2005 American Control Conference
June 8-10, 2005. Portland, OR, USA

Minimum Entropy Filtering for Multivariate Stochastic Systems with Non-Gaussian Noises

Lei Guo and Hong Wang

Abstract— **In this paper, a minimum entropy filtering algorithm is presented for a class of multivariate dynamic stochastic systems. The concerned systems are represented by a set of time-varying difference equations with multiple non-Gaussian stochastic inputs, and with nonlinearity in the measurement output. Several new concepts including hybrid random vectors, hybrid probability and hybrid entropy are introduced to describe the probabilistic property and randomness of the stochastic estimation errors. New relationships are established between the probability density functions (PDFs) of the multivariate stochastic input and output for different mapping cases. Recursive algorithms are then proposed to design the real-time optimal filters such that hybrid entropy of the estimation error is minimized.**

I. INTRODUCTION

Research on the state estimation theory has been regarded as an very important aspect following the development of the Kalman filtering theory, where the variance of estimation error is minimized, suppose that the estimated system is linear and the random inputs are Gaussian white noises (see, *e.g.* [2], [3], [6]). For the stochastic systems with nonlinearities and complex disturbances, many feasible methods such as the extended Kalman filtering (EKF) have also been proposed for various nonlinear systems (see [1], [4], [10], [12], [13] and references therein). The confinement of the most EKF methods is that the investigated noises are still supposed to be Gaussian and the properties of Gaussian noises are used. However, it is noted that even for a nonlinear system with Gaussian inputs, the system output can be non-Gaussian, or even obey a multiple-peak and asymmetric probability density function (PDF) due to the nonlinearity. This means that the current EKF theory, where only expectation and variance of the state estimation error are concerned, is not efficient for nonlinear stochastic systems with non-Gaussian noises (see also [14], [16]).

Entropy has been widely used in information, thermodynamics, communication and control theories as a measure for the average information contained in a given PDF of a stochastic variable [5], [9], [15]. By minimizing the entropy, all higher order moments (not only the second one) can be minimized. Since generally PDFs are positive nonlinear functions, one key task is to find the relationships between the PDFs of input and the concerned output, and optimize the entropy of the output PDFs. In probability theory, the Bayesian theorem has been used to carry out the computation for the PDFs of non-Gaussian signals, where several numeral algorithms such as the Monte Carlo approach are involved [2], [9]. In the statistic approaches, the approximation for the conditional probability is generally complicated and its accuracy and convergence are difficult to verify. In [14], [15], B-spline expansions were used to model the stochastic processes between the input and the measured PDFs of the output such that the problem was transferred to optimize the weights corresponding to the basic functions. Indeed, for either SISO or MIMO systems with non-Gaussian signals, less feasible synthesis methods have been seen for the stochastic filtering problem, where the stability and satisfactory performance of the estimation error dynamics should be guaranteed.

In this paper, a novel approach is presented to generalize the classical optimal filtering theory for a class of multivariate stochastic systems with non-Gaussian inputs. For this purpose, the concepts of *hybrid entropy* and *hybrid probability* are established to describe the multivariate random output vectors and their randomness (*Section 2*). Using these concepts, the relationships between the PDFs of the multivariate stochastic input and output are formulated explicitly (*Section 3*). *Minimum entropy filters* are then designed such that the hybrid entropy of the stochastic estimation error is minimized. In order to design the filter with guaranteed stability, the weighting matrices in the performance index are tuned recursively to ensure both optimality and stability (*Section 4*). In the following, if not stated, matrices and vectors are assumed to have compatible dimensions. The identity and zero matrices are denoted by I and 0 respectively. For a square matrix M, its inverse and determinant are denoted by M^{-1} and $\det M$ respectively. For two real vectors v_1 and v_2, the notation $v_1 \succeq v_2$ is used to denote that every element of v_1 is no less than the corresponding one of v_2. For a random vector z, $P\{z \preceq \tau\}$ represents the joint probability of event $\{z \preceq \tau\}$, $F_z(\tau)$ and $\gamma_z(\tau)$ denote that the joint probability distribution function and PDF of z, respectively.

II. PRELIMINARIES

A. Plant Model

Consider the following system

$$\begin{cases} x_{k+1} = A_k x_k + G_k w_{k+1} \\ y_k = H(x_k) \end{cases} \tag{1}$$

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

L. Guo and H. Wang are with Control Systems Centre, Manchester University, M60 1QD, UK. Email: hong.wang@manchester.ac.uk
L. Guo is presently with Institute of Automation, Southeast University, Nanjing 210096, P. R. China. Email:l.guo@seu.edu.cn

w_k can be an arbitrary bounded independent random vector rather than a Gaussian one used in the classical EKF theory. The random output y_k can be non-Gaussian because of the nonlinear function $H(\cdot)$.

The following assumptions are required in this paper, which can be satisfied by many practical cases (see *e.g.* [14]).

Assumption A.1: The random variables w_k $(k = 0, 1, 2, \cdots)$ are bounded, stationary and mutually independent with a known PDF as denoted by $\gamma_w(\tau)$ which is defined on $[a, b]^n$.

To model the PDF of w_k, besides a direct measurement using some advanced instruments (*i.e.*, a laser particle size distribution measure), the *kernel estimation technique* based on an open loop test and other identification methods can be used [11], [14].

Assumption A.2: $H(\cdot)$ is a known Borel measurable and smooth vector-value nonlinear function of its arguments.

B. Filter and Estimation Error

For the dynamic system given by (1), a filter can be described by

$$\widehat{x}_{k+1} = A_k \widehat{x}_k + U_k(y_k - H(\widehat{x}_k)) \tag{2}$$

where $U_k \in R^{m\times l}$ is a gain to be determined. The resulting estimation error $e_k = x_k - \widehat{x}_k$ satisfies

$$e_{k+1} = A_k e_k - U_k \left(H(x_k) - H(\widehat{x}_k)\right) + G_k w_{k+1} \tag{3}$$

where $e_{k+1} \in [\alpha, \beta]^m$, and α and β can also be respectively chosen as $\pm\infty$. A desired filter should make the measure of e_k either converge to zero or be minimized. In (3), e_{k+1} can be represented by a sum of two independent random vectors $A_k e_k$ and $G_k w_{k+1}$, as well as a measurable term $-U_k \left(y_k - H(\widehat{x}_k)\right)$. 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, y_k, $\widehat{y}_k$ and U_k. For simplicity, $\gamma_{e_k}(\cdot)$ is used to represent the *conditional joint PDF* of e_k.

In this paper the system matrices are supposed to satisfy the following condition.

Assumption A.3: There exists a matrix P_k which is invertible at each sample time k, such that

$$\begin{aligned} P_k^{-1} A_k P_k &= \begin{bmatrix} A_{1k} & A_{2k} \\ 0 & A_{3k} \end{bmatrix} := \widetilde{A}_k, \\ P_k^{-1} G_k &= \begin{bmatrix} G_{1k} \\ 0 \end{bmatrix} := \widetilde{G}_k \end{aligned} \tag{4}$$

where A_{1k} is invertible and satisfies rank condition $rank(A_{1k}) = rank(G_{1k}) = r(\leq n)$ for each sample time k.

Denote $\widetilde{e}_k := P_k^{-1} e_k$, then (3) becomes

$$\widetilde{e}_{k+1} = \widetilde{A}_k \widetilde{e}_k - \widetilde{U}_k \left(H(x_k) - H(\widehat{x}_k)\right) + \widetilde{G}_k w_{k+1} \tag{5}$$

where $\widetilde{A}_k$ and $\widetilde{G}_k$ are denoted by (4) and $\widetilde{U}_k := P_k^{-1} U_k$. (5) will be used instead of (3) in the following sections, for which the operations on the joint PDFs of the estimation errors will be simplified.

Remark 1: Although *Assumption A.3* can cover a large class of the estimated plants, including those satisfying $n \geq m$ in (3), it should be pointed out that the design procedures will be more complicated technically if it is unsatisfied. For the limitation of space, in this paper *Assumption A.3* is used to simplify the following explicit design procedure.

The purpose of filtering is to use available information of the systems input and output to estimate x_k. The criteria that can be used to assess the accuracy of such a filtering algorithm relies on the statistic nature of the estimation error e_k, which is comprehensively embedded in the PDF of the estimation error vector $\widetilde{e}_k$. In this context, it is important to formulate the PDF of the estimation error using equation (3) or (5). From equation (5), it is shown that the first key issue to be addressed is how to formulate the PDFs of $\widetilde{A}_k \widetilde{e}_k$ and $\widetilde{G}_k w_{k+1}$. To provide unified notations, at sample time $k+1$ we consider multivariate mapping

$$\theta_{k+1} = D_k \pi_{k+1} \in [\alpha, \beta]^m \tag{6}$$

where $\pi_{k+1} \in [a, b]^n$ is a non-Gaussian continuous random vector with a given joint PDF defined as $\gamma_\pi(\cdot)$ and $D_k \in R^{m\times n}$ is a known matrix.

C. Hybrid Probability and Hybrid PDFs

In order to study the stochastic behavior of the output of a multivariate stochastic system, the existing theory on multivariate random vectors and their joint probability are required to be extended. The following definitions on hybrid random vectors and their probabilities generalize some conventional concepts. Since a deterministic variable can also be regarded as a special discrete random one with only one sample point, we use the unified notation–*hybrid random vector* to represent the case which contains continuous-time, discrete-time and deterministic variables. This leads to the following definition.

Definition 1: If after re-arranging the sequence of elements, a random vector $\widetilde{z} \in [\alpha, \beta]^m$ can be transferred into $z = \begin{bmatrix} z_1^T & z_2^T \end{bmatrix}^T$, where $z_1 \in [\alpha, \beta]^{m_1}$ is a continuous random sub-vector and $z_2 \in [\alpha, \beta]^{m_2}$ is a discrete random sub-vector that takes finite values at $\{\sigma_1, \sigma_2, ..., \sigma_{m_2}\}$ with $m = m_1 + m_2$, then $\widetilde{z}$ or z is called as a *hybrid random vector*. The related probability for z is referred to as the *hybrid probability* and defined by $P\{z_1 \preceq \delta, z_2 = \sigma_i\}$, where $\delta \in [\alpha, \beta]^{m_1}$, $\sigma_i \in [\alpha, \beta]^{m_2}$, $i = 1, 2, \cdots, M$. Similarly, its *hybrid probability distribution function* is defined by

$$F_{z_1}(\delta, z_2 = \sigma_i) = P\{z_1 \preceq \delta, z_2 = \sigma_i\}, i = 1, 2, \cdots, M \tag{7}$$

and hybrid probability density function (hybrid PDF) is defined as

$$\gamma(\delta, z_2 = \sigma_i) = \frac{\partial F_{z_1}(\delta, z_2 = \sigma_i)}{\partial \delta}, i = 1, 2, \cdots, M \tag{8}$$

For hybrid random vector z we still denote

$$F_z(\eta) = F_{z_1}(\delta, z_2 = \sigma_i), \gamma_z(\eta) = \gamma_{z_1}(\delta, z_2 = \sigma_i) \tag{9}$$

where $\eta = \begin{bmatrix} \delta^T & \sigma^T \end{bmatrix}^T \in [\alpha, \beta]^m$, $\sigma = \sigma_i$, $i = 1, 2, \cdots, M$.

Remark 2: It is noted that the hybrid random vector described in *Definition 1* differ from either the so-called mixed type random variables (see Chapter 4.2 of [9]), or the ones studied in [8], where the measures of uncertainty for fuzzy variables and stochastic ones are compared and combined.

Definition 2: If after re-arranging the sequence of elements, a random vector $\widetilde{z} \in [\alpha, \beta]^m$ can be transferred into $z = \begin{bmatrix} z_1^T & z_2^T \end{bmatrix}^T \in [\alpha, \beta]^m$, where $z_1 \in [\alpha, \beta]^{m_1}$ is a continuous random sub-vector and $z_2 \in [\alpha, \beta]^{m_2}$ is a deterministic sub-vector with $m = m_1 + m_2$, then $\widetilde{z}$ or z is called a *system-output-type hybrid random vector (SOTH random vector)*. If two SOTH random vectors with the same dimensions have the same length of continuous random variables at same locations, then we say that they have the same structure.

Definition 3: For a multivariate continuous random vector $z_0 \in \Omega := [\alpha, \beta]^{m_1}$, the *joint entropy* is defined by

$$E_n(z_0) = \begin{cases} -\int_\Omega \gamma_{z_0}(\tau) \ln(\gamma_{z_0}(\tau))\, d\tau & \gamma_{z_0}(\tau) > 0 \\ 0 & \gamma_{z_0}(\tau) = 0 \end{cases} \tag{10}$$

Definition 4: For a hybrid random vector $z = \begin{bmatrix} z_1^T & z_2^T \end{bmatrix}^T \in [\alpha, \beta]^m$, where $z_1 \in \Omega := [\alpha, \beta]^{m_1}$ is a continuous random sub-vector and $z_2 \in [\alpha, \beta]^{m_2}$ is a discrete random sub-vector that takes finite values at $\{\sigma_1, \sigma_2, ..., \sigma_{m_2}\}$ with $m = m_1 + m_2$, the *hybrid entropy* is defined by

$$E_n(z) = \begin{cases} -\sum_{i=1}^{M} \int_\Omega \gamma_{z_1} \ln(\gamma_{z_1})\, d\tau & \gamma_{z_1} > 0 \\ 0 & \gamma_{z_1} = 0 \end{cases}, \tag{11}$$

where $\gamma_{z_1} := \gamma_{z_1}(\tau, z_2 = \sigma_i)$.

Remark 3: *Definition 3* is a natural generalization of entropy for a multivariate random variable z_0 (see [9]), while *Definition 4* is a generalization of *Definition 3* corresponding to hybrid PDFs, which will play an important role in the concerned multivariate filtering problem.

III. FORMULATION FOR THE ERROR PDFS

Under *Assumptions A.1, A.2* and *A.3*, from (5) it can be shown that

$$\widetilde{e}_{k+1} = v_k + s_k - \widetilde{U}_k \left(H(x_k) - H(\widehat{x}_k)\right) \tag{12}$$

where

$$s_k = \widetilde{G}_k w_{k+1}, v_k = \widetilde{A}_k \widetilde{e}_k \tag{13}$$

The third term $(y_k - H(\widehat{x}_k))$ can be measured on-line at sample time k. Under *Assumption A.3*, it can be guaranteed that at every sample time k, if $m > n(\geq r)$, $\widetilde{A}_k e_k$ and $\widetilde{G}_k w_{k+1}$ are two SOTH random vectors. The next task is to calculate the joint PDF of $\widetilde{e}_{k+1}$ in terms of those of $\widetilde{e}_k$ and w_{k+1}, where the PDF of s_k and v_k will be formulated firstly. To provide unified notations, we consider the mapping $\theta_{k+1} = D_k \pi_{k+1}$ denoted in (6), where $\pi_{k+1} \in [a, b]^n$ is a non-Gaussian continuous random vector with given joint PDF $\gamma_\pi(\cdot)$, and $D_k \in R^{m \times n}$ is a known matrix. The proofs are omitted to save space.

A. *Case I:* $m = n$, $rank(D_k) = m$

If D_k is an invertible matrix, it is relatively simple to determine $\gamma_{\theta_{k+1}}(\tau)$ based on $\gamma_{\pi_{k+1}}(\tau)$. This case corresponds to an invertible $\widetilde{G}_k$ in (13).

Lemma 1: If D_k is invertible, then the following relationship holds

$$\gamma_{\theta_{k+1}}(\tau) = \gamma_{\pi_{k+1}}\left(D_k^{-1}\tau\right)\left|\det D_k^{-1}\right| \tag{14}$$

B. *Case II:* $m < n$, $rank(D_k) = m$

In this case, we can assume that D_k is with a full row rank at sample time k, and its first m columns have full rank (through their re-arrangement). Thus, there exist a low-triangle invertible matrix T_1 and an upper-triangle invertible matrix T_2 such that

$$T_1 D_k T_2 = \begin{bmatrix} I_m & 0 \end{bmatrix} \tag{15}$$

where I_n represents an $m-$dimensional identical matrix and T_2 can be selected as

$$T_2^{-1} := \begin{bmatrix} T_{21} & T_{22} \\ 0 & T_{23} \end{bmatrix}$$

To simplify the presentation, we denote

$$\widetilde{\tau} := T_1 \tau = \begin{bmatrix} \widetilde{\tau}^{(1)} \\ \widetilde{\tau}^{(2)} \end{bmatrix}, \widetilde{\theta}_{k+1} := \begin{bmatrix} \widetilde{\theta}_{k+1}^{(1)} \\ \widetilde{\theta}_{k+1}^{(2)} \end{bmatrix} = T_1 \theta_{k+1}, \tag{16}$$

and

$$\widetilde{\pi}_{k+1} := T_2^{-1} \pi_{k+1} = \begin{bmatrix} \widetilde{\pi}_{k+1}^{(1)} \\ \widetilde{\pi}_{k+1}^{(2)} \end{bmatrix}$$

with compatible dimensions. This implies

$$\widetilde{\theta}_{k+1} = \begin{bmatrix} I_m & 0 \end{bmatrix} \widetilde{\pi}_{k+1} = \widetilde{\pi}_{k+1}^{(1)} \tag{17}$$

Lemma 2: If D_k is full row rank at its first n columns, then the following relationship holds

$$\gamma_{\theta_{k+1}}(\tau) = \left[\int_a^b \cdots \int_a^b \gamma_{\pi_{k+1}}(\eta) \left|\det T_1\right| \left|\det T_2\right| d\widetilde{\tau}^{(2)}\right] \tag{18}$$

with $\eta := T_2 \widetilde{\tau}^{(2)}$ and $\widetilde{\tau} := T_1 \tau$.

C. *Case III:* $m > n$, $rank(D_k) = n$

It is noted that $m > n$ occurs in many practical cases. Based on *Assumption A.3*, we only need consider the special case of $D_k = \begin{bmatrix} D_{1k}^T & 0 \end{bmatrix}^T$, where $D_{1k} \in R^{r \times n}$ $(r \leq n)$ is supposed to have full row rank.

Denote $\tau = \begin{bmatrix} \tau^{(1)} \\ \tau^{(2)} \end{bmatrix}$ and

$$\widetilde{\pi}_{k+1} := D_{1k} \pi_{k+1}, \theta_{k+1} := \begin{bmatrix} \theta_{k+1}^{(1)} \\ \theta_{k+1}^{(2)} \end{bmatrix} = \begin{bmatrix} \widetilde{\pi}_{k+1} \\ 0 \end{bmatrix}. \tag{19}$$

It is noted that different from *Subsection 3.2*, in this case $\tau \in [a, b]^m$, $\tau^{(1)} \in [a, b]^r$.

Lemma 3: If $D_k = \left[\begin{array}{cc} D_{1k}^T & 0 \end{array}\right]^T$and D_{1k} is full row rank, then the following relationship holds

$$\gamma_{\theta_{k+1}}(\tau) = \left\{ \begin{array}{ll} \gamma_{\widetilde{\pi}_{k+1}}\left(\tau^{(1)}\right) & \tau^{(2)} \succeq 0 \\ 0 & otherwise \end{array} \right. \quad (20)$$

where $\widetilde{\pi}_{k+1}$ is denoted by (19) and $\gamma_{\widetilde{\pi}_{k+1}}\left(\tau^{(1)}\right)$ can be computed in terms of $\gamma_{\pi_{k+1}}\left(\tau^{(1)}\right)$ based on *Lemma 2*.

Remark 4: $P\left\{\theta_{k+1}^{(1)} \preceq \tau^{(1)}, \theta_{k+1}^{(2)} = 0\right\}$ can be considered as the probability of a hybrid vector defined in *Section 2* with $M = 1$ for given τ. In this case, θ_{k+1} is exactly an SOTH random vector and *Lemma 3* implies that the hybrid entropy of θ_{k+1} is equal to the entropy of $\widetilde{\pi}_{k+1}$.

D. PDFs of the Sum of Two Hybrid Random Vectors

Based on $Lemmas$ $1 \sim 3$, we can formulate the PDFs of v_k and s_k in terms of the PDFs of $\widetilde{e}_k$ and w_{k+1} if $\widetilde{e}_k$ and w_{k+1} are both continuous random vectors with known PDFs. To discuss the algebraic sum operation v_k+s_k, where v_k and s_k are possible hybrid random vectors, *Definitions* $1 \sim 3$ and $Assumption\ A.3$ will be applied in the following formulation.

Initially we suppose $x(0)$ and $\widehat{x}(0)$ are given, then $\widetilde{e}_0$ can be regarded as a deterministic vector. In most cases, we can suppose $\widetilde{e}_0 = 0$ and $v_0 = 0$. Thus, (3) can be reduced to

$$\widetilde{e}_1 = s_0 = \widetilde{G}_0 w_1 = \left[\begin{array}{cc} G_{10}^T & 0 \end{array}\right]^T w_1, \gamma_{\widetilde{e}_1}(\tau) = \gamma_{s_0}(\tau) \quad (21)$$

where s_0 or $\widetilde{e}_1$ can be considered as SOTH random vectors and $\gamma_{\widetilde{e}_1}(\tau)$ or $\gamma_{s_0}(\tau)$ can be obtained by $Lemmas$ $1 \sim 3$. From

$$v_1 = \widetilde{A}_1 \widetilde{e}_1, \widetilde{A}_1 = \left[\begin{array}{cc} A_{11} & A_{21} \\ 0 & A_{31} \end{array}\right], \widetilde{e}_1 = \left[\begin{array}{c} G_{10} w_1 \\ 0 \end{array}\right]$$

it can be seen that v_1 is also an SOTH random vector with the same structure as s_1. At the sample time $k = 2$, (5) reduces to

$$\widetilde{e}_2 = (v_1 + s_1) - \widetilde{U}_1 \left((H(x_1) - H(\widehat{x}_1))\right) \quad (22)$$

Based on *Definition 1*, we can claim that $\gamma_{v_1}(\tau)$ or $\gamma_{s_1}(\tau)$ also can be represented by $\gamma_{\widetilde{e}_1}(\tau)$ or $\gamma_{w_1}(\tau)$, respectively, where v_1 and s_1 may be SOTH random vectors. At this stage, the task is to formulate the PDF of the sum of SOTH random vectors recursively. The following result can be given.

Lemma 4: Under *Assumptions A.1* $\sim$ *A.3*, at every sample time $k+1$ $(k = 1, 2, \cdots)$, the hybrid PDF of $\widetilde{e}_{k+1}(\tau | y_k, \widehat{y}_k, U_k)$ can be formulated recursively by

$$\int_\alpha^\beta \cdots \int_\alpha^\beta \gamma_{v_k}(\sigma)\, \gamma_{s_{k+1}}\left(\tau - \sigma + \widetilde{U}_k(y_k - \widehat{y}_k)\right) d\sigma \quad (23)$$

where $\gamma_{v_k}(\sigma)$ and $\gamma_{v_k}(\sigma)$ can be calculated using *Lemmas 1–3*.

IV. Minimum Entropy Filtering

A. Problem Formulation

Since $\widetilde{U}_k$ is an $m \times l$ matrix, in order to use conventional optimization techniques, we denote

$$\widetilde{U}_k = \left[\begin{array}{ccc} U_{k1}^T & \cdots & U_{km}^T \end{array}\right]^T, u_k = \left[\begin{array}{ccc} U_{k1}, & \cdots, & U_{km} \end{array}\right]^T \quad (24)$$

where $u_k \in R^{ml \times 1}$ is a stretched column vector and U_{ki} is the ith row vector of $\widetilde{U}_k$.

Based on *Definitions 1* $\sim$ 4, the (hybrid) entropy $E_n(\widetilde{e}_{k+1})$ of a (hybrid) random vector $\widetilde{e}_{k+1}$ with (hybrid) PDF $\gamma_{\widetilde{e}_{k+1}}(\tau)$ can be represented as $-\int_\alpha^\beta \cdots \int_\alpha^\beta \phi_{k+1}(\tau) d\tau$, where $\tau \in [\alpha, \beta]^n$ and

$$\phi_{k+1}(\tau) := R_1 \gamma_{\widetilde{e}_{k+1}}(\tau | y_k, \widehat{y}_k, u_k) \ln\left[\gamma_{\widetilde{e}_{k+1}}(\tau | y_k, \widehat{y}_k, u_k)\right].$$

Definition 5: If there exists a filter such that $E_n(e_k)$ is minimized for every sample time k, then it is called to be a *minimum entropy (ME) filter*.

To design the ME filter, the performance index J_N is desired, where

$$J_N = \sum_{k=0}^{N} \left[-\int_\alpha^\beta \cdots \int_\alpha^\beta R_1 \phi_k(\tau) d\tau + \frac{1}{2} u_k^T R_2 u_k \right] \quad (25)$$

and $R_1 > 0$ and $R_2 \geq 0$ are weighting matrices. In (25), the first term is the hybrid entropy of the estimation errors and the second term means that we want the elements of U_k to be small.

Remark 5: If $m > n$, based on (5), (13) and (23) it can be claimed that only the first n rows of $\widetilde{U}_k$ are related to $E_n(\widetilde{e}_{k+1})$ while the last $m - n$ rows are irrelevant to the hybrid entropy. It means that in this case, $U_{k,n+1}, \cdots, U_{k,m}$ in (24) are redundant vectors for the optimization, where $U_{k,n+1} = \cdots = U_{k,m} = 0$ can be selected under this circumstance. To simplify the presentation, we will still use notations (24) and (23) in the following design procedures.

B. Optimal Filter Design Strategy

Equation (25) can be rewritten as $(k = 0, 1, 2, \cdots, N, \cdots, +\infty)$

$$J_k = J_{k-1} + \left[\int_\alpha^\beta \cdots \int_\alpha^\beta \Psi(\tau, y_k, \widehat{y}_k, u_k) d\tau + \frac{1}{2} u_k^T R_2 u_k \right], \quad (26)$$

where $\Psi(\tau, y_k, \widehat{y}_k, u_k) = \Psi(\tau, u_k) = -R_1 \phi_k(\tau)$. In order to provide the filters with simple structure, the instantaneous cost function is considered for the design strategy. Based on

$$\frac{\partial \left[\int_\alpha^\beta \cdots \int_\alpha^\beta \Psi(\tau, y_k, \widehat{y}_k, u_k) d\tau + \frac{1}{2} u_k^T R_2 u_k \right]}{\partial u_k} = 0 \quad (27)$$

an explicit function for u_k will be determined in the following. To simplify the filter structure, it is denoted that

$$u_k = u_{k-1} + \Delta u_k, k = 1, 2, \cdots, N, \cdots, +\infty \quad (28)$$

which should be a function of y_k, $\widehat{y}_k$, u_k and τ. It can be approximated to give

$$\begin{aligned}\Psi(\tau, y_k, \widehat{y}_k, u_k) = \quad & h_{k0}(\tau) + h_{k1}(\tau)\Delta u_k \\ & + \tfrac{1}{2}\Delta u_k^T h_{k2}(\tau)\Delta u_k + o(\Delta u_k^T \Delta u_k)\end{aligned} \tag{29}$$

where

$$h_{k0}(\tau) = \left.\Psi(\tau, u_k)\right|_{u_k=u_{k-1}}, h_{k1}(\tau) = \left.\frac{\partial \Psi(\tau, u_k)}{\partial u_k}\right|_{u_k=u_{k-1}}, \tag{30}$$

$$h_{k2}(\tau) = \left.\frac{\partial^2 \Psi(\tau, u_k)}{\partial u_k^2}\right|_{u_k=u_{k-1}}.$$

Theorem 1: Under *Assumptions A.1$\sim$A.3*, the gain matrix of the ME filtering strategy for J_∞ subjected to nonlinear error model (5) is given by

$$\begin{aligned}\Delta u_k^* = \quad & -\left[\int_\alpha^\beta \cdots \int_\alpha^\beta h_{k2}(\tau)d\tau + R_2\right]^{-1} \\ & \left[\int_\alpha^\beta \cdots \int_\alpha^\beta h_{k1}(\tau)d\tau + R_2 u_{k-1}\right],\end{aligned} \tag{31}$$

for a weight matrix R_2 satisfying

$$\left(\int_\alpha^\beta \cdots \int_\alpha^\beta h_{k2}(\tau)d\tau + R_2\right) > 0 \tag{32}$$

Proof: From (28) we have

$$u_k^T R_2 u_k = u_{k-1}^T R_2 u_{k-1} + 2u_{k-1}^T R_2 \Delta u_k + \Delta u_k^T R_2 \Delta u_k \tag{33}$$

Substituting (29) and (33) into (27) yields recursive strategy (31) for all $k = 0, 1, 2, \cdots, N, \cdots, +\infty$, under (32). To guarantee its sufficiency, the following condition on the second-order derivative should also be satisfied

$$\frac{\partial^2 \left[\int_\alpha^\beta \cdots \int_\alpha^\beta \Psi(\tau, u_k)dx + \frac{1}{2}u_k^T R_2 u_k\right]}{\partial \Delta u_k^2} > 0$$

which is equivalent to (32), and holds if R_2 is selected sufficiently large. ■

The real-time suboptimal ME filtering algorithm can be summarized.

C. Optimal Stabilization Filtering Strategy

Stability of stochastic systems is used to focus on mean or variance of the output, which is also insufficient for non-Gaussian variables (see [7], [10] and references therein). In this subsection, an improved suboptimal filtering strategy is proposed, with which the error system can be guaranteed to be locally stable. In this context, system (1) is approximated to read

$$\widetilde{e}_{k+1} = \widetilde{A}_k \widetilde{e}_k + \widetilde{U}_k B_k \widetilde{e}_k + \widetilde{G}_k w_{k+1} \tag{34}$$

where $B_k := \left[\left.\frac{\partial H(\cdot)}{\partial x_k}\right|_{x_k=\widehat{x}_k}\right]$. Denote

$$\Delta e_k = \widetilde{e}_k - \widetilde{e}_{k-1}, \Delta U_k = \widetilde{U}_k - \widetilde{U}_{k-1}, \Delta w_k = w_k - w_{k-1}$$

then (34) can be further linearized to give

$$\Delta e_{k+1} = \widetilde{A}_k \Delta e_k + \widetilde{U}_{k-1} B_k \Delta e_k + \Delta U_k B_k \widetilde{e}_{k-1} + \widetilde{G}_k \Delta w_{k+1} \tag{35}$$

Different from (29), here Δu_k is a function of $\widetilde{e}_k$, as such we shall consider the following expansion

$$\begin{aligned}\Psi(\tau, u_k) \quad = \quad & \alpha_k + \alpha_{k1}\Delta u_k + \alpha_{k2}\Delta e_k \\ & + \frac{1}{2}\Delta u_k^T \delta_{k1}\Delta u_k + \Delta e_k^T \delta_{k2}\Delta u_k \\ & + \frac{1}{2}\Delta e_k^T \delta_{k3}\Delta e_k + o(\Delta u_k^2, \Delta e_k)\end{aligned} \tag{36}$$

To enhance the flexibility of the algorithm design, we replace the term $u_k^T R_2 u_k$ in cost function $J_N(\cdot)$ by a time varying term $u_k^T R_{2k} u_k$ here, where R_{2k} is tuned at k step. Substituting (36) into (27) and removing the higher order terms lead to

$$\Delta u_k = \Lambda_{k1}\Delta e_k + \Lambda_{k2} \tag{37}$$

Corresponding to (24), it can be verified that

$$\Delta U_k = \begin{bmatrix} \Delta e_k^T \Lambda_{k11}^T \\ \cdots \\ \Delta e_k^T \Lambda_{k1m}^T \end{bmatrix} + \begin{bmatrix} \Lambda_{k21}^T \\ \cdots \\ \Lambda_{k2m}^T \end{bmatrix}$$

In this equation $\Lambda_{kij} \in R^{l\times m}$ is denoted as the sub-matrix of Λ_{ki} which includes the $(mj+1)th \sim [m(j+1)]th$ columns of Λ_{ki}, $i = 1, 2,\ j = 1, 2, \cdots, l$. Since

$$\Delta e_k^T \Lambda_{k11}^T B_k \widetilde{e}_{k-1} = \widetilde{e}_{k-1}^T B_k \Lambda_{k11}\Delta e_k,$$

it can be obtained that

$$\begin{aligned}\Delta U_k B_k \widetilde{e}_{k-1} = \quad & \Theta_1(R_{2k}, u_{k-1}, \widetilde{e}_{k-1})\Delta e_k \\ & + \Theta_2(R_{2k}, u_{k-1}, \widetilde{e}_{k-1})\end{aligned} \tag{38}$$

where

$$\Theta_1(R_{2k}, u_{k-1}, \widetilde{e}_{k-1}) := \begin{bmatrix} \widetilde{e}_{k-1}^T B_k \Lambda_{k11} \\ \cdots \\ \widetilde{e}_{k-1}^T B_k \Lambda_{k1m} \end{bmatrix},$$

$$\Theta_2(R_{2k}, u_{k-1}, \widetilde{e}_{k-1}) := \begin{bmatrix} \Lambda_{k21} B_k \widetilde{e}_{k-1} \\ \cdots \\ \Lambda_{k2m} B_k \widetilde{e}_{k-1} \end{bmatrix}$$

Theorem 2: Under *Assumptions A.1$\sim$A.3*, if there exits $R_{2k} > 0$ such that the following inequality

$$\left(\int_\alpha^\beta \cdots \int_\alpha^\beta \delta_{k1}d\tau + R_{2k}\right) > 0$$

holds and the matrix

$$\Xi(k) = A_k + \widetilde{U}_{k-1}B_k + \Theta_1(R_{2k}, u_{k-1}, \widetilde{e}_{k-1})$$

is Schur stable at each sample time k, then the ME filtering strategy for J_N subjected to nonlinear error model (5) is given by (31), with which the local stability of the estimation error system can be guaranteed.

Proof: is omitted for simplicity.

V. SIMULATION RESULTS

To demonstrate the obtained filtering algorithm, we consider the simple model described by

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+2} \end{bmatrix} = \begin{bmatrix} 0.7 & 0.3 \\ 0 & a_0(k) \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + \begin{bmatrix} 1 \\ b_0(k) \end{bmatrix} w_{k+1} \tag{39}$$

with the measurement equation $y_k = x_{1,k} + x_{1,k}x_{2,k}$, where $a_0(k) = 0.7 + 0.03\arctan(1+k)^{-1}$ and $b_0(k) = 0.2((k+1)^{-1/2}+1)$. Random variables w_k $(k = 0, 1, 2, \cdots)$ are assumed to be mutually independent, and their PDFs are defined by

$$\gamma_w(x) = \begin{cases} -\frac{3000}{4}(x^2 - 0.01), & x \in [-0.1, 0.1] \\ 0, & x \in (-\infty, -0.1) \cup (0.1, +\infty) \end{cases}$$

which is a triangular Kernel function. The filter can be conducted by the following format

$$\begin{bmatrix} \widehat{x}_{1,k+1} \\ \widehat{x}_{2,k+2} \end{bmatrix} = \begin{bmatrix} 0.7 & 0.3 \\ 0 & a_0(k) \end{bmatrix} \begin{bmatrix} \widehat{x}_{1,k} \\ \widehat{x}_{2,k} \end{bmatrix} + (y_k - \widehat{y}_k) \begin{bmatrix} u_k \\ 0 \end{bmatrix} \tag{40}$$

where $\widehat{y}_k = \widehat{x}_{1,k} + \widehat{x}_{1,k}\widehat{x}_{2,k}$.

At first we can transform the error system to the form of (5). Without further confusions, we also use e_k instead of $\widetilde{e}_k$ in the following. Denote $e_k := \begin{bmatrix} e_k^{(1)} & e_k^{(2)} \end{bmatrix}^T$, then for sample value $\tau := \begin{bmatrix} \tau^{(1)} & \tau^{(2)} \end{bmatrix}^T$, based on $Lemmas$ $1 \sim 4$ it can be shown that

$$\begin{aligned} \gamma_{e_{k+1}}(\tau^{(1)}) = \; & \int_\alpha^\beta \frac{10}{7}\gamma_{e_k^{(1)}}\left(\frac{10}{7}\sigma^{(1)}\right) \times \\ & \gamma_{w_{k+1}}\left(\tau^{(1)} - \sigma^{(1)} + u_k(y_k - \widehat{y}_k)\right) d\sigma^{(1)} \end{aligned}$$

To compute the required filter gain, according to (30) and (31), we have

$$h_{k1}(\tau) = -\left[\frac{\partial \gamma_{e_{k+1}^{(1)}}(\tau^{(1)})}{\partial u_k}\left(\ln \gamma_{e_{k+1}^{(1)}}(\tau^{(1)}) + 1\right)\right]_{u_k = u_{k-1}}$$

and

$$\begin{aligned} h_{k2}(\tau) = \; & -\left[\frac{\partial^2 \gamma_{e_{k+1}^{(1)}}(\tau^{(1)})}{\partial u_k^2}\left(\ln \gamma_{e_{k+1}^{(1)}}(\tau^{(1)}) + 1\right)\right]_{u_k = u_{k-1}} \\ & + \left[\left(\frac{\partial \gamma_{e_{k+1}^{(1)}}(\tau^{(1)})}{\partial u_k}\right)^2 \left(\gamma_{e_{k+1}^{(1)}}(\tau^{(1)})\right)^{-1}\right]_{u_k = u_{k-1}} \end{aligned}$$

with which (31) can be used to provide suboptimal filtering laws. In the simulation, it has been selected that $R_1 = 1$ and $R_2 = 100$. Figures can be provided to demonstrate the dynamical responses of estimation errors.

VI. CONCLUSIONS

In this paper, a new solution is presented for the optimal filtering design of multivariate stochastic systems subjected to non-Gaussian noises. To effectively characterize the stochastic property of the system output, the concepts of hybrid random vectors, hybrid random probabilities and hybrid entropies are introduced. The relationships between the PDFs of multivariate stochastic input and output are firstly established, with which the PDFs and its hybrid entropy of the estimation errors are represented in terms of known information including the measurement output and the PDFs of the stochastic input. Using the formulations for the error PDFs and the minimum entropy performance index, we established an optimal algorithm recursively for the filter gain such that the hybrid entropy of the estimation errors is minimized. Furthermore, an improved method is provided to guarantee the local stability by tuning the weighting matrices.

VII. ACKNOWLEDGMENT

This work is jointly supported by *The Leverhulme Trust, UK under Grant No. F00038/D* and *National Science Foundation of China under Grant No. 60472065 and 60474050*. These are gratefully acknowledged.

REFERENCES

[1] A. Alessandri, M. Baglietto, G. Battistelli, and T. Parisini. New convergence conditions for receding-horizon state estimation of nonlinear discrete time systems,. In *Proceedings of 43rd IEEE CDC, 2004*, Atlantis, Paradise Island, Bahamas, 2004.

[2] Y. Bar-Shalom and X. R. Li. *Estimation and tracking: principles, techniques, and software*. Artech House, Norwood, MA, 1996.

[3] F. Carravetta, A. Germani, and M. Raimondi. Polynomial filtering for linear discrete time non-gaussian systems. *SIAM J. Control Optim.*, 34:1666–1690, 1996.

[4] C. D. Charalambous and R. J. Elliott. Information states in stochastic control and filtering: a lie algebraic theoretic approach. *IEEE Trans. on Automatic Control*, 45:653–674, 2000.

[5] X. B. Feng, K. A. Loparo, and Y. Fang. Optimal state estimation for stochastic systems: An information theoretic approach. *IEEE Trans. on Automatic Control*, 42:771–786, 1997.

[6] G. C. Goodwin and K. S. Sin. *Adaptive filtering, Prediction and control*. Prentice-Hall, Englewood Cliffs, NJ, 1984.

[7] X. Mao. Exponential stability of stochastic delay interval systems with Markovian switching. *IEEE Trans. on Automatic Control*, 47:1604–1612, 2002.

[8] N. R. Pal and S. K. Pal. Entropy, a new definition and its applications. *IEEE Trans. on Systems, Man and Cybernetics*, 21:1260–1270, 1991.

[9] A. Papoulis. *Probability, Random variables and stochastic processes, 3rd*. McGraw-Hill, New York, USA, 1991.

[10] K. Reif, S. Gunther, E. Yaz, and R. Unbehauen. Stochastic stability of the discrete-time extended Kalman filter. *IEEE Trans. on Automatic Control*, 44:714–728, 1999.

[11] B. W. Silverman. *Density Estimation for statistics and data analysis*. Chapman and Hall, London, 1986.

[12] V. A. Ugrinovskii and I. R. Petersen. Robust filtering of stochastic uncertain systems on an infinite time horizon. *Int. J. Control*, 75:614–626, 2002.

[13] F. Wang and V. Balakrishnan. Robust Kalman filters for linear time-varying systems with stochastic parametric uncertainties. *IEEE Trans. on Signal Processing*, 50:803–813, 2002.

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

[15] H. Wang. Minimum entropy control of non-Gaussian dynamic stochastic systems. *IEEE Trans. on Automatic Control*, 47:398–403, 2002.

[16] Q. Zhang. Hybrid filtering for linear systems with non-Gaussian disturbances. *IEEE Trans. on Automatic Control*, 45:50–61, 2000.