2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

PDF Tracking Filter Design Using Hybrid Characteristic Functions

Jinglin Zhou, Hong Wang, *Senior Member IEEE*, and Donghua Zhou, *Senior Member IEEE*

***Abstract*— A new tracking filtering algorithm for a class of multivariate dynamic stochastic systems is presented. The system is represented by a set of time-varying discrete systems with non-Gaussian stochastic input and nonlinear output. New concept such as hybrid characteristic functions is introduced to describe the stochastic nature of the dynamic conditional estimation errors, where the key idea is to ensure the distribution of the conditional estimation error to follow a target distribution. For this purpose, the relationships between the hybrid characteristic functions of the multivariate stochastic input and the outputs, and the properties of the hybrid characteristic function are established. A new performance index of the tracking filter is then constructed based on the form of the hybrid characteristic function of the conditional estimation error. An analytical solution, which guarantees the filter gain matrix to be an optimal one, is then obtained.**

***Index Terms*— Dynamic stochastic systems; characteristic functions; optimal tracking control; hybrid random vectors; optimal filtering.**

I. Introduction

To reduce the effect of noises, research into filtering design has been carried out for many years. Many approaches have been developed and widely used successfully in real applications following the development of the Kalman filtering (KF) algorithm [1], [2], [10]. Further examples are the extended Kalman filters (EKF), the H_∞ filters or robust filters. The Extended Kalman filters [8] or the Unscented Kalman Filters (UKF)[21] focus on nonlinear systems or uncertain systems with Gaussian white noise random disturbances. However, due to the presence of the nonlinearity in the system dynamics, the system output and the estimation error can generally be non-Gaussian. On the other hand, the robust estimation includes the minimum variance filtering [19], [20], the H_∞ filtering [22] and the admissible variance constraint filtering [27], etc. However, all the performance indices in the robust filtering design are still the mean and variance of the estimation errors.mprove

An alternative group of filtering approaches are based upon the use of probability density functions (PDFs), where the conditional PDFs of the system state vector are numerically calculated using the system measurements and the estimated state vector can then be formulated using the obtained conditional PDFs [3], [7], [14].

Based upon the above discussions, it can be concluded that most of the filter design techniques have used the mean and variance criterion for optimizing their estimation errors and residuals under the assumption that the estimation errors can be treated as Gaussian random processes. However, this assumption is strict for many real systems and therefore limits the use of the filter to non-Gaussian and nonlinear systems because the mean and variance cannot capture all features of non-Gaussian noise and determine the shape of its PDF. In this paper, we aim to develop a new filtering algorithm by shaping the conditional PDFs of the estimation error signals ([13], [23]).

In the following, except for specially pointing out, matrices are assumed to have appropriate dimensions. For two real vectors v_1 and v_2, the notation $v_1 \preceq v_2$ is used to denote that every element of v_2 is no less than the corresponding one of v_1. $v_1 \prec v_2$ and $v_1 \succeq v_2$ have the similar meanings. $\mathrm{diag}\{\cdot\}$ is used to denote a diagonal matrix. $\mathrm{E}(\cdot)$ and $\mathrm{Var}(\cdot)$ represent mathematical expectation and variance of random variables. $\varphi_Z(\cdots)$ represent the (hybrid) characteristic function of the (hybrid) vector Z. $\overline{f(\cdot)}$ is the conjugate function of $f(\cdot)$. I_n represents an $n \times n$ dimensions identity matrix. M^{T} is the transpose of the matrix M.

II. Preliminaries

A. System model and estimation error

Consider the following stochastic nonlinear systems,

$$\begin{aligned} x_{k+1} &= A_k x_k + G_{k+1} w_{k+1} \\ y_k &= H(x_k) + v_k, \end{aligned} \tag{1}$$

where $x_k \in R^m$ is the state, $y_k \in R^l$ is the output, $w_k \in R^n$ and $v_k \in R^l$ are the random disturbance. A_k and G_k are two known time-varying system matrices.

w_k can be an arbitrary bounded independent random vector(or hybrid random vector, see the following definitions) rather than Gaussian input. The following assumptions are required in this paper.

Assumption 2.1: $\{w_k\}$ and $\{v_k\}$ are bounded, stationary processes. $\{w_k\}$, $\{v_k\}$ and x_0 are mutually independent. w_k has a known distribution function denoted by $F_w(x)$ with $|\mathrm{E}(w_k)| < +\infty, \mathrm{Var}(w_k) < +\infty$. v_k has a known bounded mean value $|\mathrm{E}(v_k)| < +\infty$.

Assumption 2.2: [11] $H(\cdot)$ is a known Borel measurable and smooth vector-value nonlinear function of its arguments.

For the dynamic system given by (1), a full order ”observer type” filter can be adopted as follows,

$$\begin{aligned} \hat{x}_{k+1} &= A_k \hat{x}_k + U_k(y_k - \hat{y}_k) \\ \hat{y}_k &= H(\hat{x}_k) + \mathrm{E}(v_k). \end{aligned} \tag{2}$$

where $U_k \in R^{m\times l}$ is a gain matrix to be determined. Denote $e_k = x_k - \hat{x}_k$, then the conditional estimation error equation

J. Zhou is with the Department of Automation, Tsinghua University, Beijing 100084, P. R. China

Professor H. Wang is with the Control Systems Centre, University of Manchester, U.K, and also affiliated with the Northeastern Univ. and Inst. of Automation, P. R. China

Professor D. Zhou is with the Department of Automation, Tsinghua University

978-1-4244-2079-7/08/$25.00 ©2008 AACC.

satisfies

$$e_{k+1} = A_k e_k + G_{k+1} w_{k+1} - U_k(y_k - \hat{y}_k) \quad (3)$$

where $e_k \in R^m$. A desired filter should ensure that a measure of e_{k+1} be minimized.

As Guo and Wang [11] pointed out, under a set of given $A_k, G_{k+1}, y_k, \hat{y}_k$ and U_k, e_{k+1} can be represented by a sum of two independent random vectors $A_k e_k$ and $G_{k+1} w_{k+1}$, as well as a measurable term $-U_k(y_k - \hat{y}_k)$. Thus, the probability of e_{k+1} is a conditional probability related to the probabilities of both e_k and w_{k+1} for given $A_k, G_{k+1}, y_k, \hat{y}_k$ and U_k. For simplicity, $\gamma_{e_k}(\cdot)$ and $\varphi_{e_k}(\cdot)$ are used to represent the conditional joint PDF and the conditional characteristic function of e_k, respectively.

The criteria that can be used to assess the accuracy of such a filtering algorithm relies on the statistic nature of the conditional estimation error e_k, which is comprehensively embedded in the PDF (or characteristic function) of the conditional estimation error e_k. Thus, the filter design can be performed by minimizing the following performance function at every sampling time k.

$$J = \int g(x) \log \frac{g(x)}{\gamma(x)} dx + U_k^{\mathrm{T}} R_k U_k \quad (4)$$

where $g(x)$ is a pre-specified PDF for the conditional estimation error PDF $\gamma_{e_k}(x)$ to follow. In practice $g(x)$ can be selected as a narrowly distributed Gaussian PDF. This means that the filtering design should be such that the error PDF is made as narrow and as Gaussian as possible [23]. The characteristic function is employed to formulate $\gamma_{e_k}(x)$ using equation (3).

B. Hybrid characteristic functions

The following definitions on hybrid random vectors, hybrid probabilities and hybrid characteristic functions are introduced to generalize some conventional concepts in probability theory.

Definition 2.1: [16] If a random variable z only has a value $P(z = c) = 1$, then it is called as a degenerated distribution and the variable is referred to as a degenerated variable.

Definition 2.2: A random vector $Z \in R^m$ is called as a hybrid random vector [5], [16], [17] if it contains both continuous-valued and discrete-valued random variables. Let $z_C \in R^{m_1}$ and $z_D \in R^{m_2}$ be a continuous-valued random vector and a discrete-valued random vector, respectively, with $m = m_1 + m_2$. Then their related probability, which is referred to as a hybrid probability [11], is defined as

$$P(z_C \preceq \delta, z_D = \sigma_i) \triangleq P(z_C \preceq \delta) P(z_D = \sigma_i), \quad (5)$$

where $\delta \in R^{m_1}$, $\sigma_i \in R^{m_2}, i = 1, 2, \cdots, N$. Similarly, its hybrid probability distribution function [11] is defined as

$$F(\delta, z_D \preceq \sigma_i) = \sum_{z_D \preceq \sigma_i} P(z_C \preceq \delta) P(z_D \preceq \sigma_i). \quad (6)$$

In this context, the corresponding PDF is called as a hybrid probability density function (HPDF) [11].

Definition 2.3: The characteristic function of a hybrid random vector $Z = [z_C^{\mathrm{T}} \in R^{1\times m_1}, z_D^{\mathrm{T}} \in R^{1\times m_2}]^{\mathrm{T}}$ is called as a hybrid characteristic function, which is defined by

$$\begin{aligned}\varphi_Z(t_1, \cdots, t_m) &= \mathrm{E}\{\exp(jtZ)\} = \\ &= \mathrm{E}\{\exp(jt_C z_C)\}\mathrm{E}\{\exp(jt_D z_D)\} \\ &= \int_\Omega \exp\left(j \sum_{k=C1}^{Cm_1} t_k z_k\right) dF(\delta, z_D \prec +\infty) \\ &\times \left(\sum_{k=1}^{N} (\exp(jt_D \sigma_k) p_k)\right)\end{aligned} \quad (7)$$

where $j = \sqrt{-1}$ is the imaginary number unit, $t_C \in R^{1\times m_1}, t_D \in R^{1\times m_2}, \sigma_k \in R^{m_2}, P(z_D = \sigma_k) = p_k, k = 1, 2, \cdots, N$, $C1$ and Cm_1 represent the first and the last variable of the continuous-valued sub-vector.

In the third equation, $z_D \prec +\infty$ rather than $z_D \preceq \sigma_i$ as shown in equation (6) is given because the discrete-valued sub-vector can include both finite and infinite components in the hybrid characteristic function (7).

Definition 2.4: If a variable contains a continuous-valued part and a discrete-valued part then the variable is referred as a mixed random variable denoted by z_M. Similarly, the related probability is taken as the mixed probability denoted by $P(z_{MC} \preceq \delta, z_{MD} = \sigma_i)$. A hybrid random vector Z which contains some mixed random variables ($z_M \in R^{m_3}$) is called as system-output-type hybrid (SOTH) random vector. The corresponding characteristic function is therefore referred to as an SOTH Characteristic Function.

Definition 2.5: If a random vector Z only contains continuous-valued variables and degenerated variables then it is called as a strict system-output-type hybrid (SSOTH) random vector. The corresponding characteristic function is therefore referred to as an SSOTH Characteristic Function. Based upon the definition of the hybrid characteristic function, the SSOTH characteristic function is given by

$$\begin{aligned}\varphi_Z(t_1, \cdots, t_m) &= \int_\Omega \exp\left(j \sum_{k=C1}^{Cm_1} t_k z_k\right) dF(\delta, z_D \prec +\infty) \\ &\times \exp(j \sum_{i=D1}^{Dm_2} \sigma_i t_i)\end{aligned} \quad (8)$$

where $D1$ and Dm_2 represent the first and the last variable of the degenerated variables.

Remark 2.1: In this paper, the SOTH or the SSOTH is refer to that the hybrid are brought together by the linear mapping [11] or (and) the exogenous inputs.

In order to simplify the descriptions and technical formulation procedures, in the following expressions only SSOTH random vectors will be considered.

III. Conditional Characteristic function of e_k

Under Assumptions 1 and 2, from (3) it can be noticed that one key task is to calculate the conditional characteristic function of e_{k+1} using the conditional characteristic function of $A_k e_k$ and the characteristic function of $G_{k+1}w_{k+1}$. In this regard, the computation of hybrid characteristic functions of the linear transformation and the algebraic sum operation among hybrid random vectors will be described. For this purpose, the following propositions is given.

Proposition 3.1: Let $\varphi_Z(x) = \varphi_Z(x_1, x_2, \cdots, x_n)$ be an SSOTH characteristic function of SSOTH random vector $Z \in R^n$, and denote $A \in R^{m\times n}$ and $b = [b_1, b_2, ..., b_m]^{\mathrm{T}}$ as two constant matrices. Then m dimensional random vector $Y = AZ + b$ is still an SSOTH random vector with its characteristic function being given by

$$\varphi_Y(t_1, t_2, \cdots, t_m) = e^{jtb}\varphi_Z(tA) \tag{9}$$

where $\varphi_Y(t_1, t_2, \cdots, t_m)$ is the hybrid characteristic function of Y.

Proof: Using the definition of the hybrid characteristic functions, it can be shown that

$$\begin{aligned}\varphi_Y(t) &= \mathrm{E}[e^{jtY}] = \mathrm{E}[e^{jt(AZ+b)}]\\ &= e^{jtb}\mathrm{E}[e^{j(tA)Z}] = e^{jtb}\varphi_Z(tA)\end{aligned}$$

which is also an SSOTH random vector. ∎

Proposition 3.2: Let an SSOTH random vector Z be represented by $Z = Z_1 + Z_2$, where Z_1 and Z_2 are two independent SSTOH random vectors with the same dimension. Then its SSTOH characteristic function can be represented as follows

$$\varphi_Z(t) = \varphi_{Z_1}(t)\varphi_{Z_2}(t) \tag{10}$$

where $\varphi_{Z_1}(t)$ and $\varphi_{Z_2}(t)$ are the hybrid characteristic functions of Z_1 and Z_2, respectively.

Proof: Again, using the definition of the hybrid characteristic functions, it can be shown that

$$\begin{aligned}\varphi_Z(t) &= \mathrm{E}[\exp(jt(Z_1 + Z_2))] = \mathrm{E}[\exp(jtZ_1)\exp(jtZ_2)]\\ &= \mathrm{E}[\exp(jtZ_1)]\mathrm{E}[\exp(jtZ_2)] = \varphi_{Z_1}(t)\varphi_{Z_2}(t)\end{aligned}$$

∎

Proposition 3.1 and 3.2 provide a way to compute the hybrid characteristic function of the algebraic sum of any two SSOTH random vectors. Thus, the following result can be readily obtained,

Lemma 3.1: Under assumptions 1 and 2, the conditional hybrid characteristic function of e_{k+1} can be formulated recursively by

$$\varphi_{e_{k+1}}(t) = \exp(-jU_k(y_k - \hat{y}_k)t)\varphi_{s_k}(t)\varphi_{q_{k+1}}(t) \tag{11}$$

where $s_k = A_k e_k$, $q_{k+1} = G_{k+1}w_{k+1}$. $\varphi_{s_k}(t)$ and $\varphi_{q_{k+1}}(t)$ are the corresponding (conditional) hybrid characteristic functions which can be calculated by using proposition 3.1.

Lemma 3.1 plays a key role in the filtering design of this paper.

IV. Filtering design

A. PDF tracking filtering

For the required filtering algorithm design such as the minimum entropy filter [11] and the filter designed using the performance index (4), the conditional PDF of the e_{k+1} need to be obtained as a starting point. If the conditional hybrid characteristic function of e_{k+1} has been formulated using Lemma 3.1, then its corresponding conditional hybrid probability density function can be obtained by the inverse transformation and the filter can be readily designed. However, the different linear transformations in (11) can result in different PDF forms with which the filter design should be treated differently. Furthermore, the design of the filter based on the conditional PDF shaping needs another transformation which will possibly increase the computation time and thus may not be suited for the real time filtering. On the other hand, the filter design can directly use the conditional characteristic function simply because controlling the shape of conditional hybrid probability density function is equivalent to the shape control of its conditional hybrid characteristic function. Thus, in the rest of the paper, we will only consider the filter design based on the conditional hybrid characteristic function by selecting a filtering gain matrix U_k. As a result, the aim of the filter design is to select U_k such that $\varphi_{e_k}(t)$ is made as close as possible to $\varphi_g(t)$ ($\varphi_g(t)$ is the given target characteristic function).

B. Reselection of performance indices

The direct use of performance index (4) for the filter design can be too complicated to be used in practice. An alternative performance index should therefore be formed which measures directly the difference between $\varphi_{e_k}(t)$ and $\varphi_g(t)$. In addition, since the SSOTH characteristic function can still exhibit the same basic properties of normal characteristic functions such as $|\underline{\varphi(t)}| \leq 1$ (where $|\cdot|$ is the complex modulus), and $\varphi(-t) = \overline{\varphi(t)}$, we can use the characteristic functions for SSOTH random variables to define the distance between two characteristic functions. The following new performance index

$$\begin{aligned}J_1 &= \left\{\int_\Omega \mathcal{K}(t)\varphi_g(t)\log\frac{\varphi_g(t)}{\varphi_{e_k}(t)}dt\right\}^2 + U_k^{\mathrm{T}}R_kU_k\\ &= J_0 + U_k^{\mathrm{T}}R_kU_k\end{aligned} \tag{12}$$

can be defined, where $J_0 = \left\{\int_\Omega \mathcal{K}(t)\varphi_g(t)\log\frac{\varphi_g(t)}{\varphi_{e_k}(t)}dt\right\}^2$, and $\mathcal{K}(t)$ in equation (12) is a weighting function which should be selected so as to make J_0 a real number and to guarantee the boundness of J_0. In this index, the first term is similar to the well known Kullback-Leibler distance widely used in the information functional measure. The use of such a new performance index allows the transfer of the multiplication operations in equation (11) into a simple algebraic sum by using a 'logarithm' operator. Simultaneously, by minimizing this term, $\varphi_{e_k}(t)$ can be made as close as possible to $\varphi_g(t)$. The first term is zero for $\varphi_g(t) = \varphi_{e_k}(t)$ (almost surely) and infinite if there is a set of a positive Lebesgue measure on which $\varphi_{e_k}(t) \equiv 0$ [13]. The second term in equation (12) is again the soft constraint on the filter gain matrix with $R_k > 0$ being a pre-specified weighting matrix.

The effect of $\mathcal{K}(t)$ is similar to that in [25]. However, it only needs to guarantee the boundness of its performance functions. Since $|\varphi(t)| \leq 1$, it can be shown that

$$\begin{aligned} J_0 &= \left\{\int_\Omega \mathcal{K}(t)\varphi_g(t)\log\frac{\varphi_g(t)}{\varphi_{e_k}(t)}dt\right\}^2 \\ &= \left\{\left|\int_\Omega \mathcal{K}(t)\varphi_g(t)\log\frac{\varphi_g(t)}{\varphi_{e_k}(t)}dt\right|\right\}^2 \\ &\leq \left\{\int_\Omega \left|\mathcal{K}(t)\log\frac{\varphi_g(t)}{\varphi_{e_k}(t)}\right|dt\right\}^2 \\ &\leq \left\{\int_\Omega |\mathcal{K}(t)|\sqrt{\log^2\left|\frac{\varphi_g(t)}{\varphi_{e_k}(t)}\right| + \frac{\pi^2}{4}}dt\right\}^2 \end{aligned} \quad (13)$$

To guarantee the uniform boundness of J_0, it is sufficient to select the weighting function $|\mathcal{K}(t)| > 0$ such that the following inequality

$$|\mathcal{K}(t)|\sqrt{\log^2\left|\frac{\varphi_g(t)}{\varphi_{e_k}(t)}\right| + \frac{\pi^2}{4}} \leq M_1\exp(-tM_2t^{\mathrm{T}}) \quad (14)$$

holds, where $t^{\mathrm{T}} \in R^m$, $M_1 > 0$ and $M_2 = M_2^{\mathrm{T}} \in R^{m\times m}$ are pre-specified numbers and positive definite matrix, respectively. Using (14), it can be further shown that if the following inequality

$$0 < |\mathcal{K}(t)| \leq M_1'\exp(-tM_2t^{\mathrm{T}}), M_1' \leq \frac{2}{\pi}M_1 \quad (15)$$

holds, then the boundedness of J_0 can be guaranteed. Moreover, if $\mathcal{K}(-t) = \overline{\mathcal{K}(t)}$, then J_0 is a real number. To summarize, we have the following theorem.

Theorem 4.1: Suppose the weighting function $\mathcal{K}(t)$ has been selected so that $0 < |\mathcal{K}(t)| \leq M_1'\exp(-tM_2t^{\mathrm{T}})$ and $\mathcal{K}(-t) = \overline{\mathcal{K}(t)}$, then J_0 is a real number and $J_0 \leq M_1^2\pi^n / \prod_{i=1}^n \xi_i$, where $\xi_i, (i = 1, 2, \cdots, n)$ are the diagonal elements of M_2 .

C. Optimal filter gain matrix

Once the performance index J_1 is selected, the filter design can be readily carried out by directly minimizing the selected performance index. In order to provide the filter with a simple structure, the instantaneous performance index J_1 in (12) is considered firstly in the design. For this purpose, U_k should be calculated from

$$\frac{\partial J_1}{\partial U_k} = 0 \quad (16)$$

which leads to

$$2R_kU_k + \frac{\partial J_0}{\partial U_k} = 0 \quad (17)$$

To obtain an analytical solution, J_0 should be analyzed first. Indeed, it can be shown that

$$\begin{aligned} J_0 &= \left\{\int_\Omega \mathcal{K}(t)\varphi_g(t)\log\frac{\varphi_g(t)}{\varphi_{e_k}(t)}dt\right\}^2 \\ &= a_k^2 + 2a_kb_kU_k + U_k^{\mathrm{T}}b_k^{\mathrm{T}}b_kU_k \end{aligned} \quad (18)$$

where it has been defined that

$$\begin{aligned} a_k &= \int_\Omega \mathcal{K}(t)\varphi_g(t)\log\varphi_g(t)dt \\ &- \int_\Omega \mathcal{K}(t)\varphi_g(t)\log\varphi_{s_k}(t)dt \\ &- \int_\Omega \mathcal{K}(t)\varphi_g(t)\log\varphi_{q_{k+1}}(t)dt \\ b_k &= \int_\Omega \mathcal{K}(t)\varphi_g(t)jt(y_k - \hat{y}_k)dt. \end{aligned} \quad (19)$$

Substituting equation (18) into equation (17) yields

$$a_kb_k + \left(b_k^{\mathrm{T}}b_k + R_k\right)U_k = 0 \quad (20)$$

and the filter gain matrix can be obtained immediately as follows.

$$U_k = -\left(b_k^{\mathrm{T}}b_k + R_k\right)^{-1}(a_kb_k) \quad (21)$$

From equation (20), the following condition on the second-order derivative of J_1 should also be satisfied at every sampling time in order to guarantee the minimization result.

$$\frac{\partial^2 J_1}{\partial U_k^2} = b_k^{\mathrm{T}}b_k + R_k > 0 \quad (22)$$

It shows that the solution of equation (21) is a global optimal solution.

D. Filter analysis and synthesis

In this subsection, we will study the conditions under which the filter error is stochastically, exponentially and ultimately bounded in the mean square sense. This will assure the practical use of the obtained filter. For this purpose, the following definition will be required.

Definition 4.1: [26] For system (3), the solution is stochastically, exponentially and ultimately bounded in the mean square sense if there exist constants $\alpha_1 \geq 0, \alpha_2 > 0$ and $\beta > 0$ such that $\mathrm{E}\|e_k\|^2 \leq \alpha_1 + \alpha_2e^{-\beta k}$, where $\|e_k\| \triangleq \sqrt{e_k^{\mathrm{T}}e_k}$.

To simplify the expression, we introduce the following concept on hybrid cumulate generation functions.

Definition 4.2: Let $\varphi_Z(t)$ be a hybrid characteristic function of hybrid stochastic vector Z, the corresponding hybrid cumulate generation function is $\log\varphi_Z(t)$. The corresponding property of hybrid cumulate generation function can be referred to the property of cumulate generation function.

Using the above definitions, the following result can be obtained.

Theorem 4.2: Under the assumptions 1 and 2, if $\|A_k\| \triangleq \rho(A_k) < 1$ at each sampling time, then the error system (3) is exponentially and ultimately bounded in the mean square sense when the filtering strategy in (21) is applied.

Proof: Firstly, assume that $\varphi_{e_0}(t) = 1$, then from (11) it can be obtained that

$$\log\varphi_{e_k}(t) = \sum_{i=1}^{k-1}\log\varphi_w(t\tilde{A}_iG_i) - j\sum_{i=1}^{k-1}t\tilde{A}_iU_i\Delta y_i \quad (23)$$

where $\varphi_w(\cdot)$ is the characteristic function of the random vectors w_k. Denote

$$\begin{cases} \tilde{A}_i = \prod_{l=i+1}^{k-1} A_l \ , & i = 1, \cdots, k-2 \\ \tilde{A}_i = I & , \ otherwise \end{cases}$$

and $\Delta y_i = y_i - \hat{y}_i, i = 1, \cdots, k-1$. thus by denoting

$$\theta_i(t) = \frac{\varphi_w''(t\tilde{A}_i G_i)\varphi_w(t\tilde{A}_i G_i) - \varphi_w'(t\tilde{A}_i G_i)[\varphi_w'(t\tilde{A}_i G_i)]^{\mathrm{T}}}{\varphi_w^2(t\tilde{A}_i G_i)}$$

it can be shown that

$$\begin{aligned} \mathrm{E}\left\|e_k\right\|^2 &= (-j)^2 \frac{\partial^2 \log \varphi_{e_k}(t)}{\partial t \partial t^{\mathrm{T}}} \\ &+ \left\| \left((-j) \frac{\partial \log \varphi_{e_k}(t)}{\partial t} \right) \right\|^2 |_{t=0} = -\sum_{i=1}^{k-1} \theta_i(0) \\ &- \sum_{i=1}^{k-1} \left(\frac{\varphi_w'(t\tilde{A}_i G_i)}{\varphi_w(t\tilde{A}_i G_i)} - j(\tilde{A}_i U_i \Delta y_i)^{\mathrm{T}} \right) \\ &\times \left(\frac{\varphi_w'(t\tilde{A}_i G_i)}{\varphi_w(t\tilde{A}_i G_i)} - j(\tilde{A}_i U_i \Delta y_i)^{\mathrm{T}} \right)^{\mathrm{T}} |_{t=0} \end{aligned} \quad (24)$$

where $t = [t_1, \cdots, t_m]$ and

$$\begin{aligned} \frac{\partial^2 \log \varphi_{e_k}(t)}{\partial t \partial t^{\mathrm{T}}} &\triangleq \sum_{l=1}^{m} \frac{\partial^2 \log \varphi_{e_k}(t)}{\partial t_l^2}, \\ \varphi_w''(t\tilde{A}_i G_i) &\triangleq \sum_{l=1}^{m} \frac{\partial^2 \varphi_w(t\tilde{A}_i G_i)}{\partial t_l^2}, \\ \varphi_w'(t\tilde{A}_i G_i)) &\triangleq \frac{\partial \varphi_w(t\tilde{A}_i G_i)}{\partial t}. \end{aligned}$$

The first equality can be proved by using the property of the cumulate generation function. Denote $\mathrm{E}(w_k) = \mu \in R^m$ and $\mathrm{Var}(w_k) = \delta \in R^{m \times m}$, then it can be observed that

$$\begin{aligned} -\sum_{i=1}^{k-1} \theta_i(0) &= \sum_{i=1}^{k-1} \mathrm{Tr}\{\tilde{A}_i G_i \delta G_i^{\mathrm{T}} \tilde{A}_i^{\mathrm{T}}\} + \mu^{\mathrm{T}} G_i^{\mathrm{T}} \tilde{A}_i^{\mathrm{T}} \tilde{A}_i G_i \mu \\ &\leq \sum_{i=1}^{k-1} \rho_m^{2i} \rho_n^2 (\mathrm{Tr}\{\delta\} + \mu^{\mathrm{T}} \mu) \end{aligned}$$

where Tr represents the trace of a matrix, $\rho_m = \max_{1 \leq i \leq k-1} \rho(A_i)$ and $\rho_n = \max_{1 \leq i \leq k-1} \rho(G_i)$, $\rho(G_i) \triangleq \|G_i\|$. Similarly, if $\varphi_{e_0}(t) \neq 1$, then we have

$$\begin{aligned} & -\sum_{i=1}^{k-1} \frac{\varphi_w''(t\tilde{A}_i G_i)\varphi_w(t\tilde{A}_i G_i) - \varphi_w'(t\tilde{A}_i G_i)[\varphi_w'(t\tilde{A}_i G_i)]^{\mathrm{T}}}{\varphi_w^2(t\tilde{A}_i G_i)} \Bigg|_{t=0} \\ & \leq \sum_{i=1}^{k-1} \rho_m^{2i} \rho_n^2 (\mathrm{Tr}\{\delta\} + \mu^{\mathrm{T}} \mu) \\ & + \rho_m^{2k} \rho_n^2 (\mathrm{Tr}(\mathrm{Var}(e_0) + \mathrm{E}(e_0)^{\mathrm{T}} \mathrm{E}(e_0)) \end{aligned}$$

Therefore, if $\|A_k\| < 1$, then the boundedness of $-\sum_{i=1}^{k-1} \theta_i(0)$ can be guaranteed. Denote

$$\pi_i(t) = \left(\frac{\varphi_w'(t\tilde{A}_i G_i)}{\varphi_w(t\tilde{A}_i G_i)} - j(\tilde{A}_i U_i \Delta y_i)^{\mathrm{T}} \right)$$

In the following we will prove that $-\sum_{i=1}^{k-1} \pi_i(0)\pi_i^T(0)$ is also bounded. For this purpose and from the above proof procedure, we only need to prove that $\|U_i \Delta y_i\|$ is bounded. Indeed, it can be shown that

$$\begin{aligned} \|U_i \Delta y_i\| &= \left\| a_i (b_i^{\mathrm{T}} b_i + R_i)^{-1} b_k^{\mathrm{T}} \Delta y_i \right\| \\ &= \left\| a_i (b_0^{\mathrm{T}} \Delta y_i^{\mathrm{T}} \Delta y_i b_0 + R_i)^{-1} b_0^{\mathrm{T}} \Delta y_i^{\mathrm{T}} \Delta y_i \right\| \\ &\leq \left\| a_i (b_0^{\mathrm{T}} \Delta y_i^{\mathrm{T}} \Delta y_i b_0 + R_i)^{-1} \right\| \\ &\times \frac{\left\| (b_0^{\mathrm{T}} \Delta y_i^{\mathrm{T}} \Delta y_i b_0 + R_i) \right\| \left\| b_0^{\mathrm{T}} \right\|}{b_0 b_0^{\mathrm{T}}} \\ &\leq \frac{\|a_i\|}{\|b_0\|}, b_0 \neq 0. \end{aligned}$$

because otherwise one should come to the conclusion that

$$\|U_i \Delta y_i\| = 0$$

where $b_i = \Delta y_i b_0, b_0 = \int_\Omega \mathcal{K}(t)\varphi_g(t) jtdt$. From (25), it can finally be concluded that

$$\mathrm{E}\left\|e_k\right\|^2 < +\infty$$

■

The results in theorem 4.2 is a large-scale exponentially and ultimately bounded in the mean square sense. In the sequel, a local exponentially ultimately bounded filtering error in the mean square sense will be formulated. At first, it can be observed that system (3) can be approximated to read

$$e_{k+1} = (A_k - U_k B_k) e_k + U_k \sum_{i \geq 2} s_i e_k^i + G_k w_k \quad (25)$$

where $B_k \triangleq \left.\frac{\partial H(\cdot)}{\partial x_k}\right|_{x_k = \hat{x}_k}$, $s_i \triangleq [s_{i1}, \cdots, s_{in}]$ and $e_k^i \triangleq [e_{k1}^{i1_1} \cdots e_{km}^{i1_m}, \cdots, e_{k1}^{in_1} \cdots e_{km}^{in_m}]^{\mathrm{T}}$. In which, $s_{ip} \triangleq \left.\frac{\partial^i H(\cdot)}{\partial x_{k1}^{ip_1} \cdots \partial x_{km}^{ip_m}}\right|_{x_k = \hat{x}_k}$ and $\sum_{l=1}^{m} ip_l = i, ip_l \geq 0, p = 1, \cdots, n$. Denote $\hat{A}_k = A_k - U_k B_k$ and $S(e_k) = U_k \sum_{i \geq 2} s_i e_k^{i-1}$, then we have the following local result on the exponentially and ultimately bounded error term in the mean square sense.

Theorem 4.3: If there exists a constants $C > 0$ and U_k such that $\left\|\hat{A}_k\right\| + \left\|U_k \sum_{i \geq 2} s_i C^{i-1}\right\| = \varrho < 1$ and $\mathrm{E}(\|G_k w_k\|^2) \leq (1 - \varrho)^2 C^2$, then $\forall \mathrm{E}\left\|e_0\right\|^2 \leq C^2$, the solution of the system (3) is exponentially and ultimately bounded in the mean square sense when the filtering strategy in (21) is applied, where ϱ is called as a convergence exponent and $1 - \varrho$ is regarded as a noise damp exponent.

Proof: From (25), it can be computed that

$$\begin{aligned} e_{k+1}^{\mathrm{T}} e_{k+1} &= e_k^{\mathrm{T}} (\hat{A}_k^{\mathrm{T}} \hat{A}_k + S^{\mathrm{T}}(e_k) S(e_k) + 2\hat{A}_k^{\mathrm{T}} S(e_k)) e_k \\ &+ w_k^{\mathrm{T}} G_k^{\mathrm{T}} G_k w_k + 2 e_k^{\mathrm{T}} \hat{A}_k^{\mathrm{T}} G_k w_k \\ &+ 2 e_k^{\mathrm{T}} S^{\mathrm{T}}(e_k) G_k w_k \end{aligned} \quad (26)$$

When k=0, we should have the following

$$\begin{aligned} \mathrm{E}(e_1^{\mathrm{T}} e_1) &\leq \varrho^2 C^2 + (1-\varrho)^2 C^2 + 2\mathrm{E}\left\|G_0 w_0\right\| \varrho C \\ &< \varrho^2 C^2 + (1-\varrho)^2 C^2 + 2(1-\varrho)\varrho C^2 \\ &\leq \theta_1^2 C^2, (0 < \theta_1 < 1) \end{aligned} \quad (27)$$

because $\mathrm{E}\|G_0 w_0\| < \sqrt{\mathrm{E}(\|G_0 w_0\|^2)}$ As a result, it can be further obtained that

$$\begin{aligned}\mathrm{E}(e_2^{\mathrm{T}} e_2) &< \varrho^2\theta_1^2 C^2 + (1-\varrho)^2 C^2 + 2(1-\varrho)\varrho\theta_1 C^2 \\ &\le \theta_2^2(\varrho\theta_1 + 1 - \varrho)^2 C^2, (0 < \theta_2 < 1) \qquad (28)\end{aligned}$$

$$\begin{aligned}\mathrm{E}(e_3^{\mathrm{T}} e_3) &< \theta_3^2(\varrho\theta_2(\varrho\theta_1 + 1 - \varrho) + 1 - \varrho)^2 C^2 \\ &= \theta_3^2(\varrho^2\theta_2\theta_1 + \varrho\theta_2 - \varrho^2\theta_2 + 1 - \varrho)^2 C^2 \qquad (29)\end{aligned}$$

Denote $\tilde{\theta}_i = \varrho^i \prod_{l=1}^{i} \theta_{n-l}$.Thus it can be conjectured that

$$\mathrm{E}(e_k^{\mathrm{T}} e_k) \le \theta_k^2 \left(\sum_{i=1}^{k-1} \tilde{\theta}_i - \varrho \sum_{i=1}^{k-2} \tilde{\theta}_i + 1 - \varrho \right)^2 C^2,$$

$$(0 < \theta_k < 1, k \ge 2).$$

Indeed, this conjecture is certainly true when $k = 2$. Thus, if we assume that it is also true for $k = n, n \ge 2$; that is, $\mathrm{E}(e_n^{\mathrm{T}} e_n) \le \theta_n^2 \left(\sum_{i=1}^{n-1} \tilde{\theta}_i - \varrho \sum_{i=1}^{n-2} \tilde{\theta}_i + 1 - \varrho \right)^2 C^2$. Then for $k = n+1$, it can be shown that

$$\mathrm{E}(e_{n+1}^{\mathrm{T}} e_{n+1}) \le \theta_{n+1}^2 \left(\sum_{i=1}^{n} \tilde{\theta}_i - \varrho \sum_{i=1}^{n-1} \tilde{\theta}_i + 1 - \varrho \right)^2 C^2 \qquad (30)$$

Therefor the identity also holds for $k = n+1$, and thus by the principle of mathematical induction, the identity is valid for all $k \ge 2$. ■

Although the SSOTH random vectors are considered, all the propositions, lemmas, and theorems can be applied to SOTH random vectors.

V. Conclusions

Using the concept of PDF shaping, a new optimal tracking filter design for multivariate stochastic systems subjected to non-Gaussian noise is presented in this paper, where the key idea is to select the filtering gain so that the PDFs of the filtering error can be made to follow a target distribution shape. This has therefore extended the existing minimum variance based filtering algorithms. Indeed, if the targeted distribution is a narrowly distributed Gaussian PDF, then the proposed filter aims at obtaining a state estimation error whose PDFs is made as close as possible to a Gaussian shape.

Acknowledgements

This work was supported in by NSFC (Grant No. 60574084, 60534010), the national 863 project (Grant No. 2006AA04Z428), national 973 program of China (Grant No. 2002CB312200) and part by China Postdoctoral Science Foundation (Grant No. 20060390071).

References

[1] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*, Prentice Hall, Englewood Cliffs N. J., 1979.

[2] K. J. Astrom. *Introduction to Stochastic Control Theory*, Academin Press, New York, 1970.

[3] J. Carpenter, P. Clifford, and P. Fearnhead. Improved particle filter for nonlinear problems. *IEE Proc. Radar, Sonar Navig.*, Vol. 146, pp. 27, 1999.

[4] S. Challa and Y. Bar-Shalom. Nonlinear filter design using Fokker-Planck-Kolmogonov probability density evolutions. *IEEE Trans. Aerospace Electron. Systems*, Vol.36, pp.309315, 2000. 19

[5] T. Cham and J.M. Rehg. A multiple hypothesis approach to figure tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 239245, 1999.

[6] C. E. de Souza, U. Shaked, and M. Y. Fu. Robust H1 filtering for continuous time varying uncertain systems with deterministic input signals. *IEEE Trans. Signal Processing*, Vol. 43, pp. 709719, 1995.

[7] P. M. Djuric and J. H. Kotecha etc. Particle filtering: a review of the theory and how it can be used for solving problems in wirelees communications. *IEEE Signal Processing Magazine*, Vol. 9, pp.1938, 2003.

[8] C. A. Einicke and L. B. White. Robust extended Kalman filtering. *IEEE Trans. Signal Processing*, Vol. 47, pp.25962599, 1999.

[9] E. Gershon, U. Shaked D. J. N., Limebeer, and I. Yaesh. Robust H1 filtering of stationary continue time linear systems with stochastic uncertainties. *IEEE Trans. Automat. Contr.* , Vol. 46, pp.17881793, 2001.

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

[11] L. Guo and H. Wang. Minimum entropy filtering for multivariate stochastic systems with non-Gaussian noises. *IEEE Trans. Automat. Contr.*, Vol. 51, pp.695700, 2006.

[12] S. Julier. The scaled unscented transformation. *Proceedings of the 2002 American Control Conference*, Vol. 6, pp. 4555 4559, 2002.

[13] M. Karny. Towards fully probabilitic control design. *Automatica*, Vol. 32, pp. 17191722, 1996.

[14] C. Kwok, D. Fox, and M. Meila. Real time particle filters. *Proceedings of the IEEE*, Vol. 92, pp. 469484, 2004.

[15] H. Z. Li and M.Y. Fu. A linear matrix inequality approach to robust H1 filtering. *IEEE Trans. Signal Processing*, Vol. 45, pp. 2338 2350, 1997.

[16] Z. H. Ma and etc. Handbook of Modern Application Mathmatical: The Volume of Probability Statistics and Stochastic Process. Tsighua University Publishing Company, Beijing, 2001.

[17] E. Mazor, A. Averbuch, and J. Dayan. Interacting multiple model methods in target tracking: a survey. *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 34, pp. 103123, 1998.

[18] I. R. Peterson and V. A. Ugrinovskii. Minimax optimal control and filtering for stochastic uncertain systems. *Proc. ARO Workshop 2000 on Intelligent Systems*, Australian National University, Australia, 2000.

[19] U. Shaked and C. E. de Souza. Robust minimum variance filtering. *IEEE Trans. Signal Processing*, Vol.43, pp.24742483, 1995.

[20] U. Shaked, L. H. Xie, and Y. C. Soh. New approaches to robust minimum variance filtering design. *IEEE Trans. Signal Processing*, Vol. 49, pp. 26202629, 2001.

[21] E. Wan and R. van der Merwe. The unscented kalman filter. In S. Haykin, editor, Kalman Filtering and Neural Networks, pages 221269. John Wiley and Sons, 2001.

[22] F. Wang and V. Balakrishnan. Robust steady state filtering for systems with deterministic and stochastic uncertainties. *IEEE Trans. Signal Processing*, Vol. 51, pp.25502558, 2003.

[23] H. Wang. *Bounded Dynamic Stochastic Systems: Modelling and Control*. Springer Verlag Ltd., London, 2000.

[24] L. Wang and Y. Zhang. Stability of nonlinear analytic difference systems. *ACTA Mathmatica Sinica*, Vol. 38, pp. 355361, 1995,(in Chinese).

[25] Y. Wang and H. Wang. Output PDF control of linear stochastic systems with arbitrarily bounded random parameters, a new application of the laplace transforms. *Proc. of the American Control Conference*, Vol.5, pp.4262 4267, 2002.

[26] Z. D. Wang and W. C. H. Daniel. Filtering on nonlinear time delay stochastic systems. *Automatica*, Vol. 39, pp.101109, 2003.

[27] Z. D. Wang and B. Huang. Robust H2/H1 filtering for linear systems with error variance constraints. *IEEE Trans. Signal Processing*, Vol. 48, pp. 24632467, 2000.

[28] L.Guo and H.Wang, Minimum entropy filtering for multivariate stochastic systems with non-Gaussian noises, *IEEE Transactions on Automatic Control*, Vol 51, No.4, pp. 695-700,2006.