

Additive Change Detection in Nonlinear Systems With Unknown Change Parameters

Namrata Vaswani

Abstract—We study the change detection problem in partially observed, nonlinear systems [which satisfy the hidden Markov model (HMM) property]. The change parameters are assumed unknown, and the changes can be slow or sudden. A partially observed system needs to be tracked first before changes can be detected. Sudden changes result in significant loss of track. These can be detected easily using the increase in tracking error (TE) or observation likelihood (OL) or using a CUSUM-type method applied to either of these. However, slow changes (which result in small loss of track) often get missed. We propose here a statistic that uses the tracked component of the change to detect it and, hence, detects slow changes faster than TE or OL. We show, both analytically and through simulations, that this statistic complements OL and TE for change detection.

Index Terms—Abnormality detection, change detection, particle filtering, tracking, unknown change parameters.

I. INTRODUCTION

CHANGE detection is required in many practical problems arising in quality control, flight control, fault detection [1]–[3], in target tracking problems where the motion model may change over time [4, Ch. 23], [16], as well as in surveillance problems like abnormal activity detection [5], [6] or activity recognition [7]. We study the *online change detection problem in partially observed linear or nonlinear systems [also called a general hidden Markov model (HMM) [8]], when i) the changed system parameters are unknown and ii) the change can be slow or sudden*. A linear, Gaussian state-space model can be tracked using a Kalman filter (KF) [9], while, in other situations, a particle filter (PF) [4], [10], [11] can be used. Sudden changes, which result in significant loss of track, are easily detected using existing statistics such as tracking error (TE) [12] or observation likelihood (OL)¹ [2], [13]. We propose a statistic called the “negative expected log likelihood of state” (ELL, and its generalization, gELL), which uses the tracked component of the change to detect it. Hence it detects slow changes (which result in small loss of track) faster than TE or OL. We show, both

analytically and experimentally, that ELL complements OL [2], [13], i.e., it detects changes that OL misses and vice versa.

The fact that ELL can detect a change before significant loss of track is useful in target(s) tracking problem where the target(s) dynamics might change over time. If the change in target motion dynamics is slow, it may not immediately result in loss of track. If one can detect the change before loss of track, one can try to learn its parameters on the fly (or at least increase the system noise variance), before completely losing track of the target. For example, a target may be moving with a constant velocity model in a given direction, and this may slowly (or suddenly) change to a changing velocity model in a different direction, e.g., [6] or [4, Ch. 23]. Change detection in a bearings-only target tracking problem [14] is demonstrated in Section VII-E. Abnormality detection while tracking groups of moving persons (landmarks) [5] is discussed in Section VII-F. We have used ELL to detect changes in landmark-shape dynamical models [5]–[7]. This has applications in abnormal action detection [6] and in activity segmentation (segmenting a long activity sequence into piecewise stationary elementary activities) [6] and recognition [7]. Other applications of ELL are in neural signal processing (detecting changes in response of animals’ auditory neurons [15] to changes in stimuli provided to them), and in acoustic tracking of targets with changing dynamics.

We now discuss existing work. Cumulative sum (CUSUM) [16], [17] is the classical algorithm for change detection in time series. The CUSUM statistic is the maximum (taken over all previous time instants) of the likelihood ratio assuming that the change occurred at time $\tau \leq t$, i.e.,

$$\text{CUSUM}_t \triangleq \max_{1 \leq \tau \leq t} LR(\tau : t) \quad (1)$$

$$LR(\tau : t) = \frac{p_Y^{\theta_1}(Y_\tau, Y_{\tau+1} \dots Y_t | Y_1, Y_2 \dots Y_{\tau-1})}{p_Y^{\theta_0}(Y_\tau, Y_{\tau+1} \dots Y_t | Y_1, Y_2 \dots Y_{\tau-1})} \quad (2)$$

If the changed system parameters are unknown, the non-null hypothesis is composite (i.e., $\theta_1 \in \{\theta : \theta \neq \theta_0\}$). CUSUM can then be replaced by generalized CUSUM where another maximization is performed over all possible values of θ_1 .

For nonlinear partially observed systems with known change parameters, [18] proposed an efficiently computable modification of the CUSUM statistic using a PF. Some other approaches are discussed in [19]. When the change parameters are unknown and the system is nonlinear, generalized CUSUM is, in general, intractable. For the special case of the parameter belonging to a finite set of cardinality M , [20] and [21] propose to run one particle filter for each possible value of the parameter. When the unknown parameter does not belong to a finite set, one can

Manuscript received August 29, 2005; revised May 7, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Fredrik Gustafsson.

The author is with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: namrata@iastate.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2006.887111

¹In this paper, OL refers to the $-\log$ of the likelihood of current observation (Y_t) given all past observations ($Y_{1:t-1}$), see Definition 9. It should not be confused with the other frequently used meaning of OL in PF literature where it sometimes refers to the conditional likelihood of the observation at t , Y_t , given state at t , X_t .

try to use approximations to maximum likelihood (ML) for parameter estimation, such as the expectation-maximization (EM) algorithm, to perform generalized CUSUM [19]. However, the EM algorithm (or any other ML technique) requires estimates of the smoothing distributions. It is well known that a particle filter with a fixed number of particles N does not give reliable estimates of the smoothing distributions [19]. We have implemented this method for a simple example. We compare its performance with ELL in Section VII-A, where we refer to it as CUSUM-ML-OL. Recently proposed modifications of EM, such as online EM [22] or recursive ML [23], can potentially be used to improve its performance. A model validation approach to change detection using a PF is presented in [24].

There is a huge amount of literature on fault (change with unknown parameters) detection and isolation [1]–[3], [25]–[27]. References [25]–[27] tackle deterministic nonlinear systems. Some statistical approaches are presented in [1]–[3]. A common choice for the “residual” [1] is the innovation error, which is a white-noise sequence under no fault conditions. The squared sum of innovation errors, also referred to as the TE [12], is a commonly used detection statistic. This can be computed only when the observation model can be defined as in (5), i.e., the observation noise is additive. In general, it may be replaced by the logarithm of the likelihood of the current observation given past observations (OL), defined in Definition 9, or its sum over a sliding window [2]. OL (or its sum over all t) is a commonly used statistic for model validation using finite-state HMMs [13]. In fact, for additive white Gaussian observation noise, when the variance of the posterior is small, OL is approximately equal to TE.²

Another fault detection statistic is the score function [1], which is the sensitivity of the log likelihood to change in the parameter and is computed as $(\partial \text{OL} / \partial \theta)|_{\theta=\theta_0}$. Its expected value over unchanged observations is zero [1]. For certain types of partially observed systems, the score function can be efficiently calculated using a particle filter [28]. However, for the general case, computing it using a particle filter is computationally very expensive [19].

Both TE [12] and OL [2] use loss of track to detect change and hence are able to detect only sudden changes which result in large loss of track. We discuss this in detail in Section V. For the same reason, the score function, which is the sensitivity of OL to the change parameter, will also only detect sudden changes (increase in sensitivity of OL due to a slow change will also be insignificant). Our goal is to also detect slow changes for which loss of track is small. We propose the statistic called ELL, which uses the tracked part of the change to detect it. As explained in Section III-A, ELL can be understood as a measure of inaccuracy [29] between the posterior at t and the t -step-ahead prediction distribution of the unchanged state. We propose a generalization of ELL called gELL, where we replace the t -step-ahead prediction distribution by a $\Delta < t$ -step-ahead prediction that improves its sensitivity. Also, gELL can be used to detect multiple changes.

²It follows by writing Taylor series expansion of $\psi_{t,Y_t}(X)$ about $\mathbb{E}_{\pi_{t|t-1}}[X]$ (notation defined in Sections II-A and II-B) and taking $\mathbb{E}_{\pi_{t|t-1}}[\cdot]$.

II. NOTATION AND PROBLEM FORMULATION

A. Notation and Definitions

X_t denotes the state at t , and Y_t denotes the observation at t . $Y_{1:t}$ denotes all observations from time 1 to t . E denotes the state space. The notation $(\mu, \varphi) \triangleq \int_E \varphi(x)\mu(dx)$. \mathbb{E}_μ (and Var_μ) denotes expectation (and variance) under the measure, μ . \mathbb{E}_Y (and Var_Y) denotes expectation (and variance) under the distribution of random variable Y . $\mathbf{h}(\cdot)$ denotes differential entropy [30]. $\mathcal{N}(x; \mu, \Sigma)$ denotes a Gaussian probability density function (pdf) of x with mean μ , covariance Σ .

Change start and end times are denoted by t_c and t_f . Observations generated by the nominal (unchanged) system model are denoted by $Y_{1:t}^0$. Observations generated by the changed system model are denoted by $Y_{1:t}^c$. Thus,³ $Y_{1:t}^c = (Y_{1:t_c-1}^0, Y_{t_c:t}^c)$ $\forall t \leq t_f$ and $Y_{1:t}^c = (Y_{1:t_c-1}^0, Y_{t_c:t_f}^c, Y_{t_f+1:t}^0)$ $\forall t > t_f$. The posterior distribution⁴ at time t is $\pi_t(dx) \triangleq \Pr(X_t \in dx | Y_{1:t})$. The Δ -step-ahead prediction distribution is $\pi_{t|t-\Delta}(dx) = \Pr(X_t \in dx | Y_{1:t-\Delta})$.

Definition 1 (Prior State pdf): At time t is the pdf of X_t without conditioning on any observations (denoted $p_t(x)$). It can also be understood as the t -step-ahead prediction pdf of X_t , i.e., $\pi_{t|0} = p_t$. For the nominal state, it is denoted by $p_t^0(x)$, and for changed state, it is denoted by $p_t^c(x)$.

Definition 2 (Weak Norm): The weak norm between two probability distributions π_1 and π_2 is defined as [8]

$$\|\pi_1 - \pi_2\|_W \triangleq \sup_{\phi \in \mathcal{B}: \|\phi\|_\infty \leq 1} |(\pi_1 - \pi_2, \phi)| \quad (3)$$

where \mathcal{B} denotes the class of Borel measurable functions on the state space E and $\|\phi\|_\infty \triangleq \sup_{x \in E} \phi(x)$.

The superscript c is used to denote any variable, parameter, or distribution related to the changed system and 0 for the nominal (unchanged) system. The superscript $c,0$ denotes posterior or prediction distributions (or statistics computed using them) obtained when observations coming from the changed system model are tracked by a filter designed for unchanged system. Superscripts c,c and $0,0$ are similarly defined. These are often replaced by a single c or 0 when the meaning is clear from context, e.g., $\pi_t^{0,0}$ is denoted by π_t^0 . N denotes number of particles. The N -particle PF estimates of π_t^0 is denoted $\pi_t^{0,N}$, and so on.

B. State-Space Model: General HMM

We assume a discrete-time general HMM model [8] with an \mathbb{R}^{n_x} valued state process $\{X_t\}$ and an \mathbb{R}^{n_y} valued observation process $\{Y_t\}$. Thus, the state-space $E = \mathbb{R}^{n_x}$. We use the subscript “ t ” to denote the discrete-time instants with time beginning at $t = 0$. The state transition model (system model) for the nominal (unchanged) system can be written in the form

$$X_0 \sim p_0^0(x), \quad X_t = f_t^0(X_{t-1}) + n_t, n_t \sim p_{n,t}(\cdot) \quad (4)$$

where n_t is the system noise (models uncertainty in the state space description of the system) with pdf $p_{n,t}$ and

³Note that if $t_2 < t_1$, then $[t_1 : t_2]$ denotes an empty set.

⁴Misusing notation, we denote the pdf corresponding to $\pi_t(dx)$ also by π_t .

is independent of X_{t-1} . Thus, the state transition pdf is $q_t^0(x_{t-1}, x_t) = p_{n,t}(x_t - f_t^0(x_{t-1}))$ and the state transition kernel [8] is $Q_t^0(x_{t-1}, dx_t) = q_t^0(x_{t-1}, x_t)dx_t$. The prior state pdf at t , $p_t^0(x)$ is the unconditional pdf of X_t . Given X_t , Y_t is independent of all past observations and states. The conditional distribution of Y_t given X_t , $\Pr(Y_t \in dy_t | X_t = x_t) \triangleq G_t(x_t, dy_t) = g_t(x_t, y_t)dy_t$. We denote the pdf $g_t(x_t, y_t) \triangleq \psi_{t,y_t}(x_t)$. In many cases, the observation noise is additive, i.e.,

$$Y_t = h_t(X_t) + w_t, \quad w_t \sim p_{w,t}(\cdot) \quad (5)$$

where $h_t(x)$ is a continuous function of x and w_t is observation noise with pdf $p_{w,t}$. Thus $g_t(x, Y_t) \triangleq \psi_{t,Y_t}(x) = p_{w,t}(Y_t - h_t(x))$. w_t is independent of X_t .

C. Problem Statement and Formulation

A change in the system model begins to occur at some finite time t_c and lasts until a finite final time t_f . During $[t_c, t_f]$, the system model changes to $X_t = f_t^c(X_{t-1}) + n_t$.

Definition 3 (Change Model): The change in system model is completely characterized by the change start/end times, t_c, t_f and by $b_t(X_{t-1}) = f_t^c(X_{t-1}) - f_t^0(X_{t-1}), \forall t_c \leq t \leq t_f$. For an additive change, b_t is not a function of X_{t-1} .

The change model $(t_c, t_f, b_t(X_{t-1}), \forall t \in [t_c, t_f])$ is assumed unknown. The goal is to detect the change, with minimum delay.

Existing detection statistics, such as TE [12] or OL [2], [13], use the loss of track of changed observations to detect change. The loss of track is quantified by the weak norm of the error⁵ in the posterior $LoT_t \triangleq \|\pi_t^{c,0,N} - \pi_t^{c,c}\|_W$. LoT_t depends on $LoT_{t|t-1}$ and on the observation model and on N (in case of PF). $LoT_{t|t-1}$ depends on LoT_{t-1} and on the magnitude of $b_t(X_{t-1})$, in comparison to the standard deviation (in general, spread) of $\pi_t^{c,c}$. Given a $\pi_{t-1}^{c,c}$, the variance of $\pi_t^{c,c}$ is proportional to the system noise variance, $\Sigma_{n,t}$. Thus, we have Definition 4.

Definition 4 (Slowness of Change): Define slowness (change magnitude per unit time) of change at t , as

$$r_t^2 \triangleq \mathbb{E} [b_t(X_{t-1})^T \Sigma_{n,t}^{-1} b_t(X_{t-1}) | Y_{1:t-1}]. \quad (6)$$

For an additive change, b_t is not a function of X_{t-1} and in such cases $\mathbb{E}[\cdot]$ can be removed from (6).

Remark 1 (Loss of Track Depends on Slowness): For a given observation model, the loss of track, LoT_t , is directly proportional to r_t , $\tau = t_c, \dots, t$. This can be seen rigorously for a KF tracking a scalar state with an additive change. $\pi_t^{c,0}(x) = \mathcal{N}(x; \hat{X}_t^{c,0}, \hat{\sigma}_t^2)$ and $\pi_t^{c,c}(x) = \mathcal{N}(x; \hat{X}_t^{c,c}, \hat{\sigma}_t^2)$. Replace the weak norm by the total variation norm, $\|\cdot\|$. Then, it can be shown that

$$LoT_t = 4\Phi \left(\frac{|\hat{X}_t^{c,c} - \hat{X}_t^{c,0}|}{2\hat{\sigma}_t} \right) - 2$$

⁵Weak norm is the only norm that can be computed for Monte Carlo estimates of a distribution.

where Φ is the cumulative distribution function of a standard normal. Given the observation model, $\hat{\sigma}_t^2$ is directly proportional to $\sigma_{n,t}^2$. Also, given the observation model, $|\hat{X}_t^{c,c} - \hat{X}_t^{c,0}|$ is a positive linear function of b_t (see Example 2). Thus, given an observation model and LoT_{t-1} , LoT_t is directly proportional to $r_t = b_t/\sigma_{n,t}$.

Thus TE and OL fail (or take longer) to detect slower changes (smaller r_{t_c}, \dots, r_t). To address this problem, we propose a statistic (ELL) that uses the tracked part of the change to detect it. In the next section, we define ELL and its generalization, gELL, and study their detection performance (if computed exactly using $\pi_t^{c,c}$). In Section IV, we discuss how to compute ELL and gELL statistics using a KF and using a PF. Since changed system is unknown, these are computed using $\pi_t^{c,0,N}$. In Section V, we explain the errors due to this and why ELL works better than OL when loss of track is small (slow change) and vice versa. An example illustrating this is given for a KF tracker. We prove a qualitative result demonstrating the complementarity of ELL and OL for change detection using a PF in Section VI. Simulation results are presented in Section VII. Conclusions are given in Section VIII.

III. NEGATIVE EXPECTED LOG LIKELIHOOD OF STATE

Definition 5 [Negative Expected Log Likelihood of State (ELL)]: ELL⁶ is defined as

$$\text{ELL}(Y_{1:t}) \triangleq \mathbb{E} [-\log p_t^0(X_t) | Y_{1:t}] = \mathbb{E}_{\pi_t} [-\log p_t^0(X_t)]. \quad (7)$$

ELL is the expectation of $[-\log p_t^0(X_t)]$ under the posterior π_t and so can be understood as the minimum mean-square error (MMSE) estimate of $[-\log p_t^0(X_t)]$ given observations until t .

Taking expectation of ELL($Y_{1:t}$) over observation sequences generated from the unchanged system, we get

$$\begin{aligned} \mathbb{E}_{Y_{1:t}^0} [\text{ELL}(Y_{1:t})] &= \mathbb{E}_{Y_{1:t}^0} \mathbb{E}_{\pi_t} [-\log p_t^0(X_t)] \\ &= \mathbb{E}_{p_t^0} [-\log p_t^0(X_t)] = \mathbf{h}(p_t^0) \end{aligned}$$

where $\mathbf{h}(\cdot)$ denotes differential entropy [30]. Thus, the detection statistic is as follows.

Definition 6 [ELL-Based Detection Statistic (Estat)]: Is

$$\text{Estat}(Y_{1:t}) \triangleq \text{ELL}(Y_{1:t}) - \mathbb{E}_{Y_{1:t}^0} [\text{ELL}] = \text{ELL}(Y_{1:t}) - \mathbf{h}(p_t^0).$$

Estat is computed at each t and a change is declared if $\text{Estat} > \kappa$, where $\kappa > 0$ is the detection threshold. Value of κ can be decided based on allowed level of false alarms.

IV. RELATION TO KERRIDGE INACCURACY

Definition 7 (Kerridge Inaccuracy [29], [31]): Between two pdf's p_1 and p_2 is $K(p_1 : p_2) \triangleq -\int_E p_1(x) \log p_2(x) dx$. It is used as a measure of inaccuracy between distributions.

Fact 1: ELL is the Kerridge inaccuracy between π_t and p_t^0 , i.e., $\text{ELL}(Y_{1:t}) = \mathbb{E}_{\pi_t} [-\log p_t^0(X_t)] = K(\pi_t : p_t^0) \triangleq K_t$. We denote $\text{ELL}(Y_{1:t}^0) = K(\pi_t^0 : p_t^0) \triangleq K_t^0$ and $\text{ELL}(Y_{1:t}^c) = K(\pi_t^c : p_t^0) \triangleq K_t^c$.

⁶This term may be a bit confusing since the word "likelihood" usually refers to the probability of observation given state.

Fact 2: Expectation of ELL over changed system observations is $\mathbb{E}_{Y_{1:t}^c}[\text{ELL}(Y_{1:t})] = K(p_t^c : p_t^0) \triangleq EK_t^c$ and over unchanged system observations is $\mathbb{E}_{Y_{1:t}^0}[\text{ELL}(Y_{1:t})] = \mathbf{h}(p_t^0) = K(p_t^0 : p_t^0) \triangleq EK_t^0$.

Implication: The Kullback–Leibler divergence [30] between p_t^c and p_t^0 is $D(p_t^c || p_t^0) = K(p_t^c : p_t^0) - \mathbf{h}(p_t^c)$. By its positivity [30], we get that $K(p_t^c : p_t^0) > \mathbf{h}(p_t^c)$. Assuming that $\mathbf{h}(p_t^c) = \mathbf{h}(p_t^0)$ (this holds, for example, if the two distributions differ only in the location parameter), this implies that $EK_t^c = K(p_t^c : p_t^0) > \mathbf{h}(p_t^0) = EK_t^0$, i.e., the average value of ELL of changed observations is greater than the average value of ELL of unchanged observations. This gives an indication why ELL can detect changes.

A. Detectability of a Change Using ELL

Change is declared when $\text{Estat} = K_t - EK_t^0$ exceeds the detection threshold κ . Define $VK_t^0 \triangleq \text{Var}_{Y_{1:t}^0}[K_t]$. Apply Chebyshev's inequality [32] to $K_t^0 = \text{ELL}(Y_{1:t}^0)$ to bound the false-alarm probability, $P_{\text{fa},t} = \Pr(K_t^0 - EK_t^0 > \kappa) \leq \Pr(|K_t^0 - EK_t^0| > \kappa) \leq (VK_t^0/\kappa^2)$. Also, if

$$EK_t^c - EK_t^0 - \kappa \geq k_2 \sqrt{VK_t^c}, \quad VK_t^c \triangleq \text{Var}_{Y_{1:t}^c}[K_t] \quad (8)$$

then detection probability [33] at t is

$$\begin{aligned} P_{\text{detect},t} &= \Pr(K_t^c - EK_t^0 > \kappa) \\ &\geq \Pr\left(K_t^c > EK_t^c - k_2 \sqrt{VK_t^c}\right) \quad (\text{using (8)}) \\ &\geq \Pr\left(|K_t^c - EK_t^c| \leq k_2 \sqrt{VK_t^c}\right) > 1 - 1/k_2^2 \end{aligned}$$

Fact 3 (Sufficient Condition for Detectability Using ELL): Given a threshold κ , the probability of false alarm at t $P_{\text{fa},t} \leq (VK_t^0/\kappa^2)$. Also, if at t , (8) is satisfied for some k_2 , then with probability $P_{\text{detect},t} > 1 - 1/k_2^2$, the change will get detected at⁷ t .

The above gives an upper bound on the detection delay, which can be stated as the following corollary.

Fact 4 (Bounding Detection Delay): To ensure $P_{\text{fa},t} \leq 1/k^2$, $\forall t$, the threshold should be chosen as $\kappa \geq k \sup_t \sqrt{VK_t^0}$. For a given κ , w.p. $> 1 - 1/k_2^2$, the detection time will be less than or equal to the smallest value of t (denoted t_{detect}) that satisfies (8), and hence the detection delay will be less than or equal to $t_{\text{detect}} - t_c + 1$.

1) *Computing the Variances:* Computing VK_t^0 or VK_t^c analytically is not possible without having an analytical expression for π_t^0 or π_t^c . But we can use the conditional variance identity ([33, Theorem 4.4.7], $\text{Var}_Y[\mathbb{E}[Z|Y]] + \mathbb{E}_Y[\text{Var}[Z|Y]] = \text{Var}_Z[Z]$) to upper bound VK_t^0 and VK_t^c . Set $Z \triangleq [-\log p_t^0(X_t)]$, use $Y \triangleq Y_{1:t}^0$ and $Y \triangleq Y_{1:t}^c$ to get

$$\begin{aligned} VK_t^0 &= \text{Var}_{Y_{1:t}^0}[\mathbb{E}[Z|Y_{1:t}]] \leq \text{Var}_{p_t^0}[-\log p_t^0(X_t)] \\ VK_t^c &= \text{Var}_{Y_{1:t}^c}[\mathbb{E}[Z|Y_{1:t}]] \leq \text{Var}_{p_t^c}[-\log p_t^0(X_t)] \end{aligned}$$

⁷Since this is a sufficient condition, the change may get detected earlier than t also.

2) An Example:

Example 1 (computing detection delay using Fact 4): Assume scalar state and observation, known initial state and the following system model:

$$X_0 = 0, X_t = X_{t-1} + b_t + n_t, \quad n_t \sim \mathcal{N}(0, \sigma_{\text{sys}}^2) \quad (9)$$

where $b_t = 0, \forall t$ for the unchanged system. For the changed system, $b_t = b, \forall t_c \leq t \leq t_f$ and $b_t = 0$ otherwise. Thus, $p_t^0(x)$ and $p_t^c(x)$ are defined as

$$p_t^0(x) = \mathcal{N}(x; 0, \sigma_t^2), \quad p_t^c(x) = \mathcal{N}(x; a_t, \sigma_t^2) \quad (10)$$

$$\text{where } \sigma_t^2 = t\sigma_{\text{sys}}^2 \quad (11)$$

$$a_t = \begin{cases} 0, & t < t_c \\ (t - t_c + 1)b, & t_c \leq t \leq t_f \\ (t_f - t_c + 1)b, & t > t_f \end{cases} \quad (12)$$

The observation model is $Y_t = h_t(X_t) + w_t$ with $h_t(x) = x^3$. We let w_t be truncated⁸ Gaussian observation noise with variance σ_{obs}^2 and truncation parameter, $B < \infty$. Then,

$$EK_t^c - EK_t^0 = 0.5 \frac{a_t^2}{\sigma_t^2} \quad (13)$$

$$VK_t^0 \leq 0.5, \quad VK_t^c \leq 0.5 + \frac{a_t^2}{\sigma_t^2} \quad (14)$$

We use Fact 4 with $k = 3, k_2 = 3$ (3-sigma rule). Using (14), we set $\kappa = 3\sqrt{0.5} = 2.12$. This ensures $P_{\text{fa},t} \leq 1/9 = 0.11$. Also, w.p. $> 1 - 1/k_2^2 = 0.89$, the detection time is smaller than/equal to the minimum t that solves

$$\gamma_t \triangleq EK_t^c - EK_t^0 - \kappa = 0.5 \frac{a_t^2}{\sigma_t^2} - 2.12 > k_2 \sqrt{VK_t^c} \quad (15)$$

Now, consider $t \leq t_f$. Using (14), this simplifies to $0.5\alpha - 2.12 > 3\sqrt{0.5 + \alpha}$ with $\alpha \triangleq a_t^2/\sigma_t^2$. It is easy to see that this inequality is satisfied for $\alpha \geq 45$. First, assume t_c small, i.e., $t - t_c + 1 \approx t$. Then, $\alpha \approx (tb^2/\sigma_{\text{sys}}^2) = tr^2$, where $r = b/\sigma_{\text{sys}}$. Let $r = 2$. The inequality is satisfied for $t \geq 11.25$, i.e., with probability at least 0.89, the change will get detected in less than (or equal to) 10.25 time units, if the threshold is set to ensure $P_{\text{fa}} < 0.11$. This computation assumes that the true $\pi_t^{c,c}$ is used to compute ELL. Also, it uses the loose upper bounds on variance and the loose Chebyshev inequality bound. With 100 simulations for $r = 2$ and using $\pi_t^{c,0,N}$ with $N = 100$ to approximate ELL, (i.e., using a PF with 100 particles) instead of $\pi_t^{c,c}$, it was observed that for a threshold of $\kappa = 2.12$, 89% of the times, the detection delay was less than or equal to four time units.

If t_c small does not hold, then the detection time, t_{detect} , is the least t that solves $\alpha \geq 45$ with $\alpha = a_t^2/\sigma_t^2$. This simplifies to $(t - t_c + 1)^2 \geq 11.25t$. It is easy to see that the maximum delay $(t_{\text{detect}} - t_c + 1)$ increases with t_c . For $t_c = 10$, it is computed to be 17 time units.

B. gELL: A Generalization of ELL

In many cases, such as Example 1 above, the variance of $p_t^0(x)$ may keep increasing with t . For such systems, ELL will

⁸This is required for stability of the PF [34], [35].

take longer to detect a change of given magnitude, that occurs at a later time, as shown above in Example 1. For such problems, we propose a generalization of ELL.

ELL is the Kerridge inaccuracy between π_t and p_t^0 . $p_t^0 = \pi_{t|0}^0$ is the t -step-ahead prediction pdf of the nominal state. If we have prior information that no change has occurred until a certain time⁹ t_{nc} , then we can replace p_t^0 by a $(t - t_{nc})$ -step ahead prediction distribution $\pi_{t|t_{nc}}^0$, whose average variance (average taken over $Y_{1:t_{nc}}^0$) will be smaller than or equal to that of $p_t^0 = \pi_{t|0}^0$ (follows from [33, Theorem 4.4.7]). We define **generalized ELL** as ELL with regard to $\pi_{t|t_{nc}}^0$. Another way to look at this is that we move the time origin to t_{nc} , i.e., we assume that tracking starts at t_{nc} with initial state distribution given by $\pi_{t_{nc}}^0$.

The prior information about t_{nc} can either come from expert knowledge or one can adaptively choose $t_{nc} = t - \Delta$, where Δ is chosen large enough so that with high probability no change has started to occur until $t - \Delta$ (if it has not been detected until t). We define $gELL_{\Delta}$ as ELL with regard to $\pi_{t|t-\Delta}^0$. A more robust and practical solution is to maximize $gELL_{\Delta}$ minus its average value over all possible value of Δ , i.e., define $gEstat$ statistic.

Definition 8 [Generalized ELL Statistic (gEstat)]: Is

$$gEstat \triangleq \max_{1 \leq \Delta \leq t} gEstat_{\Delta}, \text{ where} \\ gEstat_{\Delta} \triangleq \mathbb{E}_{\pi_t} \left[-\log \pi_{t|t-\Delta}^0(X_t) \right] - \mathbf{h} \left(\pi_{t|t-\Delta}^0 \right) \quad (16)$$

and $\mathbf{h}(\pi_{t|t-\Delta}^0) = \mathbb{E}_{\pi_{t|t-\Delta}^0} [-\log \pi_{t|t-\Delta}^0(X)]$ is the differential entropy [30] of $\pi_{t|t-\Delta}^0$. Note $\mathbf{h}(\pi_{t|t-\Delta}^0)$ is not the conditional differential entropy [30] (we are not averaging over $Y_{1:t-\Delta}^0$).

Note that $gEstat \geq gEstat_t = Estat$. Also, the best (least average variance¹⁰) prediction distribution of the unchanged state is $\pi_{t|t_c-1}^0$ (since Y_{t_c-1} is the last unchanged observation). We do not know t_c , but we know that $gEstat \geq gEstat_{t-t_c+1}$. So, for a given threshold, the detection probability of $gEstat$ at t will be greater than or equal to that of $Estat$ or $gEstat_{t-t_c+1}$. Note, this does not say anything about false-alarm probability, which may also be larger.¹¹

A second use of gEstat is that it can be used to detect multiple changes in the system model, by maximizing Δ over $1 \leq t < t - t_f^{\text{prev}}$ where t_f^{prev} is end time of the previous change. This will be studied in future work.

C. Detectability of a Change Using $gELL_{\Delta}$

We can get a result similar to Fact 3 for $gELL_{\Delta}$. The analysis assumes $\Delta > t - t_c$, i.e., $t_{nc} < t_c$ and that $\pi_{t|t-\Delta}^0$ is known exactly. First, note that $gELL_{\Delta} = K(\pi_t : \pi_{t|t-\Delta}^0)$. Let $\mathbb{E}_{Y_{t-\Delta+1:t}^c} [gELL_{\Delta}(Y_{1:t})] = K(\pi_{t|t-\Delta}^c : \pi_{t|t-\Delta}^0) \triangleq EK_t^c$ and $\mathbb{E}_{Y_{t-\Delta+1:t}^0} [gELL_{\Delta}(Y_{1:t})] = K(\pi_{t|t-\Delta}^0 : \pi_{t|t-\Delta}^0) \triangleq EK_t^0$.

⁹As pointed out by a reviewer, this is a commonly used assumption in sliding-window-based change detection methods, which assume no change before start of the window.

¹⁰This follows from [33, Theorem 4.4.7].

¹¹False alarm probability of $gEstat$ will not be too much larger than that for $Estat$ because for the nominal system, there is not much difference in the values of $gEstat_{\Delta}$ and $Estat$. Rigorous analysis of $gEstat$ will be presented in future work.

$\pi_{t|t-\Delta}^0) = \mathbf{h}(\pi_{t|t-\Delta}^0) \triangleq EK_t^0$. Similarly, define $VK_t^c = \text{Var}_{Y_{t-\Delta+1:t}^c} [gELL_{\Delta}(Y_{1:t})]$ and VK_t^0 .

Fact 5: With these new definitions of EK_t^0 , EK_t^c , VK_t^c , and VK_t^0 , Fact 3 and Fact 4 hold. Also, for a given threshold, the detection probability for $gEstat$, $P_{\text{detect},gEstat} \geq \max_{1 \leq \Delta \leq t} P_{\text{detect},gEstat_{\Delta}}$.

D. When Is $gELL$ Not Required?

In many applications, for example the problem discussed in Section VII-F (details in [5]) or in Example 3 of Section VII-C, the nominal state can be modeled to be stationary or asymptotically stationary. In these cases, the variance of p_t^0 is bounded. If the state is stationary, $p_t^0(x)$ is constant for all t and thus ELL detects a change equally well for all t_c . If asymptotically stationary, this holds for large t_c . Thus, defining $gELL$ by approximating $\pi_{t|t-\Delta}^0$ (which introduces more approximation error) does not give much extra benefit for such applications. We show the ELL ($Estat$) and $gELL$ ($gEstat$) receiver operating characteristics (ROC) plots for Example 3 in Fig. 5(a) and (b).

The same is true for systems for which the nominal state pdf varies slowly with t . For such systems, the nominal state sequence can be modeled as being piecewise stationary (with non-stationary transitions). A common example is switched linear dynamic systems, e.g. [36] or piecewise stationary shape activities [6], [7]. If one training state sequence is available, a parametric form for the piecewise constant $p_t^0(x)$ can be learned.

V. COMPUTING ELL AND $gELL$

Computing ELL or $gELL$ requires computing the posterior, $\pi_t(dx)$ at each t . The transition from π_{t-1} to π_t is defined using the Bayes recursion as follows [8]:

$$\pi_{t-1}(dx) \xrightarrow{\text{Predict}} \pi_{t|t-1}(dx) \xrightarrow{\text{Update}} \pi_t(dx) \\ \pi_{t|t-1}(dx) \triangleq Q_t(\pi_{t-1})(dx), \pi_t(dx) \\ \triangleq \frac{\psi_{t,Y_t}(x)\pi_{t|t-1}(dx)}{(\pi_{t|t-1}, \psi_{t,Y_t})}. \quad (17)$$

If the system and observation models are linear Gaussian, the posteriors are also Gaussian and can be computed using a KF [9]. We discuss ELL and $gELL$ computation using a KF in Section IV-A. For nonlinear or non-Gaussian system or observation models, linearization methods such as extended KF [9] or the Gaussian sum filter [37] can be used, but these are often unstable and the large filtering error may wrongly be detected as a change. Instead, we use a PF [4], [10], [11], which converges (under mild assumptions) to the true filter for large enough number of Monte Carlo samples (or “particles”) [38]. This is discussed in Section IV-B.

A. Computing Using a Kalman Filter (KF)

A linear, Gaussian state-space model can be tracked using a KF [9] designed for the unchanged system. The posterior is given by $\pi_t(x) = \mathcal{N}(x; \hat{X}_t, \hat{\Sigma}_t)$, where $\hat{X}_t = \mathbb{E}[X_t|Y_{1:t}]$, $\hat{\Sigma}_t = \text{Var}[X_t|Y_{1:t}]$. Also, $\pi_{t|t-\Delta}^0 = \mathcal{N}(x; \hat{X}_{t|t-\Delta}^0, \hat{\Sigma}_{t|t-\Delta}^0)$ and can be computed exactly by applying the system model Δ times

to $\pi_{t-\Delta}^0$. Recall $\text{dimension}(X_t) = n_x$. Let the eigenvalue decomposition of $\hat{\Sigma}_{t|t-\Delta}$ be

$$\hat{\Sigma}_{t|t-\Delta} = U\Lambda U^T, \Lambda = \text{diag}(\lambda_i)$$

$$\text{Let } \hat{X}_{U,t} \triangleq U^T(\hat{X}_t - \hat{X}_{t|t-\Delta}), \hat{\Sigma}_{U,t} \triangleq U^T \hat{\Sigma}_{t|t-\Delta} U$$

$$\text{Then, } \text{gEstat}_\Delta = 0.5 \sum_{i=1}^{n_x} \left[\frac{(\hat{X}_{U,t})_i^2 + (\hat{\Sigma}_{U,t})_{i,i}}{\lambda_i} - 1 \right],$$

$$\text{gEstat} = \max_{1 \leq \Delta \leq t} \text{gEstat}_\Delta, \text{Estat} = \text{gEstat}_t. \quad (18)$$

For changed observations, the KF computes $\pi_t^{c,0}(x) = \mathcal{N}(x; \hat{X}_t^{c,0}, \hat{\Sigma}_t)$ and so the above estimates will be $\text{Estat}^{c,0}$ and $\text{gEstat}^{c,0}$. If the state is a scalar, the above simplifies to

$$\text{gEstat}_\Delta = 0.5 \left[\frac{(\hat{X}_t - \hat{X}_{t|t-\Delta}^0)^2 + \hat{\sigma}_t^2}{\hat{\sigma}_{t|t-\Delta}^2} - 1 \right]. \quad (19)$$

B. Computing Using a Particle Filter

We first explain the basic PF algorithm, then give the expressions for computing ELL, gELL statistics. Finally, we discuss how to define p_t^0 for nonlinear systems and how to approximate $\pi_{t|t_{\text{nc}}}^0$ (required for computing gELL).

1) *Particle Filtering*: A PF [4], [38] is a recursive algorithm that produces at each time t a cloud of N particles $\{x_t^{(i)}\}$ whose empirical measure, π_t^N , closely “follows” π_t . The PF starts with generating N Monte Carlo samples (called “particles”) from π_0 to approximate it by $\pi_0^N(dx) \triangleq (1/N) \sum_{i=1}^N \delta_{x_0^{(i)}}(dx)$, where $\delta_{x_0}(dx)$ denotes the Dirac delta distribution at x_0 . Then, for each time step t , the PF approximates the Bayes recursion (given in (17)), which can be summarized as follows:

$$\pi_{t-1}^N(dx) \xrightarrow{\text{Step } 1} \pi_{t|t-1}^N(dx) \xrightarrow{\text{Step } 2} \tilde{\pi}_t^N(dx) \xrightarrow{\text{Step } 3} \pi_t^N(dx)$$

$$\pi_{t|t-1}^N \triangleq \frac{1}{N} \sum_{i=1}^N \delta_{\tilde{x}_t^{(i)}}(dx), \tilde{x}_t^{(i)} \sim q_t(x_{t-1}^{(i)}, \cdot)$$

$$\tilde{\pi}_t^N \triangleq \sum_{i=1}^N w_t^{(i)} \delta_{\tilde{x}_t^{(i)}}(dx)$$

$$w_t^{(i)} \triangleq \frac{\psi_{t,Y_t}(\tilde{x}_t^{(i)})}{\sum_{j=1}^N \psi_{t,Y_t}(\tilde{x}_t^{(j)})}$$

$$\pi_t^N \triangleq \frac{1}{N} \sum_{i=1}^N \delta_{x_t^{(i)}}(dx), x_t^{(i)} \sim \left\{ \tilde{x}_t^{(i)}, w_t^{(i)} \right\}_{i=1}^N.$$

Step 1 above is the *Importance Sampling Step*, which approximates the prediction step, Step 2 is the *Weighting or the Update Step*, and Step 3 is the *Resampling Step*. We refer the interested reader to [4], [11], and [38] for details about each step and for alternate more efficient importance sampling densities.

2) *ELL, gELL Estimate Using PF*: The estimate of Estat and gEstat using the PF described above are given by

$$\text{Estat}^N = \frac{1}{N} \sum_{i=1}^N \left[-\log p_t^0(x_t^{(i)}) \right] - \mathbf{h}(p_t^0) \quad (20)$$

$$\text{gEstat}_\Delta^N = \frac{1}{N} \sum_{i=1}^N \left[-\log \pi_{t|t-\Delta}^{0,N,\text{par}}(x_t^{(i)}) \right] - \mathbf{h}(\pi_{t|t-\Delta}^{0,N,\text{par}})$$

$$\text{gEstat}^N = \max_{1 \leq \Delta \leq t} \text{gEstat}_\Delta^N, \text{Estat}^N = \text{gEstat}_t^N \quad (21)$$

where $\pi_{t|t-\Delta}^{0,N,\text{par}}$ denotes a parametric approximation of $\pi_{t|t-\Delta}^{0,N}$. For changed observations, the PF gives $\pi_t^{c,0,N}(x)$ and so the above estimates will be $\text{Estat}^{c,0,N}$ and $\text{gEstat}^{c,0,N}$. If the state is scalar and $\pi_{t|t-\Delta}^{0,N}$ is approximated by a Gaussian, denoted $\mathcal{N}(x; \hat{X}_{t|t-\Delta}^0, \hat{\sigma}_{t|t-\Delta}^2)$, the above simplifies to

$$\text{gEstat}_\Delta = 0.5 \left[\frac{\sum_{i=1}^N (x_t^{(i)} - \hat{X}_{t|t-\Delta}^0)^2}{\hat{\sigma}_{t|t-\Delta}^2} - 1 \right] \quad (22)$$

3) *Defining $p_t^0, \pi_{t|t-\Delta}^0$* : If p_0^0 is Gaussian and the system model is linear Gaussian, then p_t^0 will also be Gaussian and can be defined in closed form by recursively applying the system model to $p_{\tau-1}^0$ for $\tau = 1, 2, \dots, t$. If the system model is nonlinear, Gaussian, it can be recursively linearized (first-order Taylor series about predicted state at $\tau - 1$) for each $\tau = 1, 2, \dots, t$ and applied to the estimate of $p_{\tau-1}^0(x)$. Computing gELL requires approximating the PF estimate of $\pi_{t|t_{\text{nc}}}^0$ (with $t_{\text{nc}} = t - \Delta$) by a parametric pdf. This can be done as follows: Approximate $\pi_{t_{\text{nc}}}^N$ by a Gaussian (or by a mixture of Gaussians) about the significant mode(s). Significant modes of $\pi_{t_{\text{nc}}}^N$ can be learned using any mode-finding technique, e.g., mean shift [39]. If the nominal state dynamics is linear Gaussian, then each mode can be propagated through it to get a Gaussian (or mixture of Gaussians) approximation to $\pi_{t|t_{\text{nc}}}^{0,N}$. If the state dynamics is nonlinear Gaussian, it can be linearized for $\tau = t_{\text{nc}} + 1, \dots, t$ and applied to each mode.

VI. DETECTING SLOW AND SUDDEN CHANGES

We track changed observations using a PF (or a KF) optimal for the unchanged system and so it approximates (or computes exactly) $\pi_t^{c,0}$ instead of $\pi_t^{c,c}$. We refer to the error $\|\pi_t^{c,0} - \pi_t^{c,c}\|_W$ as the “exact filtering error.” In addition, the PF uses a finite number of particles, N , to approximate $\pi_t^{c,0}$, i.e., it computes $\pi_t^{c,0,N}$. We refer to $\|\pi_t^{c,0,N} - \pi_t^{c,0}\|_W$ as the “PF approximation error.” We use these errors to explain why OL fails for slow changes while ELL fails for sudden changes. The discussion below and in the next section is for tracking using a PF.¹² We first precisely define OL and CUSUM-OL (maximum over partial sums of OL).

Definition 9 (OL, Ostat): OL [2] is the negative log likelihood of the current observation conditioned on past observations under the no change hypothesis, i.e.,

$$\text{OL}(Y_{1:t}) = -\log \mathbb{E}_{\pi_{t|t-1}}[\psi_{t,Y_t}(X_t)]. \quad (23)$$

The OL change detection statistic *Ostat* is

$$\text{Ostat}(Y_{1:t}) = \text{OL}(Y_{1:t}) - \mathbb{E}_{Y_{1:t}^0}[\text{OL}(Y_{1:t})] \quad (24)$$

¹²It can be applied to KF by replacing $c,0,N$ by $c,0$, but one may be able to get much stronger results by studying a linear Gaussian model.

where $\mathbb{E}_{Y_{1:t}^0}[\text{OL}(Y_{1:t})] = \mathbf{h}(Y_t^0|Y_{1:t-1}^0)$ is the conditional differential entropy of Y_t^0 given past observations. In general, it needs to be computed empirically by simulating and tracking the unchanged system many times and computing a Monte Carlo estimate.

One can use a sum of the OL values over a set of past time steps for robustness, i.e., define CUSUM-OL.

Definition 10 (CUSUM-OL): Is the maximum over partial sums of Ostat (a modification of the CUSUM idea [1], [40]):

$$c\text{OL}(Y_{1:t}) \triangleq \max_{1 \leq p \leq p_{\max}} \left[\sum_{\tau=t-p+1}^t \text{Ostat}(Y_{1:\tau}) \right]. \quad (25)$$

For changed observations, we compute $\text{OL}_t^{c,0,N} = -\log \mathbb{E}_{\pi_{t|t-1}^{c,0,N}}[q|Y_t^c(X_t)]$. As explained in Section II-C, a slow change implies small loss of track, and hence $\text{OL}_t^{c,0,N} \approx \text{OL}_t^{c,c}$. Now assuming that $\mathbf{h}(Y_t^c|Y_{1:t-1}^c) = \mathbf{h}(Y_t^0|Y_{1:t-1}^0)$, this implies that $P_{\text{detect}} = \Pr(\text{OL}_t^{c,0,N} - \mathbf{h}(Y_t^0|Y_{1:t-1}^0) > \kappa_{\text{OStat}}) \approx \Pr(\text{OL}_t^{c,c} - \mathbf{h}(Y_t^c|Y_{1:t-1}^c) > \kappa_{\text{OStat}}) \leq \Pr(|\text{OL}_t^{c,c} - \mathbf{h}(Y_t^c|Y_{1:t-1}^c)| > \kappa_{\text{OStat}}) \leq (VO_t^c/\kappa_{\text{OStat}}^2)$, where κ_{OStat} is the detection threshold and $VO_t^c = \text{Var}_{Y_{1:t}^c}[\text{OL}(Y_{1:t})]$. If VO_t^c is of the order of VO_t^0 and $\kappa_{\text{OStat}} = k\sqrt{VO_t^0}$, then the detection probability is as small as the false-alarm probability. CUSUM-OL is more robust, and it can detect some changes that OL misses (sum of small errors due to the slow change may be detectable). On the other hand, slow change (small loss of track) also means that $\text{ELL}_t^{c,0,N} \approx \text{ELL}_t^{c,c}$ and Fact 4 (which holds for $\text{ELL}_t^{c,c}$) can be used here. Thus, the change will get detected using ELL as soon as it is large enough to be detectable by $\text{ELL}_t^{c,c}$ (using Fact 4).

For a sudden change, the loss of track is large, and hence $\text{ELL}_t^{c,0,N}$ is significantly smaller (closer to unchanged ELL) than the true ELL $\text{ELL}_t^{c,c}$. This leads to significantly larger detection delay (than what it would be if using the true ELL). In fact, the PF error increases nonlinearly with increasing loss of track and hence the PF may completely stop following the change and hence ELL could completely miss such a change. However, large ELL approximation error also means that at least one of $\text{OL}_{t_c}^{c,0,N}, \dots, \text{OL}_t^{c,0,N}$ is large (from Theorem 1 in the next section), and hence OL detects immediately.

The above discussion extends directly to $g\text{ELL}_\Delta$ assuming $\Delta > t - t_c$ and negligible error in approximating $\pi_{t|t_{nc}}^0$.

Thus, if the type of change (slow or sudden) is not known, one should use both ELL (or $g\text{ELL}$) and OL (or CUSUM-OL) and declare a change if either exceeds its threshold. We summarize this in Fig. 1.

Example 2 (Computing and Comparing Estat and Ostat): We consider a scalar linear Gaussian state-space model, for which Estat and Ostat can be computed in closed form. Let the unchanged and changed system model be as in Example 1. Let the observation model be $Y_t = HX_t + w_t$, $w_t \sim \mathcal{N}(0, \sigma_{\text{obs}}^2)$ with $\sigma_{\text{obs}}^2 = \gamma\sigma_{\text{sys}}^2$. Let $H = 1$. For simplicity of notation, assume that steady state of the KF has been reached. The steady-state prediction variance $\hat{\sigma}_p^2$ is the unique stabilizing solution to the

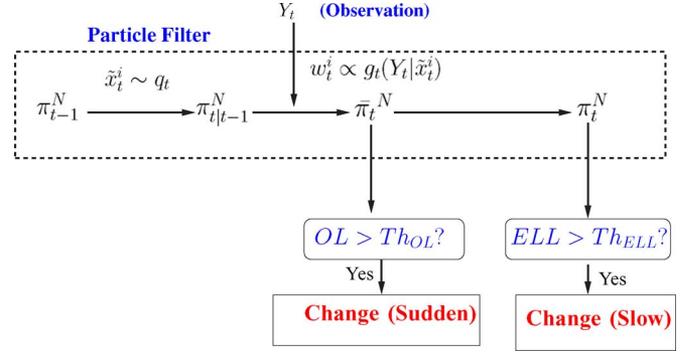


Fig. 1. Change Detection Algorithm: Note that ELL refers to Estat, OL refers to Ostat. Th_{ELL} and Th_{OL} are the detection thresholds for Estat and Ostat, respectively.

discrete-time algebraic Riccati equation [9]: i.e., $\hat{\sigma}_p^2 = P$, where P solves $P = P + \sigma_{\text{sys}}^2 - P^2 H^2 (\sigma_{\text{obs}}^2 + H^2 P)^{-1}$. The solution is $\hat{\sigma}_p^2 = (1 + \sqrt{1 + 4\gamma/2})\sigma_{\text{sys}}^2$. The steady-state Kalman gain is [9]: $K = \hat{\sigma}_p^2 / \gamma\sigma_{\text{sys}}^2 + \hat{\sigma}_p^2$. The steady state of $\hat{\sigma}_t^2$ (updated variance) is $\hat{\sigma}^2 = (1 - K)^2 \hat{\sigma}_p^2$. Also, the steady-state value of observation prediction error variance (variance of $p(Y_t|Y_{1:t-1})$) is $\hat{\sigma}_{Y_{\text{pred}}}^2 = \hat{\sigma}_p^2 + \gamma\sigma_{\text{sys}}^2$.

Computing Estat, Ostat, and the Thresholds:

$$\text{Estat}(Y_{1:t}) = 0.5 \frac{\hat{X}_t^2 + \hat{\sigma}^2}{t\sigma_{\text{sys}}^2} - 0.5 \triangleq \text{Estat}_t. \quad (26)$$

As before, we have $VK_t^0 \leq 0.5$. Thus, once again, we set the threshold $\kappa_{\text{Estat}} = 2.12$, which ensures that

$$\Pr(\text{Estat}_t^0 > \kappa_{\text{Estat}}) < 0.11. \quad (27)$$

For the linear Gaussian system, OL is computed as $\text{OL}(Y_{1:t}) = 0.5((Y_t - H\hat{X}_{t-1})^2 / \hat{\sigma}_{Y_{\text{pred}}}^2) + \text{const} \triangleq \text{OL}_t$, and so we have

$$\mathbb{E}_{Y_{1:t}^0}[\text{OL}_t] = \mathbb{E}_{Y_{1:t-1}^0}[\text{OL}_t] = 0.5 + \text{const} \quad (28)$$

$$\text{Ostat}(Y_{1:t}) = 0.5 \frac{(Y_t - H\hat{X}_{t-1})^2}{\hat{\sigma}_p^2 + \gamma\sigma_{\text{sys}}^2} - 0.5 \triangleq \text{Ostat}_t \quad (29)$$

$$VO_t^0 = \mathbb{E}_{Y_{1:t-1}^0}[(\text{Ostat}_t)^2] = 0.5. \quad (30)$$

Here again, we can set $\kappa_{\text{OStat}} = 2.12$, which ensures that

$$\Pr(\text{Ostat}_t^0 > \kappa_{\text{OStat}}) < 0.11. \quad (31)$$

Comparing Detection Delays: We compute $\text{Estat}_t^{c,0}$ and $\text{Ostat}_t^{c,0}$ at $t = t_c + \Delta$. Applying the KF [9] to changed observations, $Y_t^c = Y_t^0 + (\Delta + 1)Hb$, we have

$$\hat{X}_t^{c,0} = \hat{X}_t^0 + C_\Delta, \quad \text{where}$$

$$C_\Delta \triangleq \sum_{j=0}^{\Delta} (\Delta + 1 - j)(1 - KH)^j KHb \quad (32)$$

$$\text{Estat}_{t_c+\Delta}^{c,0} = 0.5 \frac{(\hat{X}_{t_c+\Delta}^0 + C_\Delta)^2 + (1-K)^2 \sigma_p^2}{(t_c + \Delta)\sigma_{\text{sys}}^2} - 0.5. \quad (33)$$

Now, (26) and (27) imply that at any t , w.p. ≥ 0.89

$$0.5 \frac{(\hat{X}_t^0)^2 + (1-K)^2 \sigma_p^2}{t \sigma_{\text{sys}}^2} \leq 0.5 + \kappa_{\text{Estat}} = 2.62.$$

Thus, $(\hat{X}_t^0)^2 < 5.24 t \sigma_{\text{sys}}^2$. (34)

Thus, at $t = t_c + \Delta$, w.p. ≥ 0.89

$$\text{Estat}_{t_c+\Delta}^{c,0} > 0.5 \left(-\sqrt{5.24} + \frac{C_\Delta}{\sqrt{t_c + \Delta} \sigma_{\text{sys}}} \right)^2 + 0.5 \frac{(1-K)^2 \sigma_p^2}{(t_c + \Delta) \sigma_{\text{sys}}^2} - 0.5. \quad (35)$$

The least Δ for which $\text{Estat}_{t_c+\Delta}^{c,0}$ will exceed κ_{Estat} w.p. ≥ 0.89 is obtained by finding the smallest Δ that satisfies

$$0.5 \left(-\sqrt{5.24} + \frac{C_\Delta}{\sqrt{t_c + \Delta} \sigma_{\text{sys}}} \right)^2 + 0.5 \frac{(1-K)^2 \sigma_p^2}{(t_c + \Delta) \sigma_{\text{sys}}^2} \geq 2.62. \quad (36)$$

We denote the solution by $\Delta_{\text{delay,max,ELL}}$, i.e., w.p. ≥ 0.89 the delay will be less than/equal to this value. Also

$$\text{Ostat}_{t_c+\Delta}^{c,0} = 0.5 \frac{(d_{Y,t_c+\Delta}^0 + (\Delta + 1)Hb - HC_{\Delta p})^2}{\sigma_p^2 + \gamma \sigma_{\text{sys}}^2} - 0.5$$

$$\text{where } d_{Y,t}^0 \triangleq (Y_t^0 - H\hat{X}_{t-1}^0). \quad (37)$$

Here, (29), (31), and (37) imply that w.p. ≥ 0.89

$$d_{Y,t}^0{}^2 \leq 5.24 (\sigma_p^2 + \gamma \sigma_{\text{sys}}^2). \quad (38)$$

Thus, the least Δ for which $\text{Ostat}_{t_c+\Delta}^{c,0}$ will exceed κ_{OStat} w.p. ≥ 0.89 is obtained by finding the smallest Δ that satisfies

$$0.5 \left(-\sqrt{5.24} + \frac{(\Delta + 1)Hb - HC_{\Delta-1}}{\sqrt{\sigma_p^2 + \gamma \sigma_{\text{sys}}^2}} \right)^2 > 2.62. \quad (39)$$

We denote the solution by $\Delta_{\text{delay,max,OL}}$, i.e., w.p. ≥ 0.89 the delay will be less than/equal to this value.

We used $\gamma = 1/5$, $\gamma = 5$ and varied $r = b/\sigma_{\text{sys}}$ and computed $\Delta_{\text{delay,max,ELL}}$, $\Delta_{\text{delay,max,OL}}$ by solving (36) and (39) in MATLAB. We used $t_c = 5$. The plots are shown in Fig. 2. For $\gamma = 5$ (observations less reliable than the system model), $K = 0.3583$. Thus, C_Δ is smaller for $\gamma = 5$ than for $\gamma = 1/5$ (observations more reliable than the system model), where $K = 0.8541$. Thus, maximum detection delay using ELL is smaller for $\gamma = 1/5$. Also, using OL, for $r < 5$, the maximum delay is ∞ while for $r \geq 5$, the maximum delay is 2 or less. When the change is tracked, OL is using only a single time instant value of the change magnitude, and hence is able to detect only large magnitude per unit time changes. ELL, on the other hand, uses the entire tracked part of the change, and hence

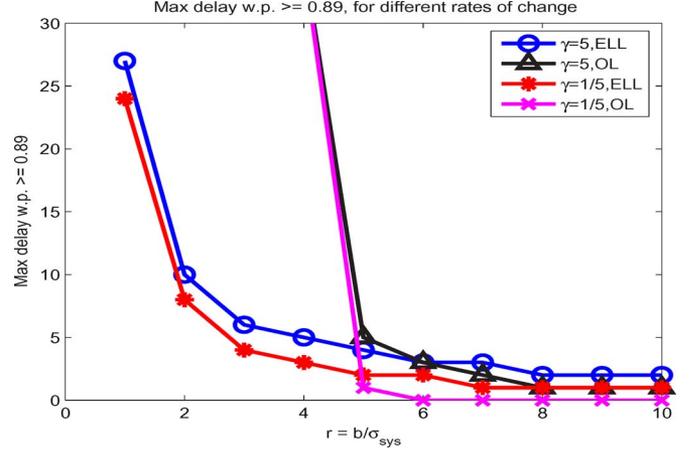


Fig. 2. $\Delta_{\text{delay,max,ELL}}$ and $\Delta_{\text{delay,max,OL}}$ as a function of r for $\gamma = 1/5$ and $\gamma = 5$. For $r < 5$, the maximum delay using OL is ∞ .

detects slower changes faster than OL. We have compared only ELL and OL here. Both cOL and gELL will perform better.

ELL-OL Complementarity: Now, the true value of ELL is

$$\text{Estat}_{t_c+\Delta}^{c,c} = 0.5 \frac{(\hat{X}_\Delta^0 + (\Delta + 1)b)^2 + \hat{\sigma}^2}{(t_c + \Delta) \sigma_{\text{sys}}^2} - 0.5 \geq \text{Estat}_{t_c+\Delta}^{c,0}. \quad (40)$$

From (33) and (40), the ELL error $e_{t_c+\Delta}^{c,0} = \text{Estat}_{t_c+\Delta}^{c,c} - \text{Estat}_{t_c+\Delta}^{c,0}$ is directly proportional to $(\Delta + 1)b - C_\Delta$. Also, $\text{Ostat}_{t_c+\Delta}^{c,0}$ is directly proportional to $(\Delta + 1)b - C_\Delta$. Thus, if $(\Delta + 1)b - C_\Delta$ decreases, ELL error decreases (ELL detects change as soon as detectable) and OL also decreases (OL fails to detect), and vice versa. This complementarity is proved for tracking using a PF in the next section. The PF error increases nonlinearly with loss of track [41], resulting in faster loss of track.

VII. COMPLEMENTARITY OF ELL AND OL

We prove here that the error in approximating ELL of changed observations using a PF optimal for the unchanged system is upper bounded by an increasing function of OL. We later discuss how this implies complementarity of ELL and OL. First we define some terms and notation.

A. Notation and Definitions

For any nonnegative measure μ , defined on state space E , $\mu(E) \triangleq \int_{x \in E} \mu(dx)$. For any finite nonnegative measure μ , the normalized measure is $\bar{\mu}(dx) \triangleq \mu(dx)/\mu(E)$. A nonnegative state transition kernel J , defined on the state space E , operates on a nonnegative measure μ , as $J(\mu)(dx') \triangleq \int_{x \in E} J(x, dx') \mu(dx)$ [8]. The normalized nonnegative operator \bar{J} is defined by $\bar{J}(\mu)(dx) \triangleq J(\mu)(dx)/J(\mu)(E)$ [8]. Inner product is defined by $(\mu, f) \triangleq \int_{x \in E} f(x) \mu(dx)$.

Definition 11 (Mixing, Mixing Parameter): A nonnegative kernel J defined on E is mixing [8] if there exists a constant,

$0 < \epsilon < 1$ (called the mixing parameter), and a nonnegative measure λ s.t. $\epsilon\lambda(A) \leq J(x, A) \leq (1/\epsilon)\lambda(A) \forall x \in E$ and for any Borel subset $A \subset E$.

Definition 12 (Unnormalized Filter Kernel): The unnormalized filter kernel [8], for a system with state transition kernel Q_t and probability of observation given state $\psi_{t,Y_t}(x)$, is $R_t(x, dx') = Q_t(x, dx')\psi_{t,Y_t}(x')$. So $R_t^{c,0} = Q_t^0\psi_{t,Y_t^c}$ is the unnormalized filter kernel for changed system observations tracked using unchanged system model Q_t^0 . Similarly, $R_t^{0,0} = Q_t^0\psi_{t,Y_t^0}$, $R_t^{c,c} = Q_t^c\psi_{t,Y_t^c}$.

Definition 13 (Posterior State Space E_{Y_t}): For an observation Y_t , the posterior state space E_{Y_t} is defined as $E_{Y_t} \triangleq \{x \in E : \psi_{t,Y_t}(x) > 0\}$. If observation noise is bounded (e.g., truncated Gaussian), then E_{Y_t} is compact.

B. Complementarity Result

Assuming that the unnormalized filter kernels $R_k^{c,0}$ and $R_k^{0,0}$ are mixing (i.e., their respective mixing parameters $\epsilon_k^{c,0} < 1$, $\epsilon_k^{0,0} < 1$) and that the posterior state space E_{Y_t} is a compact and proper subset of E , the ELL approximation error for $t > t_c$, $e_t^{c,0,N} \triangleq |\text{ELL}_t^{c,c} - \text{ELL}_t^{c,0,N}|$, obeys [34], [35]

$$e_t^{c,0,N} \leq M_t \left\| \pi_t^{c,c} - \pi_t^{c,0,N} \right\|_W \leq M_t \left(\theta_t^{c,0} + \frac{\beta_t^{c,0}}{\sqrt{N}} \right) \quad (41)$$

where $M_t = \sup_{x \in E_{Y_t}} [-\log p_t^0(x)]$, $\theta_t^{c,0}$ (the exact filtering error bound) and $\beta_t^{c,0}$ (the PF approximation error bound coefficient) are defined as follows [8], [34], [35]:

$$\theta_t^{c,0} \triangleq \delta_t^{c,0} + \frac{2\delta_{t-1}^{c,0}}{\epsilon_t^{c,0}{}^2} + \frac{4}{\log 3} \times \sum_{k=t_c}^{t-2} \tilde{\tau}_{t:k+3}^{c,0} \frac{\delta_k^{c,0}}{\epsilon_{k+1}^{c,0} \epsilon_{k+2}^{c,0}} \quad (42)$$

$$\beta_t^{c,0} \triangleq 2 \left(\rho_t^{c,0} + \frac{2\rho_{t-1}^{c,0}}{\epsilon_t^{c,0}{}^2} + \frac{4}{\log 3} \times \sum_{k=t_c}^{t-2} \tilde{\tau}_{t:k+3}^{c,0} \frac{\rho_k^{c,0}}{\epsilon_{k+1}^{c,0} \epsilon_{k+2}^{c,0}} \right) + K \quad (43)$$

$$\text{where } K \triangleq \frac{8}{\log 3} \sum_{k=1}^{t_c-1} \tilde{\tau}_{t:k+3}^{0,0} \frac{\rho_k^{0,0}}{\epsilon_{k+1}^{0,0} \epsilon_{k+2}^{0,0}}, \quad (44)$$

$$\delta_k^{c,0} \triangleq \sup_{\phi: \|\phi\|_\infty \leq 1} \left| \left(\pi_k^{c,0} - \bar{R}_k^{c,c}(\pi_{k-1}^{c,0}), \phi \right) \right| \leq 2 \quad (45)$$

$$\rho_k^{c,0} \triangleq \frac{S}{\inf_{\mu \in \mathcal{P}(E)} \left(Q_k^0(\mu), \psi_{k,Y_k^c} \right)} \quad (46)$$

$$\tilde{\tau}_k^{c,0} \triangleq \frac{1 - \epsilon_k^{c,0}{}^2}{1 + \epsilon_k^{c,0}{}^2} < 1 \text{ and } S \triangleq \sup_{x \in E} \psi_{k,Y_k^c}(x) \quad (47)$$

are defined in [8]. Note for $t < t_c$, $\rho_k^{c,0} = \rho_k^{0,0}$, $\epsilon_k^{c,0} = \epsilon_k^{0,0}$, $\tilde{\tau}_k^{c,0} = \tilde{\tau}_k^{0,0}$ and $\delta_t^{c,0} = 0$. This result is based on Theorems 4.8 and 5.8 of [8]. Also, $\text{OL}_k^{c,0}$ can be expressed as

$$e^{-\text{OL}_k^{c,0}} = R_k^{c,0} \left(\pi_{k-1}^{c,0} \right) (E) = \left(Q_k^0 \left(\pi_{k-1}^{c,0} \right), \psi_{k,Y_k^c} \right). \quad (48)$$

First, consider $\beta_t^{c,0}$. It depends on past values of $\rho_k^{c,0}$ and $\epsilon_k^{c,0}$. Using Remark 5.11 of [8] and (48) above, we have the following bounds on $\rho_k^{c,0}$:

$$\rho_k^{c,0} \leq \frac{S}{\left(\epsilon_k^{c,0} \right)^2 \left(Q_k^0 \left(\pi_{k-1}^{c,0} \right), \psi_{k,Y_k^c} \right)} = \frac{S e^{\text{OL}_k^{c,0}}}{\left(\epsilon_k^{c,0} \right)^2}. \quad (49)$$

Next, consider $\theta_t^{c,0}$. It depends on past values of $\delta_k^{c,0}$ and $\epsilon_k^{c,0}$. We use inequality (6) of [8], which states that

$$\|\bar{\mu} - \mu'\| \leq \frac{\|\mu - \mu'\|}{\mu(E)} + \frac{|\mu(E) - \mu'(E)|}{\mu(E)}. \quad (50)$$

Also note that $\delta_k^{c,0} \leq \|\pi_k^{c,0} - \bar{R}_k^{c,c}(\pi_{k-1}^{c,0})\|$ and that $\pi_k^{c,0} = \bar{R}_k^{c,0}(\pi_{k-1}^{c,0})$. Applying (50) with $\mu = R_k^{c,0}(\pi_{k-1}^{c,0})$ and $\mu' = R_k^{c,c}(\pi_{k-1}^{c,0})$ and using (48) to replace the denominator of both terms, we get that

$$\delta_k^{c,0} \leq e^{\text{OL}_k^{c,0}} \left[\left\| R_k^{c,0} \left(\pi_{k-1}^{c,0} \right) - R_k^{c,c} \left(\pi_{k-1}^{c,0} \right) \right\| + \left| R_k^{c,0} \left(\pi_{k-1}^{c,0} \right) (E) - R_k^{c,c} \left(\pi_{k-1}^{c,0} \right) (E) \right| \right]. \quad (51)$$

Applying inequalities (7) and (8) from [41, Appendix] to bound the two terms inside the square brackets, we get

$$\delta_k^{c,0} \leq 2D_{Q,k} e^{\text{OL}_k^{c,0}} \quad (52)$$

where $D_{Q,k} \triangleq \int_{x \in E} \psi_{t,Y_t}(x') \sup_x |q_t^c(x, x') - q_t^0(x, x')| dx'$ is a distance between the state transition kernels Q_t^c and Q_t^0 [41]. Now, it is easy to see that $D_{Q,k} \leq 2S$, where S is defined in (47). We also know that $\delta_k^{c,0} \leq 2$ (from (45)). Combining these two facts, we get that

$$\delta_k^{c,0} \leq \min \left[2, 4S e^{\text{OL}_k^{c,0}} \right]. \quad (53)$$

Thus, combining (41), (42), and (43) with (49) and (53), we have the following result.

Theorem 1 (ELL-OL Complementarity): Assuming that the unnormalized filter kernels $R_k^{c,0}(x, dx')$, $R_k^{0,0}(x, dx')$ are mixing (with mixing parameters $\epsilon_k^{c,0}$, $\epsilon_k^{0,0}$) and that the posterior state space E_{Y_t} is a compact and proper subset of E for all Y_t , the ELL approximation error for $t > t_c$, $e_t^{c,0,N}$ is upper bounded by an increasing function of $\text{OL}_k^{c,0}$, $k = t_c, \dots, t$, as given below:

$$e_t^{c,0,N} \leq M_t \left\| \pi_t^{c,c} - \pi_t^{c,0,N} \right\|_W \leq M_t \sum_{k=t_c}^t \left[\left(\frac{2S e^{\text{OL}_k^{c,0}}}{\sqrt{N} \left(\epsilon_k^{c,0} \right)^2} + \min \left(2, 4S e^{\text{OL}_k^{c,0}} \right) \right) \omega_{k,t} \right] + K$$

where

$$\omega_{k,t} \triangleq \begin{cases} 1, & k = t, \\ \frac{2}{\epsilon_t^{c,0}{}^2}, & k = t-1 \\ \frac{4}{\log 3} \tilde{\tau}_{t:k+3}^{c,0} \frac{1}{\left(\epsilon_{k+1}^{c,0} \right)^2 \left(\epsilon_{k+2}^{c,0} \right)^2}, & k < t-1 \end{cases} \quad (54)$$

and $M_t = \sup_{x \in E_{Y_t}} [-\log p_t^0(x)]$, S , and $\tau_k^{c,0}$ are defined in (47), and K is defined in (44).

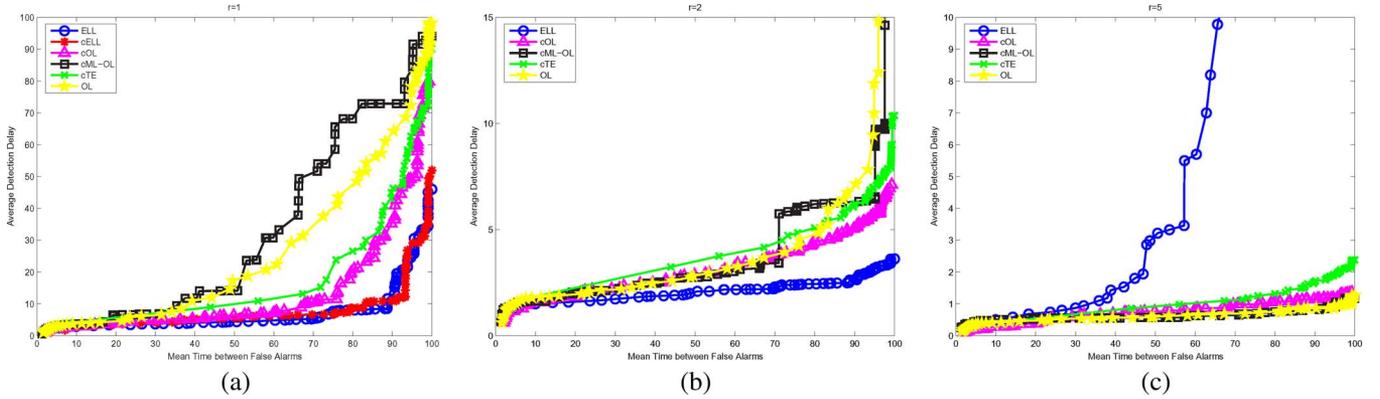


Fig. 3. Example 1, Change time $t_c = 5$: ROC curves comparing performance of ELL and CUSUM-ELL with CUSUM-ML-OL (cML-OL), OL, CUSUM-OL (cOL), and CUSUM-TE (cTE). Note that we mean Estat and Ostat wherever we say ELL and OL. (a) Slow ($r = 1$), $t_c = 5$, $t_f = 15$; (b) faster ($r = 2$), $t_c = 5$, $t_f = 15$; and (c) sudden ($r = 5$), $t_c = 5$, $t_f = 15$.

C. Implication and Extension to CUSUM-OL and gELL

If a change is not detected by OL until time t , it means that all values of OL , $OL_{t_c}^{c,0}, \dots, OL_k^{c,0}, \dots, OL_t^{c,0}$ are below the detection threshold and, hence, small. By the above theorem, this implies that the loss of track is small (slow change) and that the bound on ELL approximation error $e_t^{c,0,N}$ is also small. This implies that the estimated ELL ($ELL_t^{c,c}$) behaves like the true ELL ($ELL_t^{c,c}$), i.e., the change gets detected when true ELL is large enough to be “detectable” (using Fact 3).

If estimated, ELL does not detect a change that is “detectable”; it means that ELL approximation error is large. By the above theorem, this implies that loss of track is large (sudden change) and that at least one of $OL_{t_c}^{c,0}, \dots, OL_k^{c,0}, \dots, OL_t^{c,0}$ is large. Hence, OL will detect it.

Similarly, if CUSUM-OL (max over partial sums of OL) fails to detect until t , it also means that OL_{t_c}, \dots, OL_t are small and so ELL will eventually detect. If ELL fails, at least one of OL_{t_c}, \dots, OL_t are large and so CUSUM-OL will be large. The above theorem also carries over to gELL versus OL, if we assume negligible error in approximating $\pi_{t_{nc}}^0$ and choice of $t_{nc} < t_c$. The only difference is that, in the definition of K in (44), the summation starts at $k = t_{nc}$ (instead of $k = 1$).

Thus, if OL (or CUSUM-OL) fails to detect, ELL (or gELL) will eventually detect. If ELL (or gELL) fails to detect, OL (or CUSUM-OL) will detect. If loss of track is large enough to be detectable by OL, but small enough for ELL error to not be too large, both will detect.

VIII. SIMULATION AND EXPERIMENTAL RESULTS

We simulated the system defined in Example 1, a nonadditive change (a change from asymptotically stationary system to a random walk model), a nonlinear system model problem from [10] and the bearings-only tracking problem [10] (with additive change in x direction velocity). Performance of ELL, gELL, and CUSUM-ELL (cELL) (defined exactly like CUSUM-OL in Definition 10) was compared against that of OL, CUSUM-OL (cOL), CUSUM-TE (cTE) (defined exactly like CUSUM-OL in Definition 10) and CUSUM-ML-OL (cML-OL), described in Section I. CUSUM-ML-OL involves running t PFs, with the maximum-likelihood estimate of the changed parameter evaluated starting at $\tau = t - 1, t - 2, \dots, 1, 0$, respectively, for each

PF, where $\tau = 0$ corresponds to the unchanged system. It then evaluates CUSUM on the observation likelihood ratio as given in (1) (see [19] for details).

Performance was compared by plotting ROC curves. An ROC curve for a change detection problem [16] is a plot of the average detection delay against the average mean time between false alarms (MTBFA). It is obtained by varying the detection threshold (in an appropriate range) and evaluating average detection delay and average time between false alarms for each value of threshold.¹³ We simulated the nominal system 100 times and used the maximum and half the mean value of the detection statistic to define the limits of the range in which to vary the threshold. The average MTBFA is computed by simulating the nominal system 100 times, and average delay is computed by simulating the changed system 100 times.

Finally, we also show time plots of ELL, OL, and TE for detecting abnormal activity from a video sequence using the framework of [5].

A. Example 1, Small t_c

We simulated Example 1 discussed in Section III-B-1) with $t_c = 5$, $t_f = 15$. We used system noise variance $\sigma_{\text{sys}}^2 = 0.04$, observation noise variance $\sigma_{\text{obs}}^2 = 0.2$, and truncation parameter $B = 10\sigma_{\text{obs}}$. We tested detection performance for increasing magnitudes of r . We show ROC plots for $b = r\sigma_{\text{sys}}$ with $r = 1$ (slow), $r = 2$ (faster), and $r = 5$ (sudden) in Fig. 3. We simulated the system 100 times for 50 time steps. Computing cML-OL takes very long ($O(t^2)$ time to simulate t time steps) and, hence, its ROC was computed with only 40 simulations instead of 100 simulations for computing all other statistics. We used $N = 100$ particles for each PF simulation. As can be seen from Fig. 3, for the slow ($r = 1$) and the “faster” ($r = 2$) changes, ELL (blue “o”) has smaller detection delay (for the same level of MTBFA) than cOL (magenta “-Δ”), OL (yellow pentagons), and cTE (green “-x”). Also, cML-OL has the worst performance for the slow change, and we noticed from experiments that this was because the change parameter b was overestimated most of the time. For the sudden change ($r = 5$), all of cML-OL, cOL, OL, and cTE detect changes much faster

¹³Please note that the threshold is constant for all t .

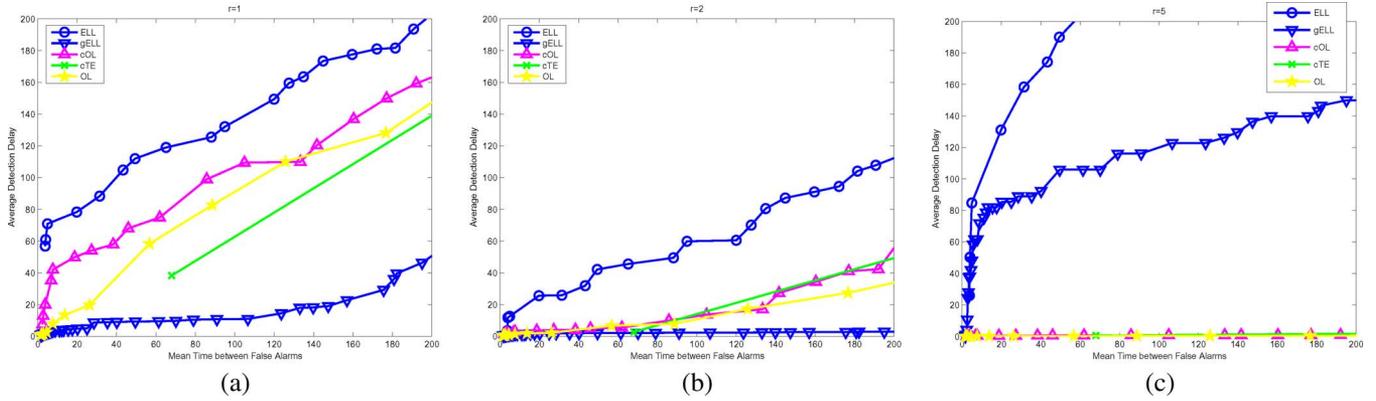


Fig. 4. Example 1, Change time $t_c = 100$: ROC curves comparing performance of gELL with ELL and with OL, CUSUM-OL (cOL), and CUSUM-TE (cTE). Note that we mean Estat, gEstat and Ostat wherever we say ELL, gELL, and OL. (a) Slow ($r = 1$), $t_c = 100, t_f = 110$; (b) faster ($r = 2$), $t_c = 100, t_f = 110$; and (c) sudden ($r = 5$), $t_c = 100, t_f = 110$.

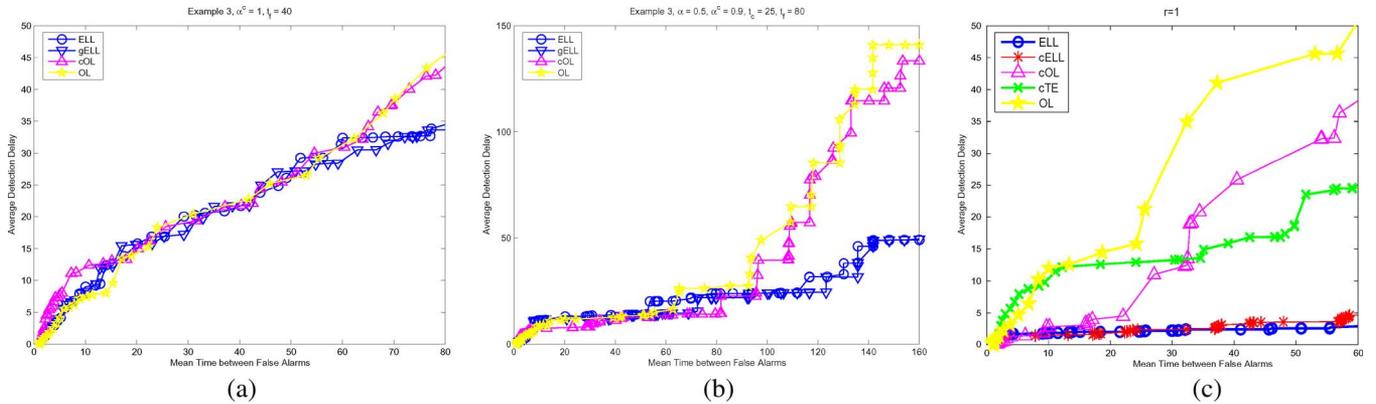


Fig. 5. Change detection ROCs for Example 3 and the nonlinear system example of Section VII-D. (a) Example 3, $\alpha^c = 1, t_c = 25, t_f = 40$; (b) Example 3, $\alpha^c = 0.9, t_c = 25, t_f = \infty$; and (c) nonlinear system (slow change, $t_c = 5$).

than ELL or cELL. In particular, cML-OL had the best performance. Note that cOL performs much better than OL, while cELL (red “-”) and ELL have similar performance. This is because ELL already uses the total magnitude of change until t , while OL uses only the change at the current time instant (when the change is tracked). cELL is shown only in Fig. 3(a). Also, gELL is not shown since its performance is similar to that of ELL for small t_c .

B. Example 1, Large t_c

We also simulated Example 1 100 times, with $t_c = 100$ and $t_f = 110$. ROC plots are shown in Fig. 4 for $r = 1, r = 2$, and $r = 5$. Here, gELL performs significantly better than ELL for reasons discussed earlier in Section III-C. Also, gELL significantly outperforms OL, cOL, and cTE for $r = 1, r = 2$, while it fails for $r = 5$.

C. Asymptotically Stationary Nominal State, Nonadditive Change, Large t_c

Example 3: (Asymptotically Stationary Nominal State, Non-additive Change): Assume scalar state and observation $h_t(x) = x^3$ and additive truncated Gaussian observation noise. Let the unchanged system dynamics be $X_0 = 0, X_t = \alpha X_{t-1} + n_t, n_t \sim \mathcal{N}(0, \sigma_{\text{sys}}^2)$ with $\alpha = 0.5$. Thus, $p_t^0(x) = \mathcal{N}(x; 0, \sigma_t^0)$,

with $\sigma_t^0 \rightarrow \sigma_0^2 = \sigma_{\text{sys}}^2 / (1 - \alpha^2)$ as $t \rightarrow \infty$ (asymptotically stationary).

Consider a change to $\alpha^c = 1$ (change from stable system to random walk) for $t \in [t_c, t_f]$. After $t_f, \alpha = 0.5$ again. Thus, the changed system dynamics is $X_t = X_{t-1} + n_t$. For $t_c \leq t \leq t_f$, the change $b_t(X_{t-1}) = (1 - \alpha)X_{t-1} = 0.5X_{t-1}$ and so $\mathbb{E}_{Y_{1:t-1}^c}[r_t^2] \approx 0.25[0.75 + t - t_c + 1]$ if $\sigma_t^0 \approx \sigma_0$ (system reached steady state before change). Also, $EK_t^c - EK_t^0 = 0.5(\sigma_t^c^2 - \sigma_t^0^2 / \sigma_t^0^2)$. Thus, for $t_c \leq t \leq t_f, EK_t^c - EK_t^0 \geq (t - t_c + 1)(1 - \alpha^2)$ and decreases slowly to zero after t_f . In this case, the increase in $EK_t^c - EK_t^0$ is due to the increase in average X_t^2 (due to increase in α). Thus, there is more variability for different realizations of X_t : in some cases, the change is easily detected; in others, it will not be detectable even if X_t were fully observed. This is observed as larger average detection delays [see Fig. 5(a)]. Also, since the nominal state is asymptotically stationary, variance of p_t^0 is bounded, and so ELL and gELL have similar performance (seen in the figure).

The ROC plots of Fig. 5(a) were generated by simulating the above example 100 times with $\alpha^c = 1$ between $t_c = 25$ to $t_c = 35$. We used $\sigma_{\text{sys}}^2 = 0.2$, observation noise variance, $\sigma_{\text{obs}}^2 = 1$, and $B = 10\sigma_{\text{obs}}$. Note that we track the observations with a random walk model, i.e., in the PF, $\alpha = 1$ for all t . This introduces a small extra error in tracking the unchanged

system but tracks the changed system much better. ELL (and gELL) outperform OL and cOL for smaller thresholds (larger MTBFA).

Now, consider a change from $\alpha = 0.5$ to $\alpha^c = 0.9$, but now let¹⁴ $t_f = \infty$. This change is also detected using ELL and gELL faster than OL or cOL, as can be seen from Fig. 5(b).

D. Nonlinear State Dynamics

We study here an example from [10] (a univariate nonstationary growth model) with $X_t = X_{t-1} + 25(X_{t-1}/1 + X_{t-1}^2) + 8 \cos(1.2(t-1)) + n_t$. The observation model is $Y_t = (X_t^2/20) + w_t$. Here, n_t is Gaussian system noise with $\sigma_{\text{sys}}^2 = 10$, and w_t is a truncated Gaussian observation noise with $\sigma_{\text{obs}}^2 = 1$. Initial state is taken to be known at zero. We introduce a change by adding a bias $b = r\sigma_{\text{sys}}$ to the state equation for 11 time instants beginning at $t_c = 5$. By using an approximate linearization of the nominal system model,¹⁵ we take $p_t^0(x)$ to be a Gaussian with mean $\sum_{\tau=1}^t 8 \cos(1.2(\tau-1))$ and variance $t\sigma_{\text{sys}}^2$. We show in Fig. 5(c), the ROC plots comparing the various statistics for detecting a slow change ($r = 1$) in the above setup. Here again, ELL and cELL detect the slow change much faster than OL, cOL, or cTE.

E. Bearings-Only Tracking

Bearings-only tracking is a common technique for many applications using radar, sonar, or infra red (IR) sensor information in the passive mode [14]. Here, one assumes that the target moves on the x - y plane according to the standard second order motion model (a constant velocity model with Gaussian system noise) [10]. The state vector consists of the x - y location and x - y velocities. The observation Y_t is a noisy measurement of the target bearing $Y_t = \tan^{-1}(x_{2,t}/x_{1,t}) + w_t$, where $x_{2,t}$ denotes the y location and $x_{1,t}$ denotes the x location. Here, the system model is linear Gaussian and hence $p_t^0(x)$ can be defined in closed form. We attempt to detect a change in the state dynamics where the change is due to an additive bias added to only the x direction velocity, for ten time steps beginning at $t = 5$. We show ROC plots for $r = 1$ (slow change) in Fig. 6.

F. Abnormal Activity Detection

We now show application to the problem of abnormal activity detection from video sequences [5], where we defined a general HMM to represent the normal activity. We proposed in [5] a stochastic landmark shape [42] dynamical model for modeling the changing configuration of a group of moving and interacting objects (here persons) treated as point objects. The shape and scaled Euclidean motion at t constituted the state vector, i.e., $X_t = [c_t, s_t, \theta_t]$. Here, c_t denotes the tangent coordinate of the shape at t , z_t , in the tangent space at the mean shape, μ ,

¹⁴If t_f is finite, the system will return to the stationary value of unchanged system soon after t_f , and it may not get detected before t_f .

¹⁵Note that one does not need an accurate characterization of p_t^0 . Knowing its mean at each time instant and defining an appropriate distance function from the mean, to replace $[-\log p_t(x)]$ is sufficient in practice.

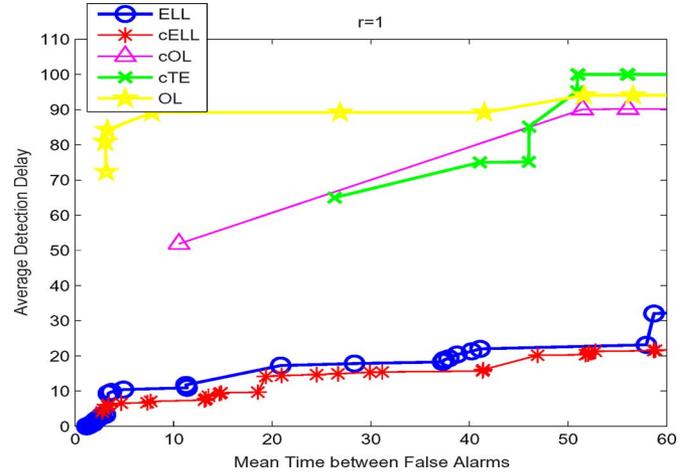


Fig. 6. Change detection ROC for slow change detection in the bearings-only tracking example (Section VII-E), $t_c = 5$, $t_f = 15$.

s_t is global scale and θ_t is global rotation. The noisy measurements of objects' configuration (after translation normalization) formed the observation vector Y_t . The observation model was¹⁶

$$Y_t = h(X_t) + w_t, \quad h(X_t) = z_t(c_t, \mu) s_t e^{-j\theta_t} \quad (55)$$

where $z_t(c_t, \mu) = (1 - c_t^* c_t)^{1/2} \mu + U(\mu) c_t$, and $U(\mu)$ is an orthogonal basis for the tangent space at μ . The system model was

$$\begin{aligned} c_t &= A_c c_{t-1} + n_t, & n_t &\sim \mathcal{N}(0, \Sigma_{n,c,t}) \\ \log s_t &= \alpha_s \log s_{t-1} + (1 - \alpha_s) \mu_s + n_{s,t}, & n_{s,t} &\sim \mathcal{N}(0, \sigma_r^2) \\ \theta_t &= \alpha_\theta \theta_{t-1} + (1 - \alpha_\theta) \mu_\theta + n_{\theta,t}, & n_{\theta,t} &\sim \mathcal{N}(\mu_\theta, \sigma_\theta^2) \end{aligned}$$

We studied the problem of detecting the change in shape introduced by one person walking away from his normal path. Abnormality was defined as a change in the shape dynamics (change in c_t) with change parameters unknown and the change being slow or sudden. The dynamics of c_t is linear Gaussian and asymptotically stationary. So p_t^0 is defined in closed form and has bounded variance. Thus, here again, gELL is not needed. The speed at which the person walked away decided the slowness of the change. We show in Fig. 7(a)–(c), the plots of ELL, OL, and TE as a function of time for the no-change case (blue “-o”) and for increasing walk-away velocities, as follows: velocity = 1 (red “-*”), velocity = 4 (magenta “-Δ”), and velocity = 4 (cyan “-pentagon”). Abnormality began at $t_c = 5$. Velocity = 1 was a slow change that was detected by ELL (at $t = 13$) much faster than by OL or TE (at $t = 23$). For velocity = 32, OL or TE detected immediately. Also note that the TE plot is very similar to that of OL since TE is approximately equal to OL (to first order) for white Gaussian observation noise, which is what was used here.

IX. CONCLUSION

We have studied the change detection problem in partially observed nonlinear systems (satisfying the HMM property) when change parameters are unknown. We have proposed a statistic called ELL that uses the tracked part of the change to detect

¹⁶Complex number notation was used to simplify writing equations.

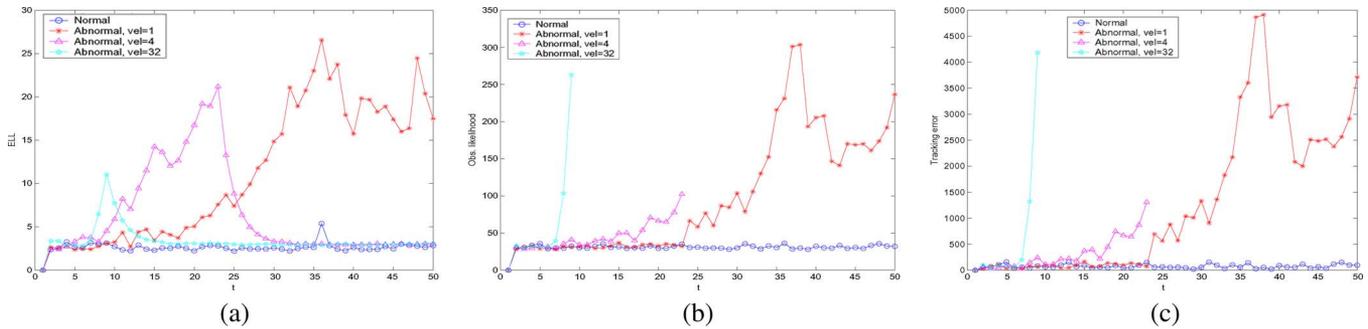


Fig. 7. Abnormal activity detection: ELL, OL, and TE plots for different rates of change. Abnormality begins at $t_c = 5$. ELL detects the slow abnormality much faster than either OL or TE. Also notice the similarity in the OL and TE plots. (a) ELL; (b) OL; and (c) TE.

it and, hence, detects changes for which loss of track is small. ELL can be understood as a measure of Kerridge inaccuracy [31] between the posterior at t and the t -step ahead prediction pdf of the unchanged state. We have proposed a generalization of ELL called gELL, where we replace the t -step ahead prediction pdf by a $\Delta < t$ -step-ahead prediction pdf. This improves sensitivity of ELL for large change times and also makes it possible to detect multiple changes in an observation sequence. We have shown, both analytically and experimentally, that ELL complements existing detection statistics such as OL, which are designed to detect the loss of track that occurs due to sudden changes.

Ongoing work includes a study of the assumptions required for stability of the errors in approximating ELL of changed observations using a PF designed for the unchanged system and of practical examples of systems satisfying these assumptions [43]. We have shown in [41] that the ELL error is upper bounded by an increasing function of change magnitude per unit time with all increasing derivatives. We are currently studying the implications of this result on the design of PF.

ACKNOWLEDGMENT

The author would like to thank R. Chellappa, A. Papamarcou, E. Slud, M. Salapaka, and A. Ramamoorthy for valuable suggestions and comments at various times in improving the writing of this paper.

REFERENCES

- [1] M. Basseville, "Model-based statistical signal processing and decision theoretic approaches to monitoring," in *Proc. 5th IFAC/IMACS Symp. Fault Detection, Supervision, Safety for Technical Processes (SAFE-PROCESS)*, Jun. 2003, pp. 1–12.
- [2] Q. Zhang, F. Campillo, F. Cerou, and F. LeGland, "Nonlinear system fault detection and isolation based on bootstrap particle filters," in *Proc. IEEE Conf. Decision Control (CDC)*, Dec. 2005, pp. 3821–3826.
- [3] P. Li and V. Kadiramanathan, "Fault detection and isolation in nonlinear stochastic systems—A combined adaptive Monte Carlo filtering and likelihood ratio approach," *Int. J. Control* vol. 77, pp. 1101–1114, 2004.
- [4] A. Doucet, N. deFreitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. New York: Springer, 2001.
- [5] N. Vaswani, A. RoyChowdhury, and R. Chellappa, "'Shape Activity': A continuous state HMM for moving/deforming shapes with application to abnormal activity detection," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1603–1616, Oct. 2005.
- [6] N. Vaswani and R. Chellappa, "Nonstationary shape activities," presented at the IEEE Conf. Decision Control (CDC), Seville, Spain, Dec. 2005.
- [7] B. Song, N. Vaswani, and A. Roy-Chowdhury, "Summarization and indexing of human activity sequences," in *Proc. IEEE Int. Conf. Image Processing*, 2006.
- [8] F. LeGland and N. Oudjane, "Stability and uniform approximation of nonlinear filters using the Hilbert metric, and application to particle filters," *Ann. Appl. Probab.*, vol. 14, no. 1, pp. 144–187, Feb. 2004.
- [9] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 2000.
- [10] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/nongaussian bayesian state estimation," *Proc. Inst. Elect. Eng.—F (Radar Signal Process.)*, vol. 140, no. 2, pp. 107–113, 1993.
- [11] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.
- [12] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*. New York: Academic, 1988.
- [13] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–285, Feb. 1989.
- [14] R. Karlsson and F. Gustafsson, "Recursive bayesian estimation—bearings-only applications," in *Proc. Inst. Elect. Eng.—Radar, Sonar, Navig. (Special Issue on Target Tracking: Algorithms and Applications)*, Oct. 2005, vol. 152, no. 5, pp. 305–313.
- [15] "Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex," *Nature Neurosci.*, vol. 6, no. 11, pp. 1216–1223, 2003.
- [16] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [17] F. Gustafsson, *Adaptive Filtering and Change Detection*. New York: Wiley, 2001.
- [18] B. Azimi-Sadjadi and P. S. Krishnaprasad, "Change detection for nonlinear systems: a particle filtering approach," in *Proc. Amer. Control Conf.*, 2002.
- [19] C. Andrieu, A. Doucet, S. S. Singh, and V. B. Tadic, "Particle methods for change detection, system identification, and control," *Proc. IEEE*, vol. 93, pp. 423–438, Mar. 2004.
- [20] P. Li and V. Kadiramanathan, "Particle filtering based likelihood ratio approach to fault diagnosis in nonlinear stochastic systems," *IEEE Trans. Syst., Man, Cybern. C*, vol. 31, pp. 337–343, Aug. 2001.
- [21] V. Kadiramanathan, P. Li, M. H. Jaward, and S. G. Fabri, "Particle filtering based fault detection in nonlinear stochastic systems," *Int. J. Syst. Sci.*, vol. 33, pp. 259–265, 2002.
- [22] C. Andrieu and A. Doucet, "Online expectation-maximization type algorithms for parameter estimation in general state space models," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2003, pp. VI-69–VI-72.
- [23] A. Doucet and V. B. Tadic, "Parameter estimation in general state-space models using particle methods," *Ann. Inst. Stat. Math.*, vol. 55, pp. 409–422, 2003.
- [24] J. Vermaak, C. Andrieu, A. Doucet, and S. J. Godsill, "Particle methods for bayesian modeling and enhancement of speech signals," *IEEE Trans. Speech Audio Process.*, vol. 10, pp. 173–185, Mar. 2002.
- [25] H. Hammouri, M. Kinnaert, and E. ElYaagoubi, "Observer-based approach to fault detection and isolation for nonlinear systems," *IEEE Trans. Autom. Control*, vol. 44, no. 10, pp. 1879–1884, Oct. 1999.
- [26] C. De Persis and A. Isidori, "A geometric approach to nonlinear fault detection and isolation," *IEEE Trans. Autom. Control*, vol. 46, no. 6, pp. 853–865, Jun. 2001.

- [27] A. Xu and Q. Zhang, "Nonlinear system fault diagnosis based on adaptive estimation," *Automatica*, pp. 1181–1193, 2004.
- [28] F. Cerou and F. Le Gland, "Efficient particle filters for residual generation in partially observed sde's," in *Proc. IEEE Conf. Decision Control (CDC)*, Dec. 2000.
- [29] D. F. Kerridge, "Inaccuracy and inference," *J. Roy. Statist. Soc., Ser. B*, vol. 23, 1961.
- [30] T. Cover and J. Thomas, *Elements of Information Theory*, ser. Wiley Series in Telecommunications. New York: Wiley, 1991.
- [31] R. Kulhavy, "A geometric approach to statistical estimation," in *Proc. IEEE Conf. Decision Control (CDC)*, Dec. 1995.
- [32] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 1991.
- [33] G. Casella and R. Berger, *Statistical Inference*, 2nd ed. Pacific Grove, CA: Duxbury Thomson Learning, 2002.
- [34] N. Vaswani, "Change detection in partially observed nonlinear dynamic systems with unknown change parameters," in *Proc. Amer. Control Conf. (ACC)*, 2004.
- [35] —, "Change detection in stochastic shape dynamical models with applications in activity modeling and abnormality detection," Ph.D. dissertation, Elect. Comput. Eng. Dept., Univ. Maryland, College Park, MD, Aug. 2004.
- [36] T.-J. Cham and J. M. Rehg, "A multiple hypothesis approach to figure tracking," in *Proc. IEEE Conf. Computer Vision Pattern Recognition (CVPR)*, 1999.
- [37] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
- [38] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering for practitioners," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 736–746, Mar. 2002.
- [39] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, 2002.
- [40] N. Vaswani, "The modified CUSUM algorithm for slow and drastic change detection in general HMMs with unknown change parameters," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2005.
- [41] —, "Bound on errors in particle filtering with incorrect model assumptions and its implication for change detection," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2004.
- [42] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*. New York: Wiley, 1998.
- [43] N. Vaswani, "Particle filtering under system model error and implications for change detection," in preparation.



Namrata Vaswani received the B.Tech. degree in electrical engineering from the Indian Institute of Technology (I.I.T.), Delhi, India, in August 1999 and the Ph.D. degree in electrical and computer engineering from the University of Maryland, College Park, in August 2004. Her Ph.D. dissertation was on change detection in stochastic shape dynamical models and applications to activity modeling and abnormal activity detection.

She was a Postdoctoral Fellow in the School of Electrical and Computer Engineering at the Georgia Institute of Technology in 2004–2005. She has been an Assistant Professor in the Department of Electrical and Computer Engineering at Iowa State University, Ames, since August 2005. Her research interests are in detection and estimation problems in statistical signal and video processing, computer vision and in biomedical image analysis. In particular, she is interested in particle filtering theory and applications in tracking and change detection and in shape analysis and filtering.