTRACT

In the unscented Kalman filter (UKF), the state vector is typically augmented with process and measurement noise in order to approximate the joint predictive distribution of state and observation. For that, the unscented transform is used. As its point selection mechanism changes the higher order moments between the random variables, statistical independence is not preserved. In this work, we show how statistical independence can be preserved by representing independent variables by separate point-sets. In addition to that, we show how the Kalman filter (KF) can be derived based on a particular type of linear transform that allows for a more uniform treatment of KF and UKF.

Index Terms— Kalman filter, conditional Gaussian distribution, unscented transform

1. INTRODUCTION

Usually, the *Kalman filter* (KF) [1, 2] is viewed as consisting of two steps: a prediction step that predicts the distribution of the system state; and an update step that corrects the predicted state distribution on arrival of a new observation. The Bayesian formulation of the KF [3, 4], however, suggests that – from a probabilistic point of view – it might make more sense to reinterpret these two steps as first constructing the joint, predictive Gaussian distribution of both state and observation; and then conditioning that distribution on the observation actually received (the realization of the random variable). This point of view is explained in more detail in Section 3, where, in contrast to [3, 4], we augment the state vector with process and measurement noise, as it is done in the *unscented Kalman filter* (UKF) [5, 6]. This allows for a unified treatment of both KF and UKF; and it directly yields generalizations such as the KF for correlated process and measurement noise [2].

For possibly nonlinear system and observation models, prediction of the joint distribution in general requires nonlinear transforms of random variables. In this case, the Kalman filter framework might not be optimal. However, it can still be applied if the joint predictive distribution is, at each time, approximated as a Gaussian. The UKF achieves that by using the *unscented transform* (UT) [5, 6], which approximates the initial Gaussian distribution by a point-mass representation; then maps all the points according to the potentially nonlinear function; and finally reestimates the mean and covariance in order to reobtain a Gaussian fit. For that, the state vector is typically augmented with process and measurement noise in order to obtain a joint Gaussian distribution of the random variables. Setting the cross-covariance terms to zero preserves uncorrelatedness, which – in case of a joint Gaussian distribution – implies statistical independence. Unfortunately, the latter is not valid for the UT as its point selection mechanism changes the higher-order moments between the random variables. That is the motivation behind introducing the *extensive unscented transform* (XUT) in Section 4. It preserves statistical independence by representing statistically independent random variables as separate point-sets and then considering the Cartesian product thereof.

This work was supported by the Federal Republic of Germany through the German Research Foundation (DFG) under the research training network IRTG 715 "Language Technology and Cognitive Systems".

2. TRANSFORMING GAUSSIAN RANDOM VARIABLES

2.1. Linear Transforms

Let A be a full-rank $n \times n$ matrix, so that A is invertible. Furthermore, let X be an n-dimensional Gaussian random variable with distribution $p_X(x) = \mathcal{N}(x; \mu, \Sigma)$, where $\mathcal{N}(; \mu, \Sigma)$ denotes a Gaussian distribution with mean μ and covariance matrix Σ. Then the distribution of the linear transform $Y = A \cdot X$ of the random variable X can be obtained with the fundamental transformation law of probabilities [7]: $p_Y(y) = \mathcal{N}(y; A\mu; A\Sigma A^T)$.

2.2. Kalman Filter-Type Linear Transforms

In order to construct the joint predictive distribution of state and observation for the Kalman filter we introduce a particular type of linear transform: the *KF-type linear transform* specified in (1). Let X and U be n- and m-dimensional Gaussian random variables, that have a joint Gaussian distribution

$$p_{X,U}(x,u) = \mathcal{N}\left(\begin{bmatrix} x \\ u \end{bmatrix}; \begin{bmatrix} \mu_X \\ \mu_U \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XU} \\ \Sigma_{UX} & \Sigma_{UU} \end{bmatrix}\right).$$

Let g be a linear function of the form $g(x,u) = Bx + u$ with an $m \times n$ matrix B. Furthermore, let Z be the random variable resulting from the linear transform

$$Z = \underbrace{\begin{bmatrix} I_{n,n} & 0_{n,m} \\ B & I_{m,m} \end{bmatrix}}_{\triangleq A} \begin{bmatrix} X \\ U \end{bmatrix} = \begin{bmatrix} X \\ g(X,U) \end{bmatrix} \tag{1}$$

with $n \times n$ and $m \times m$ identity matrices $I_{n,n}$ and $I_{m,m}$. Then A obviously has full rank. Hence we can make use of section 2.2, which yields:

$$p_Z(z) = \mathcal{N}\left(z; A\begin{bmatrix} \mu_X \\ \mu_U \end{bmatrix}, A\begin{bmatrix} \Sigma_{XX} & \Sigma_{XU} \\ \Sigma_{UX} & \Sigma_{UU} \end{bmatrix} A^T\right).$$

978-1-4244-2710-9/09/$25.00 ©

Now, substituting A by its definition from (1) and defining $Y \triangleq g(X,U)$, $p_Z(z)$ can be written as joint distribution of X and Y:

$$p_{X,Y}(x,y) = \mathcal{N}\left(\begin{bmatrix}x\\y\end{bmatrix}; \begin{bmatrix}\mu_X\\\mu_Y\end{bmatrix}, \begin{bmatrix}\Sigma_{XX} & \Sigma_{XY}\\\Sigma_{YX} & \Sigma_{YY}\end{bmatrix}\right) \quad (2)$$

with $\Sigma_{X,Y} = \Sigma_{Y,X}^T$ and

$$\begin{aligned}\mu_Y &= B\mu_X + \mu_U, \quad \Sigma_{Y,X} = B\Sigma_{XX} + \Sigma_{UX},\\ \Sigma_{Y,Y} &= B\Sigma_{XX}B^T + B\Sigma_{XU} + \Sigma_{UX}B^T + \Sigma_{UU}.\end{aligned} \quad (3)$$

2.3. The Unscented Transform

The unscented transform was introduced by Julier and Uhlmann [5] in order to approximate a nonlinear transform $Y = g(X)$ of an n-dimensional Gaussian random variable X, based on a point-mass representation that, by design, captures the first two moments of the distribution. Let $p_X(x)$ denote the distribution of X,

$$p_X(x) = \mathcal{N}(x; \mu_X, \Sigma_X)$$

with mean μ_X and covariance matrix Σ_X. Further, let $R^T R$ be the Cholesky decomposition of Σ_X. Then, denoting the rows of R by R_i and defining $\lambda = n + \kappa$ for an arbitrary $\kappa \in \mathcal{R}$, the distribution of X can be represented by the weighted empirical distribution

$$\tilde{p}_X(x) = \sum_{i=0}^{2n} W_i \delta(x - \mathcal{X}_i) \quad (4)$$

where δ is the Dirac delta and where the points and weights, $\mathcal{X}_i$ and W_i, are given by

$$\begin{array}{lcl|lcl}\mathcal{X}_0 & = & \mu_X & W_0 & = & \kappa/\lambda\\ \mathcal{X}_{2i+1} & = & \mu_X + \sqrt{\lambda}R_i & W_{2i+1} & = & 1/(2\lambda)\\ \mathcal{X}_{2i+2} & = & \mu_X - \sqrt{\lambda}R_i & W_{2i+2} & = & 1/(2\lambda)\end{array} \quad (5)$$

$i = 0, \ldots, (n-1)$. Note that κ specifies how much weight is placed on the mean, $\mathcal{X}_0$. Setting it to $1/2$ results in a weight of $1/n$ for each of the points. Setting it to $3 - n$ minimizes the error in the fourth moment [6]. Similar as in Monte Carlo methods [7], the weighted points $\mathcal{X}_i$ can be instantiated through the function g, $\mathcal{Y}_i = g(\mathcal{X}_i)$, which then in turn yields a weighted empirical distribution of Y:

$$\tilde{p}_Y(y) = \sum_{i=0}^{2n} W_i \delta(y - \mathcal{Y}_i). \quad (6)$$

Consequently, a Gaussian approximation $\hat{p}_Y(y) = \mathcal{N}(Y; \hat{\mu}_Y, \hat{\Sigma}_Y)$ of the distribution can be obtained by estimating the mean and covariance of Y in a maximum likelihood fashion:

$$\hat{\mu}_Y = \sum_{i=0}^{2n} W_i \mathcal{Y}_i, \quad \hat{\Sigma}_Y = \sum_{i=0}^{2n} W_i (\mathcal{Y}_i - \mu_Y)(\mathcal{Y}_i - \mu_Y)^T. \quad (7)$$

That is the unscented transform. It is exact for linear transforms. For nonlinear transforms its mean and covariance estimates are accurate up to the second order term of the Taylor series expansion [6].

3. THE DISCRETE-TIME KALMAN FILTER REVISITED

The discrete-time Kalman filter is based on a discrete-time *dynamic state-space model* (DSSM) model, where a system state x_t evolves according to a process equation

$$x_t = f(x_{t-1}, w_t), \quad (8)$$

driven by Gaussian process noise w_t. The system state itself is considered unknown, but related to observations y_t through a measurement equation

$$y_t = h(x_t, v_t), \quad (9)$$

where v_t denotes Gaussian measurement noise. The process and measurement noise distributions are given by

$$p_{W_t}(w_t) = \mathcal{N}(w_t; \mu_{W_t}, \Sigma_{W_t W_t}), \quad (10)$$

$$p_{V_t}(v_t) = \mathcal{N}(v_t; \mu_{V_t}, \Sigma_{V_t V_t}), \quad (11)$$

where μ_{W_t} and $\Sigma_{W_t W_t}$ denote the mean and covariance matrix of W_t, μ_{V_t} and $\Sigma_{V_t V_t}$ the mean and covariance matrix of V_t, respectively. For the Kalman filter, the functions f and h are required to be linear, as only then the filtering density – that is $p(x_t|y_{1:t})$ with $y_{1:t} \triangleq \{y_1, \ldots, y_t\}$ – remains Gaussian while being propagated in time. Based on this DSSM, the Kalman filter is typically derived by using the orthogonality principle, linear regression, the matrix inversion lemma, et cetera. The result is a set of equations, which are commonly used by engineers, but which can obscure a probabilistic understanding of what the filter is actually doing. Nevertheless, there is a very simple explanation to the KF, which directly follows from the Bayesian formulation [3, 4]: it does nothing else but

1. construct the joint Gaussian distribution of the next state and observation X_t and Y_t, given the process and measurement models – specified by (8) and (9) – as well as the observation history $y_{1:t-1}$: $p(x_t, y_t|y_{1:t-1})$.
2. condition that joint Gaussian distribution on $Y_t = y_t$, on arrival (realization) of a new observation y_t, in order to obtain $p(x_t|y_{1:t})$.

Alternating between these two steps propagates the filtering density $p(x_t|y_{1:t})$ in time. The construction step is explained in more detail in Section 3.1 where we make use of the KF-type transform from Section 2.2 in order to obtain the joint predictive distribution. Conditioning on a new observation is described in Section 3.2.

3.1. Constructing the Joint Distribution (Prediction)

For the KF, it is required that the process and measurement equations can be written $f(x_{t-1}, w_t) = Fx_{t-1} + w_t$ and $h(x_t, v_t) = Hx_t + v_t$ with matrices F and H. Then, given $X_{t-1}|y_{1:t-1}$ is a Gaussian Random variable with mean $\mu_{X_{t-1}}$ and covariance $\Sigma_{X_{t-1}}$, the KF-type transform from Section 2.2 can be applied with $X = X_{t-1}$, $U = W_t$, $Y = X_t$, $g = f$ and $B = F$, in order to obtain the joint distribution $p(x_{t-1}, x_t|y_{1:t-1})$ of last and current state (X_{t-1}, X_t) given the observation history $y_{1:t-1}$:

$$\mathcal{N}\left(\begin{bmatrix}x_{t-1}\\x_t\end{bmatrix}; \begin{bmatrix}\mu^-_{X_t}\\\mu_{X_{t-1}}\end{bmatrix}, \begin{bmatrix}\Sigma_{X_{t-1}X_{t-1}} & \Sigma^-_{X_{t-1}X_t}\\\Sigma^-_{X_t X_{t-1}} & \Sigma^-_{X_t X_t}\end{bmatrix}\right),$$

where the parameters are calculated according to (3):

$$\begin{aligned}\mu^-_{X_t} &= F\mu_{X_{t-1}} + \mu_{W_t}, \quad \Sigma^-_{X_t X_{t-1}} = F\Sigma_{X_{t-1}X_{t-1}},\\ \Sigma^-_{X_t X_t} &= F\Sigma_{X_{t-1}X_{t-1}}F^T + \Sigma_{W_t W_t}.\end{aligned} \quad (12)$$

In (12), we made the assumption that X_{t-1} and W_t are uncorrelated, i.e. $\Sigma_{X_{t-1}W_t} = 0$. Marginalizing over x_{t-1} (see [8] on marginal Gaussian distributions) gives the distribution $p(x_t|y_{1:t-1})$ of the predicted state X_t conditioned on $y_{1:t-1}$:

$$p(x_t|y_{1:t-1}) = \mathcal{N}(x_t; \mu^-_{X_t}, \Sigma^-_{X_t X_t}).$$

Now, the joint predictive distribution of state and observation (X_t, Y_t) can be constructed by again using the KF-type transform from 2.2,

but this time with $X = X_t$, $U = V_t$, $Y = Y_t$, $g = h$ and $B = H$. That gives

$$p(x_t, y_t|y_{1:t-1}) = \mathcal{N}\left(\begin{bmatrix} x_t \\ y_t \end{bmatrix}; \begin{bmatrix} \mu^-_{X_t} \\ \mu_{Y_t} \end{bmatrix}, \begin{bmatrix} \Sigma^-_{X_tX_t} & \Sigma_{X_tY_t} \\ \Sigma_{Y_tX_t} & \Sigma_{Y_tY_t} \end{bmatrix}\right). \tag{13}$$

Assuming uncorrelated process and measurement noise, i.e. $\Sigma_{X_tV_t} = 0$, the parameters of the distribution are:

$$\begin{aligned} &\mu_{Y_t} = H\mu^-_{X_t} + \mu_{V_t}, \quad \Sigma_{X_tY_t} = H\Sigma^-_{X_tX_t}, \\ &\Sigma_{Y_tY_t} = H\Sigma^-_{X_tX_t}H^T + \Sigma_{V_tV_t}. \end{aligned} \tag{14}$$

Not making that assumption directly gives the Kalman filter for correlated process and measurement noise [2].

3.2. Conditioning on the Observation (Update)

On arrival of a new observation y_t, the joint Gaussian distribution of $(X_t, Y_t)|y_{1:t-1}$ can be conditioned on $Y_t = y_t$, resulting in the conditional Gaussian distribution (see [8] on conditional Gaussian distributions):

$$p(x_t|y_{1:t}) = \mathcal{N}(x_t; \mu^+_{X_t}, \Sigma^+_{X_t,X_t}) \tag{15}$$

with conditional mean and covariance

$$\begin{aligned} \mu^+_{X_t} &= \mu^-_{X_t} + \Sigma_{X_t,Y_t}\Sigma^{-1}_{Y_t,Y_t}(y_t - \mu_{Y_t}), \\ \Sigma^+_{X_tX_t} &= \Sigma^-_{X_tX_t} - \Sigma_{X_tY_t}\Sigma^{-1}_{Y_tY_t}\Sigma_{Y_tX_t}, \end{aligned} \tag{16}$$

where the $\mu^-_{X_t}$, $\Sigma^-_{X_tX_t}$, $\Sigma_{X_tY_t}$, $\Sigma_{Y_tY_t}$ and μ_{Y_t} are defined as in (12) and (14). These are the famous Kalman filter update formulae. A superscript "$-$" denotes a parameter is conditioned on $y_{1:t-1}$, a "$+$" denotes a parameter is conditioned on $y_{1:t}$.

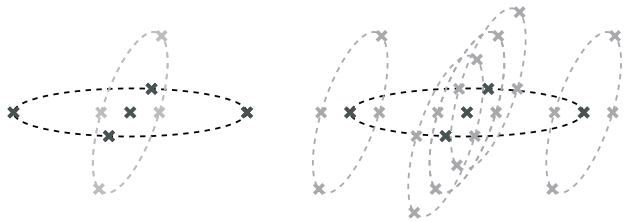
4. THE EXTENSIVE UNSCENTED TRANSFORM

Both the extended and unscented Kalman filter are based on the same principle as the original Kalman filter. They still construct a joint Gaussian distribution for $(X_t, Y_t)|y_{1:t-1}$ and then condition it on the new observation y_t. The point where they differ is in how the joint Gaussian distribution with generally nonlinear functions f and h is approximated. The extended Kalman filter uses a first-order Taylor series approximation around the current state estimates [2] in order to "locally" linearize f and h. The unscented Kalman filter [5] replaces the KF-type transform by the unscented transform (UT) from Section 2.3 with

$$\tilde{X} = \begin{bmatrix} X \\ U \end{bmatrix}, \quad \tilde{g}(x,u) = \begin{bmatrix} x \\ g(x,u) \end{bmatrix}, \quad \tilde{Y} = \begin{bmatrix} X \\ Y \end{bmatrix},$$

where the tilde-terms are terms from the UT, the non-tilde-ones are terms from the KF-type transform. We call this the augmented unscented transform (AUT) in contrast to the extensive unscented transform (XUT), which we introduce here. As mentioned before, the motivation is to obtain a point-mass representation that preserves statistical independence of the individual variables. Let X and U be n and m dimensional Gaussian random variables as in the KF-type transform from section 2.2. Moreover, let X and U be statistically independent as assumed for the Kalman filter. Then, making use of the point-mass representation from the unscented transform, the distributions of X and U can be approximated by

$$\tilde{p}_X(x) = \sum_{i=0}^{2n} W_i^{(X)}\delta(x - \mathcal{X}_i), \quad \tilde{p}_U(u) = \sum_{j=0}^{2m} W_j^{(U)}\delta(u - \mathcal{U}_j)$$



(a) augmented *(b) extensive*

Fig. 1. Augmented versus Extensive Unscented Transform. *The pictures to the left and to the right portray the points (crosses) used by the augmented and extensive unscented transforms along with covariance ellipses (dashed lines). The black crosses indicate the $2n + 1 = 5$ points chosen for X, augmented with the mean of U.*

with points and weights

$$\begin{aligned} &\mathcal{X}_i,\ W_i^{(X)} \quad i = 1, \ldots, 2n+1 \\ &\mathcal{U}_j,\ W_j^{(U)} \quad j = 1, \ldots, 2m+1 \end{aligned} \tag{17}$$

where $\mathcal{X}_i$, $W_i^{(X)}$ and $\mathcal{U}_j$, $W_j^{(U)}$ are defined as in (5), respectively. Consequently, the joint distribution of the statistically independent variables X and U can be approximated as

$$\tilde{p}_{X,U}(x,u) = \sum_{i=0}^{2n}\sum_{j=0}^{2m} W_i^{(X)}W_j^{(U)}\delta(x - \mathcal{X}_i)\delta(u - \mathcal{U}_j). \tag{18}$$

Transforming all the points $[\mathcal{X}_i^T\ \mathcal{U}_j^T]^T$ gives a weighted empirical distribution for the nonlinear transform $Y = g(X,U)$:

$$\tilde{p}_Y(y) = \sum_{i=0}^{2n}\sum_{j=0}^{2m} W_{i,j}^{(Y)}\delta(y - \mathcal{Y}_{i,j})$$

with $\mathcal{Y}_{i,j} = g(\mathcal{X}_i, \mathcal{U}_j)$ and $W_{i,j}^{(Y)} = W_i^{(X)}W_j^{(U)}$. Then, estimating the mean and covariance in analogy to (7) yields a Gaussian approximation to the distribution of Y. That is what we call the extensive unscented transform. The difference to the augmented unscented transform is portrayed in Figure 1: The AUT augments each of the $(2n + 1)$ points chosen for X with the mean of U and, conversely, each of the $(2m + 1)$ points chosen for U with the mean of X, resulting in the points on the two ellipses portrayed in Figure 1-(a). The XUT, in contrast, uses all possible combinations of points that the AUT considers for X and U individually, i.e. it considers $(2n + 1) \cdot (2m + 1)$ points instead of $(2n + 2m + 1)$. That results in an increase in computation time, which, however, is still low compared to full Gauss-Hermite quadrature [9] whose computational cost grows exponentially with the dimension.

As the point-set of the XUT correctly captures the mean and covariance of $p_{X,U}(x,u)$, the XUT is exact up to the second order term of the Taylor series expansion [6], just as the UT is. In the case where $n = m = 1$ and κ from Section 2.3 is set to 2, the XUT coincides with the 3-point Gauss-Hermite quadrature rule. This is a direct consequence of Ito and Xiong's work [9]. In addition to that, the XUT point-set guarantees statistical independence of X and U as $\tilde{p}_{X,U}(x,u)$ is, by definition, equal to $\tilde{p}_X(x)\tilde{p}_U(u)$. This does not hold for the UT whose point selection mechanism changes the higher-order moments between X and U, as shown in Appendix II of [6]. Especially, for $n = m = 1$ and even integers k and l we have

$$\mathrm{E}[x^k u^l] \approx W_0\mu_X^k\mu_U^l + \sum_{i=1}^{2n} W_i\mathcal{X}_i^k\mu_U^l + \sum_{i=2n+1}^{2(n+m)} W_i\mu_X^k\mathcal{U}_i^l, \tag{19}$$

Table 1. *MSE / track averaged over 50 runs*

Measurements	UKF		PF		
	AUT	XUT	5K	10K	25k
Polar	927	904	1098	1037	958
Cartesian	644	644	791	746	703

Table 2. *MSE / track in dependence on κ*

κ	-0.5	0	0.5	1	1.5	2
AUT	1098	1206	1713	2267	2533	2856
XUT	904	926	946	962	986	1009

which clearly violates $\mathrm{E}[x^k u^l] = \mathrm{E}[x^k]\mathrm{E}[u^l]$ and which directly translates to higher dimensions. That might be detrimental if the nonlinear transform is sensitive to these moments.

5. EXPERIMENTS

In order to evaluate the performance of the extensive unscented transform, we performed a series of simulations, in which a maneuvering object was tracked based on sensor measurements. In these simulations, the object moved along the synthetic trajectory portrayed in Figure 2,

$$x_t = 10\left[\sin(s_t)+1 \quad (\cos(s_t)+1)\sin(\tfrac{s_t}{2}) \quad \cos(2s_t)s_t\right]^T,$$

$s_t = \frac{4\pi t}{500}$, $t = 1,\ldots,500$, and was observed by virtual sensors located at $[0\ 0\ 0]^T$ providing measurements in Cartesian or polar coordinates, respectively. Additive measurement noise was simulated from a Gaussian, whose means and covariances were chosen at random – once for each of the 50 experiments performed. For polar measurements, the average variance was 0.073 for the distance, 0.0031 and 0.0044 for the angles. For Cartesian measurements it was 0.45, 0.87 and 1.23, respectively. As a process model we used $x_t = x_{t-1} + w_t$ with zero-mean Gaussian process noise w_t, whose covariance matrix was estimated on the synthetic trajectory and further scaled by a factor of two in order to increase stability of the UKF. At time $t = 1$ the filters were initialized with a Gaussian distribution around the true state x_1 with process noise covariance.

Table 1 shows the mean squared error (MSE) for a UKF using the AUT or XUT, respectively, as well as for a *sampling importance resampling* (SIR) *particle filter* (PF) with 5000, 10000 and 25000 particles. All the numbers are averaged over 50 simulations and correspond to 500 point trajectories. The upper half of the table shows MSEs for measurements in Cartesian coordinates, the lower half for measurements in polar coordinates. In the Cartesian (linear) case the AUT and XUT-based UKFs gave the same results – as should be expected. In the polar (nonlinear) case, the XUT-based UKF performed slightly better than the AUT-based one, if the parameter κ from section 2.3 was chosen optimally. For the AUT the optimal value was $\kappa = -3$, for the XUT it was $\kappa = -0.5$. Lower values resulted in complete divergence due to indefinite covariance matrices. Table 2 shows the MSEs obtained with different settings. The PF performed slightly worse, but got close to the UKF results if a great numbers of particles was used. With 25K particles the PF's computation time was over 120 times higher than that of the XUT-based UKF.

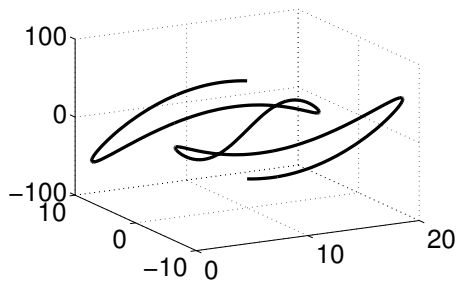


Fig. 2. *Trajectory used in the simulations.*

6. CONCLUSIONS

We have shown how the Kalman filter can be derived based on a particular type of linear transform that directly motivates the augmentation of the state vector with noise vectors in the UKF. In addition to that, we have shown how the point-set of the augmented unscented transform can be extended in such a way that statistical independence of the random variables is preserved, which might slightly improve accuracy in nonlinear tracking problems.

7. ACKNOWLEDGEMENTS

We would like to thank Steve Renals and John McDonough for making good suggestions on an early version of this paper as well as Peter Bell from CSTR for correcting the English. Special thanks to the anonymous reviewers, some of whom greatly helped to improve the quality of this paper by posing critical questions and pointing us to usefull references.

8. REFERENCES

[1] S. S. Haykin, *Adaptive Filter Theory*, Prentice Hall International, fourth edition, 2001.

[2] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*, Wiley, 2006.

[3] Y. C. Ho and R. C. K. Lee, "A Bayesian approach to problems in stochastic estimation and control," *IEEE Trans. Automatic Control*, vol. 9, no. 4, pp. 333–339, Oct. 1964.

[4] R. J. Meinhold and N. D. Singpurwalla, "Understanding the Kalman filter," *The American Statistician*, vol. 37, no. 2, pp. 123–127, May 1983.

[5] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," *Proc. AeroSense: The 11th Int. Symp. on Aerospace/Defense Sensing, Simulation and Controls*, 1997.

[6] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.

[7] C.P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer Texts in Statistics. Springer, second edition, 2004.

[8] C. M. Bishop, *Pattern Recognition and Machine Learning*, Information Science and Statistics. Springer, 2006.

[9] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Trans. Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000.