ILC-Based Minimum Entropy Filter Design and Implementation for Non-Gaussian Stochastic Systems

Puya Afshar, *Member, IEEE*, Fuwen Yang, *Senior Member, IEEE*, and Hong Wang, *Senior Member, IEEE*

Abstract—A new filtering approach based on the idea of iterative learning control (ILC) is proposed for linear and non-Gaussian stochastic systems. The objective of filtering is to estimate the states of linear systems with non-Gaussian random disturbances so that the entropy of output error is made to monotonically decrease along the progress of *batches* of process operation. The term *Batch* is referred to a period of time when the process repeats itself. During a batch, the filter gain is kept fixed and state estimation is performed. Between any two adjacent batches, the filter gain is updated so that the entropy of closed-loop output error is reduced for the next batch. Analysis is carried out to explicitly determine the learning rates which lead to convergence of the overall algorithm. Experiments have been implemented on a laboratory-based process test rig to demonstrate the effectiveness of proposed filtering method.

Index Terms—Iterative learning control (ILC), minimum entropy filtering, non-Gaussian linear systems, process control rig.

I. INTRODUCTION

STATE estimation is an important aspect of many research areas where a knowledge about the internal dynamical behavior of the system is required while it is not possible to directly measure them. The main purpose of estimating the states in control engineering is to perform an accurate state-feedback control when at least one of the system states are not measurable [1]. Perhaps the most significant progress in the theory of state estimation was made when Kalman filtering was introduced in 1960's, where the variance of estimation error is minimized provided that the system is known, linear and is subjected to Gaussian white noises [2]. Since then, the idea of filtering has been widely extended and applied in many engineering and information systems, such as signal processing, communication, and control engineering [3]. Different filtering ideas are different in terms of the assumptions made for the system dynamics (e.g., linear or nonlinear), noise structure (e.g., additive, multiplicative), or the methods used to solve the filtering problem. For instance, the idea of robust filtering has been developed for linear time varying systems [4], or for uncertain systems [5],[6]. Extended Kalman filters (EKF) were also developed to deal with state estimation in systems with complex nonlinearities [7]–[11], or systems with multiplicative noises [12]. Apart from the above basic assumptions made for the system or noise structure and also the methods used to solve the filtering problem (e.g., game-theoretic methods [6], linear matrix inequality (LMI) [13], [14], and the Riccati equation approach [15]), all above mentioned works share the assumption on the statistical characteristics of the noise. That is, regardless of the noise model, they assume Gaussian noises corrupting the system dynamics. This leads to a probabilistic state estimation by use of mean and variance to describe the distribution of the states. However, there are the following two main issues that limit the application of mean and variance as the only statistical properties to describe the state distribution.

- It is known that even for a nonlinear system with Gaussian inputs, the system output can be a non-Gaussian variable with a multiple-peak and asymmetric probability density function (PDF) (see, e.g., [16] and [17]).
- If the stochastic inputs are non-Gaussian, the above case holds and state distributions cannot be precisely characterized by using only mean and variances [18]. In these cases, the general measure PDF must be considered.

The case of non-Gaussian noises has led to the development of stochastic distribution control (SDC) based on which a number of applications, namely PDF tracking control [18]–[20], fault diagnosis of general stochastic systems [21], [22] have been developed. Also, as a general measure of uncertainty, entropy (which is originated in information theory and communication) has been applied to the control of non-Gaussian stochastic systems [4], [23].

The filtering problem has also been solved for non-Gaussian stochastic systems. For instance, minimum entropy filtering is studied where an instantaneous cost function of the hybrid entropy for the stochastic estimation error is concerned and an LMI approach is used to determine the filter gains [24]. Also, a robust filter with error squared constraint has been developed where the estimation is performed based on the current observation and known deterministic input signals. A time-varying LMI approach is used to derive an upper bound of the state estimation error square and optimal bound is obtained by solving a convex optimization problem via semi-definite programming (SDP) approach [25].
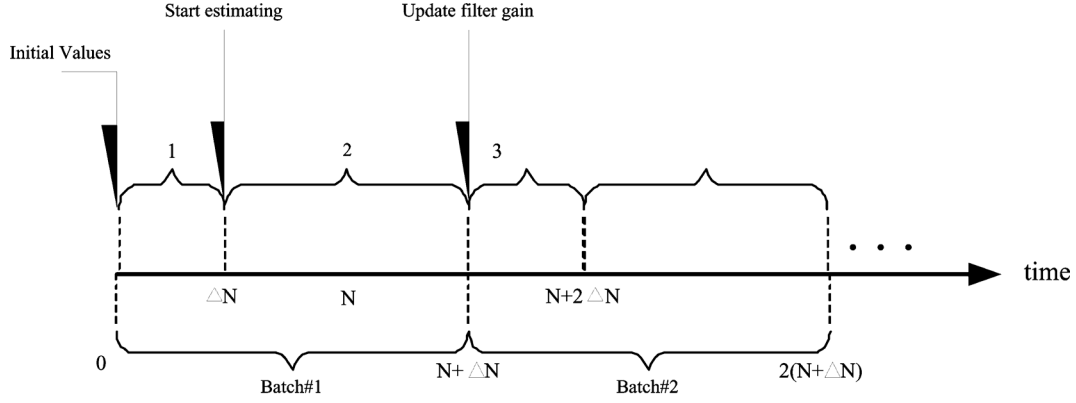
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

2 IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY

Stage 1: Choosing an initial filter gain and a set of system states

Stage 2: Estimating the system states using given filter gain

Stage 3: Updating the filter gain for less error and entropy of estimation error

Fig. 1. ILC-based filter design scheme.

The above mentioned works have provided the preliminary results for designing the filters for non-Gaussian stochastic systems. The idea of iterative learning has been already applied to SDC for non-Gaussian systems [26]–[30] and also to minimum entropy control of non-Gaussian systems [28],[31]–[33]. In this paper, a novel filtering method for linear, non-Gaussian stochastic systems is proposed which is based on the idea of iterative learning control (ILC). The developed technique is also applied to a laboratory-based process control test rig. In addition, necessary conditions for convergence of the algorithm have been developed. Compared to previous works on filtering for non-Gaussian systems, it is shown in this paper that the ILC-based filter design can be solved and applied to practical systems efficiently. This paper is organized as follows. In Section II, the problem of ILC-based minimum entropy filtering is defined. Section III presents the ILC-based solution for filtering problem and also provides a set of necessary conditions for the convergence of the algorithm. In Section IV, the experimental results from a process control test rig are proposed to demonstrate the effectiveness of method. Concluding remarks are made in Section V.

II. MINIMUM ENTROPY FILTERING PROBLEM

The objective, as mentioned in Section I, is to employ the idea of ILC to state estimation problem so that the entropy of closed loop output error becomes monotonically decreasing along with the progress of batches. To achieve this, the estimation horizon is divided into K identical intervals (batches) specified as $[(k-1)(N+\Delta N), k(N+\Delta N)](k = 1, 2, \ldots, K)$, where N is considered as the batch length, during which the filter gains are fixed and ΔN is the time period known as *between adjacent batches*. Within each batch, i.e., $[(k-1)(N+\Delta N), (k-1)N+k\Delta N]$, a fixed-gain filter is used to perform the state estimation. Between adjacent batches, i.e., during the time specified by $[(k-1)N+k\Delta N, k(N+\Delta N)]$ the filter gains are updated so that the entropy of the output error takes a smaller value within the following batch. During ΔN while the optimization takes place,

the filter gains remain fixed on their previous values. This enables tuning to be focused on the filter gain. The above design approach is illustrated in Fig. 1.

Consider a linear system with n_s states, n_u inputs, n_y outputs, and the following state-space model in which each output is defined as $y_p \in [a_p, b_p]$, $p = 1, \ldots, n_y$:

$$x_k(i+1) = Ax_k(i) + Bu_k(i) + B_w w_k(i) \quad (1)$$

$$y_k(i) = Cx_k(i) + Du_k(i) + v_k(i) \quad (2)$$

where $x_k(i) \in \mathbb{R}^{n_s}$, $u_k(i) \in \mathbb{R}^{n_u}$ are the state vector and the control signal measured at the ith time sample within the kth batch, respectively. In addition, $w_k(i)$ and $v_k(i)$ are non-Gaussian, yet bounded scalar additive noises. The objective is to design a filter to estimate the system states by minimizing the entropy of output error. Firstly, the filter dynamics is expressed as follows:

$$e_k(i) = y_k(i) - \hat{y}_k(i) \quad (3)$$

$$\hat{x}_k(i+1) = A\hat{x}_k(i) + Bu_k(i) + L_k e_k(i) \quad (4)$$

$$\hat{y}_k(i) = C\hat{x}_k(i) + Du_k(i) \quad (5)$$

where $L_k \in \mathbb{R}^{n_s \times n_y}$ is the filter gain to be designed so that the entropy of output error is minimized batch by batch. For this purpose, the α-order Renyi's entropy is chosen. Based on Renyi's entropy measure, the entropy of the output error can be written as follows:

$$H(e_k) = \frac{1}{1-\alpha} \ln \sum_{i=1}^{N} (\gamma^\alpha(e_k(i))) \quad (6)$$

where $\gamma(e_k(i))$ is the probability density function (PDF) of the output error, mathematically defined as

$$P(a \le e_k(i) < \beta, L_k) = \int_a^\beta \gamma(e_k(i)) de \quad (7)$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

AFSHAR *et al.*: ILC-BASED MINIMUM ENTROPY FILTER DESIGN AND IMPLEMENTATION FOR NON-GAUSSIAN STOCHASTIC SYSTEMS 3
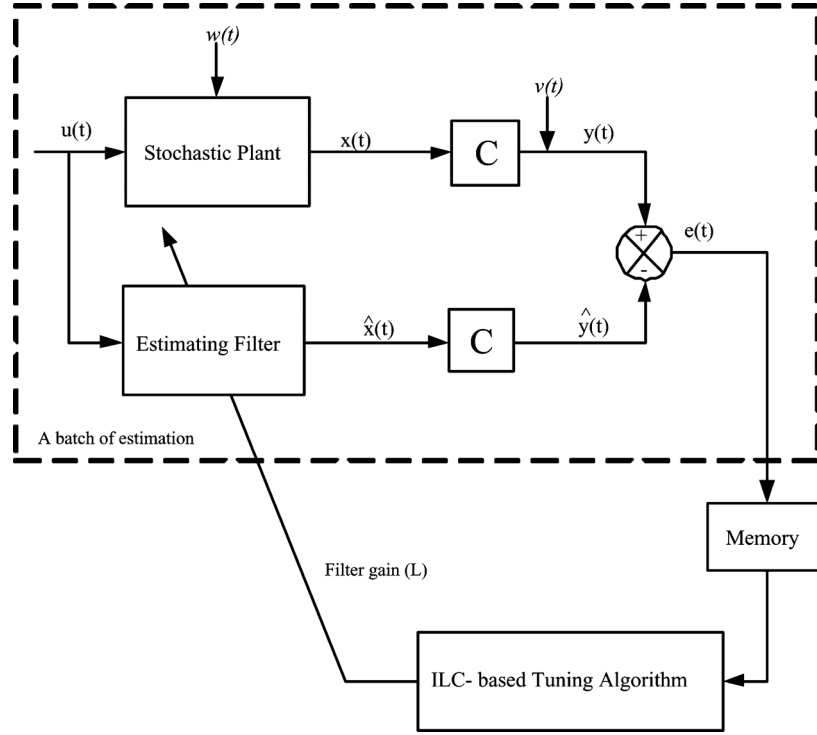

Fig. 2. Overall scheme of the minimum entropy filter design.

where $P(a \leq e_k(i) < \beta, L_k)$ represents the probability of output error $e_k(i)$ lying in interval $[a, \beta)$ when filtering gain L_k is applied. The PDF is estimated by kernel density estimation (KDE) denoted by

$$\gamma(e_k) \approx \hat{\gamma}(e_k) = \frac{1}{Nh} \sum_{j=1}^{N} K_\sigma \left(\frac{e_k - e_k(j)}{h} \right) \qquad (8)$$

where K_σ is a Gaussian Kernel function and h is a bin width [34]. This way, the entropy of output error can be rewritten as follows:

$$H(e_k) = \frac{1}{1-\alpha} \ln \left(\frac{1}{(Nh)^\alpha} \sum_{i=1}^{N} \left[\sum_{j=1}^{N} K_\sigma \left(\frac{e_k(i) - e_k(j)}{h} \right) \right]^\alpha \right) \qquad (9)$$

It is desired to determine the filter gain L_k in (4) so that entropy calculated in (9) is minimized batch by batch. The key idea to ILC-based estimation is to divide the estimation time (i.e., *batches*). Within each batch, the filter gain is kept fixed and the estimation is carried out. Then the filter gain is updated between batches so that output error entropy takes less value within the following batch. This way, the filtering problem can be solved in a time-space domain where the estimation performance is improved batch-by-batch. Fig. 2 illustrates the proposed idea.

The proposed technique is fundamentally different and hence non-comparable to commonly-used filtering algorithms such as Kalman filtering. The reason for this is that the present algorithm assumes presence of non-Gaussian noises which prevent characterizing the output error by its mean/variance due to their unsymmetrical PDF shapes. The proposed filter estimates the unmeasurable system states by batch-by-batch reduction of the output error uncertainty (or randomness). However, since the noises are non-Gaussian, the output error's uncertainty is characterized by its entropy which directly depends on the overall probability density shape of the measurable output error. Indeed, if the noises are assumed Gaussian, the entropy is collapsed to output error's variance (i.e., the common formulation of the uncertainty) to which well-established techniques such as Kalman filtering could be effectively applied. In other words, the proposed algorithm can be considered as a generalization to standard filtering techniques when noises are non-Gaussian.

It is also worthwhile to mention that the concept of minimum-entropy estimation proposed in this paper is essentially different than minimum-entropy control/estimation approaches commonly used in information theory literature such as [35] and [36]. To be more specific, the original definition of the entropy used in H_∞ control for linear time invariant systems at $s = \infty$ is defined in terms of the closed loop transfer function as follows:

$$I_d(G; l; \infty) = -\frac{l^2}{2\pi} \int_{-\infty}^{+\infty} \ln | \det(I - l^{-2} G^*(j\omega) G(j\omega)) | d\omega. \qquad (10)$$

In the above definition, the transfer function G is supposed to be strictly proper satisfying $\|G\|_\infty < l$, G^* denotes that the adjoint of G and I_d is finite. This interpretation of entropy measure associates to the mutual information maximization, as the entropy refers to the average information content of a signal in the context of information theory [35], [36]. As opposed to the interpretation based on information theory, in this paper the entropy is studied from a probabilistic point of view and is used to quantify the uncertainty and/or randomness of a general stochastic control system. Therefore, the knowledge base and the view point

of the filtering method proposed in this paper is essentially different than other entropy-based filtering methods based on mutual information principle. In Section III, an ILC-based tuning law is proposed to solve the introduced filtering problem.

III. ILC-BASED SOLUTION AND CONVERGENCE ANALYSIS

In this section, the ILC-based solution to the problem introduced in Section II is proposed. Here, a gradient descent algorithm is used between batches.

A. ILC-Based Filter Tuning

Considering the kth batch's entropy as cost function (i.e., $J_k = H(e_k)$), the following gain update law can be considered:

$$L_{k+1} = L_k - \mu \frac{\partial J_k}{\partial L_k} \tag{11}$$

where L_k is the filter gain within the kth batch and μ is a pre-specified learning rate. Using the well known chain rule, the gradient term can be further formulated as follows:

$$\frac{\partial J_k}{\partial L_k} = \frac{\alpha}{h(1-\alpha)} \frac{1}{\sum_{i=1}^{N} \left[\sum_{j=1}^{N} K_\sigma \left(\frac{e_k(i) - e_k(j)}{h} \right) \right]^\alpha}$$
$$\times \sum_{i=1}^{N} \left\{ \left[\sum_{j=1}^{N} K_\sigma \left(\frac{e_k(i) - e_k(j)}{h} \right) \right]^{\alpha-1} \right.$$
$$\times \left. \left[\sum_{j=1}^{N} \dot{K}_\sigma \left(\frac{e_k(i) - e_k(j)}{h} \right) \left(\frac{\partial e_k(i)}{\partial L_k} - \frac{\partial e_k(j)}{\partial L_k} \right) \right] \right\}. \tag{12}$$

Using (3)–(5), the partial differential term in (12) can be written as follows:

$$\frac{\partial e_k(i)}{\partial L_k} = -C \frac{\partial \hat{x}_k(i)}{\partial L_k}. \tag{13}$$

By employing the dynamic state evolution equation of the filter (4) and pre-multiplying it by C, the following forward differential equation can be determined:

$$C \frac{\partial \hat{x}_k(i)}{\partial L_k}$$
$$= C \left\{ A \frac{\partial \hat{x}_k(i-1)}{\partial L_k} + e_k(i-1) + L_k \frac{\partial e_k(i-1)}{\partial L_k} \right\}$$
$$= C \left\{ (A - L_k C) \frac{\partial \hat{x}_k(i-1)}{\partial L_k} + e_k(i-1) \right\}$$
$$= C \left\{ A \left[(A - L_k C) \frac{\partial \hat{x}_k(i-2)}{\partial L_k} + e_k(i-2) \right] + e_k(i-1) \right\}. \tag{14}$$

By continuing the substitution of differential terms with their equivalents in previous samples, the following dynamics can be achieved from (14), shown in (15) at the bottom of the page.

Therefore, the filter gain is updated by ILC-based laws (11), (12), and (15).

B. Analysis on Convergence

As the filter gain tuning rule incorporates the learning rate μ, analysis should be carried on to ensure the convergence of ILC-based tuning laws by a proper choice of the learning rate. The result of such analysis should give a range of acceptable learning rates which result in decrease of the squared sum of output error in a batch-by-batch manner. This means that one needs to find out a range for μ-s so that the convergence condition is satisfied. Under known errors $e_{k-1}(i)$, the following conditional expectation needs to be satisfied for convergence:

$$J_{\text{ILC}} = \mathbb{E}\left[e_k^T(i) e_k(i) | e_{k-1}(i) \right] - e_{k-1}^T(i) e_{k-1}(i) < 0. \tag{16}$$

Performing the analysis after the end of $(k-1)$th batch would mean that error signals are known and would have no effect on the analysis. As such, here the analysis is focused on the values corresponding to the kth batch. Using the filter error dynamics, the following can be expressed:

$$e_k^T(i) e_k(i) = [y_k^T(i) - \hat{x}_k^T(i)C^T - u_k^T(i)D^T]$$
$$\times [y_k(i) - C\hat{x}_k(i) - Du_k(i)]$$
$$= y_k^T(i)y_k(i) - y_k^T(i)C\hat{x}_k(i) - y_k^T(i)Du_k(i)$$
$$- \hat{x}_k^T(i)C^T y_k(i)$$
$$+ \hat{x}_k^T(i)C^T C\hat{x}_k(i) + \hat{x}_k^T(i)C^T Du_k(i)$$
$$- u_k^T(i)D^T y_k(i) + u_k^T(i)D^T C\hat{x}_k(i)$$
$$+ u_k^T(i)D^T Du_k(i). \tag{17}$$

In a similar way, it can be formulated that

$$e_{k-1}^T(i)e_{k-1}(i) = [y_{k-1}^T(i) - \hat{x}_{k-1}^T(i)C^T - u_{k-1}^T(i)D^T]$$
$$\times [y_{k-1}(i) - C\hat{x}_{k-1}(i) - Du_{k-1}(i)]. \tag{18}$$

Also, from the state evolution equation of the filter, the following dynamics can be written:

$$\hat{x}_k(i) = A\hat{x}_k(i-1) + Bu_k(i-1) + L_k e_k(i-1). \tag{19}$$

Furthermore, from the ILC-based update law, it can be seen that

$$L_k = L_{k-1} - \mu \frac{\partial J_{k-1}}{\partial L_{k-1}}. \tag{20}$$

Substituting (19) and (20) into (17), and using (16) and (18) can be written as shown in (21) at the bottom of the next page.

$$-C\frac{\partial \hat{x}_k(i)}{\partial L_k} = -C\left[(A - L_k C)^{n-1} e_k(1) + (A - L_k C)^{n-2} e_k(2) + \cdots + (A - L_k C)e_k(i-2) + e_k(i-1) \right] \tag{15}$$

It can be shown that in order for the ILC-based tuning convergence condition to hold, the following second order inequality with respect to learning rate μ must be satisfied

$$a\mu^2 - 2b\mu + c < 0. \tag{22}$$

In (22), the constant parameters a, b, and c are defined as follows:

$$a = e_k^T(i-1)\frac{\partial J_{k-1}^T}{\partial L_{k-1}^T}C^T C \frac{\partial J_{k-1}}{\partial L_{k-1}}e_k(i-1) \tag{23}$$

and

$$\begin{aligned}
b = & -y_k^T(i)C\frac{\partial J_{k-1}}{\partial L_{k-1}}e_k(i-1) \\
& + \hat{x}_k^T(i-1)A^T C^T C\frac{\partial J_{k-1}}{\partial L_{k-1}}e_k(i-1) \\
& + u_k^T(i-1)B^T C^T C\frac{\partial J_{k-1}}{\partial L_{k-1}}e_k(i-1) \\
& + e_k^T(i-1)L_{k-1}^T C^T C\frac{\partial J_{k-1}}{\partial L_{k-1}}e_k(i-1) \\
& + e_k^T(i-1)\frac{\partial J_{k-1}^T}{\partial L_{k-1}}C^T Du_k(i) \tag{24}
\end{aligned}$$

and also

$$\begin{aligned}
c = & y_k^T(i)y_k(i) - 2y_k^T(i)CA\hat{x}_k(i-1) \\
& - 2y_k^T(i)CBu_k(i-1) \\
& - 2y_k^T(i)CL_{k-1}e_k(i-1) \\
& + \hat{x}_k^T(i-1)A^T C^T CA\hat{x}_k(i-1) \\
& + 2\hat{x}_k^T(i-1)A^T C^T CBu_k(i-1) \\
& + 2\hat{x}_k^T(i-1)A^T C^T CL_{k-1}e_k(i-1) \\
& + 2u_k^T(i-1)B^T C^T CL_{k-1}e_k(i-1) \\
& + u_k^T(i-1)B^T C^T CBu_k(i-1) \\
& + e_k^T(i-1)L_{k-1}C^T CL_{k-1}e_k(i-1) \\
& - 2y_k^T(i)Du_k(i) \\
& + 2\hat{x}_k^T(i-1)A^T C^T Du_k(i) \\
& + 2u_k^T(i-1)B^T C^T Du_k(i) \\
& + 2e_k^T(i-1)L_{k-1}^T C^T Du_k(i) \\
& + u_k^T(i)D^T Du_k(i) - e_{k-1}^T(i)e_{k-1}(i). \tag{25}
\end{aligned}$$

Solving (22) for μ, the following values are determined:

$$\mu_1 = \frac{-b + \sqrt{b^2 - ac}}{a} \tag{26}$$

$$\mu_2 = \frac{-b - \sqrt{b^2 - ac}}{a}. \tag{27}$$

Therefore, in order to satisfy the convergence condition at the end of each batch, the learning rate should be selected so that

$$\mu_1 < \mu < \mu_2. \tag{28}$$

In Section IV, the experimental results from the implementation of proposed method on a laboratory-based process is proposed.

IV. EXPERIMENTAL RESULTS

In this section, the details of the experimental results performed on the laboratory-based process control test rig located at the University of Manchester are presented. The objective is to design an ILC-based filter for online estimation of the process control unit (PCU) so that the entropy of the output error is minimized batch-by batch. Therefore, it is required to formulate the filtering and estimation problem. Also, it is required to determine the dynamical model of the considered system to provide the filer with the approximated system matrices. First, a brief overview of the process is given as follows.

A. Process Description

The overview PCU is shown in Fig. 3. In the designed experiment, at each batch of the process, the water is pumped out of the sump and depending on its temperature it is diverted either through the cooling system or the main pipe by the manual/solenoidal diverting valve. The diverting valve causes the measurement noise $v_k(i)$. The flow of the water is measured by the turbine flow meter before being sent to the main processing tank where it is heated, stirred, and sent back to the sump by either manual or the solenoidal drainage valve which impose the process noise $w_k(i)$. The tank level is considered as the unmeasurable system state as the system is not equipped with a level meter.

The main control loops include the water flow and the processing tank temperature. However, for simplicity, only the water flow control loop is considered in this paper. The control

$$\begin{aligned}
J_{\text{ILC}} = \mathbb{E}\Bigg\{ & y_k^T(i)y_k(i) - 2y_k^T(i)C\left[A\hat{x}_k(i-1) + Bu_k(i-1) + \left(L_{k-1} - \mu\frac{\partial J_{k-1}}{\partial L_{k-1}}\right)e_k(i-1)\right] \\
& - 2y_k^T(i)Du_k(i) + \left[\hat{x}_k^T(i-1)A^T + u_k^T(i-1)B^T + e_k^T(i-1)\left(L_{k-1}^T - \mu\frac{\partial J_{k-1}^T}{\partial L_{k-1}}\right)\right]C^T C \\
& \times \left[A\hat{x}_k(i-1) + Bu_k(i-1) + \left(L_{k-1} - \mu\frac{\partial J_{k-1}}{\partial L_{k-1}}\right)e_k(i-1)\right] + 2u_k^T(i)D^T C \\
& \left[A\hat{x}_k(i-1) + Bu_k(i-1) + \left(L_{k-1} - \mu\frac{\partial J_{k-1}}{\partial L_{k-1}}\right)e_k(i-1)\right] + u_k^T(i)D^T Du_k(i)\Bigg\} - e_{k-1}^T(i)e_{k-1}(i) < 0 \tag{21}
\end{aligned}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

6   IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY

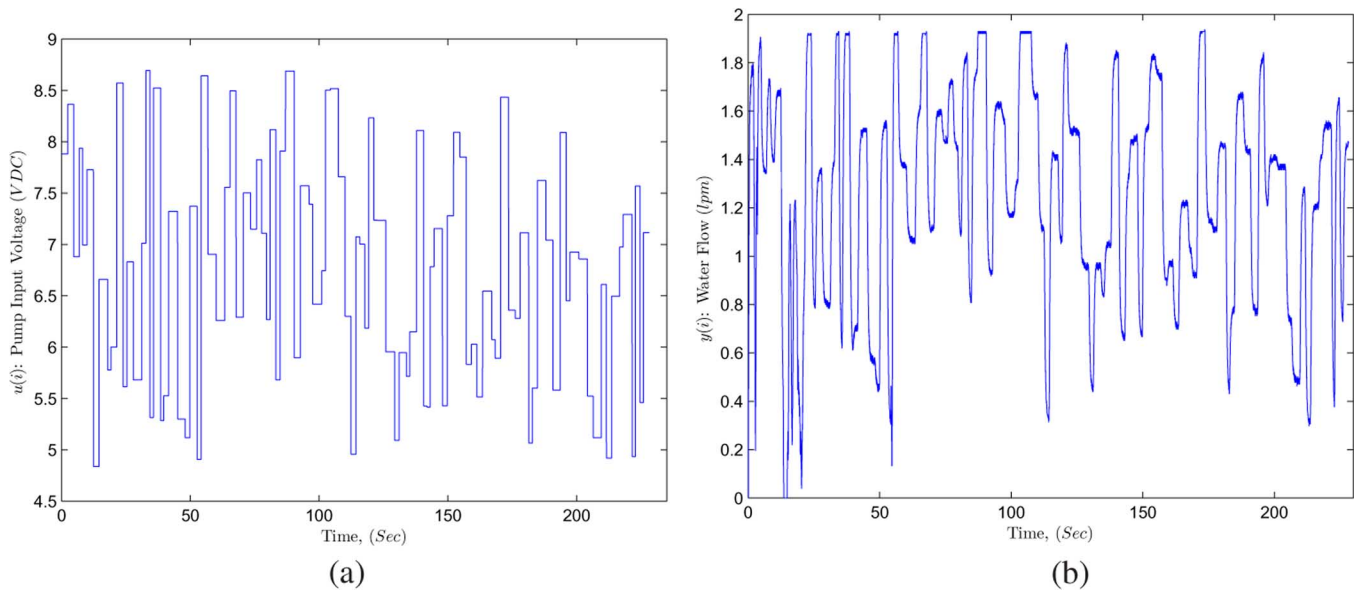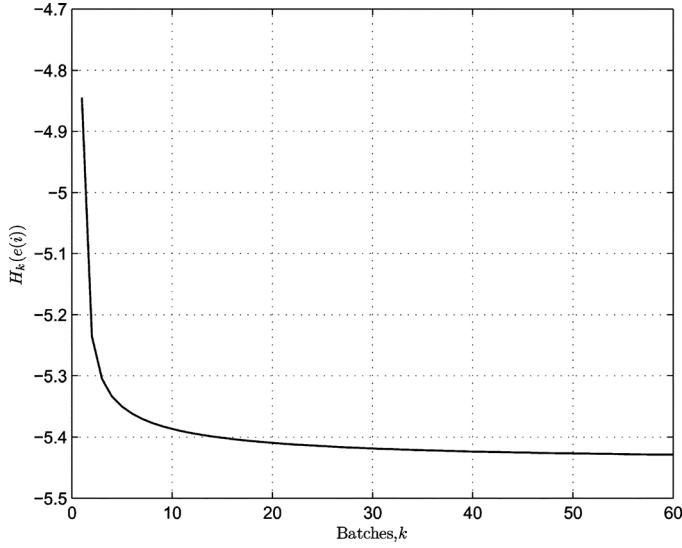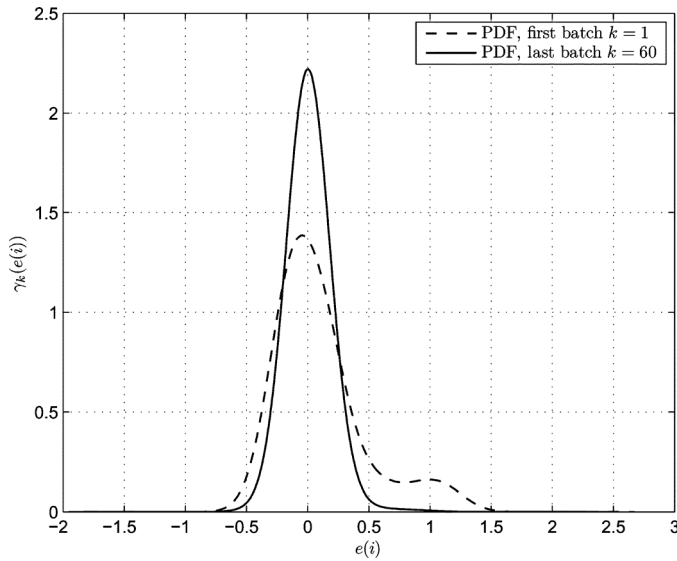Fig. 3.   Process control test rig at the University of Manchester.

Fig. 4.   Test input and output sequences. (a) Input sequence applied to the pump. (b) The recorded flow from the PCU system.
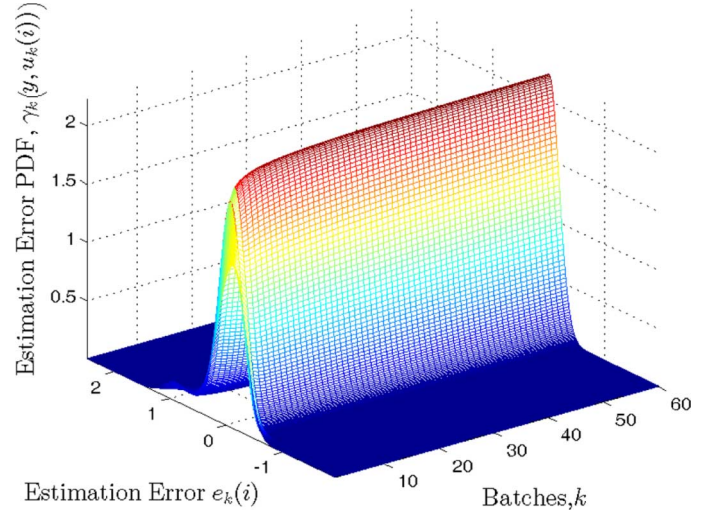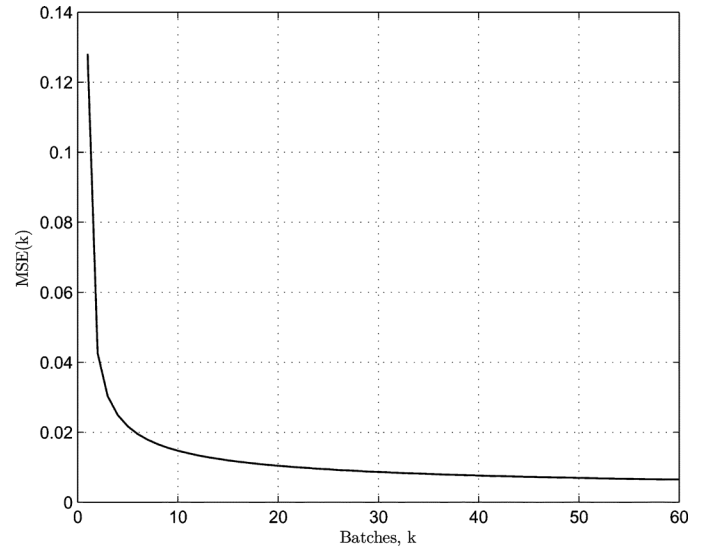
system has been recently upgraded at the Control System Centre (CSC) to facilitate control and monitoring via MATLAB or LabVIEW [33], [37]. The aim is to design a filter to estimate an unmeasurable system state (in this case, tank level) by using the available information so that the entropy of the output error is minimized batch-by-batch.

B. Offline System Identification

In order to implement the filter, a dynamic model of the process in state-space form is required. Assuming a black-box linear state-space model, a series of experiments were set to collect a sequence of input and output signals suitable for system identification. The selected input/output sequences are shown in Fig. 4(a) and (b), respectively.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

AFSHAR *et al.*: ILC-BASED MINIMUM ENTROPY FILTER DESIGN AND IMPLEMENTATION FOR NON-GAUSSIAN STOCHASTIC SYSTEMS

7

Fig. 5. Monotonic decrease of output entropy by choosing μ_1.

Fig. 7. Historical changes of the output error PDF when μ_1 is chosen.

Fig. 6. PDF of the output error at the first batch $k = 1$ and $k = 60$.

Fig. 8. Batch-by-batch decrease of the output error MSE when μ_1 is chosen.

By setting the system sampling time to 0.0119 s, 19290 samples are collected which are persistently exciting of degree 50 and satisfy the D-optimality criterion [38]. Without the valves operating, the identification task could be performed by assuming Gaussian noises. It can be argued that the assumption of Gaussian noises for system identification can result in inaccurate model as the actual filtering noises are assumed non-Gaussian. In fact, this cannot be considered as a disadvantage, as it will be shown that in spite of such restriction in the statistics of model, the minimum entropy filter can effectively estimate the flow signal and result in negligible estimation error as well as minimized entropy. Furthermore, preliminary data analysis suggested a relative skewness in the flow signal which is caused by the flow meter inaccuracy in reading flows as small as 0.4 lpm. Also, no significant nonlinearity or time delay could be highlighted in system dynamics. Experiment confirms that the data is best represented by a 2-state system. It can be shown that the other state can represent the potential

energy level caused by the height difference of the pump and processing tank and also the fluid level in the processing tank. Using the N4SID algorithm, after 200 iterations, the following parameters are obtained:

$$A = \begin{bmatrix} 0 & 1 \\ -0.94335 & 1.9421 \end{bmatrix}, \ B = \begin{bmatrix} -0.0010986 \\ -0.00049215 \end{bmatrix}$$
$$C = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

and

$$D = -0.00037581, \ B_w = \begin{bmatrix} 0.64673 \\ 0.79116 \end{bmatrix}$$
$$x(0) = \begin{bmatrix} -1.289 \\ -1.295 \end{bmatrix}.$$

It must be noted that the obtained initial values correspond to the de-trended signal values and must be re-scaled so that the actual values (i.e., process flow and tank level) are obtained. After the convergence of N4SID, the model shows to fit to 87% of the validation data; however, the T-test expressed by T_{SS}
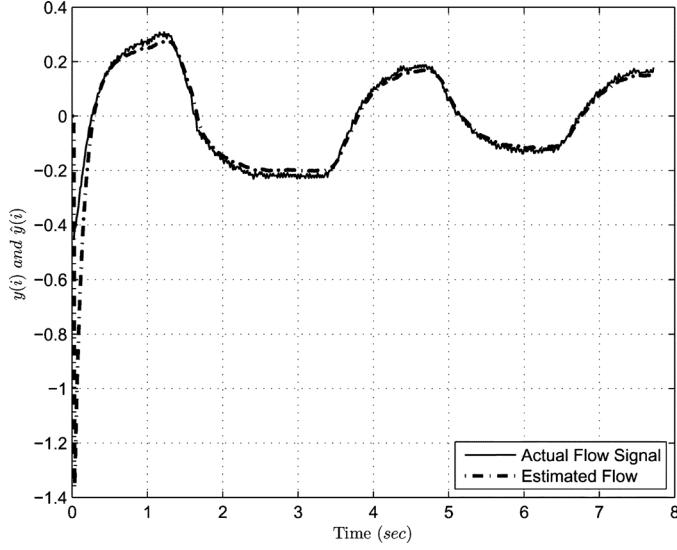
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8   IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY

Fig. 9.   Estimated and measured flow signals by choosing mu_1 within the last batch $k = 60$.

Fig. 11.   Strictly decreasing entropy performance by choosing μ_2.

Fig. 10.   Convergence of the filter gains when μ_1 is chosen.

Fig. 12.   PDF of the output error at the first batch $k = 1$ and $k = 60$ under μ_2.

as follows, confirms that matrices B and D are not statistically significant, as shown in the equation at the bottom of the page.

Further analysis on the model residuals as well as the parameter uncertainty test, collectively confirm the validity of the model obtained by N4SID. After the approximate model is obtained, the ILC-based minimum entropy estimation is performed.

C. Minimum Entropy State Estimation

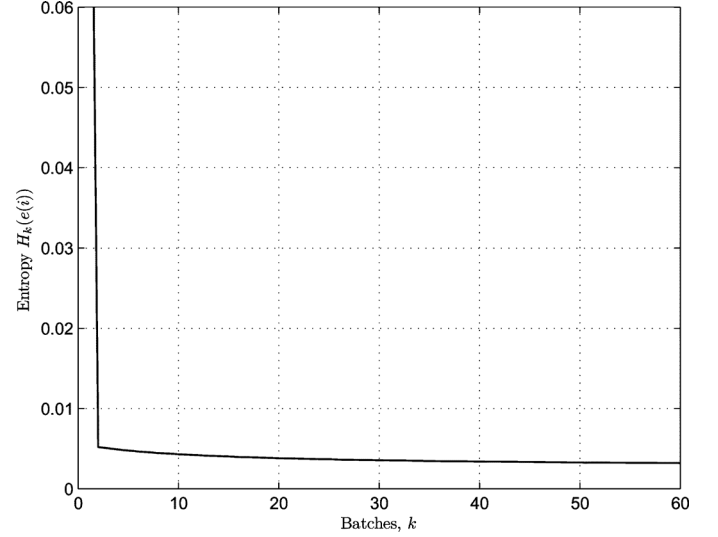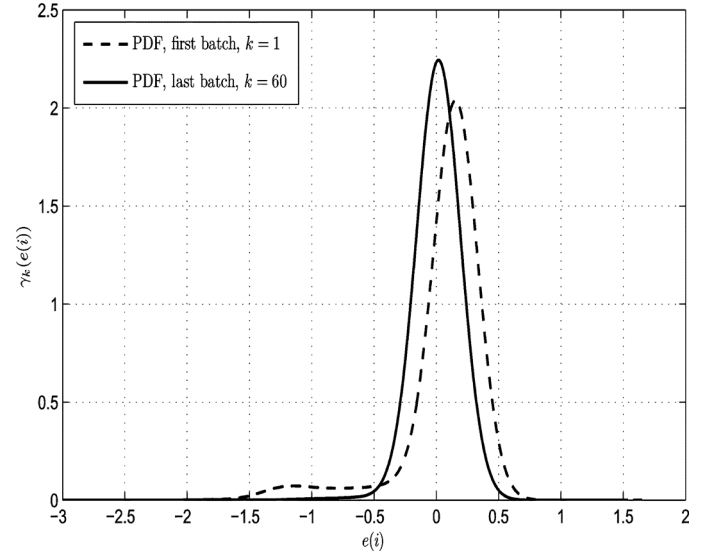Here, as mentioned in Section III, given an initial value for filter gain, the ILC-based tuning laws (11), (12), and (15) are applied to tune the filter batch by batch. Also, as formulated in Section III-B, the learning rate μ must be chosen so that (28) is satisfied. Based on the value of the learning rate chosen, three modes can be observed, namely, slow convergence, fast convergence and divergence. If the learning rate is chosen too small, the convergence will be slow and more batches are required in order that the algorithm converges. On the other hand, if the learning rate is chosen very close to the boundary of convergence, the convergence will be fast and less batches are required to achieve the desired performance. Obviously, learning rates outside the convergence interval result in divergence of the ILC-based tuning. In this test, 60 batches are chosen and

$$T_{SS} = \begin{bmatrix} 306.7646, & 623.6688, & 3.8698, & 1.8662, & 0.6660, & 72.5638, & 75.2749, & 191.9030, & 164.9697 \end{bmatrix}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

AFSHAR *et al.*: ILC-BASED MINIMUM ENTROPY FILTER DESIGN AND IMPLEMENTATION FOR NON-GAUSSIAN STOCHASTIC SYSTEMS
9
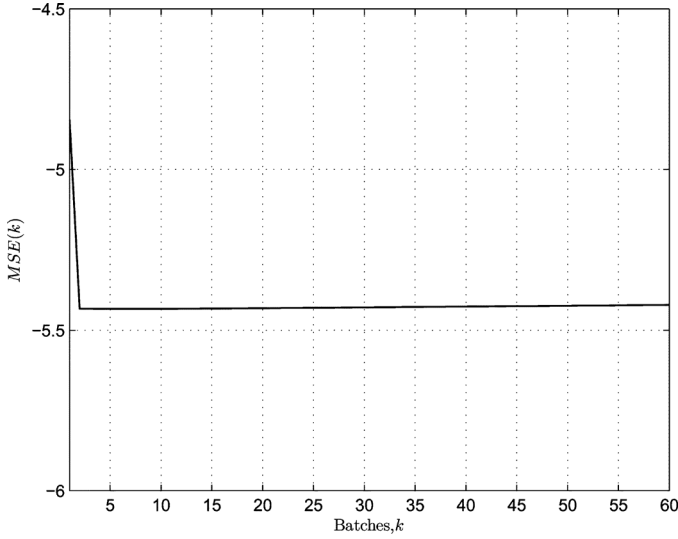
Fig. 13.   MSE of the output error under μ_2.
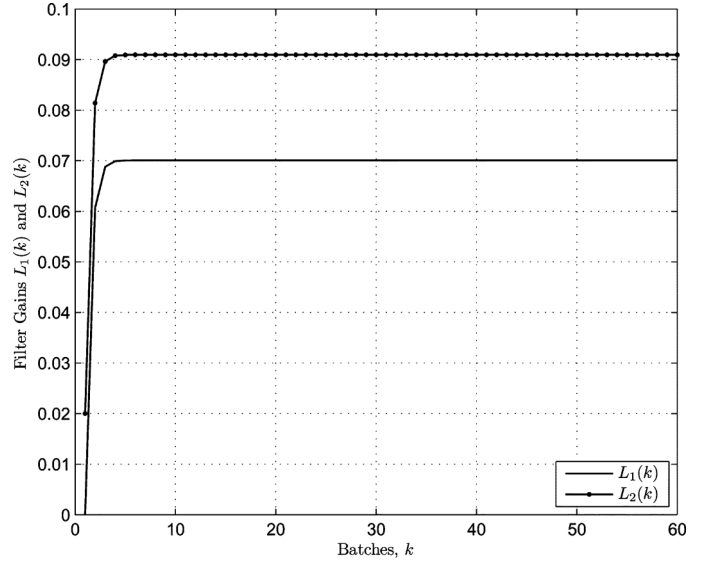
Fig. 15.   Gains of the filter when μ_2 is chosen.
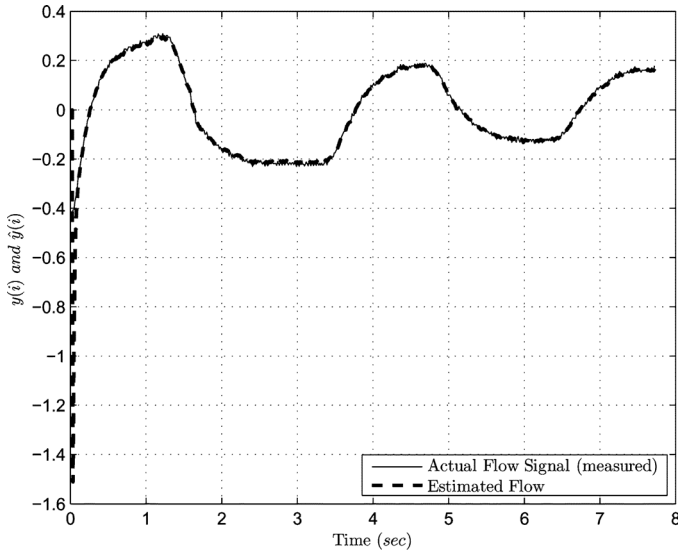
Fig. 14.   Estimation tracking performance by choosing μ_2.

the performance is demonstrated for slowly converging and fast converging learning rates.

1) Slowly-Converging Algorithm: In this first example, the learning rate is chosen $\mu_1 = 5 \times 10^{-5}$. By choosing such learning rate, the output error entropy is reduced batch by batch, as shown in Fig. 5.

As the entropy signal takes smaller values, the PDF of the output error must become more narrowly-shaped around zero, meaning that the output error is becoming more unbiased and closer to a Gaussian shaped distribution. Fig. 6 confirms the gradual change in the shape of output PDF from the first $(k = 1)$ to the last $(k = 60)$ batch.

The historical changes of the output error can be more closely monitored by investigating the 3-D shape of the PDF along with the batches. Fig. 7 illustrates the historical variations of the output error PDF under learning rate μ_1.

The chosen filter must also improve the tracking performance of the estimation which can be checked by examining the MSE

of the closed-loop output error. Fig. 8 confirms that the drift between the measured and estimated output is strictly decreasing when learning rate μ_1 was chosen.

This can also be shown by comparing the estimated output with the measured output (water flow signal) which is shown in Fig. 9. It is also worthwhile to check the convergence of the filter gain parameters as the batches progress. Fig. 10 shows such changes by running a slowly-converging algorithm. The figures collectively suggest an slowly-converging algorithm which is due to the selection of the learning rate. In Section V, a learning rate leading to a faster-converging algorithm is chosen and the same graphs are proposed to check the performance.

2) Fast-Converging Algorithm: In this section, the learning rate is made 100 times bigger (i.e., $\mu_2 = 5 \times 10^{-3}$. First, the entropy minimization performance is examined. Fig. 11 shows the decrease in the output error entropy when a fast-converging algorithm (i.e., learning rate μ_2) was chosen. Compared to the slowly-converging algorithm (see Fig. 5), the fast converging algorithm can find the local minimum in fewer iterations $(k = 5)$. Fig. 12 compares the initial $(k = 1)$ and final $(k = 60)$ shapes of the output error PDF when a fast-converging algorithm is chosen. Choosing a fast-converging learning rate also helps to achieve better locally optimum points. As Fig. 13 shows, choosing μ_2 results in a smaller values for MSE which suggests that the output tracking performance should have been improved. However, it shows a gradual increase in the MSE which suggests that the selected learning rate is close to the boundary of convergence. Although the algorithms can be technically considered as diverging, as the slope of divergence is very small, the other performance criteria can be effectively verified. Similar to the slow-converging filter, comparing the estimated and measured output signals can help to validate the effectiveness of the filter performance. As shown in Fig. 14, the output has been tracked effectively, although comparing these signals, the close-to-divergence performance cannot be visualized. Although not converging to a local minimum value, the filter gains also show a faster response which reflects the choice

of learning rate close to the convergence boundary. Fig. 15 shows changes in the filter gains by running a fast-converging algorithm.

V. CONCLUSION

A new filtering method is proposed for linear stochastic systems with non-Gaussian noises. A minimum entropy criterion is chosen to determine the filter gain which is updated by gradient-based ILC tuning laws. Due to the presence of non-Gaussian noises, the uncertainty/randomness cannot be sufficiently formulated if only mean and standard deviation of the output error are used to characterize them. Therefore, the proposed algorithm is fundamentally different and hence non-comparable to commonly-used filtering algorithms such as Kalman filtering. As the filter gain is determined in an iterative fashion, convergence analysis must be performed to guarantee the stability of the algorithm. For this purpose, the learning rate corresponding to the ILC law is explicitly obtained so that the algorithm is convergent. As the problem is not necessarily a convex optimization problem, the solution is only guaranteed to be locally optimal. The proposed method incorporates less complexity compared to the other non-Gaussian filtering methods and can be easily implemented in practical applications. Future work will comprise adopting different estimation criteria to assess the efficiency of entropy minimization as well as extension of the case to nonlinear systems.

REFERENCES

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

[2] R. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME—J. Basic Eng.*, vol. 82, pp. 35–45, 1960.

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

[4] F. Wang and V. Balakrishnan, "Robust kalman filters for linear time varying systems with stochastic parametric uncertainties," *IEEE Trans. Signal Process.*, vol. 50, no. 4, pp. 803–813, Apr. 2002.

[5] I. Petersen and D. McFarlane, "Optimal guaranteed cost filtering for uncertain discrete-time linear systems," *Int. J. Robust Nolinear Control*, vol. 6, pp. 267–280, 1996.

[6] U. Shake and C. E de Souza, "Robust minimum variance filtering," *IEEE Trans. Signal Process.*, vol. 43, no. 11, pp. 2474–2483, Nov. 1995.

[7] T. Song and J. Speyer, "A stochastic analysis of a modified gain extended Kalman filter with applications to estimation with bearings only measurements," *IEEE Trans. Autom. Control*, vol. 30, no. 10, pp. 940–949, Oct. 1985.

[8] J. H. Lee and N. Ricker, "Extended kalman filter based nonlinear model predictive control," *Ind. Eng. Chem. Res.*, vol. 33, no. 6, pp. 1530–1541, 1994.

[9] L. Jetto, S. Longhi, and G. Venturini, "Development and experimental validation of an adaptive extended Kalman filter for the localization of mobile robots," *IEEE Trans. Robot. Autom.*, vol. 15, no. 2, pp. 219–229, Apr. 1999.

[10] R. Reif and K. Unbehauen, "The extended Kalman filter as an exponential observer for nonlinearsystems," *IEEE Trans. Signal Process.*, vol. 47, no. 8, pp. 2324–2328, Aug. 1999.

[11] Z. Zhang and J. Zhang, "A novel strong tracking finite-difference extended Kalman filter for nonlinear eye tracking," *Sci. China Series F: Inform. Sci. Co-Published With Springer-Verlag GmbH*, vol. 52, no. 4, pp. 688–694, 2009.

[12] F. Yang, Z. Wang, and Y. S. Hung, "Robust Kalman filtering for discrete time-varying uncertain systems with multiplicative noises," *IEEE Trans. Autom. Control*, vol. 47, no. 7, pp. 1179–1183, Jul. 2002.

[13] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*.   Philadelphia, PA: SIAM Studies in Applied Mathematics, 1994.

[14] F. Wang and V. Balakrishnan, "Robust adaptive kalman filters for linear time-varying systems with stochastic parametric uncertainties," in *Proc. Amer. Control Conf.*, 1999, pp. 440–444.

[15] E. Gershon, U. Shaked, and I. Yaesh, "h_∞ control and filtering of discrete-time stochastic systems with multiplicative noise," *Automatica*, vol. 37, pp. 409–417, 2001.

[16] F. Carravetta, A. Germani, and M. Raimondi, "Polynomial filtering for linear discrete time nongaussian systems," *SIAM J. Control Opt.*, vol. 34, pp. 1666–1690, 1996.

[17] Q. Zhang, "Hybrid filtering for linear systems with non-gaussian disturbances," *IEEE Trans. Autom. Control*, vol. 1, no. 1, pp. 50–61, Jan. 2000.

[18] H. Wang, *Bounded Dynamic Stochastic Systems: Modeling and Control*.   New York: Springer-Verlag, 2000.

[19] L. Guo and H. Wang, "Applying constrained nonlinear generalized PI strategy to pdf tracking control through square root b-spline models," *Int. J. Control*, vol. 77, no. 17, pp. 1481–1492, Nov. 2004.

[20] L. Guo and H. Wang, "PID controller design for output pdfs of stochastic systems using linear matrix inequalities," *IEEE Trans. Syst., Man Cybern. B, Cybern.*, vol. 35, no. 1, pp. 65–71, Jan. 2005.

[21] L. Guo and H. Wang, "Fault detection and diagnosis for general stochastic systems using b-spline expansions and nonlinear filters," *IEEE Trans. Circuits Syst.—I, Reg. Papers*, vol. 52, no. 8, pp. 1644–1652, Aug. 2005.

[22] Y.-M. Zhang, L. Guo, and H. Wang, "Filter-based FDD using pdfs for stochastic systems with time delays," in *Proc. Amer. Control Conf.*, 2005, pp. 1660–1665.

[23] H. Yue and H. Wang, "Minimum entropy control of closed-loop tracking errors for dynamic stochastic systems," *IEEE Trans. Autom. Control*, vol. 48, no. 1, pp. 118–122, Jan. 2003.

[24] L. Guo and H. Wang, "Minimum entropy filtering for multivariate stochastic systems with non-Gaussian noises," *IEEE Trans. Autom. Control*, vol. 51, no. 4, pp. 695–700, Apr. 2006.

[25] F. Yang, Y. Li, and X. Liu, "Robust error square constrained filter design for systems with non-Gaussian noises," *IEEE Signal Process. Lett.*, vol. 15, pp. 930–933, Dec. 2008.

[26] H. Wang and P. Afshar, "Radial basis function based iterative learning control for stochastic distribution systems," in *Proc. IEEE Int. Symp. Intell. Control*, 2006, pp. 100–105.

[27] H. Wang, P. Afshar, and H. Yue, "ILC-based generalised PI control for output pdf of stochastic systems using LMI and RBF neural networks," in *Proc. 45th IEEE Conf. Decision Control*, 2006, pp. 5048–5053.

[28] P. Afshar, H. Yue, and H. Wang, "Robust iterative learning control of output PDF in non-Gaussian stochastic systems using youla parametrization," in *Proc. Amer. Control Conf. (ACC)*, 2007, pp. 576–581.

[29] H. Wang and P. Afshar, "ILC-based fixed-structure controller design for output pdf shaping in stochastic systems using LMI techniques," *IEEE Trans. Autom. Control*, vol. 54, no. 4, pp. 760–773, Apr. 2009.

[30] P. Afshar, "Intelligent model reference adaptive distribution control for non-Gaussian stochastic systems," presented at the IEEE Int. Conf. Netw., Sens., Control, Okayama, Japan, 2009.

[31] P. Afshar and H. Wang, "ILC-based adaptive minimum entropy control for general stochastic systems using neural networks," in *Proc. 46th IEEE Conf. Decision Control*, 2007, pp. 252–257.

[32] P. Afshar and H. Wang, "An ILC-based minimum entropy PI controller for unknown and non-Gaussian stochastic systems," in *Proc. 17th IFAC World Congr.*, 2008, pp. 12230–12235.

[33] P. Afshar, H. Wang, and T. Chai, "An ILC-based adaptive control for general stochastic systems with strictly decreasing entropy," *IEEE Trans. Neural Netw.*, vol. 20, no. 3, pp. 471–482, Mar. 2009.

[34] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*.   London, U.K.: Chapman and Hall/CR, 1992.

[35] D. Mustafa, K. Glover, and G. K. Paperback, *Minimum Entropy H_∞ Control*, ser. Lecture Notes in Control and Information science.   Berlin/Heidelberg, Germany: Springer, 1990.

[36] J.-W. Xu, D. Erdogmus, and J. C. Principe, "Minimum error entropy luenberger observer," presented at the Amer. Control Conf. (ACC), Portland, OR, 2005.

[37] P. Afshar, "Iterative learning-based stochastic distribution control," Ph.D. dissertation, Univ. Manchester, Sch. Elect. Electron. Eng., Manchester, U.K., 2007.

[38] M. Mossberg, "Optimal experimental design for identification of viscoelastic materials," *IEEE Trans. Control Syst. Technol.*, vol. 12, no. 4, pp. 578–582, Jul. 2004.

Puya Afshar was born in Tehran, Iran, in 1980. He received the B.Sc. degree in electrical engineering from K.N. Toosi University of Technology (KNTU), Tehran, Iran, in 2002, the M.Sc. degree in control engineering from Tarbiat Modares University (TMU), Tehran, Iran, in 2004, and the Ph.D. degree in advanced process control from The University of Manchester, Manchester, U.K., 2007.

He was an Integrated Control and Safety System (ICSS) Applications Engineer in oil, gas, and petrochemical industries from 2001 to 2005. In 2005, he joined Control Systems Centre (CSC), The University of Manchester. Since January 2008, he has been a Research Fellow with CSC, where he is in process control and optimization. His main research interests include are industrial process control and optimization, stochastic distribution control, iterative learning, hybrid product scheduling, and fault tolerant control systems.

Dr. Afshar was a recipient of the National Petrochemical Company of Iran Award for his M.Sc. thesis in fault tolerant control.

Fuwen Yang received the B.Sc. degree in electrical engineering from Anhui University of Science and Technology, Huainan, China, and the M.Sc. and Ph.D. degrees in control engineering from Huazhong University of Science and Technology, Huazhong, China.

He is currently a Professor with the School of Information Science and Engineering, East China University of Science and Technology, Shanghai, China. Before joining East China University of Science and Technology, he worked with Fuzhou University, China, King's College London, U.K., Brunel University, U.K., the University of Manchester, U.K., and the University of Hong Kong. He has published over 100 journal and conference papers in systems control and signal processing. His research interests include H-infinity control and filtering, networked control, fault detection and diagnosis, signal processing, industrial real-time control, and power electronics.

Dr. Yang was a recipient of the Teaching Excellence Award for Young Teachers in 1995 from Fok Ying Tung Education Foundation, China; four Science and Technology Development Awards in 1996, 1999, 2002, and 2006, from Fujian Province of China; and two Youth Science and Technology Awards both in 1998 from Fujian Province of China and from Yun Sheng Foundation. He was listed in Ten Outstanding Youth from Fujian Province of China in 1999 and was honored the May 4 Youth Medal from Fujian Province of China in 2000. Since 2001, he has been acting as a State Consultant receiving Special Allowance from the State Council of the People's Republic of China.

Hong Wang (M'97) was born in Beijing, China., in 1960. He received the B.S. degree from Huainan University of Mining Engineering, Huainan, China, in 1982 and the M.S. and Ph.D. degrees from Huazhong University of Science and Technology, Huazhong, China, in 1984 and 1987, respectively.

From 1988 to September 1992, he was a Research Fellow with Salford, Brunel, and Southampton Universities. He then joined University of Manchester Institute of Science and Technology, Manchester, U.K., and has been a Professor in process control since 2002 at the University of Manchester, UK. He also holds a research visiting position with the Northeastern University (China), Huazhong University of Science and Technology (China) and Institute of Automation, Chinese Academy of Sciences. His research interests include stochastic distribution control, fault detection and diagnosis, nonlinear control, and data-based modeling for complex systems. He has published 200 papers and 3 books in these area (http://www.eee.manchester.ac.uk).

Prof. Wang was an Associate Editor of the IEEE Transactions on Automatic Control and now serves as an editorial board member for the *Journal of Measurement and Control, Transactions of the Institute of Measurement and Control, IMechE Journal of Systems and Control Engineering, Automatica Sinica*, and *Journal of Control Theory and Applications*.