Computationally Efficient Kalman Filtering for a Class of Nonlinear Systems

Alexandros C. Charalampidis and George P. Papavassilopoulos

***Abstract*—This paper deals with recursive state estimation for the class of discrete time nonlinear systems whose nonlinearity consists of one or more static nonlinear one-variable functions. This class contains several important subclasses. The special structure is exploited to permit accurate computations without an increase in computational cost. The proposed method is compared with standard Extended Kalman Filter, Unscented Kalman Filter and Gauss–Hermite Kalman Filter in three illustrative examples. The results show that it yields good results with small computational cost.**

***Index Terms*—Covariance matrices, nonlinear filters, numerical methods, recursive state estimation, state space methods, unscented Kalman filtering.**

I. Introduction

In many applications the state of a dynamical system must be estimated. In practice, all measurements are noisy and all processes are affected by some kind of disturbance. For linear equations with normally distributed disturbance and measurement error stochastic processes it is well known that the Kalman Filter [1]–[4] provides an exact solution of the problem.

For nonlinear systems, even if all disturbances are normally distributed, the nonlinearities distort the distribution thus leading to non-normal distributions for the state of the system. Since it is not possible to describe an arbitrary probability density function using a finite number of parameters, the problem is infinite-dimensional.

While for low dimensional systems it is feasible to approximate the exact state distribution by partitioning the state space, even for medium scale systems this approach is infeasible. Other approximate solutions [5]–[10] have been developed, which do not suffer from the curse of dimensionality. In the Extended Kalman Filter (EKF) [2], [4], [5] the system equations are linearized and then Kalman Filter Equations are used. Although widely used, linearization yields a good approximation only when the nonlinearities are mild. Otherwise, the performance may be poor, and even instability phenomena are possible [7].

Sigma Point Kalman Filters, such as the Unscented Kalman Filter (UKF) [6]–[8] or the Gauss-Hermite Kalman Filter (GHKF) [16], provide a promising alternative. At each time step, they approximate the state distribution with a Gaussian distribution, whose mean and covariance is propagated through the system dynamics and updated by the measurements. The propagation is not done by using first-order Taylor expansions of the system equations, but instead by approximating the state distribution using a finite number of sigma-points. UKF has been used in a lot of applications [11]–[13], and the consensus is that it is superior to EKF in most cases. GHKF can yield even better results [16] but its computational cost is high except for low-dimensional systems.

Sigma Point Kalman Filters have been designed for a general nonlinear system. Thus, if it is known that the system under study is of special structure, this information can be exploited in order to make the propagation or update more accurate. In this paper a class of nonlinear discrete time systems with additive noise is considered. Specifically, for the output equation, each output is a nonlinear one-variable function of a linear combination of the system states. For the system dynamics, except for linear terms, only nonlinear one-variable functions of linear combinations of the system states are allowed. Exploiting this structure, the integration in n-dimensions can be substituted by solving a number of linear systems in n-dimensions and integration in one and two dimensions. Avoiding n-dimensional integration allows accurate computations to be made cost-effectively. The proposed technique is applied to three illustrative examples and shown to outperform the standard techniques.

The remainder of the paper is organized as follows. In Section II the problem formulation for general nonlinear system state estimation is presented, as well as some basic results from probability theory and the standard filtering techniques to be compared with the proposed one. In Section III the class of nonlinear systems studied is defined and the proposed approach is presented in detail. In Section IV it is shown how the equations of five important subclasses are formulated in the form of the general class, while Section V presents the simulation examples used to compare the performance of the techniques under consideration. Conclusions are drawn in Section VI.

II. Background

A. Formulation of the Nonlinear Filtering Problem

The general nonlinear filtering problem for dynamic systems with additive noise is to estimate the state of systems of the form

$$x_{k+1} = f(x_k) + w_k, \qquad k = 0, 1, 2, \ldots \tag{1}$$

$$y_k = h(x_k) + v_k, \qquad k = 1, 2, \ldots \tag{2}$$

where x_k is the state of the system and y_k the measured output at time k. w_k is the disturbance, also referred to as process noise, and v_k is the measurement noise. In this paper it is assumed

Manuscript received July 17, 2009; revised January 29, 2010; accepted July 08, 2010. Date of publication September 23, 2010; date of current version March 09, 2011. Recommended by Associate Editor B. Ninness.

The authors are with the School of Electrical and Computer Engineering, National Technical University of Athens, Athens 157 80, Greece (e-mail: alexchar@central.ntua.gr; yorgos@netmode.ntua.gr).

Digital Object Identifier 10.1109/TAC.2010.2078090

that the random variables x_0, w_k, $k = 0, 1, 2, \ldots$ and v_k, $k = 1, 2, \ldots$ are mutually independent and normally distributed with known parameters. Furthermore, w_k and v_k have zero mean. Measurements are available from time $k = 1$ onwards.

Suppose that $p_{X_0}(x_0)$ is the probability density function (pdf) of x_0, $p_V(v_k)$ is the pdf of the measurement noise and $p_W(w_k)$ is the pdf of the process noise. It holds $p_{Y|X}(y_k|x_k) = p_V(y_k - h(x_k))$ and $p_{X_{k+1}|X_k}(x_{k+1}|x_k) = p_W(x_{k+1} - f(x_k))$. The subscripts of probability density functions will be omitted for convenience. Let us define $y_{1:k} = \{y_1, y_2, \ldots, y_k\}$.

Then, using Bayes Rule [14], [15] the following recursive equations hold:

$$p(x_{k+1}|y_{1:k}) = \int p(x_{k+1}|x_k)p(x_k|y_{1:k})dx_k, \quad (3)$$

$$p(x_{k+1}|y_{1:k+1}) = \frac{p(y_{k+1}|x_{k+1})p(x_{k+1}|y_{1:k})}{c_k} \quad (4)$$

where

$$c_k = \int p(y_{k+1}|x_{k+1})p(x_{k+1}|y_{1:k})dx_{k+1}. \quad (5)$$

However the above integrals cannot be evaluated analytically. Numerical integration for a sufficiently dense mesh of x_k at each time step is also impractical, so (3)–(5) are mainly of theoretical interest.

B. Standard Filtering Techniques

In this subsection the standard filtering techniques on which the proposed technique is based are briefly reviewed for easy reference.

1) Kalman Filter: Since both EKF and UKF are based on Kalman Filter (KF), the equations of KF ([1]–[3]) are first presented. Suppose that a dynamical system is described by the following equations:

$$x_{k+1} = A_k x_k + B_k + w_k, \quad (6)$$

$$y_k = C_k x_k + D_k + v_k \quad (7)$$

where w_k and v_k are normally distributed with zero mean, while their covariance matrices are Q and R respectively. Suppose also that it is known that x_k follows the normal distribution with mean $\hat{x}_k$ and covariance P_{x_k}. Then, from (6) and a priori with respect to y_{k+1}, x_{k+1} follows the normal distribution with mean $\hat{x}^-_{k+1}$ and covariance $P^-_{x_{k+1}}$ given by

$$\hat{x}^-_{k+1} = A_k \hat{x}_k + B_k, \quad (8)$$

$$P^-_{x_{k+1}} = A_k P_{x_k} A_k^T + Q. \quad (9)$$

From (7), the predicted value of y_{k+1} is then

$$\hat{y}^-_{k+1} = C_{k+1}\hat{x}^-_{k+1} + D_k. \quad (10)$$

while its covariance is

$$P^-_{y_{k+1}} = C_{k+1} P^-_{x_{k+1}} C^T_{k+1} + R \quad (11)$$

and the cross covariance of x_{k+1} and y_{k+1} is equal to

$$P^-_{x_{k+1}y_{k+1}} = P^-_{x_{k+1}} C^T_{k+1}. \quad (12)$$

The value of y_{k+1} can be then used to refine the distribution. The Kalman gain is given by

$$K_{k+1} = P^-_{x_{k+1}y_{k+1}} P^{-1}_{y_{k+1}}. \quad (13)$$

This Kalman gain is used to calculate the a posteriori mean and covariance of x_{k+1}, according to

$$\hat{x}_{k+1} = \hat{x}^-_{k+1} + K_{k+1}\left(y_{k+1} - \hat{y}^-_{k+1}\right) \quad (14)$$

$$P_{x_{k+1}} = P^-_{x_{k+1}} - K_{k+1} P^-_{y_{k+1}} K^T_{k+1}. \quad (15)$$

Thus the prediction step is accomplished using (8)–(9), while the correction step is accomplished using (14)–(15) with the definitions of (10)–(13). If it is known that x_0 is normally distributed with mean $\hat{x}_0$ and covariance P_{x_0}, (8)–(15) can be applied recursively to yield the statistics of x_n when $y_{1:n}$ is known.

2) Nonlinear Kalman-Based Filters: Although the state distributions for nonlinear systems are not Gaussian, the filters considered in this paper propagate only the state mean and covariance and approximate the conditional distributions $p(x_k|y_{1:k})$ and $p(x_{k+1}|y_{1:k})$ by Gaussian distributions. The correction is made using the equations of Kalman Filter, but each filtering technique must provide a method to approximate the expected values and covariance matrices (which are also defined as expected values) involved. The expected values are of the form $\mathbb{E}[g(X)]$ where g is a nonlinear function and X is a normally distributed random variable (see, for example, [16]).

Extended Kalman Filter: In EKF the system equations (1)–(2) are linearized using first-order Taylor expansions. More specifically, for the prediction step, if $\hat{x}_k$ is the estimated mean of x_k after the correction step of time k then (1) is linearized around $\hat{x}_k$. For the correction step, (2) is linearized around $\hat{x}^-_{k+1}$, where $\hat{x}^-_{k+1}$ is the estimated mean of x_{k+1} after the prediction step of time $k+1$. Obviously, it is assumed that the Jacobian matrices of f and h are available.

Sigma Point Kalman Filters: In this category of filters the integrals associated with expectations are approximated by a finite weighted sum over a suitable set of points, called sigma points (σ-points). The approximation made is $\mathbb{E}[g(X)] \approx \sum_{i=1}^{l} w_i g(X_i)$. This approximation can be based on the intuition that the set $\{X_i, i = 1, \ldots, l\}$ with weights $\{w_i, i = 1, \ldots, l\}$ approximates the normal distribution of X [6], or it can be considered as a numerical integration scheme [16] since

$$\mathbb{E}[g(X)] = \int g(x) \frac{1}{(2\pi)^{n/2}|P|^{1/2}} e^{-(1/2)(x-\mu)^T P^{-1}(x-\mu)} dx \quad (16)$$

in the case X has mean μ and covariance P. See also [17] for a different perspective of the same approximation. Regardless of the physical insight behind the approximation, the choice of the σ-points and their weights affects the accuracy of the approximation.

One solution to the problem is provided by the unscented transformation [6], [7]. Let n be the dimension of the underlying space. Then $2n+1$ σ-points are calculated. In the basic form of the algorithm [6] these are: $X_0 = \mu$, $X_i = \mu + \sqrt{n+\kappa} S_i, i = 1, \ldots, n$ and $X_i = \mu - \sqrt{n+\kappa} S_i, i = n+1, \ldots, 2n$, where S_i is the i-th column of the matrix $S = \sqrt{P}$ (S is obtained using Cholesky decomposition and is such that $P = SS^T$). The

weights assigned to these points are $W^{(0)} = \kappa/(n+\kappa), W^{(i)} = 1/\{2(n+\kappa)\}, i = 1, \ldots, 2n$. For more details and criteria of parameter choice consult [7]. The use of the unscented transform leads to the UKF. In this paper, UKF will be applied with the choice $\kappa = 3 - n$.

Another solution is provided by the Gauss-Hermite quadrature [16], [18]. It uses the Hermite polynomials to determine the σ-points and their weights. For the one-dimensional case the approximation is made as follows [16]. Let $m \in \mathbb{N}^*$, J be the tridiagonal symmetric $m \times m$ matrix with zero diagonals and $J_{r,r+1} = \sqrt{r}, r = 1, \ldots, m-1$, x_i the i-th eigenvalue of J and w_i the square of the first component of the i-th eigenvector of J. Then the approximation for the standard normal distribution is chosen to be

$$\mathbb{E}\left[g(X)\right] \approx \sum_{i=1}^{m} w_i g(\mu + \sqrt{P} x_i) \tag{17}$$

and is exact for polynomials of degree up to $2m - 1$. In the n-dimensional case m^n points are used and the rule is

$$\mathbb{E}\left[g(X)\right] \approx \sum_{i_1=1}^{m} \ldots \sum_{i_n=1}^{m} w_{i_1} \cdots w_{i_n} g(\mu + x_{i_1} S_1 + \cdots + x_{i_n} S_n) \tag{18}$$

where S_i is the i-th column of $S = \sqrt{P}$ as previously. The rule is exact for functions of the form $x_1^{i_1} \ldots x_n^{i_n}, 0 \leq i_k \leq m$.

It must be noted that if the system dynamics (correspondingly, the output equation) is linear, then for the prediction (correspondingly, for the correction) step the KF equations can be applied directly.

III. Analysis for the Class Under Study

A. Motivation and Definition of the Class

GHKF yields better results than UKF, but this happens at a computational cost which is exponential with respect to the state space dimension, since m^n points are used. In this paper it is shown that for the specific class defined below, it is possible to compute the expected values and covariance matrices without the need of computing n-dimensional integrals and thus without using an exponential number of σ-points in order to get accurate results.

The class considered consists of systems with dynamics of the form

$$x_{k+1} = Ax_k + \sum_{i=1}^{N_c} B_i g_i(D_i^T x_k) + w_k \tag{19}$$

and output equation of the form

$$y_{k,i} = h_i(C_i^T x_k) + v_{k,i}, \qquad i = 1, \ldots, N_o \tag{20}$$

where g_i and h_i are nonlinear one-variable functions, $N_c, N_o \in \mathbb{N}$, while C_i and D_i are column vectors in $\mathbb{R}^n$.

In the next subsection the problem of recursive estimation for systems of the form (19)–(20) is analyzed and the proposed technique is presented. In Section III-C it is shown that linear transformations of the state space do not affect the filtering results.

B. Proposed Filtering Algorithm

The following proposition is the main tool used to reduce the dimensionality of the integration.

Proposition 1: Suppose that x is a normally distributed random variable with values in $\mathbb{R}^n$, mean M and covariance $P > 0$, C is a non-zero column vector in $\mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}$ is a Borel-measurable function such that $g(C^T x)$ has finite variance. Suppose also that $\{v_i, i = 1, \ldots, n-1\}$ are $n-1$ linearly independent vectors such that $v_i^T PC = 0$. Let $S = \mathbb{E}[xg(C^T x)]$. Then S is the unique solution of the linear system

$$\begin{bmatrix} C^T \\ v_1^T \\ \vdots \\ v_{n-1}^T \end{bmatrix} S = \begin{bmatrix} \mathbb{E}[C^T x g(C^T x)] \\ v_1^T M \mathbb{E}[g(C^T x)] \\ \vdots \\ v_{n-1}^T M \mathbb{E}[g(C^T x)] \end{bmatrix}. \tag{21}$$

Proof: First of all, the vectors $\{v_1, \ldots, v_{n-1}, C\}$ are linearly independent. This can be proved as follows. Suppose that

$$\lambda_1 v_1 + \cdots + \lambda_{n-1} v_{n-1} + \lambda_n C = 0. \tag{22}$$

Transposition and post-multiplying of the last equation by PC yields

$$\begin{aligned} \lambda_1 v_1^T PC + \cdots + \lambda_{n-1} v_{n-1}^T PC + \lambda_n C^T PC = 0 \Rightarrow \\ \lambda_n C^T PC = 0 \Rightarrow \lambda_n = 0. \end{aligned} \tag{23}$$

But then

$$\lambda_1 v_1 + \cdots + \lambda_{n-1} v_{n-1} = 0 \tag{24}$$

and since $\{v_i, i = 1, \ldots, n-1\}$ are linearly independent, it follows that $\lambda_1 = \cdots = \lambda_{n-1} = 0$. This completes the argument. Therefore the coefficient matrix of the linear system is non-singular thus it has a unique solution.

To prove that this solution is S, the fact that $C^T x$ and $v_i^T x$ are two uncorrelated, normally distributed and thus independent random variables is used. Indeed, $\mathrm{Cov}\left(C^T x, v_i^T x\right) = C^T P v_i = 0$. Thus $v_i^T S = \mathbb{E}[v_i^T x g(C^T x)] = \mathbb{E}[v_i^T x]\mathbb{E}[g(C^T x)] = v_i^T M \mathbb{E}[g(C^T x)]$. The equation $C^T S = \mathbb{E}[C^T x g(C^T x)]$ follows from the definition of S. ∎

Remark 1: Since P is positive definite and C is non-zero, always there exist such $\{v_i, i = 1, \ldots, n-1\}$.

Remark 2: Computing S directly from its definition, $S = \mathbb{E}[xg(C^T x)]$, is a problem of integration in n dimensions. With the above proposition, it has been reduced to the n-dimensional linear system (21) and two one-dimensional integration problems, namely the approximation of $\mathbb{E}[x_C g(x_C)]$ and $\mathbb{E}[g(x_C)]$, where $x_C = C^T x$ is a real valued random variable.

In this paper, both for the prediction and the correction step, the numerical approximation of the integrals is made using Gauss-Hermite quadrature, namely (17) and (18), although the proposed algorithm can be combined with any quadrature technique. As it will be made clear in the rest of this section, only up to two-dimensional quadrature is needed, because the proposed algorithm uses Proposition 1 to reduce the dimensionality of the integration. Thus the approximation can be accurate yet not computationally intensive.

The reader must notice the subtle point that Proposition 1 holds for normal random variables, while the true state distribution is not normal. However, even without Proposition 1, for example in the standard GHKF, the expectations are computed for a Gaussian distribution. Thus, if the approximation is accurate, the use of the proposition does not affect the results, while reduces the computational time.

For both steps, all calculations are first analyzed in detail, and then the algorithm is presented in concise form.

1) Prediction Step: Assume that x_k has mean $\hat{x}_k$ and covariance P_{x_k}. Then from (19) it holds

$$\mathbb{E}[x_{k+1}] = A\hat{x}_k + \sum_{i=1}^{N_c} B_i \mathbb{E}[g_i(D_i^T x_k)]. \tag{25}$$

Since, for each $i = 1 \dots N_c$, $D_i^T x_k$ is a real valued normally distributed random variable with mean $D_i^T \hat{x}_k$ and variance $D_i^T P_{x_k} D_i$, $\mathbb{E}[g_i(D_i^T x_k)]$ is approximated using (17) with $g = g_i$, $\mu = D_i^T \hat{x}_k$ and $P = D_i^T P_{x_k} D_i$. Let $\bar{g}_i$ be the resulting value. Then the prediction for the mean is as follows:

$$\hat{x}_{k+1}^{-} = A\hat{x}_k + \sum_{i=1}^{N_c} B_i \bar{g}_i. \tag{26}$$

Remark 3: All expectations in this paragraph are conditioned upon $\hat{x}_k$ and P_{x_k}, but the conditioning is omitted in order to simplify the notation. For example, $\mathbb{E}[x_{k+1}]$ is written instead of $\mathbb{E}[x_{k+1}|\hat{x}_k, P_{x_k}]$ and $\mathrm{V}\,[x_{k+1}]$ instead of $\mathrm{V}\,[x_{k+1}|\hat{x}_k, P_{x_k}]$.

The prediction step includes also calculation of the a priori covariance of x_{k+1}. It holds

$$\begin{aligned} P_{x_{k+1}}^{-} &= \mathrm{V}\,[x_{k+1}] = \mathrm{V}\left[Ax_k + \sum_{i=1}^{N_c} B_i g_i(D_i^T x_k) + w_k\right] = \\ &= \mathrm{V}\,[Ax_k] + \mathrm{V}\left[\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)\right] + \\ &\quad + \mathrm{Cov}\left(Ax_k, \sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)\right) + \\ &\quad + \mathrm{Cov}\left(\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k), Ax_k\right) + \mathrm{V}\,[w_k]. \end{aligned} \tag{27}$$

It is known that $\mathrm{V}\,[w_k] = Q$ and $\mathrm{V}\,[Ax_k] = AP_{x_k}A^T$.

Let us consider the second term of (27)

$$\begin{aligned} &\mathrm{V}\left[\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)\right] = \\ &= \mathbb{E}\left[\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)\left(\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)\right)^T\right] - \\ &\quad - \mathbb{E}\left[\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)\right]\mathbb{E}\left[\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)\right]^T = \\ &= \mathbb{E}\left[\sum_{i=1}^{N_c}\sum_{j=1}^{N_c} B_i g_i(D_i^T x_k) g_j(D_j^T x_k) B_j^T\right] - \\ &\quad - \sum_{i=1}^{N_c}\sum_{j=1}^{N_c} B_i \mathbb{E}\left[g_i(D_i^T x_k)\right]\mathbb{E}\left[g_j(D_j^T x_k)\right] B_j^T = \\ &= \sum_{i=1}^{N_c}\sum_{j=1}^{N_c} B_i \mathbb{E}\left[g_i(D_i^T x_k) g_j(D_j^T x_k)\right] B_j^T - \\ &\quad - \sum_{i=1}^{N_c}\sum_{j=1}^{N_c} B_i \mathbb{E}\left[g_i(D_i^T x_k)\right]\mathbb{E}\left[g_j(D_j^T x_k)\right] B_j^T. \end{aligned} \tag{28}$$

For $i = j$, $\mathbb{E}\left[g_i(D_i^T x_k) g_j(D_j^T x_k)\right]$ is approximated using (17) with $g = g_i^2$, $\mu = D_i^T \hat{x}_k$ and $P = D_i^T P_{x_k} D_i$. For $i \neq j$, $\mathbb{E}\left[g_i(D_i^T x_k) g_j(D_j^T x_k)\right]$ is approximated using (18) with $g(x_1, x_2) = g_i(x_1) g_j(x_2)$, $\mu = (D_i^T \hat{x}_k, D_j^T \hat{x}_k)^T$ and P equal to

$$\begin{bmatrix} D_i^T P_{x_k} D_i & D_i^T P_{x_k} D_j \\ D_j^T P_{x_k} D_i & D_j^T P_{x_k} D_j \end{bmatrix}. \tag{29}$$

In both cases, let us denote by $\bar{g}_{ij}$ the approximate value of $\mathbb{E}\left[g_i(D_i^T x_k) g_j(D_j^T x_k)\right]$. Then the approximation used is

$$\mathrm{V}\left[\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)\right] \approx \sum_{i=1}^{N_c}\sum_{j=1}^{N_c} B_i \bar{g}_{ij} B_j^T - \sum_{i=1}^{N_c}\sum_{j=1}^{N_c} B_i \bar{g}_i \bar{g}_j B_j^T. \tag{30}$$

The fourth term of (27) is the transpose of the third term, so it suffices to approximate the third term only. It holds

$$\begin{aligned} &\mathrm{Cov}\left(Ax_k, \sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)\right) = \\ &= \mathbb{E}\left[Ax_k\left(\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)\right)^T\right] \\ &\quad - \mathbb{E}\,[Ax_k]\mathbb{E}\left[\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)\right]^T \\ &= \sum_{i=1}^{N_c} \mathbb{E}\left[Ax_k g_i(D_i^T x_k) B_i^T\right] - A\hat{x}_k\left(\sum_{i=1}^{N_c} B_i \mathbb{E}[g_i(D_i^T x_k)]\right)^T \\ &= A\sum_{i=1}^{N_c} \mathbb{E}\left[x_k g_i(D_i^T x_k)\right] B_i^T - A\hat{x}_k \sum_{i=1}^{N_c} \mathbb{E}\left[g_i(D_i^T x_k)\right] B_i^T. \end{aligned} \tag{31}$$

function PREDICTIONSTEP($\hat{x}_k, P_{x_k}$)
  for $i \leftarrow 1 : N_c$ **do**
    Calculate $\bar{g}_i$ using (17)
  end for
  for $i \leftarrow 1 : N_c$ **do**
    for $j \leftarrow 1 : N_c$ **do**
      if $i = j$ Calculate $\bar{g}_{ij}$ using (17)
      if $i \neq j$ Calculate $\bar{g}_{ij}$ using (18)
    end for
  end for
  for $i \leftarrow 1 : N_c$ **do**
    Calculate $\overline{xg}_i$ solving (21)
  end for
  $V \leftarrow \sum_{i=1}^{N_c}\sum_{j=1}^{N_c} B_i\bar{g}_{ij}B_j^T - \sum_{i=1}^{N_c}\sum_{j=1}^{N_c} B_i\bar{g}_i\bar{g}_jB_j^T$
  $COV \leftarrow A\sum_{i=1}^{N_c}\overline{xg}_iB_i^T - A\hat{x}_k\sum_{i=1}^{N_c}\bar{g}_iB_i^T$
  $\hat{x}^-_{k+1} \leftarrow A\hat{x}_k + \sum_{i=1}^{N_c} B_i\bar{g}_i$ ▷ Prediction Mean
  $P^-_{x_{k+1}} \leftarrow AP_{x_k}A^T + V + COV + COV^T + Q$
  ▷ Prediction Covariance
end function

Fig. 1. Prediction step algorithm.

function CORRECTIONSTEP($\hat{x}^-_{k+1}, P^-_{x_{k+1}}, y_{k+1}$)
  for $i \leftarrow 1 : N_o$ **do**
    Calculate $\bar{h}_i$ using (17)
  end for
  for $i \leftarrow 1 : N_o$ **do**
    for $j \leftarrow 1 : N_o$ **do**
      if $i = j$ Calculate $\bar{h}_{ij}$ using (17)
      if $i \neq j$ Calculate $\bar{h}_{ij}$ using (18)
    end for
  end for
  for $i \leftarrow 1 : N_o$ **do**
    Calculate $\overline{xh}_i$ solving (21)
  end for
  $\hat{y}^- \leftarrow (\bar{h}_i)$
  $P^-_y \leftarrow (\bar{h}_{ij} - \bar{h}_i\bar{h}_j) + R$
  $P^-_{x,y} \leftarrow (\overline{xh}_j - \hat{x}^-_{k+1}\bar{h}_j)$
  $K \leftarrow P^-_{x,y}P_y^{-1}$
  $\hat{x}_{k+1} \leftarrow \hat{x}^-_{k+1} + K(y_{k+1} - \hat{y}^-)$ ▷ Updated Mean
  $P_{x_{k+1}} \leftarrow P^-_{x_{k+1}} - KP^-_yK^T$ ▷ Updated Covariance
end function

Fig. 2. Correction step algorithm.

$\mathbb{E}\left[x_kg_i(D_i^Tx_k)\right]$ is approximated using (21). The expected values in the right-hand side of (21) are approximated using (17); the second approximated value is $\bar{g}_i$, while the first is calculated similarly but with $g(x) = xg_i(x)$. The vectors $\{v_1, \ldots, v_{n-1}\}$ depend on i as well as on the time step k. Let us denote by $\overline{xg}_i$ the solution of (21). Then the approximation obtained for the third term of (27) is

$$\mathrm{Cov}\left(Ax_k, \sum_{i=1}^{N_c} B_ig_i(D_i^Tx_k)\right) \approx A\sum_{i=1}^{N_c}\overline{xg}_iB_i^T - A\hat{x}_k\sum_{i=1}^{N_c}\bar{g}_iB_i^T. \tag{32}$$

Using (30) and (32) an approximation of $P^-_{x_{k+1}}$ can be calculated to conclude the prediction step. The algorithm described above is summarized in Fig. 1.

2) Correction Step: For the correction step $\hat{y}^-_{k+1}$, $P^-_{y_{k+1}}$ and $P^-_{x_{k+1},y_{k+1}}$ must be computed. This task is accomplished with the same techniques used in the prediction step. Let us consider $\hat{y}^-_{k+1}$. From (20) it follows that for $i = 1 \ldots N_i$

$$\hat{y}^-_{k+1,i} = \mathbb{E}\left[h_i(C_i^Tx_{k+1})\right]. \tag{33}$$

Remark 4: All expectations in this paragraph are conditioned upon $\hat{x}^-_{k+1}$ and $P^-_{x_{k+1}}$, but the conditioning is omitted in order to simplify the notation. For example, $\mathbb{E}\left[h_i(C_i^Tx_{k+1})\right]$ is written instead of $\mathbb{E}\left[h_i(C_i^Tx_{k+1})|\hat{x}^-_{k+1}, P^-_{x_{k+1}}\right]$.

Thus $\mathbb{E}\left[h_i(C_i^Tx_{k+1})\right]$ can be approximated using (17) with $g = h_i$, $\mu = C_i^T\hat{x}^-_{k+1}$ and $P = C_i^TP^-_{x_{k+1}}C_i$. Let $\bar{h}_i$ be the resulting value.

With respect to $P_{y_{k+1}}$, (20) yields

$$\begin{aligned}
&(P_{y_{k+1}})^-_{i,j} = \\
&= \mathbb{E}[(h_i(C_i^Tx_{k+1}) + v_{k+1,i})(h_j(C_j^Tx_{k+1}) + v_{k+1,j})] - \\
&\quad - \mathbb{E}[h_i(C_i^Tx_{k+1}) + v_{k+1,i}]\mathbb{E}[h_j(C_j^Tx_{k+1}) + v_{k+1,j}] = \\
&= \mathbb{E}[h_i(C_i^Tx_{k+1})h_j(C_j^Tx_{k+1})] + R_{i,j} - \\
&\quad - \mathbb{E}[h_i(C_i^Tx_{k+1})]\mathbb{E}[h_j(C_j^Tx_{k+1})].
\end{aligned} \tag{34}$$

$\mathbb{E}[h_i(C_i^Tx_{k+1})h_j(C_j^Tx_{k+1})]$ can be approximated in the same way that $\mathbb{E}[g_i(D_i^Tx_k)g_j(D_j^Tx_k)]$ is approximated in the prediction step. Let us denote by $\bar{h}_{ij}$ the approximate value. Then the approximation made for $(P_{y_{k+1}})^-_{i,j}$ is

$$(P_{y_{k+1}})^-_{i,j} \approx \bar{h}_{ij} - \bar{h}_i\bar{h}_j + R_{ij}. \tag{35}$$

Finally, the j-th column of $P^-_{x_{k+1},y_{k+1}}$, $(P_{x_{k+1},y_{k+1}})^-_j$ is equal to

$$\begin{aligned}
&(P_{x_{k+1},y_{k+1}})^-_j \\
&\quad = \mathbb{E}[x_{k+1}h_j(C_j^Tx_{k+1})] - \mathbb{E}[x_{k+1}]\mathbb{E}[h_j(C_j^Tx_{k+1})].
\end{aligned} \tag{36}$$

$\mathbb{E}[x_{k+1}h_j(C_j^Tx_{k+1})]$ can be calculated in the same way as $\mathbb{E}[x_kg_i(D_i^Tx_k)]$ in the prediction step. If the approximated value is denoted by $\overline{xh}_j$, the approximation for $(P_{x_{k+1},y_{k+1}})^-_j$ is

$$(P_{x_{k+1},y_{k+1}})^-_j \approx \overline{xh}_j - \hat{x}^-_{k+1}\bar{h}_j. \tag{37}$$

After the calculation of $\hat{y}^-_{k+1}$, $P^-_{y_{k+1}}$ and $P^-_{x_{k+1},y_{k+1}}$, the correction step can be completed using (13)–(15) as in KF and its nonlinear variations. The correction step algorithm is summarized in Fig. 2.

C. Linear Transformation Invariance

Coordinate transformations are frequently used both in linear [19] and nonlinear [20] system theory. The Kalman Filter exhibits invariance under linear transformations, namely if the system state is transformed using $x' = Tx$ then it holds $\hat{x}'_k = T\hat{x}_k$ for all k. In this subsection it is shown that for the class under study the modified technique proposed also exhibits such an invariance. This fact has the important implication that the choice of the state space realization of a transfer function is indifferent when only the system input and output are of interest. Section V further clarifies this concept.

Let therefore T be a non-singular $n \times n$ matrix. Straightforward calculations show that under the coordinates defined by $x' = Tx$ the system is described by

$$x'_{k+1} = A'x'_k + \sum_{i=1}^{N_c} B'_i g_i(D_i'^T x'_k) + w'_k, \tag{38}$$

$$y_{k,i} = h_i(C_i'^T x'_k) + v_{k,i}, \quad i = 1, \ldots, N_o \tag{39}$$

where

$$A' = TAT^{-1}, \tag{40}$$

$$B'_i = TB_i, \tag{41}$$

$$D'_i = (T^T)^{-1} D_i, \tag{42}$$

$$C'_i = (T^T)^{-1} C_i \tag{43}$$

while

$$w'_k \sim \mathcal{N}(0, TQT^T), \tag{44}$$

$$x'_0 \sim \mathcal{N}(T\hat{x}_0, TP_{x_0}T^T). \tag{45}$$

The result of this subsection is stated in the following proposition.

Proposition 2: Consider the system described by (19)–(20) and the coordinates transformation $x' = Tx$. If $\hat{x}_k$ and P_{x_k} are the estimates produced by the proposed filter for the initial coordinates while $\hat{x}'_k$ and $P_{x'_k}$ are those for the transformed coordinates, then for every $k \in \mathbb{N}$ it holds

$$\hat{x}'_k = T\hat{x}_k, P_{x'_k} = TP_{x_k}T^T. \tag{46}$$

Proof: For $k = 0$ (46) is an elementary result of probability theory, stated also in (45). Suppose that (46) holds for the time step k. It will be proved that then it is also valid for $k+1$.

For each $i = 1 \ldots N_c$, (42) implies that $D_i'^T \hat{x}'_k = D_i^T T^{-1} T\hat{x}_k = D_i^T \hat{x}_k$ and $D_i'^T P_{x'_k} D'_i = D_i^T T^{-1} T P_{x_k} T^T (T^T)^{-1} D_i = D_i^T P_{x_k} D_i$. Thus the σ-points for the nonlinear function g_i are not affected by the transformation. This implies that $\overline{g}'_i = \overline{g}_i$ and $\overline{g}'_{ij} = \overline{g}_{ij}$.

The right-hand side of (21) is thus unchanged. The coefficient matrix is post-multiplied by T^{-1}, because this holds for the first row by (42), and $v'_i = v_i T^{-1}$ is a valid choice for the rest of the rows. Therefore $\overline{xg}'_i = T\overline{xg}_i$ is the solution of the corresponding system (21).

Using the equalities $\overline{g}'_i = \overline{g}_i$, $\overline{g}'_{ij} = \overline{g}_{ij}$ and $\overline{xg}'_i = T\overline{xg}_i$ as well as (38), (40)–(41) and (44), it can be verified that $\hat{x}'^{-}_{k+1} = T\hat{x}^{-}_{k+1}$ and $P'^{-}_{x_{k+1}} = TP^{-}_{x_{k+1}}T^T$.

For the update step, similarly with the prediction step, it holds $C_i'^T \hat{x}'^{-}_{k+1} = C_i^T T^{-1} T\hat{x}^{-}_{k+1} = C_i^T \hat{x}^{-}_{k+1}$ and $C_i'^T P'^{-}_{x_{k+1}} C'_i = C_i^T T^{-1} T P^{-}_{x_{k+1}} T^T (T^T)^{-1} C_i = C_i^T P^{-}_{x_{k+1}} C_i$, therefore again the transformation does not affect the σ-points for the nonlinear functions. Thus $\overline{h}'_j = \overline{h}_j$, $\overline{h}'_{ij} = \overline{h}_{ij}$ and $\overline{xh}'_j = T\overline{xh}_j$. Using these equalities together with $\hat{x}'^{-}_{k+1} = T\hat{x}^{-}_{k+1}$ and $P'^{-}_{x_{k+1}} = TP^{-}_{x_{k+1}}T^T$ it is easy to verify that the correction step algorithm of Fig. 2 yields $\hat{x}'_{k+1} = T\hat{x}_{k+1}$ and $P'_{x_{k+1}} = TP_{x_{k+1}}T^T$. ■

Remark 5: Using similar arguments it is easy to prove that neither UKF is affected by linear transformations.

IV. Important Subclasses

A. Systems With Linear Dynamics and Nonlinear Output

It is obvious that linear dynamics is of the form (19) with $N_c = 0$. Therefore, such a system belongs to the class under study provided that every output of the system has the form (20). This subclass is a subset of the subclass of paragraph Section IV-D.

B. SISO Linear Systems With Nonlinear Feedback

If a SISO linear system is described by

$$x_{k+1} = Ax_k + bu_k + w_k,\ y_k = cx_k \tag{47}$$

namely there is no measurement error, and an output feedback of the form $u_k = g(y_k)$ is applied, then the system dynamics is of the form (19) with $N_c = 1$, $B_1 = b$, $C_1 = c$ and $g_1 = g$. The output is linear, and thus for the correction step (10)–(15) can be used directly with $R = 0$ in (11).

C. MIMO Linear Systems With Nonlinear Decoupled Feedback

Suppose a MIMO system is described by

$$x_{k+1} = Ax_k + Bu_k + w_k,\ y_k = Cx_k \tag{48}$$

where B is a $n \times m$ matrix, while C is a $m \times n$ matrix. Let also B_i the i-th column of B and C_i the i-th row of C. If a feedback of the form $(u_k)_i = g_i((y_k)_i)$ is applied, then as previously the system dynamics is of the form (19) with $N_c = m$, and for the correction step (10)–(15) can be used directly with $R = 0$ in (11). The reader may note that (48) with feedback $(u_k)_i = g_i((y_k)_i)$ and without disturbance is the system form of the hypothesis of the Popov Criterion (see [20] or [21], and [22] for the discrete time case).

D. Cascades of Linear Systems With Nonlinear Characteristics

Assume that the output of a linear system is connected to the input of a nonlinear one-variable function whose output is connected to the input of another linear system and so on. The output of the total system is the output of the last nonlinear characteristic. If there are n_s linear systems, and n_s nonlinear functions, the mathematical description is as follows.

The linear systems are described by

$$x_{k+1}^{(m)} = A^{(m)}x_k^{(m)} + b_m u_k^{(m)} + w_k^{(m)}, \tag{49}$$

$$y_k^{(m)} = c_m x_k^{(m)}, \quad 1 \le m \le n_s - 1 \tag{50}$$

$$x_{k+1}^{(n_s)} = A^{(n_s)}x_k^{(n_s)} + b_{n_s} u_k^{(n_s)} + w_k^{(n_s)}, \tag{51}$$

$$y_k^{(n_s)} = g^{n_s}(c_{n_s} x_k^{(n_s)}) + v_k^{(n_s)} \tag{52}$$

while their interconnection by

$$u_k^{(m)} = g^{(m-1,m)}(y_k^{(m-1)}), \qquad m = 2, \ldots, n_s. \tag{53}$$

The state of the whole system is

$$x = \begin{bmatrix} x^{(1)^T} & x^{(2)^T} & \ldots & x^{(n_s)^T} \end{bmatrix}^T. \tag{54}$$

Then the whole system is of the form (19) with $N_c = n_s - 1$, $N_o = 1$

$$A = diag(A_1, A_2, \ldots, A_{n_s}) = \begin{bmatrix} A_1 & O & \cdots & O \\ O & A_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & \cdots & \cdots & A_{n_s} \end{bmatrix}, \tag{55}$$

$$B_1 = [O \quad b_2^T \quad \cdots \quad O]^T, \ldots,$$
$$B_{n_s-1} = [O \quad O \quad \cdots \quad b_{n_s}^T]^T, \tag{56}$$
$$g_i = g^{(i,i+1)}, i = 1, \ldots, n_s - 1, \tag{57}$$
$$D_1 = [c_1 \quad O \quad O \quad \cdots \quad O], \ldots,$$
$$D_{n_s-1} = [O \quad \cdots \quad O \quad c_{n_s-1}], \tag{58}$$
$$Q = diag(Q_1, Q_2, \ldots, Q_{n_s}), \tag{59}$$
$$h_1 = g^{(n_s)}, \tag{60}$$
$$C_1 = [O \quad \cdots \quad O \quad c_{n_s}], \tag{61}$$
$$R = diag(O, \ldots, R_{n_s}). \tag{62}$$

E. Arbitrary Networks of Linear Systems Interconnected With Nonlinear Characteristics

The case of a number of linear systems interconnected in a more complex topology than the series connection of the previous paragraph can also be treated using the method presented in this paper.

Suppose that there are n_s linear systems described by (49)–(51), and interconnected through the nonlinear characteristics $g_i, i = 1 \ldots N_c$, where the input of the characteristic g_i is the output y_{r_i}, while its output is driven to the input u_{t_i}. It is possible that two nonlinear characteristics are driven to the input of the same linear system, i.e., $t_i = t_j$ for $i \neq j$, which means that the input of the linear system is the sum of all the corresponding nonlinearities. Thus every interconnection topology is included in the subclass. As in the previous paragraph, the output of the total system is the output of the system with index n_s.

Then the whole system is of the form (19) where (55) and (59)–(62) hold verbatim, while for $i = 1 \ldots N_c$ the vectors B_i and D_i are given by

$$B_i = [O \quad \cdots \quad b_{t_i}^T \quad \cdots \quad O]^T,$$
$$D_i = [O \quad \cdots \quad c_{r_i} \quad \cdots \quad 0]. \tag{63}$$

V. Examples

A. Linear System With Sensor Nonlinearity (A)

Suppose that a linear system whose transfer function is

$$G_{sys}(z) = \frac{0.093258}{(z - 0.9)(z^2 - 1.559z + 0.81)} \tag{64}$$

is driven by Gaussian white noise with zero mean and variance equal to 1, and that the output of the linear system is measured but the sensor suffers from nonlinearity and noise so that if $s(k)$ is the output of the linear system and $y(k)$ is the available measurement at time k, it holds

$$y(k) = s(k)^3 + v(k) \tag{65}$$

where $v(k)$ is Gaussian white noise with zero mean and standard deviation equal to 0.3. The goal is to estimate the output of the linear system. The following matrices provide a minimal state space realization of the system

$$A = \begin{bmatrix} 0.9 & 1 & 0 \\ 0 & 0.7794 & 1 \\ 0 & -0.2025 & 0.7794 \end{bmatrix}, \; b = \begin{bmatrix} 0 \\ 0 \\ 0.25 \end{bmatrix},$$
$$c = [0.3730 \quad 0 \quad 0]. \tag{66}$$

It must be noted that although the realization is not unique, any other equivalent realization would yield the same results, as shown in Section III-C.

TABLE I
RMS Estimation Error for the Example of Section V-A

Estim. Techn.	Mean	Standard Dev.	Worst Case	Time (ms)
EKF	0.9519	0.1743	1.6149	0.09
UKF	0.2976	0.0430	0.5403	0.28
AGHKF(3)	0.2840	0.0433	0.5456	1.11
AGHKF(5)	0.2858	0.0378	0.3941	1.16
GHKF(3)	0.2840	0.0433	0.5456	0.86
GHKF(5)	0.2858	0.0378	0.3941	3.74

EKF, UKF, GHKF and the accelerated GHKF (AGHKF) as presented in this paper have been applied to this problem for comparison purposes. 1000 runs have been made, each for time $k = 1 \ldots 100$. x_0 is supposed to follow the normal distribution with zero mean and covariance $0.01 I_3$. Table I presents the statistics of the RMS estimation error for the output of the linear system, as well as the computation time, recorded in a 64-bit PC running MATLAB 7.2 for Linux. GHKF and AGHKF have been applied with two choices for the parameter m of (17) and (18). The values used are given in the parenthesis of Table I. EKF performs poorly. In this example GHKF and AGHKF have the same performance for equal m. For $m = 3$ GHKF is faster, because AGHKF is subject to time overhead due to the linear algebraic operations described in Section III-B. However, increasing m to 5 reduces significantly the worst-case error, and this reduction comes at almost no cost for AGHKF, while for GHKF the time increase is considerable. It is noted that G_{sys} is selected so that the RMS value of the system output is equal to 1.

B. Linear System With Sensor Nonlinearity (B)

Suppose now that it holds $y(k) = s(k)^3\big(1 + 0.25\cos(20s(k))\big) + v(k)$ instead of the output (65) and all other parameters are the same with the previous example. The results are presented in Table II. EKF again performs poorly. The difference between the various σ-point filters is now greater than in the previous case. This can be attributed to the

TABLE II
RMS ESTIMATION ERROR FOR THE EXAMPLE OF SECTION V-B

Estim. Techn.	Mean	Standard Dev.	Worst Case	Time (ms)
EKF	0.9502	0.1733	1.6416	0.10
UKF	0.3289	0.0841	1.3668	0.28
AGHKF(7)	0.3099	0.1028	3.2055	1.22
AGHKF(11)	0.3010	0.0342	0.4271	1.32
GHKF(4)	0.3208	0.0961	1.7011	1.94
GHKF(7)	0.3108	0.1170	3.6893	10.10

TABLE III
RMS ESTIMATION ERROR FOR THE EXAMPLE OF SECTION V-C

Estim. Techn.	Mean	Standard Dev.	Worst Case	Time (ms)
EKF	0.9458	0.1729	1.6848	0.18
UKF	0.4752	0.0447	0.6197	0.73
AGHKF(4)	0.4652	0.0427	0.5960	1.20
AGHKF(21)	0.4652	0.0427	0.5960	1.57
GHKF(3)	0.4649	0.0460	0.6095	5.79
GHKF(4)	0.4652	0.0427	0.5960	18.02

fact that now the output equation has a more complex form, and thus a finer point set for integration is more significant. AGHKF permits a value of m as high as 11 without significant computational cost.

It may be also noticed that in this case, in which the quadrature yields approximate values for the expected values in (21), the results for GHKF (7) and AGHKF (7) are not identical.

C. Linear System With Sensor Nonlinearity and Dynamics

In this subsection, it is assumed that the sensor except for its nonlinearity, is a dynamic system. This situation is described mathematically as follows. The output of the linear system described by (64) is connected to the input of a static nonlinearity with equation $y = x^3$. At the output of the static nonlinearity Gaussian white noise with variance equal to 0.2 is added, and the sum is in turn connected to the input of a linear system described by state equations of the form (47), with matrices equal to $A = 0.1,\ b = 1,\ c = 0.995$. The corresponding transfer function is

$$G_{sen}(z) = \frac{0.99499}{z - 0.1}. \tag{67}$$

The output of this transfer function is measured with the presence of additive Gaussian white noise with standard deviation 0.3. Then the problem is of the form described in Section IV-D. x_0 is supposed to follow the normal distribution with zero mean and covariance $0.01I_4$. The results are presented in Table III. The difference in computation time between GHKF and AGHKF is much greater than in the two previous cases because the system now is four-dimensional.

VI. CONCLUSION

This paper shows that, in the case of the class considered, a solution to the nonlinear filtering problem in n dimensions can be given by solving a number of linear systems in n dimensions and a number of integral approximation problems in one and two dimensions. This permits more accurate calculations than the direct approach of integral approximation in n dimensions. Three examples show that indeed the proposed technique leads to computational time reduction, which is greater for higher system dimension. The results also validate the known fact that EKF performs poorly in comparison with σ-point filters under the presence of strong nonlinearities such as a cubic nonlinear characteristic.

Further research could focus on finding more general or other system classes for which the special structure can be exploited to design class-specific filters.

REFERENCES

[1] R. E. Kalman, "A new approach to linear filtering and prediction problem," *ASME J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, Mar. 1960.

[2] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[3] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

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

[5] G. L. Smith, S. F. Schmidt, and L. A. McGee, Application of Statistical Filter Theory to the Optimal Estimation of Position and Velocity on Board a Circumlunar Vehicle NASA, Tech. Rep. TR R-135, 1962.

[6] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Trans. Autom. Control*, vol. 45, no. 3, pp. 477–482, Mar. 2000.

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

[8] R. Van der Merwe, "Sigma-Point Kalman Filters for Probability Inference in Dynamic State-Space Models," Ph.D. dissertation, Oregon Health Sci. Univ., Portland, OR, 2004.

[9] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sum," *Automatica*, vol. 7, no. 4, pp. 465–479, Jul. 1971.

[10] H. W. Sorenson and A. R. Stubberud, "Nonlinear filtering by approximation of the a posteriori density," *Int. J. Control*, vol. 8, no. 1, pp. 33–51, 1968.

[11] C. C. Qu and J. Hahn, "Computation of arrival cost for moving horizon estimation via unscented Kalman filtering," *J. Process Control*, vol. 19, no. 2, pp. 358–363, Feb. 2009.

[12] M. St-Pierre and D. Gingras, "Comparison between the unscented Kalman filter and the extended Kalman filter for the position estimation module of an integrated navigation information system," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2004, pp. 831–835.

[13] A. Romanenko and J. A. A. M. Castro, "The unscented filter as an alternative to the EKF for nonlinear state estimation: A simulation case study," *Comp. Chem. Eng.*, vol. 28, no. 3, pp. 347–355, Mar. 2004.

[14] P. Billingsley, *Probability and Measure*, 2nd ed. Hoboken, NJ: Wiley-Interscience, 1986.

[15] R. B. Ash, Basic Probability Theory Dover, Mineola, NY, 2008 [Online]. Available: www.math.uiuc.edu/ r-ash/BPT.html

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

[17] T. Lefebvre, H. Bruyninckx, and J. d. Schutter, "Kalman filters for nonlinear systems: A comparison of performance," *Int. J. Control*, vol. 77, no. 7, pp. 639–653, 2004.

[18] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*. New York: Academic Press, 1975.

[19] C. T. Chen, *Linear System Theory and Design*, 3rd ed. New York: Oxford Univ. Press, 1999.

[20] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

[21] V. M. Popov, "Criterion of quality for non-linear controlled systems," in *Preprints 1st IFAC World Congress*, 1960, pp. 173–176.

[22] G. Szego, "On the absolute stability of sampled-data control systems," *Proc. Nat. Acad. Sci.*, vol. 50, no. 3, pp. 558–560, Sep. 1963.

Alexandros C. Charalampidis was born in Greece, in 1984. He received the Diploma in electrical and computer engineering from the National Technical University of Athens, Greece, in 2007 where he is currently pursuing the Ph.D. degree.

His research interests include systems theory with emphasis on nonlinear stochastic systems as well as control applications.

George P. Papavassilopoulos received the Diploma from the Department of Mechanical and Electrical Engineering, National Technical University of Athens, Greece, in 1975 and the M.S. and Ph.D. degrees from the Department of Electrical Engineering, University of Illinois at Urbana-Champaign, in 1977 and 1979, respectively.

He was a Professor in the Department of EE-Systems, University of Southern California, Los Angeles, from 1979 to 2000 when he joined the Department of Electrical and Computer Engineering, NTUA, as Professor. He has conducted research in dynamic and stochastic game theory, optimization and control theory, parallel algorithms, Markovian learning, BMIs for robust control. He is also interested in applications in biomedical engineering, energy and telecommunication policies.