Optimal Filtering for Polynomial Measurement Nonlinearities with Additive Non–Gaussian Noise

Uwe D. Hanebeck
Institute of Automatic Control Engineering
Technische Universität München
80290 München, Germany
Uwe.Hanebeck@ieee.org

Abstract

In this paper, we consider the problem of estimating the N–dimensional state of a dynamic system based on M–dimensional discrete–time measurements. Measurements depend nonlinearly on the state and are corrupted by white non–Gaussian noise. The problem is solved by recursively calculating the complete posterior density of the state given the measurements. For that purpose, a new exponential type density is introduced, the so called pseudo Gaussian density, which is used to represent the complicated non–Gaussian posterior densities resulting from the recursion. For polynomial measurement nonlinearities and for pseudo Gaussian noise densities, it is shown that the result of the optimal Bayesian measurement update is *exactly* obtained by a Kalman Filter operating in a higher dimensional space. The resulting filtering algorithms are easy to implement and always guarantee valid posterior densities.

1 Introduction

Filtering consists of estimating parameters of one process, the system state sequence, given uncertain information from another related process, the measurement sequence. When the measurements are related nonlinearly to the system state, this estimation problem is in general difficult to solve. Usually, a linearization is performed to permit application of filtering methods derived for linear systems [2]. Of course, this only works for certain type of nonlinearities. In addition, the presence of non–Gaussian measurement noise further limits the applicability of linear methods.

More advanced methods for providing state estimates in the nonlinear case have been developed by keeping nonlinear terms in a Taylor series expansion of the nonlinearity, see [4] for an elegant derivation. However, here the focus is on calculating the *complete* posterior density of the unknown system state given all the measurements. A parametric closed–form density description is desired, which is defined by a finite number of parameters. In addition, the density representation should allow for recursive application and should not suffer from a permanently growing number of description parameters with an increasing number of available measurements.

A grid representation of densities for numerical nonlinear filtering based on quantization of the state space has been introduced in [5], but has proven to be useful only for a limited state vector dimension [3]. Monte Carlo techniques [7, 13] use stochastic samples to represent density functions in order to numerically solve the filtering problem.

Closed–form representations of densities include the Edgeworth expansion, i.e., a Gaussian density times a sum of Hermite polynomials, which has been proposed in [16]. A method for updating this type of density numerically is described in [6]. The approach has the disadvantage that truncated Edgeworth expansions are not themselves valid density functions and may give negative values [10]. A Gaussian mixture representation has been proposed in [1], which always provides valid density functions. However, each term is individually updated based on linearization, which results in a bank of parallel extended Kalman filters.

The most simple form of the measurement update seems to be obtained when using exponential type densities [12]. In addition, these densities are always positive. However, depending on the exponent function, e.g. polynomials, numerical inaccuracies during the update recursion may lead to densities that are not integrable, i.e., the integral over the density does not give a finite value.

In this paper, a new type of exponential density, the so called pseudo Gaussian density, is proposed. It is defined by a standard Gaussian function in a hyperspace S^* related to the original state space S via a nonlinear transformation. Because of its special structure, pseudo Gaussians are always valid density functions even in the presence of numerical inaccuracies. In addition, it will be shown that under certain assumptions, this type of density can be *exactly* updated by means of a Kalman filter operating in the hyperspace S^*.

Section 2 formulates the nonlinear filtering problem. In Section 3 the concept of pseudo Gaussian densities is explained in detail. The new nonlinear filtering algorithm is then derived in Section 4 and illustrated in Section 5 by means of a simple simulation example.

2 Problem Formulation

Estimating the state of a nonlinear dynamic system is considered, which may either evolve in continuous time or in discrete time steps. The system state is not directly observable, but will instead be deduced from measurements of the system output. Measurements are assumed to be taken sequentially at discrete time steps $k = 1, 2, \ldots$ and are corrupted by white non–Gaussian noise.

An M–dimensional measurement $\underline{\hat{y}}_k$ at time step k is related to the N–dimensional system state $\underline{x}_k$ via the *nonlinear time–variant* measurement equation

$$\underline{\hat{y}}_k = \underline{h}_k(\underline{x}_k) + \underline{v}_k$$

and is corrupted by *additive* white noise $\underline{v}_k$ from a possibly non–Gaussian noise density $p_v(\underline{v}_k)$. In this paper, the focus is on polynomial nonlinearities $\underline{h}_k(\underline{x}_k)$.

Instead of providing point estimates of the unknown state $\underline{x}_k$, an estimator should construct the complete conditional density of the state

$$p_e(\underline{x}_k) = p(\underline{x}_k | \underline{\hat{y}}_k, \underline{\hat{y}}_{k-1}, \ldots, \underline{\hat{y}}_1)$$

given all observations up to time step k. A recursive estimation procedure is preferred, which calculates a state estimate based on the estimate at the previous time step and hence, does not require to store all measurements. A suitable time update procedure is assumed for that purpose, which produces a predicted density

$$p_p(\underline{x}_k) = p(\underline{x}_k | \underline{\hat{y}}_{k-1}, \ldots, \underline{\hat{y}}_1)$$

by propagating the previous estimate $p_e(\underline{x}_{k-1})$ through the system model.

This paper is concerned with the filtering step (measurement update) only, i.e., how to recursively incorporate the information provided by a measurement $\underline{\hat{y}}_k$ into the prior density $p_p(\underline{x}_k)$ to construct the posterior density $p_e(\underline{x}_k)$. Observability is assumed, but not discussed here. The prediction step (time update) is outside the scope of this paper.

Although not strictly required, an initial density $p_e(\underline{x}_0)$ is assumed to be given.

3 Pseudo Gaussians

The key idea is to represent complicated probability density functions in the N–dimensional original state space S_x by simpler densities in a higher dimensional space S_x^*. Points $\underline{x}_k$ in S_x are related to points $\underline{x}_k^*$ in S_x^* via a nonlinear transformation $\underline{T}_x(.)$ according to

$$\underline{x}_k^* = \underline{T}_x(\underline{x}_k) = [T_1(\underline{x}_k), \ldots, T_{L_x}(\underline{x}_k)]^T \ ,$$

where L_x denotes the dimension of space S_x^*. Hence, the original space S_x is transformed by $\underline{T}_x(.)$ to an N–dimensional manifold U_x^* in the L_x–dimensional space S_x^*.

In S_x^*, L_x–dimensional Gaussian probability density functions are defined according to

$$p(\underline{x}_k^*) = c_k^x \exp\left\{ -\frac{1}{2}(\underline{x}_k^* - \underline{\hat{x}}_k^*)^T (\mathbf{C}_x^*)^{-1} (\underline{x}_k^* - \underline{\hat{x}}_k^*) \right\}$$

with mean $\underline{\hat{x}}_k^*$, symmetric positive definite covariance matrix $\mathbf{C}_k^{x,*}$, and normalizing constant c_k^x. Densities of this type will be called pseudo Gaussian in the following, because the elements of $\underline{x}_k^*$ are not independent.

The intersection of a pseudo Gaussian $p(\underline{x}_k^*)$ with the manifold U_x^* defines a non–Gaussian, e.g. multimodal, probability density function in the original space S_x.

Remark 3.1 A non–Gaussian density in the original space S_x is defined by *both* the transformation $\underline{T}_x(.)$ *and* the mean $\underline{\hat{x}}_k^*$ and covariance matrix $\mathbf{C}_k^{x,*}$ of the pseudo Gaussian $p(\underline{x}_k^*)$.

Example 3.1 A scalar state x_k is considered, which is related to a two–dimensional state $\underline{x}_k^*$ via

$$\underline{x}_k^* = \underline{T}_x(\underline{x}_k) = [x_k, x_k^2]^T \ .$$

An example of a pseudo Gaussian density defined in the space S_x^* with mean

$$\underline{\hat{x}}_k^* = \begin{bmatrix} 0 & 2 \end{bmatrix}^T$$

and covariance matrix

$$\mathbf{C}_k^{x,*} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

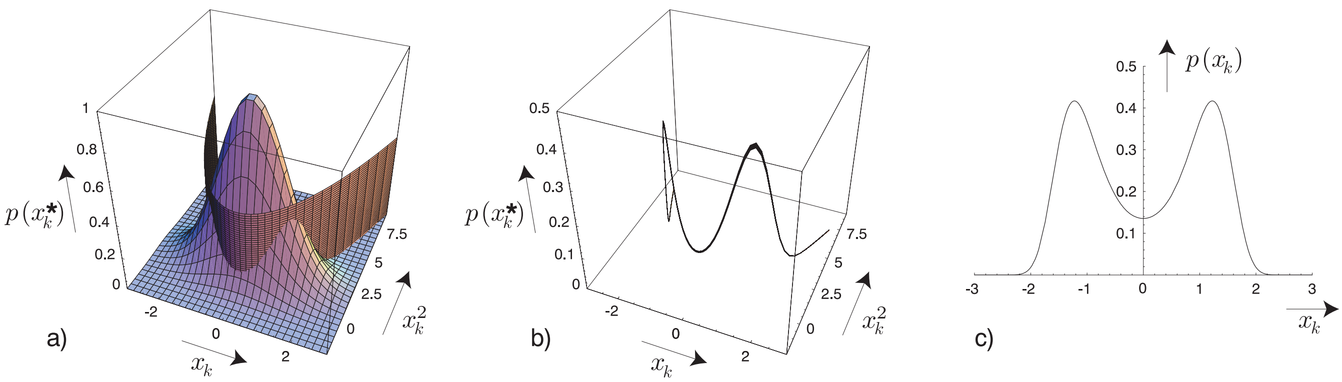

Figure 1: Example for demonstrating the concept of pseudo Gaussians with scalar state x_k and twodimensional hyperspace S_x^*. a) Pseudo Gaussian in hyperspace S^* with mean and covariance matrix according to example 3.1. b) Parts of the pseudo Gaussian density lying on the manifold U_x^*. c) Corresponding density in the original space S_x.

is shown in Fig. 1 a) together with the manifold U_x^*. Fig. 1 b) then shows that part of the pseudo Gaussian density lying on the manifold U_x^*, which defines the density in the original space shown in Fig. 1 c).

The selection of the functions $T_i(\underline{x}_k)$, $i = 1, \dots, L_x$ depends on the type of nonlinearity considered. However, multidimensional Bernstein–polynomials appear to be a good choice in many cases, e.g. polynomial nonlinearities. They are defined on the basis of one–dimensional Bernstein–polynomials, which on the interval $[l, r]$ are given by

$$H_i^n(x) = \binom{n}{i} \left(\frac{l-x}{l-r}\right)^i \left(\frac{r-x}{r-l}\right)^{n-i}$$

for $i = 0, \dots, n$. With

$$\underline{x}_k = \begin{bmatrix} x_k^1 & x_k^2 & \dots & x_k^N \end{bmatrix}^T \quad ,$$

the above transformation is defined by

$$T_i(\underline{x}_k) = \prod_{j=1}^{N} H_{i_j}^{L_j - 1}(x_k^j) \quad ,$$

for $i_j = 0, \dots, L_j - 1$, $j = 1, \dots, N$, $L_x = \prod_{j=1}^N L_j$, and $i = \sum_{j=1}^N i_j$. For example, in two dimensions this gives

$$T_i(\underline{x}_k) = H_{i_1}^{L_1-1}(x_k^1)\, H_{i_2}^{L_2-1}(x_k^2) \quad ,$$

for $i_1 = 0, \dots, L_1 - 1$, $i_2 = 0, \dots, L_2 - 1$, $L_x = L_1 L_2$, and $i = i_1 + i_2$.

4 Filtering

In the additive noise case, the posterior conditional density $p_e(\underline{x}_k)$ of the state given measurements up to time k is recursively calculated according to Bayes' law as

$$p_e(\underline{x}_k) = c_k^e\, p_p(\underline{x}_k)\, p_v\Big(\underline{\hat{y}}_k - \underline{h}_k(\underline{x}_k)\Big) \quad ,$$

where c_k^e is a normalizing constant.

Now the surprising result will be derived, that the Bayes measurement update is obtained *exactly* by a standard Kalman filter operating in a higher dimensional space S_x^* with state dimension L_x, provided the noise density $p_v(\underline{v}_k)$ is given as a pseudo Gaussian

$$p(\underline{v}_k^*) = c_k^v \exp\left\{-\frac{1}{2}(\underline{v}_k^* - \underline{\hat{v}}_k^*)^T (\mathbf{C}_k^{v,*})^{-1} (\underline{v}_k^* - \underline{\hat{v}}_k^*)\right\}$$

in a space S_v^* with dimension L_v. For that purpose, the nonlinear measurement equation is transformed according to

$$\underline{T}_v(\underline{\hat{y}}_k - \underline{v}_k) = \underline{T}_v(\underline{h}_k(\underline{x})) \quad . \tag{1}$$

The left hand side is then converted into an affine function of $\underline{v}_k^*$

$$\underline{T}_v(\underline{\hat{y}}_k - \underline{v}_k) = -\mathbf{G}_k^* \underline{v}_k^* + \underline{\hat{y}}_k^* \quad ,$$

with $\underline{v}_k^* = \underline{T}_v(\underline{v}_k)$, where the term $\underline{\hat{y}}_k^*$ does not depend on elements of $\underline{v}_k^*$. Of course, $\mathbf{G}_k^*$ and $\underline{\hat{y}}_k^*$ are polynomial functions of the measurements $\underline{\hat{y}}_k$. The right hand side of (1) is expanded into a linear function of $\underline{x}_k^*$

$$\underline{T}_v(\underline{h}_k(\underline{x})) = \mathbf{H}_k^* \underline{x}_k^*$$

with

$$\underline{x}_k^* = \underline{T}_x(\underline{x}_k)$$

and $L_x \geq \max(N, L_v)$. This expansion is exact for a polynomial measurement nonlinearity $\underline{h}_k(.)$. Finally, we obtain a linear measurement equation

$$\underline{\hat{y}}_k^* = \mathbf{H}_k^* \underline{x}_k^* + \mathbf{G}_k^* \underline{v}_k^*$$

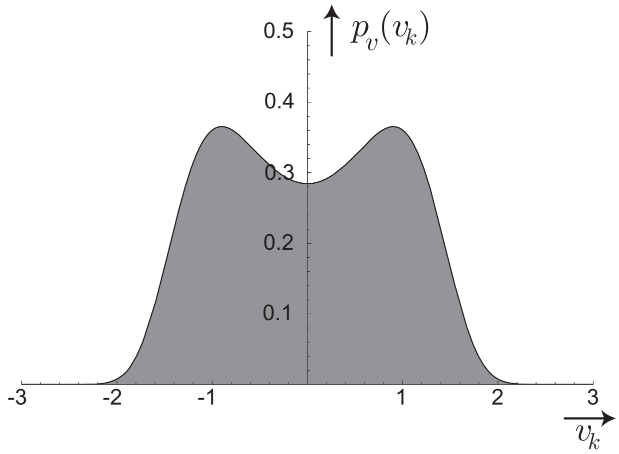

Figure 2: Pseudo Gaussian noise density $p_v(v_k)$.

in S_x^* with $\underline{\hat{y}}_k^* \in \mathrm{I\!R}^{L_v}$, $\underline{x}_k^* \in \mathrm{I\!R}^{L_x}$, $\underline{v}_k^* \in \mathrm{I\!R}^{L_v}$. Given a pseudo Gaussian prior $p_p(\underline{x}_k)$ defined by $\underline{\hat{x}}_k^{p,*}$ and $\mathbf{C}_k^{p,*}$, a Kalman filter according to

$$\begin{aligned}\underline{\hat{x}}_k^{e,*} &= \underline{\hat{x}}_k^{p,*} + \mathbf{C}_k^{p,*}(\mathbf{H}_k^*)^T \left\{\mathbf{G}_k^*\mathbf{C}_k^{v,*}(\mathbf{G}_k^*)^T \right.\\ &\quad \left. +\mathbf{H}_k^*\mathbf{C}_k^{p,*}(\mathbf{H}_k^*)^T\right\}^{-1} (\underline{\hat{y}}_k^* - \mathbf{G}_k^*\underline{\hat{v}}_k^* - \mathbf{H}_k^*\underline{\hat{x}}_k^{p,*})\\ \mathbf{C}_k^{e,*} &= \mathbf{C}_k^{p,*} - \mathbf{C}_k^{p,*}(\mathbf{H}_k^*)^T \left\{\mathbf{G}_k^*\mathbf{C}_k^{v,*}(\mathbf{G}_k^*)^T \right.\\ &\quad \left. +\mathbf{H}_k^*\mathbf{C}_k^{p,*}(\mathbf{H}_k^*)^T\right\}^{-1} \mathbf{H}_k^*\mathbf{C}_k^{p,*}\end{aligned}$$

can now be applied to perform the measurement update with the resulting posterior $p_e(\underline{x}_k)$ defined by $\underline{\hat{x}}_k^{e,*}$ and $\mathbf{C}_k^{e,*}$. However, to ensure symmetry and positive definiteness of the covariance matrix $\mathbf{C}_k^{e,*}$, square–root forms of the Kalman filter [14, 15] are a better choice.

5 Simulation Example

To illustrate the proposed filtering algorithm, the following dynamic system with scalar state x_k is considered, which evolves in discrete time steps according to the noise–free linear system equation

$$x_{k+1} = a\, x_k + u_k$$

with $a = 0.9$, $u_k = -0.25$, and initial state $x_0 = 1.5$. The evolution of the true state x_k for 15 time steps is displayed in Fig. 3 c). Measurements $\hat{y}_k$ of the system output are related to the system state x_k via the nonlinear measurement equation

$$\hat{y}_k = x_k^3 + v_k \quad .$$

The noise distribution $p_v(v_k)$ is given by a three–dimensional pseudo Gaussian $p_v(v_k^*)$, i.e., $L_v = 3$, with mean and covariance matrix

$$\underline{\hat{v}}_k^* = \begin{bmatrix}1\\0\\1\end{bmatrix} \;,\; \mathbf{C}_k^{v,*} = \begin{bmatrix}0.5 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0.5\end{bmatrix} \quad .$$

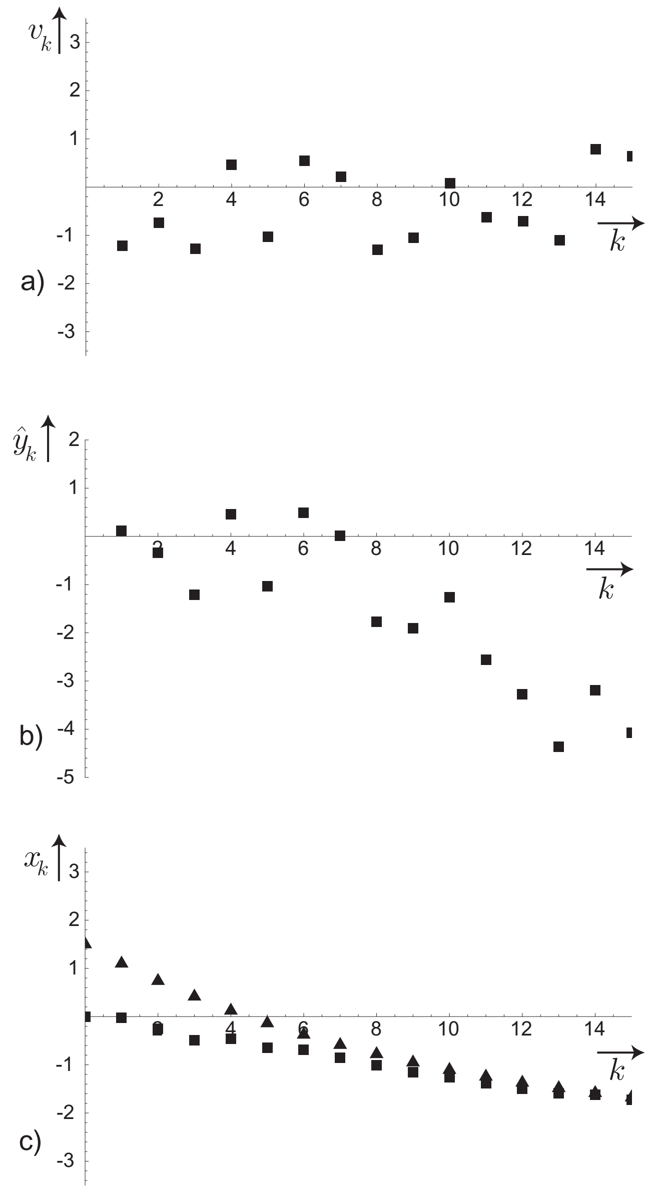

Figure 3: a) Noise sequence v_k. b) Measurement sequence $\hat{y}_k$. c) Evolution of the true state (marked by triangles) and the estimated state (marked by boxes).

$p_v(v_k)$ is visualized in Fig. 2. The sample path of v_k used in the simulation, which is not available to the estimator, is shown in Fig. 3 a). The resulting measurement sequence is given in Fig. 3 b).

The densities of the unknown state resulting from application of the new filter are shown in Fig. 4 for $k = 1, \ldots, 9$, where the density for $k = 0$ represents the a priori knowledge about the state x_0. Point estimates are produced at every time step by calculating the expected values of the posterior densities. The resulting estimates are depicted in Fig. 3 c) for $k = 0, \ldots, 15$.

6 Conclusions

By introducing pseudo Gaussians, a specific type of exponential probability density functions, the Bayesian measurement update step for a polynomial measurement equation and non-Gaussian measurement noise can be reformulated as a standard Kalman filter recursion in a higher dimensional space.

This is similar to the concept of support vector machines [17], which perform nonlinear classification by means of linear hyperplane classifiers in a higher dimensional space nonlinearly related to the input or problem space.

Complicated non–Gaussian posterior densities resulting from the update step can be conveniently handled in the higher dimensional hyperspace. In the hyperspace they are simply described by pseudo Gaussians parametrized by mean vectors and covariance matrices. Furthermore, the measurement nonlinearity is expanded into a linear relation in the hyperspace, which is *exact* for polynomial nonlinearities. Hence, a standard Kalman filter recursion can be used in the hyperspace to perform the desired Bayesian measurement update.

Besides being convenient, the proposed approach always guarantees valid probability densities to result from the update step even in the presence of numerical inaccuracies.

7 Extensions

So far, an exact expansion of the measurement nonlinearity was assumed to exist. In that case a sufficient statistic is provided by the mean vectors and the covariance matrices of the pseudo Gaussians used to represent the posterior densities. However, in many practical applications it is not possible to use an exact expansion of the nonlinearity. In addition, an approximation may be desirable to keep the dimensions of the hyperspace low even when an exact expansion is known. The resulting pseudo Gaussian densities are then approximations of the true posterior densities and are described by a nonsufficient or reduced statistic. However, an approximate expansion can be selected in such a way that a certain distance, e.g. the Kullback–Leibler distance, between the approximate and the exact posterior is minimized. This will be shown in a forthcoming paper.

The proposed technique can also be applied to more complex additive noise descriptions, for example colored noise or noise with partially known statistics. These problems can be solved analogously by applying the appropriate linear filter in the higher dimensional space, e.g. [8, 9, 11].

For the more difficult case of *nonadditive* noise problems, more advanced techniques are currently under development.

References

[1] D. L. Alspach, H. W. Sorenson, “Nonlinear Bayesian Estimation Using Gaussian Sum Approximation”, *IEEE Transactions on Automatic Control*, Vol. 17, No. 4, pp. 439–448, 1972.

[2] B. D. O. Anderson, J. B. Moore, *Optimal Filtering*, Prentice–Hall, 1979.

[3] N. Bergman, L. Ljung, F. Gustafsson, “Terrain Navigation Using Bayesian Statistics”, *IEEE Control Systems Magazine*, Vol. 19, No. 3, pp. 33–40, 1999.

[4] C. Bohn, H. Unbehauen, “The Application of Matrix Differential Calculus for the Derivation of Simplified Expressions in Approximate Non–Linear Filtering Algorithms ”, *Automatica*, Vol. 36, No. 10, pp. 1553–1560, 2000.

[5] R. S. Bucy, K. D. Senne, “Digital Synthesis of Non–linear Filters”, *Automatica*, Vol. 7, pp. 287–298, 1971.

[6] S. Challa, Y. Bar–Shalom, V. Krishnamurthy, “Nonlinear Filtering via Generalized Edgeworth Series and Gauss–Hermite Quadrature”, *IEEE Transactions on Signal Processing*, Vol. 48, No. 6, pp. 1816, 2000.

[7] A. Doucet, S. Godsill, C. Andrieu, “On Sequential Monte Carlo Sampling Methods for Bayesian Filtering”, *Statistics and Computing*, Vol. 10, No. 3, pp. 197–208, 2000.

[8] U. D. Hanebeck, J. Horn, G. Schmidt, “On Combining Statistical and Set–Theoretic Estimation”, *Automatica*, Vol. 35, No. 6, pp. 1101–1109, 1999.

[9] U. D. Hanebeck, J. Horn, “Fusing Information Simultaneously Corrupted by Uncertainties with Known Bounds and Random Noise with Known Distribution”, *Information Fusion*, Vol. 1, No. 1, pp. 55–63, 2000.

[10] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, 1970.

[11] S. J. Julier, J. K. Uhlmann, “A Non–Divergent Estimation Algorithm in the Presence of Unknown Correlations”, *Proceedings of the 1997 American Control Conference (ACC’97)*, pp. 2369-2373, 1997.

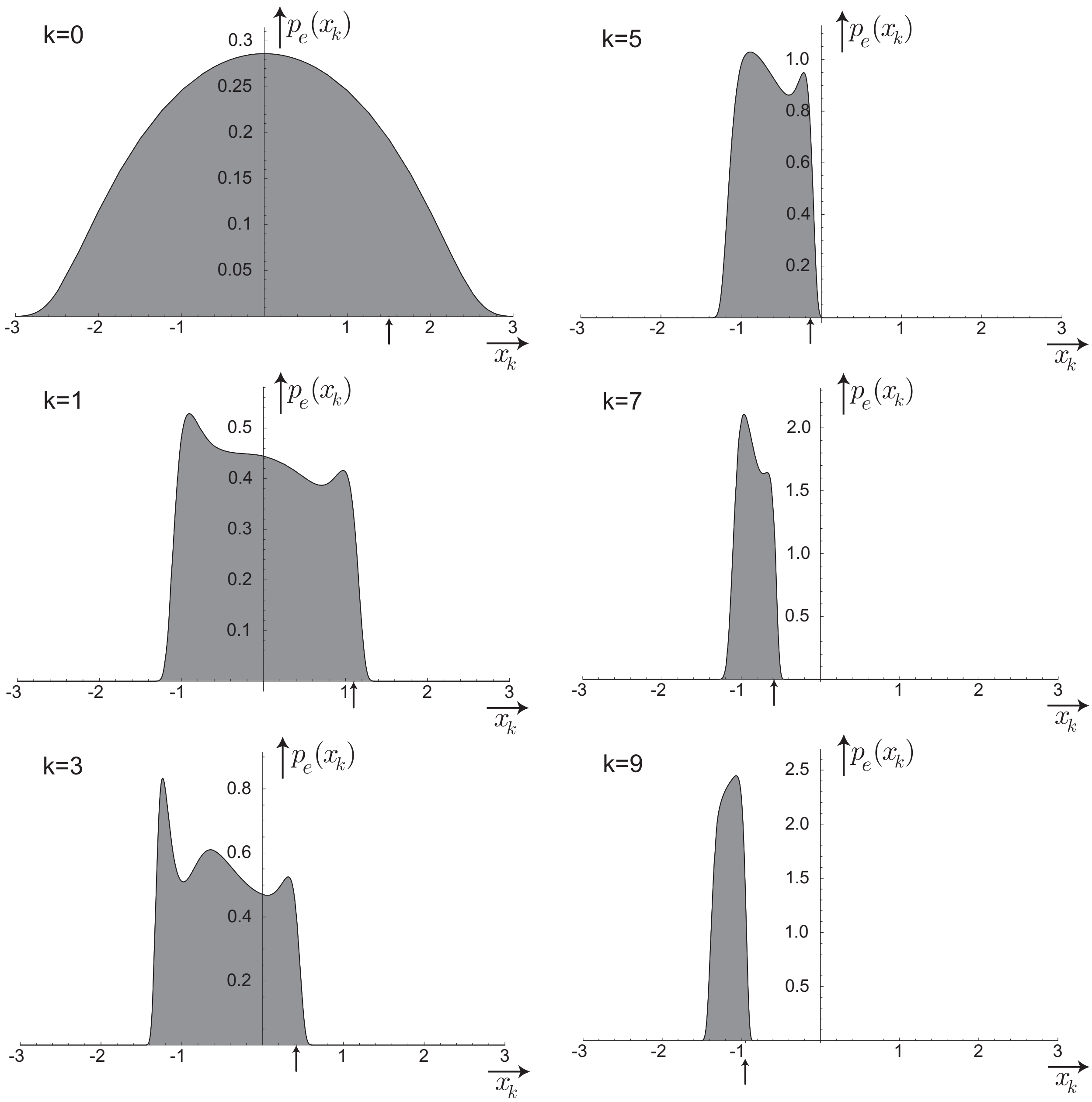

Figure 4: Posterior densities of the state x_k resulting from the application of the new filter for time step $k = 0, 1, 3, 5, 7, 9$. The true state x_k is marked by an arrow.

[12] R. Kulhavý, "Recursive Nonlinear Estimation: Geometry of a Space of Posterior Densities", *Automatica*, Vol. 28, No. 2, pp. 313–323, 1992.

[13] J. S. Liu, R. Chen, "Sequential Monte Carlo Methods for Dynamic Systems", *Journal of American Statistical Association*, Vol. 93, pp. 1032–1043, 1998.

[14] P. Park, T. Kailath, "New Square–Root Algorithms for Kalman Filtering", *IEEE Transactions on Automatic Control*, Vol. 40, No. 5, pp. 895–899, 1995.

[15] A. H. Sayed, T. Kailath "A State–Space Approach to Adaptive RLS Filtering, *IEEE Signal Processing Magazine*, Vol. 11, No. 3, pp. 18–70, 1994.

[16] H. W. Sorenson, A. R. Stubberud, "Non–Linear Filtering by Approximation of the A Posteriori Density", International Journal of Control, Vol. 8, pp. 33–51, 1968.

[17] B. Schölkopf, "Support Vector Machines – A Practical Consequence of Learning Theory", In: M. A. Hearst, B. Schölkopf, S. Dumais, E. Osuna, J. Platt. Trends and Controversies – Support Vector Machines, *IEEE Intelligent Systems*, Vol. 13, No. 4, pp. 18–28, 1998.