Modified CUSUM for Slow and Sudden Change Detection with Unknown Parameters

Namrata Vaswani

School of Electrical and Computer Engineering

Georgia Institute of Technology

The Problem

- Partially Observed and Nonlinear System: Observations Y_t are noisy nonlinear functions of the state X_t

$$Y_t = h_t(X_t) + w_t, \quad w_t\text{: observation noise}$$

- The system model (dynamics of X_t) can also be nonlinear:

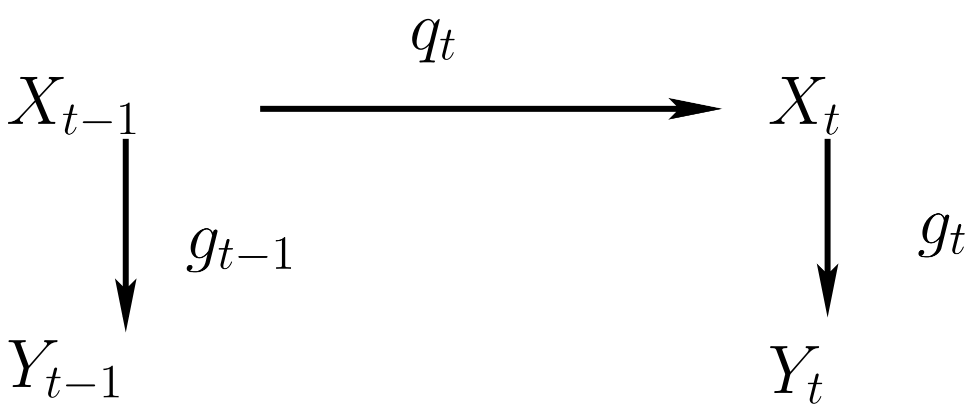
$$X_t = f_t(X_{t-1}) + n_t, \quad n_t\text{: system noise}$$

- **Given the observations $Y_1, Y_2, ...Y_t$, detect, as quickly as possible, if a change occurred in the dynamics of X_t**
  - **Parameters of changed system unknown**
  - **Change can be slow or sudden**

An Application: Detect Changes in Landmark Shape Dynamics

- **Observation:** Vector of observed object locations (Configuration)
- **State:** [Shape, Translation, Scale, Rotation, Velocities]
- **Observation model:** $h_t : \mathcal{S} \times \mathbb{R}^2 \times \mathbb{R}^+ \times S0(2) \rightarrow \mathbb{R}^{2k}$, Gaussian noise
- **System model:**
  - Gauss-Markov model on shape velocity, parallel transported to tangent space of the current shape
  - Gauss-Markov model on group action velocities
- **Detect changes in shape using posterior distribution of shape given observed object locations**

Notation



- **Prior: Given no observations,** $X_t \sim p_t(.)$
- Posterior: $X_t | Y_{1:t} \sim \pi_t(.)$
- **Superscripts:** ^{0} **(unchanged system),** ^{c} **(changed system)**
- $X_t^0 \sim p_t^0(.), \quad X_t^c \sim p_t^c(.)$

Exact Solution to Optimal Filtering

- **t=0:** Posterior of X_0 given no observations is its prior, $\pi_{0|0} = p_0$
- **Bayes' rule applied to system and observation model at t:**

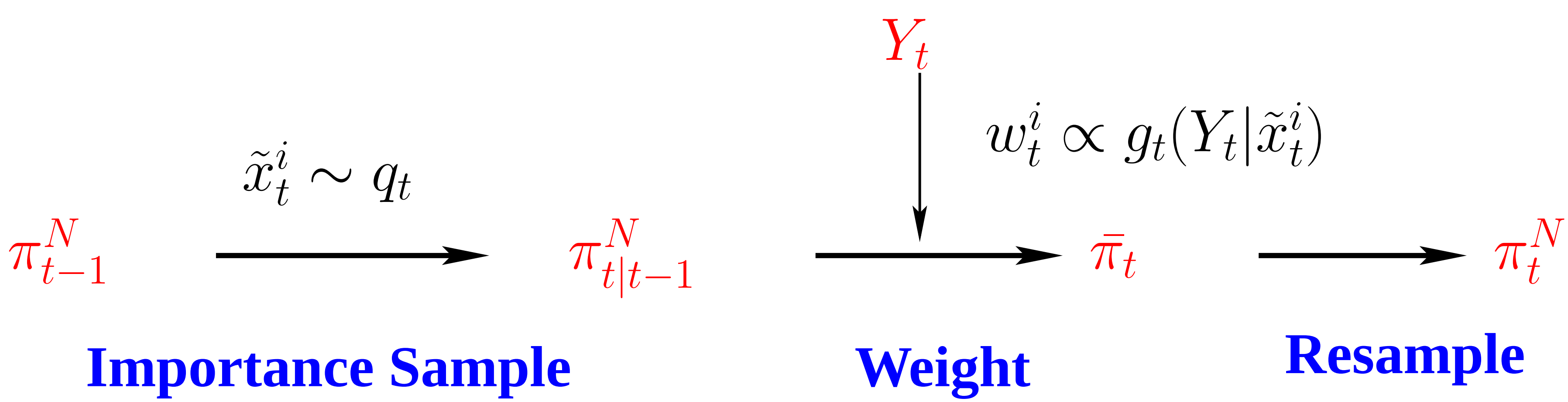
Prediction dist. $\pi_{t|t-1}(dx_t) = \int_{x_{t-1}} q_t(x_t|x_{t-1})\pi_{t-1}(dx_{t-1})dx_t$

Filtering dist. $\pi_t(dx_t) = \dfrac{g_t(Y_t|x_t)\pi_{t|t-1}(dx_t)}{\int_x g_t(Y_t|x)\pi_{t|t-1}(dx)}$

- **System & observation model linear, Gaussian: Kalman filter**
- Any general system: approx. solution using a Particle Filter

Particle Filter: Basic Idea

- **Sequential Monte Carlo method, approx. true filter as number of Monte Carlo samples ("particles"),** $N \to \infty$
- Given π^N_{t-1}, **perform importance sampling/ weighting,** followed by **resampling** to approx. the Bayes' recursion: π^N_t



Existing Work: Change Detection in Nonlinear Systems

- **Fully observed state** (no observation noise, h_t invertible)
  - CUmulative SUM, generalized CUSUM, negative log likelihood
- **Partially observed state**
  - **Known change parameters**
    - CUSUM uses $t+1$ particle filters at t [Azimi-Sadjadi et al'02]
  - **Unknown change parameters: few existing solutions**
    - **generalized CUSUM not tractable [Andrieu et al'2004]**
    - **Tracking Error** [Bar-Shalom]
    - **negative Log Likelihood of Observations (OL)**
    - **Fail to detect slow changes**

Change Detection Statistics [Vaswani, ACC'2004]

- **Expected (negative) Log Likelihood of state (ELL)**

$$\mathbf{ELL}(\mathbf{Y_{1:t}}) = \mathbf{E}[-\log \mathbf{p_t^0}(\mathbf{X_t})|\mathbf{Y_{1:t}}] = \mathbf{E}_{\pi_\mathbf{t}}[-\log \mathbf{p_t^0}(\mathbf{X})]$$

- For sudden changes, can use
  - **(negative) log of Observation Likelihood (OL)**

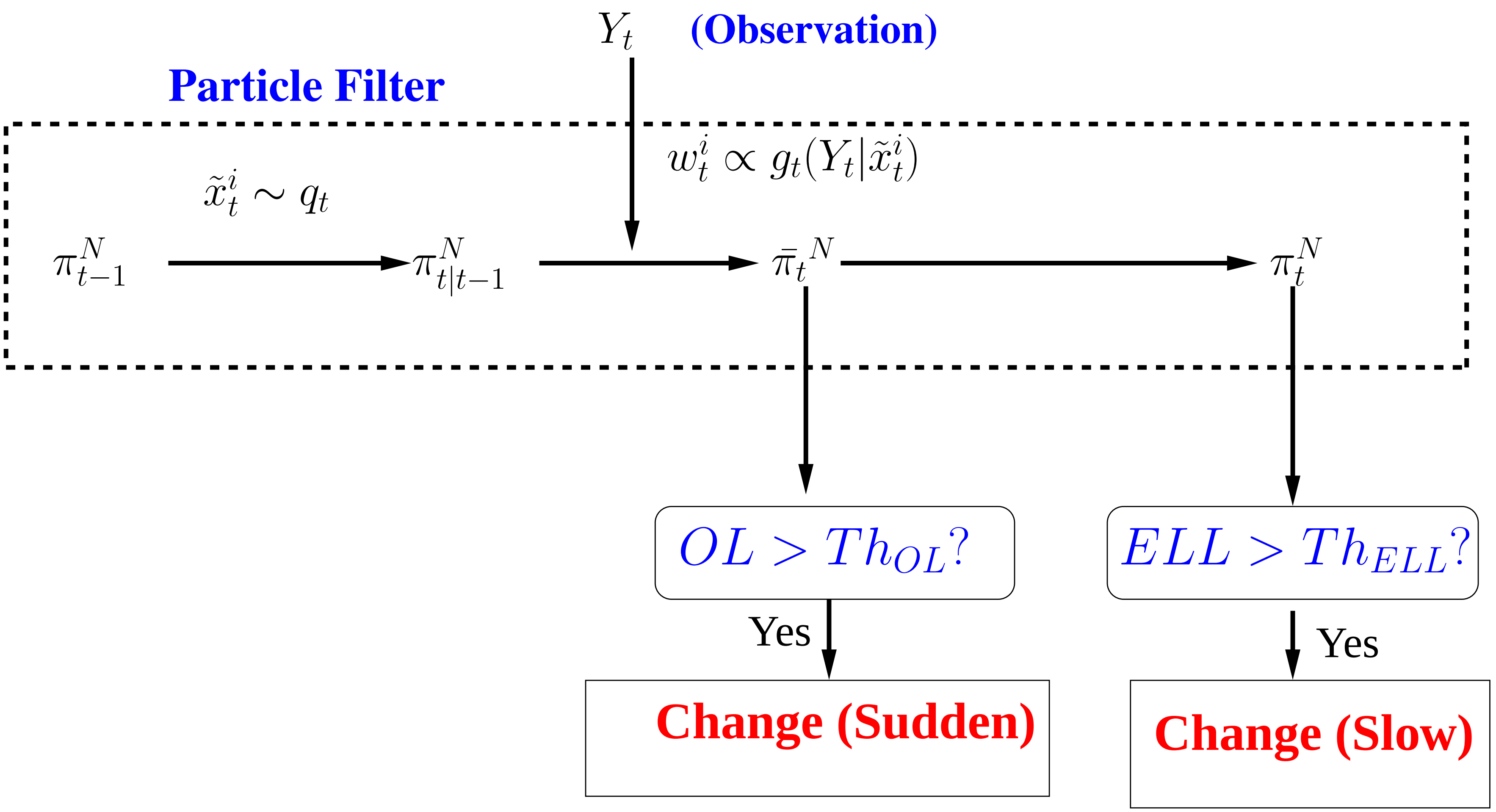
$$\mathbf{OL}(\mathbf{Y_{1:t}}) = -\log \mathbf{p_Y}(\mathbf{Y_t}|\mathbf{Y_{1:t-1}}) = -\log \mathbf{E}_{\pi_{\mathbf{t}|\mathbf{t-1}}}[\mathbf{g_t}(\mathbf{Y_t}|\mathbf{X})]$$

  - **Tracking Error** [Bar-Shalom]

$$\mathbf{TE} = ||\mathbf{Y_t} - \hat{\mathbf{Y}}_\mathbf{t}||^\mathbf{2}, \ \hat{\mathbf{Y}}_\mathbf{t} = \mathbf{E}[\mathbf{Y_t}|\mathbf{Y_{1:t-1}}] = \mathbf{E}_{\pi_{\mathbf{t}|\mathbf{t-1}}}[\mathbf{h_t}(\mathbf{X})]$$

  - OL $\approx$ TE (to first order) for white Gaussian observation noise

Change Detection Algorithm



An Example: Evaluating the statistics

Consider, $Y_t = X_t^3 + w_t,\ X_t = X_{t-1} + n_t,\ n_t \sim \mathcal{N}(0, \sigma_{sys}^2)$

- **Prior state dist.:** $p_t^0(x) = \frac{1}{\sqrt{2\pi t\sigma^2}} e^{-\frac{x^2}{2t\sigma_{sys}^2}}$
- Using particle filtering estimate of posteriors, evaluate

$$ELL_t^N(Y_{1:t}) = \frac{1}{N}\sum_{i=1}^{N} \frac{{x_t^{(i)}}^2}{t\sigma_{sys}^2} + const,$$

$$OL_t^N(Y_{1:t}) = -\log \frac{1}{N}\sum_{i=1}^{N} \exp \frac{-(Y_t - (\tilde{x}_t^{(i)})^3)^2}{2\sigma_{obs}^2},$$

$$TE_t^N(Y_{1:t}) = (Y_t - \frac{1}{N}\sum_{i=1}^{N}(\tilde{x}_t^{(i)})^3)^2$$

- Note that $OL \approx TE$ (to first order)

Evaluating p_t^0

- If state space dynamics is linear/Gaussian: easy
  - If **dynamics of the part of state space used to detect the change is linear/Gaussian**: easy
- If **the normal system is stationary**: assume a parametric form of p^0, use a training data seq. to learn parameters
  - In general, can assume piecewise stationarity of p_t^0
- If no training data is available but p_0^0 and system noise are Gaussian: linearize $f_\tau(X_\tau)$, $\forall \tau$ to approx. p_t^0 by a Gaussian

Detectable Changes Using ELL

- **Kerridge inaccuracy, $K(p_1 : p_2)$: second term of KL divergence**
- $ELL(Y_{1:t}) = -E_{\pi_t}[\log p_t^0(X)] = K(\pi_t : p_t^0)$
- Average ELL of unchanged observations: differential entropy of the state at t, $h(p_t^0)$
- **A sufficient condition for changes "detectable using ELL" (with small $P_{f.a.}$, P_{miss}) [Vaswani'2004]:**

$$\mathbf{K(p_t^c : p_t^0) - h(p_t^0) > 3\sqrt{Var(ELL^c)} + 3\sqrt{Var(ELL^0)}}$$

Modified CUSUM Algorithm

- Given a change detection statistic $stat(\tau)$, define

$$sum\text{-}stat(p,t) \triangleq \sum_{\tau=t-p+1}^{t} stat(\tau)$$

- **Define the modified CUSUM statistic,** $CUSUM\text{-}stat(t)$**, as**

$$\sup_{1\le p\le t} \left(sum\text{-}stat(p,t) - E_{Y^0_{1:t}}[sum\text{-}stat(p,t)]\right)$$

- **Change Time is**

$$t_{change} = \min\{t : CUSUM\text{-}stat(t) > \lambda\}$$

Application to ELL and OL

- Define $sum\text{-}ELL$ and $sum\text{-}OL$ as above. Then

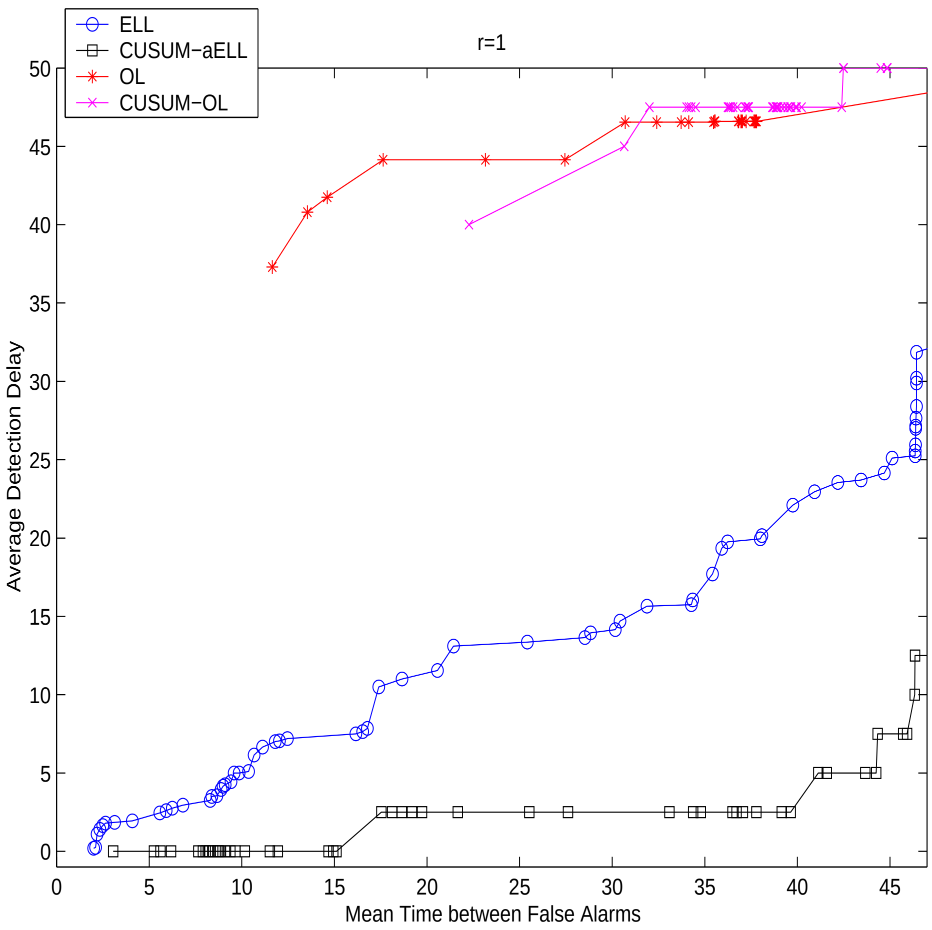
$$E_{Y^0_{1:t}}[sum\text{-}OL(p,t)] = \sum_{\tau=t-p+1}^{t} h(Y^0_\tau | Y^0_{1:\tau-1}) = h(Y^0_{t-p+1:t} | Y_{1:t-p})$$

$$E_{Y^0_{1:t}}[sum\text{-}ELL(p,t)] = \sum_{\tau=t-p+1}^{t} h(X_\tau) = \sum_{\tau=t-p+1}^{t} h(p^0_\tau)$$

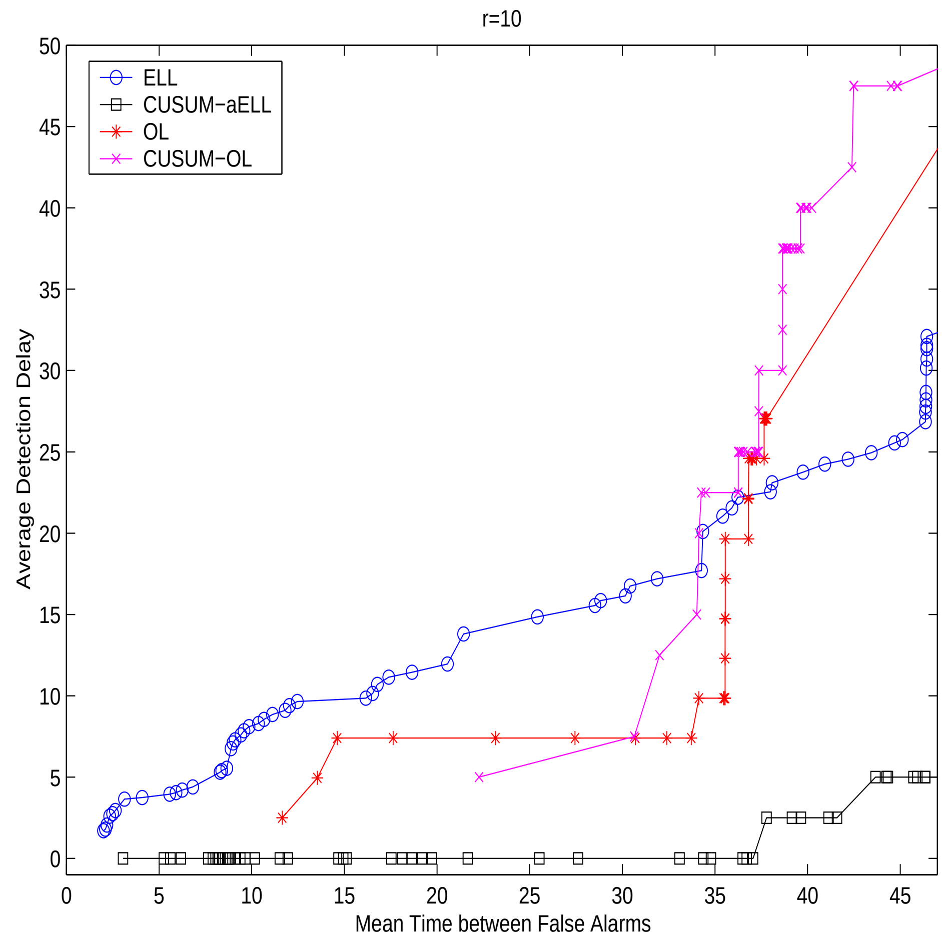
- h **denotes differential entropy**

- **Can also use** $joint\text{-}ELL(p,t) = E_{\pi_{t-p+1:t}}[-\log p_{t-p+1:t}(X)]$

ROC Plots for Bearings only Tracking

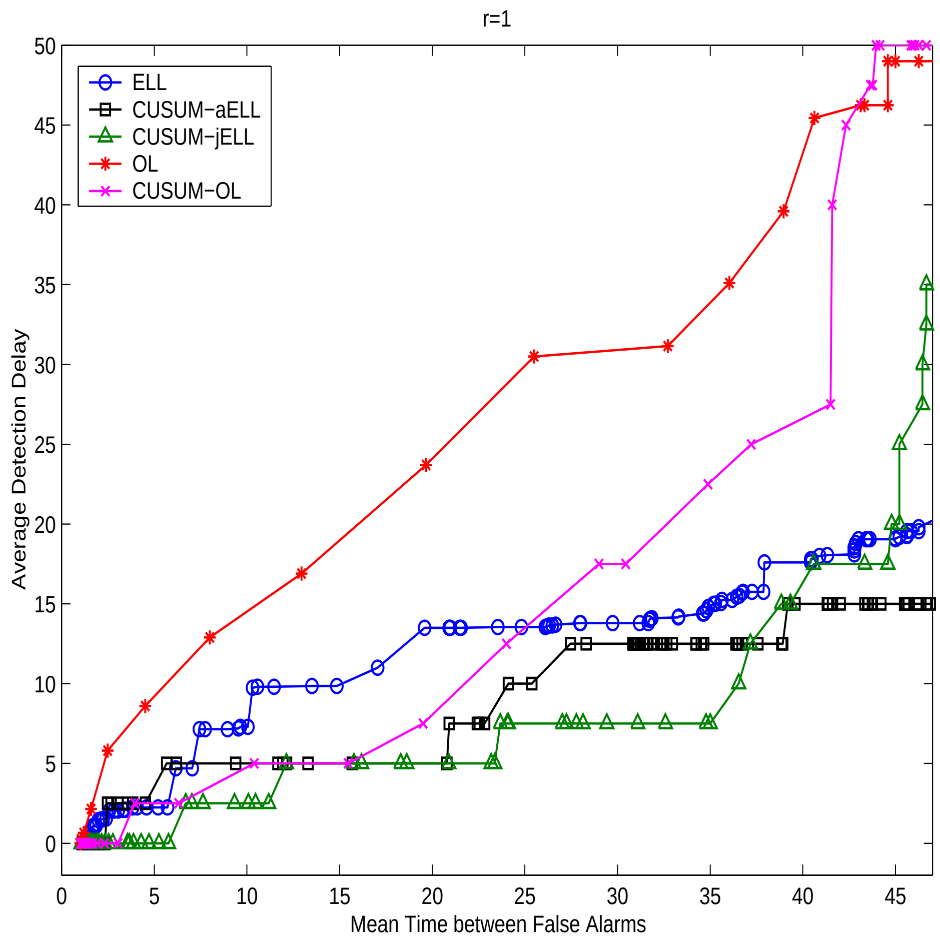


Slow Change

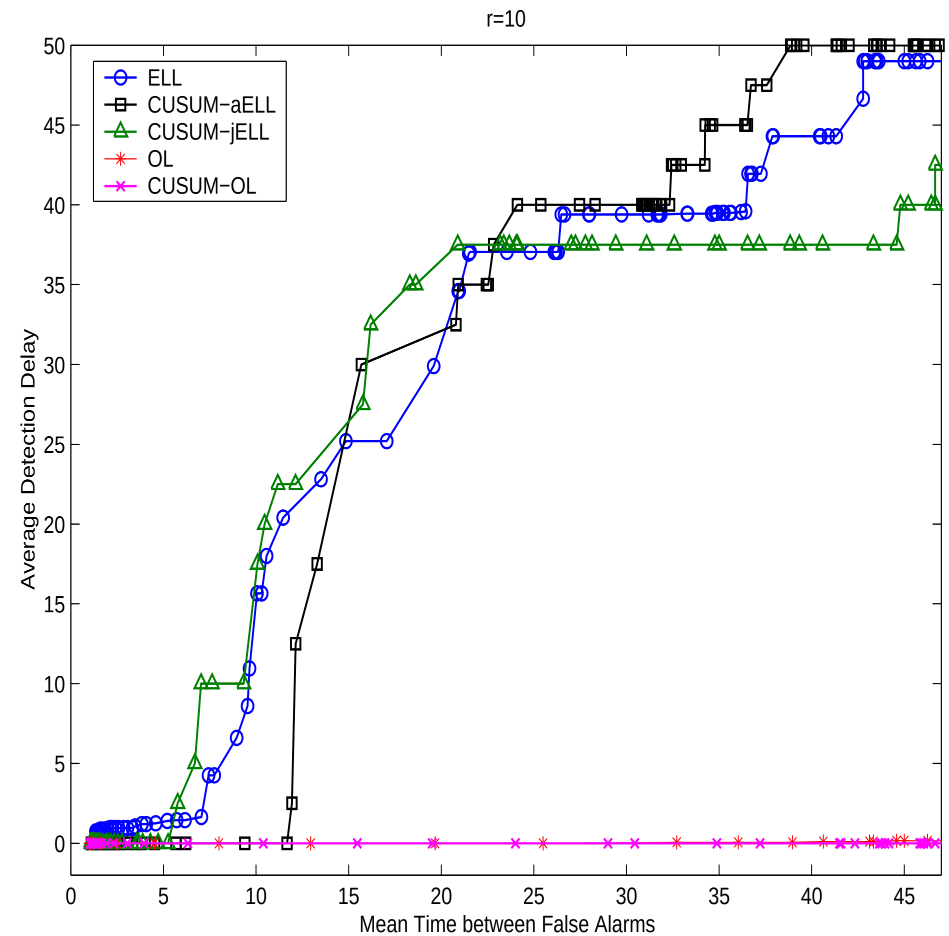


Faster Change

ROC Plots for Non-linear System Model

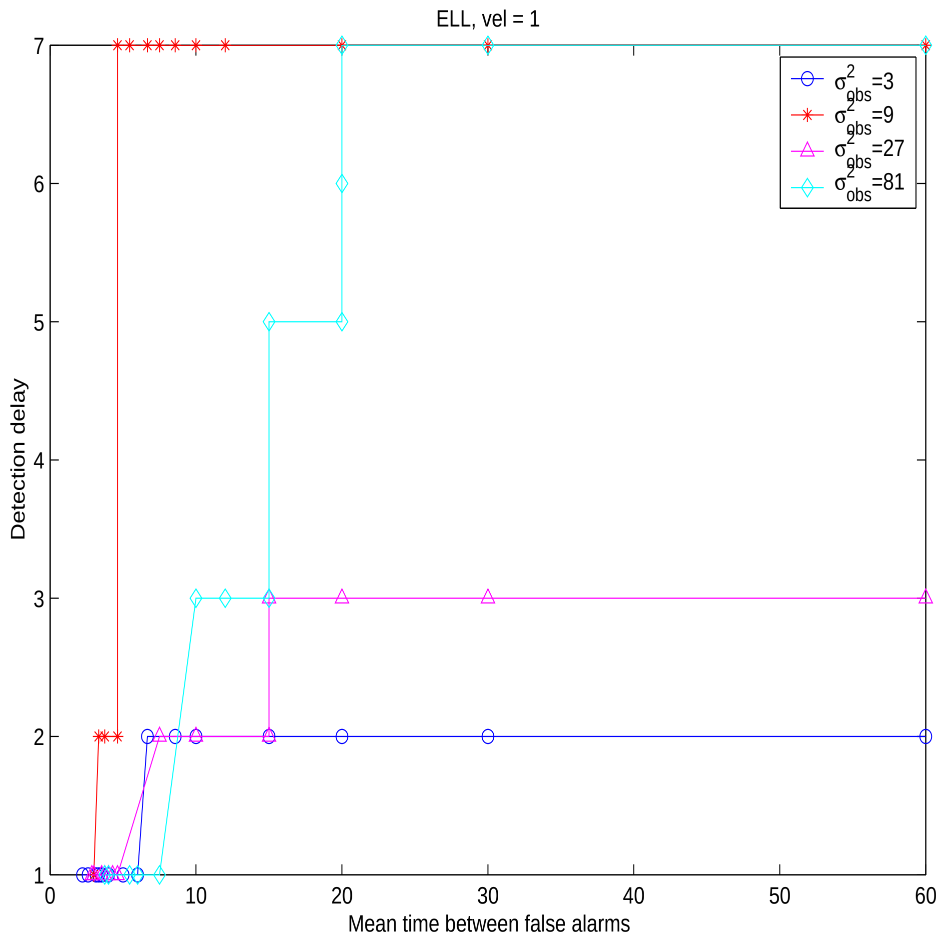


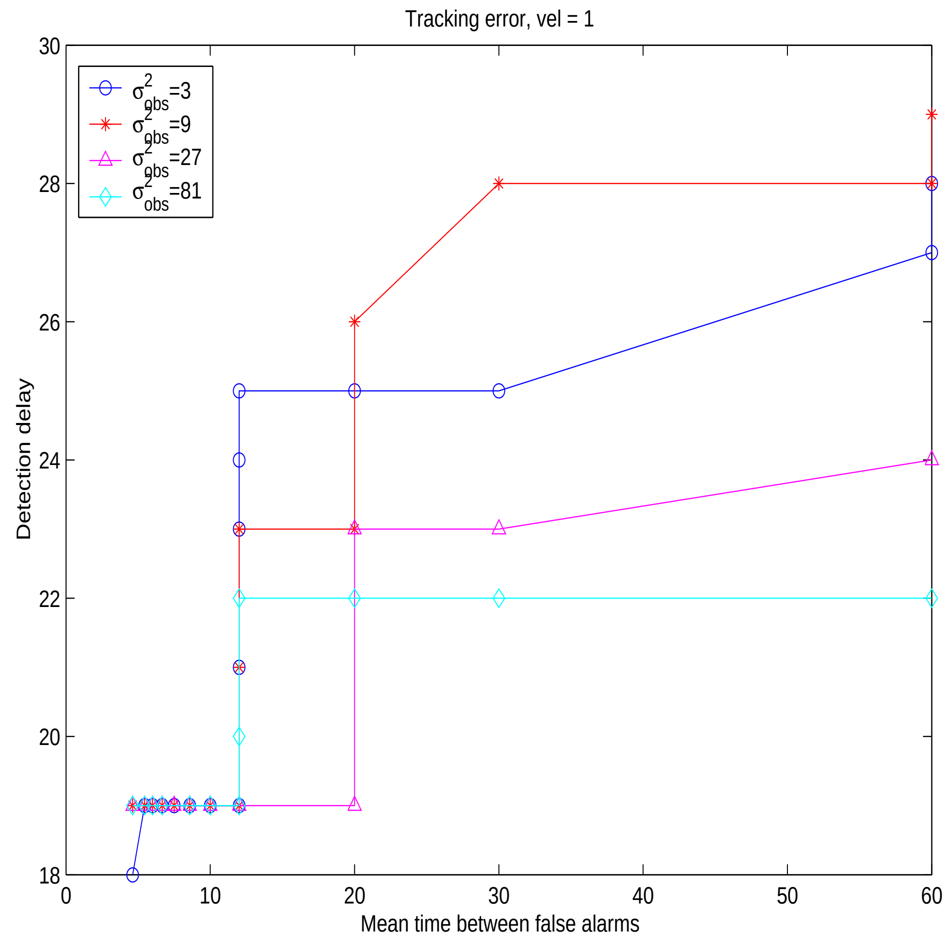
Slow Change



Faster Change

ROCs for Slow Abnormal Activity Detection





ELL Detects

Tracking Error: Takes much longer

Modified ELL

- **Variance of p_t^0 usually increases with t: miss smaller changes**

- **Solution: Replace p_t^0 by $\pi^0_{t|t_{nc}}$. It is assumed that no change has occurred until t_{nc}. Approx. $\pi^0_{t|t_{nc}}$ as:**
  - Evaluate a Gaussian or Gaussian mixture approx. to $\pi^{0,N}_{t_{nc}|t_{nc}}$
  - Linearize $f_\tau(X),\ \tau = t_{nc}, ...t$ to approx. to $\pi^0_{t|t_{nc}}$

- **Useful to detect multiple changes in a long sequence**

Slow Changes: ELL v/s OL

- **Slow Change gets tracked $\Leftrightarrow$ $\pi_{t|t}^{c,0,N} \approx \pi_{t|t}^{c,c}$, same for $\pi_{t|t-1}^{c,0,N}$**
  - Estimate of OL (ELL) close to OL (ELL) evaluated with changed system model,

$$OL_t^{c,0,N} \approx OL_t^{c,c}, \;\; ELL_t^{c,0,N} \approx ELL_t^{c,c}$$

  - Assume conditional entropies of Y_t^0 given past and of Y_t^c given past are equal, then $E[OL_t^{c,c}] = E[OL_t^{0,0}] < Th_{OL}$
  - **OL does not detect.**
  - When change becomes detectable ($ELL_t^{c,c} > Th_{ELL}$) then $ELL_t^{c,0,N} > Th_{ELL}$
  - **Change detected by ELL**

Sudden Changes: ELL v/s OL

- **Sudden Change gets filtered out**
  - Filter starts following the system model, large error in posterior
  - Large error in ELL estimate, $E[ELL_t^{c,0,N}] \approx h(p_t^0) < Th_{ELL}$
  - **ELL fails to detect**
  - Large ELL error implies large value of OL (proved)
  - **Change detected by OL**
- Most changes are in between the two extremes
- **ELL & OL complement each other for slow & sudden changes**

Complementariness of ELL and OL [Vaswani, ACC'2004]

Theorem. *ELL approx. error, $err_t^{c,0,N}$, is upper bounded by an increasing function of $OL_\tau^{c,0,N}$, $t_c \le \tau \le t$, i.e.*

$$err_t^{c,0,N} \le \sum_{\tau=t_c}^{t} e^{OL_\tau^{c,0,N}} \omega_1(\sigma_{obs}^2)\omega_2(\epsilon_\tau^{c,0}) + const$$

Implication for a **"detectable" change (true value of ELL large)**:

- OL fails to detect a change $\Longrightarrow$ ELL detects
- ELL fails to detect $\Longrightarrow$ OL detects

Stability of ELL Error [Vaswani, ACC'2004]

Theorem. *Average ELL approximation error is* **eventually monotonically decreasing (and hence stable)**, *for large enough N if*

- Change lasts for a finite time

- $f_t(X_t)$ continuous for all t

- π_0 has compact support

- $g_t(Y_t|x)$ (as a function of x) has compact support, for all Y_t

- The convergence of the bounded approx. of ELL is uniform in time

- Uses optimal filter stability results of [LeGland and Oudjane]
- Valid for any **unbounded function of state** (not just ELL)
- Error **asymptotically stable** under stronger assumptions

Extensions to $CUSUM$-ELL, $CUSUM$-OL

- **ELL error stability result extends to sum-ELL as well and hence to $CUSUM$-ELL**
- **For finite p (e.g. $p \leq 5$), can show stability of $joint$-ELL, by defining a new state space $\mathcal{X}_t = X_{t-p+1:t}$**
- **Use previous argument for slow changes: $CUSUM$-OL fails, $CUSUM$-ELL detects**
- **Complementariness result: for $CUSUM$-ELL & $CUSUM$-OL**

Contributions

- **ELL detects a change before loss of track (very useful). OL or Tracking Error detect after partial loss of track.**

- **Complementary behavior of ELL & OL for slow & sudden changes**
- **Stability of the total ELL approximation error for large N**

- **Relation to Kerridge Inaccuracy and a sufficient condition for the class of detectable changes using ELL [Vaswani, ACC'04]**

- **ELL error upper bounded by increasing function of "rate of change", increasing derivatives of all orders [Vaswani, ICASSP'04]**

Future Research

- **Changed Parameter Estimation**

- **Practical implications of the "rate of change" bound result and the stability result for particle filter design**

- **Applications and Performance Analysis**
  - Abnormal activity detection and activity segmentation
  - Neural signal processing (changes in STRFs of auditory neurons)
  - Acoustic tracking (changes in target motion model)
  - Communications applications
    - ∗ tracking slowly varying channels
    - ∗ congestion detection in networks