THE MODIFIED CUSUM ALGORITHM FOR SLOW AND DRASTIC CHANGE DETECTION IN GENERAL HMMS WITH UNKNOWN CHANGE PARAMETERS

Namrata Vaswani

ECE Dept., Georgia Institute of Technology, Atlanta, GA 30332
namrata@ece.gatech.edu

ABSTRACT

We study the change detection problem in a general HMM when the change parameters are unknown and the change can be slow or drastic. Drastic changes can be detected easily using the increase in tracking error or the negative log of observation likelihood (OL). But slow changes usually get missed. We have proposed in past work a statistic called ELL which works for slow change detection. Now single time estimates of any statistic can be noisy. Hence we propose a modification of the Cumulative Sum (CUSUM) algorithm which can be applied to ELL and OL and thus improves both slow and drastic change detection performance.

1. INTRODUCTION

Change detection is required in many practical problems arising in quality control, flight control, fault detection and in surveillance problems like abnormal activity detection [1]. In most cases, the underlying system in its normal state can be modeled as a parametric stochastic model which is usually nonlinear. The observations are usually noisy (making the system partially observed). Such a system forms a "general HMM" [2] (also referred to as a "partially observed nonlinear dynamical model" or a "stochastic state space model" in different contexts). It can be approximately tracked (estimate probability distribution of the hidden state variables given observations) using a Particle Filter (PF) [3].

We study here the change detection problem in a general HMM when the change parameters are unknown and the change can be slow or drastic. We use a PF to estimate the posterior probability distribution of the state at time t, X_t, given observations up to t, $Pr(X_t \in dx|Y_{1:t}) \triangleq \pi_t(dx)$. Even when change parameters are unknown, drastic changes can be detected easily using the increase in tracking error [4] or the negative log of observation likelihood (OL). OL is the negative log likelihood of the current observation conditioned on all past observations. But slow changes usually get missed. We have proposed in past work [5], a statistic called ELL (Expected Log-Likelihood) which is able to detect slow changes. We have also shown complementary behavior of ELL and OL for slow and drastic changes [1].

Now single time instant estimates of ELL or OL may be noisy and are prone to outliers. Hence in practice, one needs to average the statistic over a set of past time instants. A principled way of doing this is the CUSUM algorithm [6]. For known change parameters, the CUSUM algorithm finds the maximum (over all previous time instants, $t - p + 1$) of the observation likelihood ratio, $LR(p, t)$, assuming that the change occurred at time $t - p + 1$ (see equation (2) in Section 3). For linear systems, the LR is well

defined but for nonlinear systems, linearization techniques like extended Kalman filters are the main tools [6]. These are computationally efficient but are not always applicable. In [7], the authors attempt to use a particle filtering approach for evaluating the CUSUM statistic without linearization. This, in the most general case (when the PF is not asymptotically stable), would require to run $(1 + t)$ PFs to evaluate CUSUM at time t. In [7], they define a modification of the CUSUM statistic which has non-growing computational complexity with t.

For unknown changed system parameters, the LR can be replaced by Generalized Likelihood Ratio (GLR). The solution of GLR for linear systems in well known [6]. But for nonlinear systems, CUSUM applied to GLR would require to run one PF for each value of the unknown parameter. In [8], the authors considered a case where the unknown parameter belongs to a finite set with cardinality P. This required running $(1 + Pt)$ PFs at time t to evaluate CUSUM at t. Hence the authors proposed to search over a finite past of length Δ, i.e. to use $CUSUM_{t,\Delta} = \max_{0 \le p \le \Delta} LR(p, t)$. For the general case where the unknown parameter belongs to an uncountable set, ML parameter estimation procedures have been proposed, see a recent survey article [9]. But most of these use a PF estimate of the joint posterior of all states, $\pi_t^N(dx_{1:t}|Y_{1:t})$ and the error in the joint posterior estimated using a fixed number of particles increases with time [9] (unstable). Also, all of the above algorithms are based on observation LR and hence are unable to detect slow changes [1].

In this paper, we propose to modify the CUSUM algorithm in a different way. This modification is applicable to any single time instant change detection statistic. We apply it to both OL and ELL and so are able to detect drastic as well as slow changes. Also, it can be evaluated using a single PF even when the unknown change parameter belongs to an uncountable set and the CUSUM statistic estimate is stable. The paper is organized as follows: In Section 2, we discuss ELL and OL. The modified CUSUM algorithm and its application to ELL and OL is discussed in Section 3. In Section 4, we study two example applications, discuss their stability and show simulation results. Conclusions are presented in Section 5.

1.1. The General HMM Model and Problem Definition

The system (or state) process $\{X_t\}$ is a Markov process with state transition kernel $Q_t(x_t, dx_{t+1})$ and the observation process is a memoryless function of the state given by $Y_t = h_t(X_t) + w_t$ where w_t is an i.i.d. noise process and h_t is, in general, a nonlinear function. The conditional distribution of the observation given state, $G_t(dy_t, x_t)$, is assumed to be absolutely continuous and its pdf is given by $g_t(Y_t, x) \triangleq \psi_t(x)$. Now $p_0(x)$ (the prior initial state distribution), $G_t(dy_t, x_t)$, $Q_t(x_t, dx_{t+1})$ are known

and assumed to be absolutely continuous[1]. Thus the prior distribution of the state at any t is also absolutely continuous and admits a density, $p_t(x)$. Note that we use the superscript ^{c} is used to denote any parameter related to the changed system, ^{0} for the original system and ^{c,0} for the case when the observations of the changed system are filtered using a filter optimal for the original system[2]. Also the **unnormalized filter kernel** [2] is defined as $R_t(x_t, dx_{t+1}) = \psi_t(x_{t+1})Q_t(x_t, dx_{t+1})$.

We assume that the normal (original/unchanged) system has state transition kernel Q_t^0. A change (which can be slow or drastic) in the system model begins to occur at some finite time t_c and lasts till a final finite time t_f. In the time interval, $[t_c, t_f]$, the state transition kernel is Q_t^c and after t_f it again becomes Q_t^0. Both Q_t^c and the change start and end times t_c, t_f are assumed to be unknown. The goal is to detect the change, with minimum delay.

1.2. Approximate Non-linear Filtering using a Particle Filter

The problem of nonlinear filtering is to compute at each time t, the conditional probability distribution, of the state X_t given the observation sequence $Y_{1:t}$, $\pi_t(dx) = Pr(X_t \in dx | Y_{1:t})$. The transition from π_{t-1} to π_t is:

$$\pi_{t-1} \longrightarrow \pi_{t|t-1} = Q_t(\pi_{t-1}) \longrightarrow \pi_t = \frac{\psi_t \pi_{t|t-1}}{(\pi_{t|t-1}, \psi_t)}$$

A **Particle Filter** [3] is a recursive algorithm for approximate nonlinear filtering which produces at each time t, a cloud of N particles $\{x_t^{(i)}\}$ whose empirical measure, π_t^N, closely "follows" π_t. At $t = 0$, it samples N times from π_0 to approximate it by $\pi_0^N(dx) \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{x_0^{(i)}}(dx)$. Then, for each t, it runs the Bayes recursion:

$$\pi_{t-1}^N \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{x_{t-1}^{(i)}}(dx) \longrightarrow \pi_{t|t-1}^N \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{\bar{x}_t^{(i)}}(dx)$$

$$\longrightarrow \bar{\pi}_t^N \triangleq \frac{1}{N}\sum_{i=1}^N w_t^{(i)}\delta_{\bar{x}_t^{(i)}}(dx) \longrightarrow \pi_t^N \triangleq \sum_{i=1}^N \delta_{x_t^{(i)}}(dx)$$

where $\bar{x}_t^{(i)} \sim Q_t(x_{t-1}^{(i)}, dx), x_t^{(i)} \sim \text{Multinomial}(\{\bar{x}_t^{(i)} w_t^{(i)}\}_{i=1}^N)$, $w_t^{(i)} \triangleq \frac{\psi_t(\bar{x}_t^{(i)})}{(\pi_{t|t-1}^N, \psi_t)}$.

2. CHANGE DETECTION STATISTICS: ELL AND OL

2.1. The ELL statistic

"Expected (negative) Log Likelihood" or ELL [5] at time t, is the conditional expectation of the negative logarithm of the prior likelihood of the state at time t, under the no change hypothesis (H_0), given observations till time t, i.e.

$$ELL(Y_{1:t}) \triangleq E[-\log p_t^0(x)|Y_{1:t}] = E_{\pi_t}[-\log p_t^0(x)]. \quad (1)$$

Using the PF estimate π_t^N, the ELL estimate becomes $ELL^N = \frac{1}{N}\sum_{i=1}^N[-\log p_t^0(x_t^{(i)})]$. ELL as defined above is equal to the Kerridge Inaccuracy [10] between the posterior and prior state pdf.

---

[1] Note that for ease of notation, we denote the pdf either by the same symbol or by the lowercase of the probability distribution symbol

[2] At most places ^{0,0} is replaced by ^{0} and ^{c,c} by ^{c}

The **Kerridge Inaccuracy** [10] between two pdfs p and q is defined as $K(p:q) = \int p(x)[-\log q(x)]dx$. We have $ELL(Y_{1:t}) \triangleq E_{\pi_t}[-\log p_t^0(x)] = K(\pi_t : p_t^0)$.

Why ELL works: Now, taking expectation of $ELL(Y_{1:t}^0) = K(\pi_t^0 : p_t^0)$ over normal observation sequences, we get $E_{Y_{1:t}^0}[ELL(Y_{1:t}^0)] = E_{Y_{1:t}^0}E_{\pi_t^0}[-\log p_t^0(x)] = E_{p_t^0}[-\log p_t^0(x)] = H(p_t^0) = K(p_t^0 : p_t^0) \triangleq EK_t^0$ where $H(.)$ denotes entropy. Similarly, for the changed system observations, $E_{Y_{1:t}^c}[ELL(Y_{1:t}^c)] = K(p_t^c : p_t^0) \triangleq EK_t^c$, i.e. the expectation of $ELL(Y_{1:t}^c)$ taken over changed system observation sequences is actually the Kerridge Inaccuracy between the changed system prior, p_t^c, and the original system prior, p_t^0, which will be larger than the Kerridge Inaccuracy between p_t^0 and p_t^0 (entropy of p_t^0) [11].

Now, ELL will detect the change when EK_t^c is "significantly" larger than EK_t^0. Setting the change threshold to $\kappa_t \triangleq EK_t^0 + 3\sqrt{VK_t^0}$, where $VK_t^0 = Var_{Y_{1:t}}(K_t^0)$, will ensure a false alarm probability less than 0.11 (follows from the Chebyshev inequality [12]). By the same logic, if $EK_t^c - 3\sqrt{VK_t^c} > \kappa_t$ then the miss probability [12] (probability of missing the change) will also be less than 0.11.

2.2. When ELL fails: The OL Statistic

The above analysis assumed no estimation errors in evaluating ELL and if this were true ELL would work for drastic changes as well. But, the PF being used is optimal for the unchanged system. Hence when estimating π_t (required for evaluating the ELL) for the changed system, there is "exact filtering error" [1]. Also the particle filtering error is much larger in this case. The upper bound on the approximation error in estimating the ELL is proportional to the "rate of change" [1]. Hence ELL is approximated accurately for a slow change and thus it detects such a change as soon as it becomes "detectable" [1]. But ELL fails to detect drastic changes because of large estimation error in evaluating π_t. But large estimation error in evaluating π_t also corresponds to a large value of OL (Observation Likelihood) which can be used for detecting such changes (Theorem 2.4 of chapter 2 of [1]). OL as explained earlier, is the negative log likelihood of the current observation conditioned on past observations under the no change hypothesis, i.e. $OL = -\log Pr(Y_t|Y_{1:t-1}, H_0)$. It is evaluated using a PF as $OL_t^N = -\log(Q_t^0\pi_{t-1}^N, \psi_t)$. OL, on the other hand, takes longer to detect a slow change or may not detect it at all (discussed in [1]).

3. THE MODIFIED CUSUM ALGORITHM

Now single time instant estimates of ELL or OL may be noisy and prone to outliers. Hence in practice, one needs to average the statistic over a set of past time instants. Averaging OL over past p time instants gives $aOL(p,t) = (1/p)[-\log Pr(Y_{t-p+1:t}|Y_{1:t-p})]$. Average ELL is $aELL(p,t) = (1/p)\sum_{k=t-p+1}^t ELL(Y_{1:k})$. But since $ELL(Y_{1:t})$ is not stationary and hence also not ergodic, the above averaging cannot be justified for large p. For small p, one can use the standard assumption of approximating any nonstationary process by a piecewise stationary process over small time intervals (same idea as that used to justify any other moving average). Instead of $aELL$, one can evaluate joint ELL as $jELL(p,t) = (1/p)E[-\log p_{t-p+1:t}(X_{t-p+1:t})|Y_{1:t}]$, which is the Kerridge Inaccuracy between the joint posterior distribution of $X_{t-p+1:t}$ given $Y_{1:t}$ and their joint prior. If using $aELL(p,t)$, the

threshold can be set as $Th^{aELL}(p,t) = (1/p)E_{Y_{1:t}}[aELL(p,t)]$ which is the sum of individual entropies of $X_{t-p+1:t}$. If using $jELL(p,t)$, $Th^{jELL}(p,t) = (1/p)E_{Y_{1:t}}[jELL(p,t)]$ which is the joint entropy of $X_{t-p+1:t}$.

Now, for the case of known change parameters, the CUSUM statistic applied to the observation likelihood ratio is [6, 7]

$$CUSUM_t = \max_{1 \leq p \leq t} \log LR(p,t), \quad \text{where}$$

$$\log LR(p,t) \triangleq \log Pr^c(Y_{t-p+1:t}|Y_{1:t-p}) - \log Pr^0(Y_{t-p+1:t}|Y_{1:t-p})$$
$$= [aOL^c(p,t) - aOL^0(p,t)] \quad (2)$$

and change is declared if $CUSUM_t > \lambda$ for some positive threshold λ. For unknown change parameters and for any statistic, denoted by $stat(p,t)$, one can modify this as follows: Set a threshold for $stat(p,t)$ for normal observations, $Th^{stat}(p,t)$[3], and replace $\log LR(p,t)$ by $[stat(p,t) - Th^{stat}(p,t)]$, i.e. use

$$CUSUM_t^{stat} = \max_{1 \leq p \leq t} [stat(p,t) - Th^{stat}(p,t)] \quad (3)$$

where $stat$ is aOL^0, $aELL$ or $jELL$. Change is declared if $CUSUM_t^{stat} > \lambda$. Change time is given by $t - p^* + 1$ where p^* is the argument maximizing $[stat(p,t) - Th^{stat}(p,t)]$ over p.

Now the error in the PF estimate of the joint posterior of $X_{t-p+1:t}$ given $Y_{1:t}$ increases as p increases, and as a thumb rule, the estimates are meaningless for $p > 5$ [9]. Also as explained above, aELL can be justified for only small values of p. Thus in practice we use the following modification (with $\Delta = 5$):

$$CUSUM_{t,\Delta}^{stat} = \max_{1 \leq p \leq \Delta} [stat(p,t) - Th^{stat}(p,t)]. \quad (4)$$

4. TWO APPLICATIONS

4.1. Bearings-only Target Tracking

In bearings-only target tracking (details in [3]), the target moves on the x-y plane according to the standard second order model:

$$X_t = \Phi X_{t-1} + \Gamma n_t, \quad \Phi = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} 0.5 & 0 \\ 1 & 0 \\ 0 & 0.5 \\ 0 & 1 \end{pmatrix} (5)$$

where $X_t = x_{1,t}, \dot{x}_{1,t}, x_{2,t}, \dot{x}_{2,t}$. Here $x_{1,t}$, $x_{2,t}$ denote the x and y components of the target location and $\dot{x}_{1,t}, \dot{x}_{2,t}$ denote the x and y components of the target velocity. The observation, Y_t, is a noisy measurement of the target bearing, $Y_t = \tan^{-1}(x_{2,t}/x_{1,t}) + w_t$. In this case, the observation model is nonlinear, but the system model is linear Gaussian and hence $p_t(x)$ can be defined in closed form. The system noise, n_t, is zero mean i.i.d. Gaussian with $\Sigma_{sys} = 0.001I$ and the observation noise is zero mean i.i.d. truncated Gaussian with variance $\sigma_{obs}^2 = 0.005$ and truncation parameter $B = 10\sigma_{obs}$. We attempt to detect a change in the state dynamics where the change is due to an additive bias of $\Gamma[r\sigma_{sys}\ 0]^T$, for 10 time steps starting at $t = 5$. The initial state was assumed to be known (zero variance).
Stability: With the following mild assumption, the above system satisfies the stability results of [1]: Assume that for each t, the x component of the target location is bounded, i.e. there exists a $P_t < \infty$, s.t. $-P_t \leq x_{1,t} \leq P_t$. Using this assumption,

along with the fact that $h(x) = \tan^{-1}(x_{2,t}/x_{1,t})$ and that the observation noise is truncated, it is easy to see that the support set of $\psi_t(x)$, E_{x,Y_t}, compact. Using Example 3.10 of [2], this along with the fact that ΦX_{t-1} is continuous, that π_0 has compact support and that the system noise is Gaussian, makes the unnormalized filter kernels $R_t^0, R_t^{c,0}, R_t^c$ mixing. Also since E_{x,Y_t} is compact, $sup_{x \in E_{x,Y_t}} \psi_t(x)$ is finite and $M_t = sup_{x \in E_{x,Y_t}}[-\log p_t(x)]$ is finite[4]. Thus we satisfy all assumptions of Theorem 2.2 of Chapter 2 of [1]. Also using an argument similar to that in Section 2.7.1 of [1][5], assumption (iv)' of the stronger Theorem 2.1 is also satisfied and hence it holds, i.e. the error between the true ELL and its approximation averaged over PF realizations and observation sequences is eventually monotonically decreasing with time, for large N, and hence is stable. This implies that the error in CUSUM-ELL (finite sum of past ELLs) estimate is also stable.
Simulation Results: ROC plots of Figure 1 compare performance of ELL (blue -o), OL (red -*), CUSUM-ELL (black -*square*) and CUSUM-OL (majenta -x). An ROC (Receiver Operating Characteristics) plot for a change detection problem [6] is obtained by plotting average detection delay against average time between false alarms for different values of the detection threshold. We simulated 20 realizations and calculated average detection delay and average time between false alarms for different values of the detection threshold. As can be seen from the figure, CUSUM-ELL performs much better than ELL for all 3 cases. Also, CUSUM-ELL and ELL significantly outperform CUSUM-OL and OL for the slow change (r=1). In fact, in this example because of the nature of $h(x)$, even for the faster (r=10) and drastic change (r=20), CUSUM-ELL outperforms CUSUM-OL and OL.

4.2. Nonlinear State Dynamics

If the system dynamics is nonlinear, in most cases it is not possible to define $p_t(x)$ in closed form. One solution for such cases is to use prior knowledge to define $p_t(x)$ as coming from a certain parametric family for example a Gaussian or a mixture of Gaussians. We study here an example discussed in [3] which has the following nonlinear state dynamics:
$X_t = X_{t-1} + 25\frac{X_{t-1}}{1+X_{t-1}^2} + 8\cos(1.2(t-1)) + n_t$ and $Y_t = \frac{X_t^2}{20} + w_t$ where n_t is Gaussian system noise with $\sigma_{sys}^2 = 10$ and w_t is truncated Gaussian observation noise with $\sigma_{obs}^2 = 1$. Initial state is taken to be zero. We introduce a change by adding a bias $r\sigma_{sys}$ to the state equation for 10 time instants starting at $t = 5$. Based on prior knowledge, we take $p_t(x)$ to be a Gaussian with mean $\sum_{\tau=1}^t 8\cos(1.2(\tau - 1))$ and variance σ_{sys}^2.
Stability: It is easy to see that $\psi_t(x)$ has compact support, E_{x,Y_t}. Also, the system dynamics is continuous and system noise is Gaussian. This along with the fact that π_0 has compact support makes $R_t^0, R_t^{c,0}, R_t^c$ mixing. Since E_{x,Y_t} is compact, $sup_{x \in E_{x,Y_t}} \psi_t(x)$ is finite and $M_t = sup_{x \in E_{x,Y_t}}[-\log p_t(x)]$ is finite. Thus all assumptions of Theorem 2.2 of Chapter 2 of [1] are satisfied.
Simulation Results: We show ROC plots in Figure 2 for comparing performance of ELL (blue -o), OL (red -*), CUSUM-aELL (black -*square*), CUSUM-jELL (green -△) and CUSUM-OL (majenta -x). Here ELL outperforms OL and CUSUM-aELL/jELL

---

[3]Usually take $Th^{stat}(p,t) = E_{Y_{1:t}}[stat(p,t)]$

[4]Continuous functions (here $\psi_t(x)$ and $[-\log p_t(x)]$) map compact sets onto compact sets.

[5]The difference here is that since $-P_t \leq x_{1,t} \leq P_t$, the system noise is truncated Gaussian distributed instead of Gaussian

| (a) Slow Change | (b) Faster Change | (c) Drastic Change |
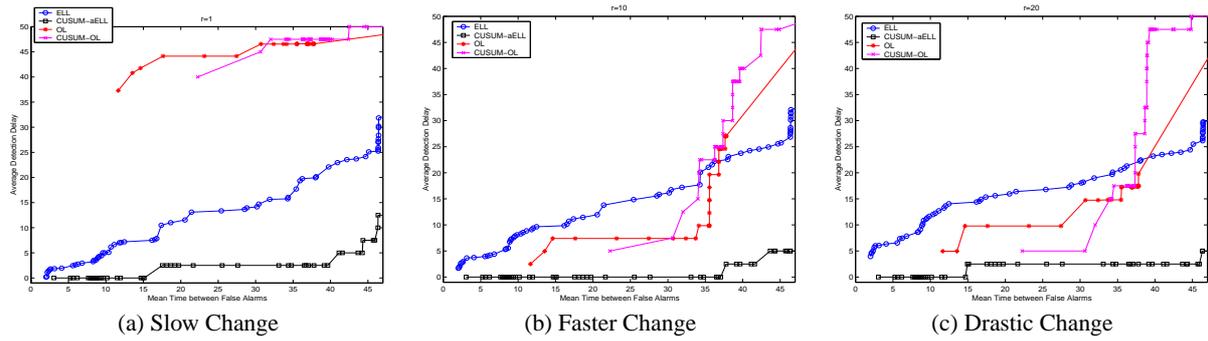
Fig. 1. Bearings-only tracking example: ROC curves comparing ELL, CUSUM on aELL, OL and CUSUM on OL

| (a) Slow Change | (b) Faster Change | (c) Drastic Change |
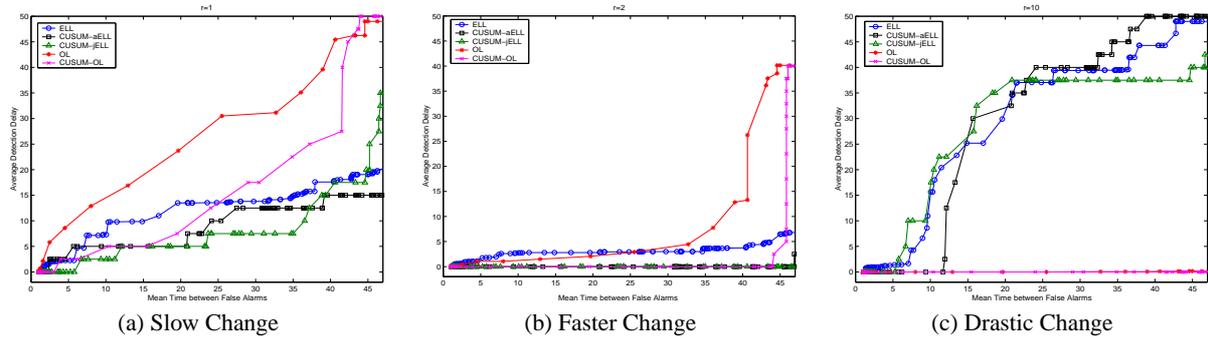
Fig. 2. Nonlinear System Model example: ROC curves comparing ELL, CUSUM on aELL, CUSUM on jELL, OL and CUSUM on OL. Note that in (b), CUSUM-aELL and CUSUM-jELL remain at zero while in (c), CUSUM-OL and OL remain at zero.

outperform CUSUM-OL for the slow (r=1) and faster (r=2) change and vice versa for the drastic (r=10) change. Infact for r=10, CUSUM-OL and OL have zero detection delay while CUSUM-ELL and ELL completely fail. Also, for all cases, CUSUM-ELL is better than ELL and CUSUM-OL is better than OL. CUSUM-aELL and CUSUM-jELL have comparable performance, jELL is better at lower thresholds while aELL is better at higher thresholds.

5. CONCLUSIONS

We have modified the CUSUM algorithm to work for change detection in nonlinear systems with unknown change parameters. This modification can be applied to the ELL statistic [5] and is the first application of CUSUM for slow change detection. Also our modified CUSUM can be evaluated using a single PF and the CUSUM statistic estimates are stable. As part of future work, we would like to use ML parameter estimation [9] to reduce the "exact filtering error" [1] in ELL (and CUSUM-ELL) approximation.

6. REFERENCES

[1] N. Vaswani, *Change Detection in Stochastic Shape Dynamical Models with Applications in Activity Modeling and Abnormality Detection*, Ph.D. Thesis, ECE Dept, University of Maryland at College Park, August 2004.

[2] LeGland F. and Oudjane N., "Stability and Uniform Approximation of Nonlinear Filters using the Hilbert Metric, and Application to Particle Filters," *Tech. report, RR-4215, INRIA*, 2002.

[3] N.J. Gordon, D.J. Salmond, and A.F.M. Smith, "Novel approach to nonlinear/nongaussian bayesian state estimation," *IEE Proceedings-F (Radar and Signal Processing)*, pp. 140(2):107–113, 1993.

[4] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*, Academic Press, 1988.

[5] N. Vaswani, "Change detection in partially observed nonlinear dynamic systems with unknown change parameters," in *American Control Conference (ACC)*, 2004.

[6] M. Basseville and I Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice Hall, 1993.

[7] B. Azimi-Sadjadi and P.S. Krishnaprasad, "Change detection for nonlinear systems: A particle filtering approach," in *ACC*, 2002.

[8] P. Li and V. Kadirkamanathan, "Particle filtering based likelihood ratio approach to fault diagnosis in nonlinear stochastic systems," *IEEE Trans. Syst., Man, Cybern. C*, vol. 31, pp. 337–343, August 2001.

[9] C. Andrieu, A. Doucet, S.S. Singh, and V.B. Tadic, "Particle methods for change detection, system identification, and control," *Proc. of the IEEE*, vol. 93, pp. 423– 438, March 2004.

[10] D.F. Kerridge, "Inaccuracy and inference," *J. Royal Statist. Society, Ser. B*, vol. 23 1961.

[11] Rudolf Kulhavy, "A geometric approach to statistical estimation," in *IEEE CDC*, Dec. 1995.

[12] G. Casella and R. Berger, *Statistical Inference*, Duxbury Thomson Learning, second edition, 2002.