Quickest Change Detection via a Stochastic Approximation Approach

A Project Report

submitted by

K NAVNEETH NAIR

in partial fulfilment of the requirements
for the award of the degree of

MASTER OF TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY MADRAS.

May 2014

THESIS CERTIFICATE

This is to certify that the thesis titled **Quickest Change Detection via a Stochastic Approximation Approach**, submitted by **K Navneeth Nair**, to the Indian Institute of Technology, Madras, for the award of the degree of **Dual Degree in Electrical Engineering**, is a bona fide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof. Srikrishna Bhashyam
Research Guide
Associate Professor
Dept. of Electrical Engineerng
IIT-Madras, 600 036

Prof. Arun Pachai Kannu
Research Guide
Assistant Professor
Dept. of Electrical Engineering
IIT-Madras, 600 036

Place: Chennai

Date: 05th May 2014

ACKNOWLEDGEMENTS

This thesis would not have been possible without the help, guidance and support of both my guides, Prof.Srikrishna Bhashyam and Prof. Arun Pachai Kannu. I wish to thank them for introducing me to this area of study and for making this thesis happen. They have been a constant source of support. They were always available whenever I was in need of help and without their valuable perspectives, this study would not have progressed.

I would also like to thank all the professors of the Department of Electrical Engineering at IIT Madras, under whom I have done my courses.The list is long and I am indebted to all of them for the sound knowledge base they helped create and which stood me in great stead while doing this project.

Furthermore,a word needs to be said about my wingmates. The seven of them (Ravi,Ashwin, Subbu,Sarak, Praneesh, Dheeraj, Prithin) have been a constant in my life at college. They have provided the fun quotient to my life here, without which I believe I wouldn't have been able to work half as well as I did.

Lastly, I would like to thank my parents. They have been my greatest pillar of support throughout my life. Their undeterred belief in me and my abilities has always given me the confidence to pursue my interests to the fullest.

ABSTRACT

This project deals with the study of algorithms that detect change in sensor networks. More particularly, we study change detection in a sensor network with an unknown number of affected sensors. Existing schemes like the Parallel Cusum algorithm and the Adaptive Cusum algorithm are considered. A stochastic approximation variation to the Adaptive Cusum algorithm is introduced. It's convergence is proven theoretically and its performance with respect to existing schemes is further studied via simulation.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Context and motivation

Many practical problems arising in quality control, recognition oriented signal processing, fault detection and monitoring in industrial plants can be modeled with the help of parametric models in which the parameters are subject to *abrupt changes at unknown time instants*.

The detection of abrupt changes thus refers to tools that help us decide whether such a change occurred in the characteristics of the considered object.

In this project, the model we study is that of sensor networks. We assume that an unknown number of sensors in this sensor network are affected suddenly at an unknown time instant. We look at algorithms that will help us identify this time instant of change and predict the number of affected sensors as well.

1.2 System Model

Consider a network of K sensors where each sensor makes discrete observations $X_{k,n}$ with discrete time index $n = 1, 2, 3....\infty$ and sensor index $k = 1, 2, 3...K$. Observations are assumed to be statistically independent across time and across sensors.

Initially the observations $X_{k,n}$ follow the distribution f_k at the k^{th} sensor. A change event happens at an unknown time $v \in N$. Only a subset of sensors $S \subset 1, 2, ...K$ are affected by the change. After the change event, the observations at the affected sensors follow the distribution g_k.The set of affected sensors S and its size $|S|$ are unknown apriori.

This can be summarized as

$$X_{k,n} = \begin{cases} f_k : n < v \\ f_k : n \geq v \text{ and } k \notin S \\ g_k : n \geq v \text{ and } k \in S \end{cases}$$

Based on their observations, sensors send signals $U_{k,n}$ to the fusion center. Let Y_n denote the signal received at the fusion center at time n.

We consider the noisy physical layer fusion model (T.Bannerjee and Jayaprakasam (2011)) where the scalar observation (Y_n) is then given by

$$Y_n = \sum_{k=1}^{K} U_{k,n} + V_n$$

The model can be diagramatically represented as shown in the figure below.

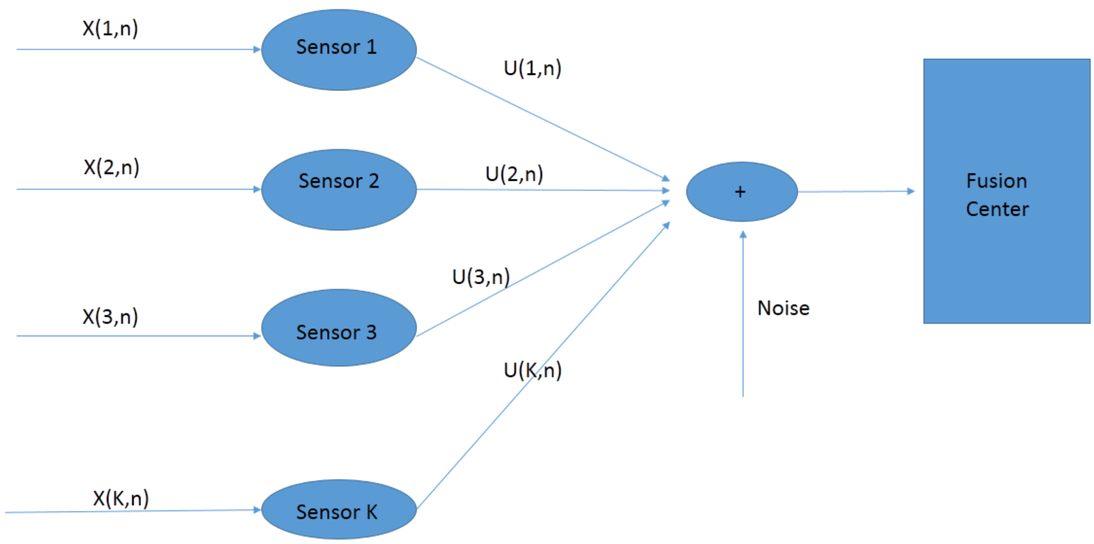


Figure 1.1: System Model

CHAPTER 2

Problem Formulation

Based on its observations say at a time T, the fusion center declares that change has happened. The general idea is to minimize the delay in declaring change for a given false alarm rate (can be visualized as an acceptable error probability.)

There are several ways in which the delay to be minimized can be formulated. We study Lorden's formulation [Lorden (1971)]. In Lorden's formulation, the objective is to minimize the supremum of the average delay conditioned on the worst possible realizations, subject to a constraint on the false alarm rate.

Let F_n denote the set of all observations upto time n. The worst case detection delay for a given change point $\tau_\omega(T)$ is then defined as

$$\tau_\omega(T) = \sup_{v \geq 1} esssup E_v[(T - v + 1)^+ / F_{v-1}] \tag{2.1}$$

where the supremum is taken over all possible change points and all possible pre change observations and expectation is taken over the post change observations.

The mean time to false alarm, is given by

$$\phi_l = E[T/v = \infty] \tag{2.2}$$

ϕ_l gives the average time taken to falsely declare the change when there was no change event.

Lorden's problem formulation is to find the T that minimizes $\tau_\omega(T)$ as given in (2.1) while satisfying the constraint on ϕ_l, defined in (2.2) .

CHAPTER 3

Existing Algorithms

3.1 Single Sensor Case

Let us first consider the single sensor case, where the problem of minimizing worst case detection delay under the average run length constraint is well studied (Basseville and Nikiforov).

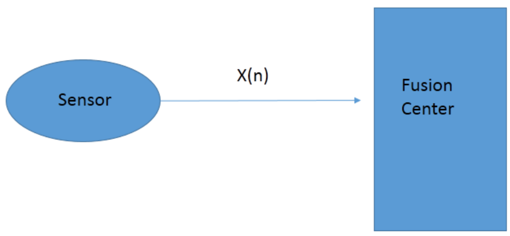


Figure 3.1: System Model Single Sensor

In this case the fusion center gets a random variable X that follows distribution f till the change point v and follows distribution g after that.

Theorem 3.1.1. *(Lorden, 1971; V.Nikiforov, 2000) Let γ represent the mean time to false alarm. There exists an asymptotic lower bound for worst case detection delay.*

$$\inf \tau_\omega(T) \sim \frac{log(\gamma)}{D(f||g)} \text{ as } \gamma \to \infty$$

where $D(f||g)$ represents the KL distance between the two distributions f and g.

The Cusum algorithm (Basseville and Nikiforov) gives the same lower bounds on the worst case detection delay as average run length approaches ∞. The Cusum algorithm is defined below. W_n is called the Cusum metric, L_n the log likelihood ratio and b is a pre-specified constant.

Algorithm 1 CUSUM Algorithm

```
Initialize n = 0 and W_n = 0

while W_n ≤ b do
    L_n = log g(X_n)/f(X_n)
    W_n = max(W_{n−1} + L_n, 0)
end while

Declare Change
```

3.2 Multiple Sensor Case

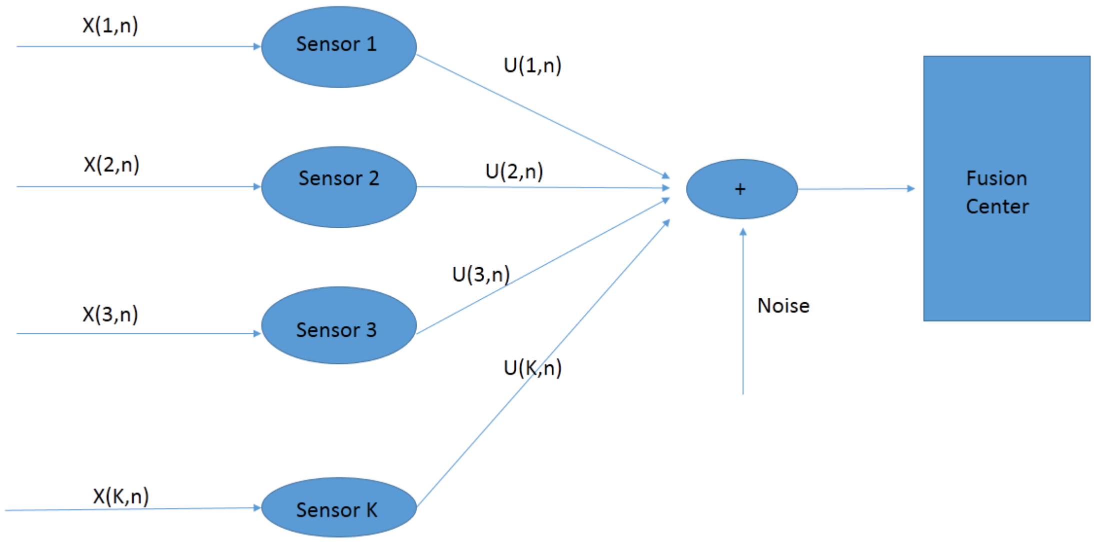


Figure 3.2: System Model

The summary messages $U_{k,n}$ sent by individual sensors at each time instant to the fusion center need to be computed. For this purpose, the Cusum metric at each sensor, at each time instant $W_{k,n}$ are computed based on the log likelihood ratios $L_{k,n}$.

Based on pre specified constant b, s the summary messages are computed by following the rule

$$U_{k,n} = \begin{cases} 0 : W_{k,n} < b \\ s : W_{k,n} \geq b \end{cases}$$

The algorithm can be summarized as shown below.

Algorithm 2 Algorithm for computing summary messages sent by each sensor to the fusion center

Initialize $n = 0$ and $W_{k,n} = 0 \forall k$

while n **do**
  for $k < K$ **do**
    $L_{k,n} = \log \frac{g_k(X_{k,n})}{f_k(X_{k,n})}$
    $W_{k,n} = \max(W_{k,n-1} + L_{k,n}, 0)$
    if $W_{k,n} \leq b$ **then**
      $U_{k,n} = 0$
    else
      $U_{k,n} = s$
    end if
    $k++$
  end for
  $n++$
end while

We further study a very specific case of the system model initially described in section 1.2. We assume that $\tau_k = \tau$ (actual change instant for each sensor)for all $k \in S$ i.e., all the affected sensors identify the change at the same time. We also assume that all sensors receive signals with Gaussian distribution $f_k = N(0, 1)$ prior to the change point. Post change, the affected sensors alone receive signals with Gaussian distribution $g_k = N(0.5, 1)$. The noise the fusion center receives is Gaussian $N(0, 1)$.

The signal received by the fusion center is given by the expression

$$Y_n = \sum_{k=1}^{K} U_{k,n} + V_n$$

Therefore, for the specific problem settings that we are studying,

$$Y_n = \begin{cases} N_0 : n < \tau \\ N_{\overline{m}} : else \end{cases}$$

where the mean $\overline{m}$ is unknown since the number of affected sensors is unknown. The mean $\overline{m}$ belongs to a finite discrete set $M = 0, s, ..., Ms$ for some M which depends on the number of sensors K.

The following algorithms have been studied in literature for the given problem.

- Parallel Cusum Algorithm
- Adaptive Cusum Algorithm

3.2.1 Parallel Cusum Algorithm

$$Y_n = \begin{cases} N_0 : n < \tau \\ N_{\overline{m}} : else \end{cases}$$

where the mean $\overline{m}$ is unknown since the number of affected sensors is unknown. The mean $\overline{m}$ belongs to a finite discrete set M = $0, s, ..., Ms$ for some M which depends on the number of sensors K.

Basing it on the system model described in section 1.2, $f_k = N(0, 1)$ and $g_k = N(\overline{m}, 1)$ where $\overline{m}$ can be from one of $0, s, ..., Ms$.

Lorden proved that the Cusum algorithm can be used to solve this problem optimally in an asymptotic sense (Lorden, 1971). His basic idea was to calculate a Cusum metric W_n^m for each possible value of $m \in M$. If any of the Cususm metrics crossed the pre-specified threshold, the change would be declared.

The fusion center detection rule is given by

$$\delta_p = \begin{cases} \text{change} & : \max_m W_n^m \geq a \\ \text{Continue} & : \text{else} \end{cases}$$

where a is a suitably chosen threshold to satisfy the average run length constraint. The algorithm can be represented as shown below.

Algorithm 3 Parallel Cusum algorithm

Initialize $n = 0$ and $W_n^m = 0 \forall m$

while n **do**
  for $m \in M$ **do**
    $L_n^m = log \frac{N_m(Y_n)}{N_0(Y_n)}$
    $W_n^m = \max(W_{n-1}^m + L_n^m, 0)$
  end for
  if $\max(W_n^m) \geq a$ **then**
    Declare Change
  else
    continue
  end if
  $n++$
end while

3.2.2 Adaptive Cusum Algorithm

This algorithm has been studied in detail in (Chengzhi Li, 2009).

It is a sub-optimal algorithm that is computationally less intensive that the optimal Parallel Cusum algorithm. Parallel Cusum computes the Cusum metric W_n^m for all $m \in M$. To reduce the computational load, the adaptive Cusum algorithm tries to estimate an $\overline{m}$ and computes only $W_n^{\overline{m}}$ at each time instant.

The algorithm can be interpreted as having two interleaved steps

- Parameter Tracking
- Cusum Test

Parameter Tracking

$$\text{Define } \; F(m) = E(L_n^m) \tag{3.1}$$

Theorem 3.2.1. *(Chengzhi Li, 2009) When the pre-change and post-change distributions belong to the exponential family, $F(m)$ is a strictly concave function and achieves its global maxima at $m = \overline{m}$.*

Proof. Exponential family of distributions is $p_\theta(x) = h(x)\exp(\theta T(x) - A(\theta))$

The hypothesis being tested is

$$\begin{cases} H_0 : \theta = 0 \\ H_1 : \theta = m \end{cases}$$

$F(\overline{m}) = E(L_n^{\overline{m}}) = log\frac{p_{\overline{m}}(Y_n)}{p_0(Y_n)}$

$\Rightarrow F(\overline{m}) = log\frac{p_m(Y_n)}{p_0(Y_n)} - log\frac{p_m(Y_n)}{p_{\overline{m}}(Y_n)}$

$\Rightarrow F(\overline{m}) = E(L_n^m) - D(p_m||p_{\overline{m}})$

where $D(p_m||p_{\overline{m}})$ is the KL distance between p_m and $p_{\overline{m}}$.

KL distance is a non negative function.$\Rightarrow$ $F(\overline{m})$ achieves maximum at $m = \overline{m}$.

$$\frac{dF(\overline{m})}{d\overline{m}} = -\frac{dD(p_m||p_{\overline{m}})}{d\overline{m}}$$

$$\Rightarrow \frac{dF(\overline{m})}{d\overline{m}} = -E[\frac{d}{d\overline{m}}((m - \overline{m})T(x) - (A(m) - A(\overline{m})))]$$

$$\Rightarrow \frac{dF(\overline{m})}{d\overline{m}} = E[T(x)] - \frac{dA(\overline{m})}{d\overline{m}}$$

$$\frac{d^2F(\overline{m})}{d^2\overline{m}} = -\frac{d^2A(\overline{m})}{d^2\overline{m}}$$

According to the differential identities of $A(m)$(Mei, 2003) , $\frac{d^2A(\overline{m})}{d^2\overline{m}} = Var(T(x)) > 0$

Therefore $F(\overline{m})$ is strictly concave. $\square$

Thus Adapative Cusum tries to find the value of m that maximizes $F(m)$ and sets that as $\overline{m}$.

Given a small value of ϵ, we can always find a p,q such that $q = p + \epsilon$ and $F(p) = F(q)$. $\overline{m}$ lies in the interval (p, q).

To find p, q we use the following recursive process.

Algorithm 4 Algorithm to find such a p and q

Choose an arbitrary p_o and set $q_o = p_o + \epsilon$

Set $n = 0$

while $F(p) \notin \delta$neigbourhood of $F(q)$ **do**

  $p_{n+1} = \max(0, p_n + \xi D^n)$

  $q_{n+1} = \min(Ms, p_{n+1} + \epsilon)$

  $n++$

end while

where ξ is the step size and $D^n = F(q^n) - F(p^n) = E[log(\frac{N_{q_n}(Y_n)}{N_{p_n}(Y_n)})]$.

Convergence Analysis. If $D^n > 0$, p_{n+1} and q_{n+1} will grow, so that D_{n+1} will decrease due to the concavity of the function. Similarly, if $D^n < 0$, p_{n+1} and q_{n+1} will move back so that D_{n+1} will increase. In both cases, D^n converges to zero surely. □

In practice, we replace the ensemble average with the time average. Therefore,

$$D^n = log(\frac{N_{q_n}(Y_n)}{N_{p_n}(Y_n)})$$

The mean value m_n at each iteration step is estimated as $\frac{p_n+q_n}{2}$. m_n converges to $\overline{m}$ with time.

Cusum Test

The Cusum metric is computed based on the current estimate of the mean (m_n)

$$W_n = \max(0, W_{n-1} + log(\frac{N_{m_n}(Y_n)}{N_0(Y_n)}))$$

The fusion center detection rule is given by

$$\delta_a = \begin{cases} \text{change} & : W_n \geq a \\ \text{Continue} & : \text{else} \end{cases}$$

where a is a suitably chosen threshold to satisfy the average run length constraint.

CHAPTER 4

Stochastic Approximation to Adaptive Cusum

4.1 Variation of Adaptive Cusum

Just like the Adaptive Cusum algorithm, this algorithm too has two interleaved steps

- Parameter Tracking
- Cusum Test

The difference lies in how the parameter tracking is performed.

4.1.1 Parameter Tracking

As shown in the previous section, $F(m) = E[L_n^m]$ is a strictly concave function. We try to estimate the value of m for which this function attains its maximum.

From (Kiefer and Wolfowitz, 1952) we know that, given two infinite sequences of positive numbers a_n and c_n that satisfy the following properties,

- $c_n \to 0$
- $\Sigma a_n = \infty$
- $\Sigma a_n c_n < \infty$
- $\Sigma a_n^2 c_n^- 2 < \infty$

For example, $a_n = n^{-1}$ and $c_n = n^{-1/3}$.

We can estimate the $\overline{m}$ that maximizes $F(m)$ using the following algorithm

---

Algorithm 5 Stochastic estimation of maximum of concave function

---

Choose an arbitrary m_0

Set $n = 0$

while n **do**

$$m_{n+1} = m_n + a_n \frac{log(\frac{N_{m_n+c_n}(Y_n)}{N_0(Y_n)}) - log(\frac{N_{m_n-c_n}(Y_n)}{N_0(Y_n)})}{c_n}$$

$n++$

end while

4.2 Proof of Convergence

Theorem 4.2.1 (Stochastic Estimation of the maximum of a regression function)**.** *(Kiefer and Wolfowitz, 1952) Let $M(x)$ be a concave function achieving its maximum at $x = \theta$. Let us say that $M(x)$ itself is unknown, but its value at various observation levels x is known. Then, starting from an arbitrary point x_1, one can successively obtain $x_2, x_3....$ such that $x_n \to \theta$ in probability as $n \to \infty$.*

In our case $F(m)$ is a concave function. Thus, starting from some arbitrary m_1, it is possible to successively find $m_2, m_3....$ such that $m_n \to \theta$ in probability as $n \to \infty$.

Given two infinite sequences of positive numbers a_n and c_n that satisfy the following properties,

- $c_n \to 0$
- $\Sigma a_n = \infty$
- $\Sigma a_n c_n < \infty$
- $\Sigma a_n^2 c_n^- 2 < \infty$

We can estimate the $\overline{m}$ that maximizes $F(m)$ using the following recursion

$$m_{n+1} = m_n + a_n \frac{log(\frac{N_{m_n+c_n}(Y_n)}{N_0(Y_n)}) - log(\frac{N_{m_n-c_n}(Y_n)}{N_0(Y_n)})}{c_n}$$

The above stochastic approximation recursion will converge to the maximum only when $F(m)$ is a *regular* function. Thus it needs to be shown that $F(m)$ is a *regular* function. This is shown in the following lemma.

Theorem 4.2.2. *When the pre change and post change distributions of the observations are Gaussian distributions varying only in their mean, then the function $F(m)$ is a regular function and satisfies the following criteria,*

1. *There exist positive β and B such that*

$$|m' - \overline{m}| + |m'' - \overline{m}| < \beta \Rightarrow |F(m') - F(m")| < B|m' - m"| \tag{4.1}$$

2. *There exist positive ρ and R such that*

$$|m' - m"| < \rho \Rightarrow |F(m') - F(m")| < R \tag{4.2}$$

3. *For every $\delta > 0$ there exists a positive $\Pi(\delta)$ such that*

$$|m - \overline{m}| > \delta \Rightarrow \inf_{\frac{1}{2}\delta > \epsilon > 0} \frac{|F(m+\epsilon) - F(m-\epsilon)|}{\epsilon} > \Pi(\delta) \tag{4.3}$$

Proof. Let the pre change distribution be $N(\lambda, \sigma^2)$ and the actual post change distribution be $N(\Psi, \sigma^2)$.

$F(m) = E(l_m(Y_n))$ where

$l_m(t) = log(\frac{N_m(Y_n)}{N_\lambda(Y_n)})$

We know that $N_m(Y_n) = \sqrt{\frac{1}{2\Pi\sigma^2}} \exp(-\frac{(Y_n - m)^2}{2\sigma^2})$

Substituting in the $l_m(t)$ expression, we get

$l_m(t) = \frac{m - \lambda}{2\sigma^2}(2Y_n - \lambda - m)$

$F(m) = E(l_m(Y_n)) = \int_{-\infty}^{\infty} l_m(Y_n)\sqrt{\frac{1}{2\Pi\sigma^2}} \exp(-\frac{(Y_n - \Psi)^2}{2\sigma^2})dY_n$

On simplification, this gives $F(m) = (\frac{m-\lambda}{2\sigma^2})[\frac{2\sigma^2}{\sqrt{2\pi\sigma^2}} - \lambda - m + 2\Psi]$

Thus $F(m)$ is a quadratic function of the form $F(m) = (m-a)(m-b)$. Furthermore, $F(m)$ is defined over a restricted domain on m.

The proof of statement 4.1 comes directly from the fact that a quadratic function is Lipschitz continuous.

Since $F(m)$ is defined only over a restricted range of values of m, $F(m)$ will also take values from a restricted real set. Hence, $|F(m') - F(m")|$ will always have an upper bound. Therefore, statement 4.2 is proved.

For a quadratic function in a restricted domain, the rate of fall outside a δ neighborhood of the maximum, will have a lower bound. In specific, when $|m-\overline{m}| > \delta$, then infimum of $\inf_{\frac{1}{2}\delta>\epsilon>0} \frac{|F(m+\epsilon) - F(m-\epsilon)|}{\epsilon}$ is given by

$$inf(\frac{F(\overline{m}-\delta/2) - F(\overline{m}-3\delta/2)}{\delta/2}, \frac{F(\overline{m}+\delta/2) - F(\overline{m}+3\delta/2)}{\delta/2})$$

This lower bound is a function of δ.
Hence,$|m-\overline{m}| > \delta$ implies $\inf_{\frac{1}{2}\delta>\epsilon>0} \frac{|F(m+\epsilon) - F(m-\epsilon)|}{\epsilon} > \Pi(\delta)$, thus proving statement 4.3. □

Therefore the maximum of $F(m)$ can be found by the recursive process detailed above.

4.3 Simulation of Performance

Number of sensors considered is $K = 100$. Pre change distribution is $N(0, 1)$. Post change distribution is $N(0.5, 1)$. Detection delay at the fusion center is used as the performance metric keeping the same average length constraint for different schemes. Simulations are performed with signal to noise ratio of 0 db and binary quantization threshold of b = 460.52. The change point is set at v=100.

The following graphs are plotted for each set of parameters:

1. Detection delay at fusion center for each algorithm

2. Convergence of estimated mean ($\overline{m}$) to actual mean (number of affected sensors)

4.4 Performance comparison between algorithms

The sub optimal behavior of the Adaptive Cusum and Stochastic Approximation Cusum algorithms is due to the time the parameter tracking part of the algorithms take. The faster the algorithms find the maximum of $F(m)$, the closer their detection delays will be to the detection delay of the Parallel Cusum algorithm.

The parameter tracking of the Adaptive Cusum algorithm as explained in subsection 3.2.2 depends on two parameters:

1. ξ

2. ϵ

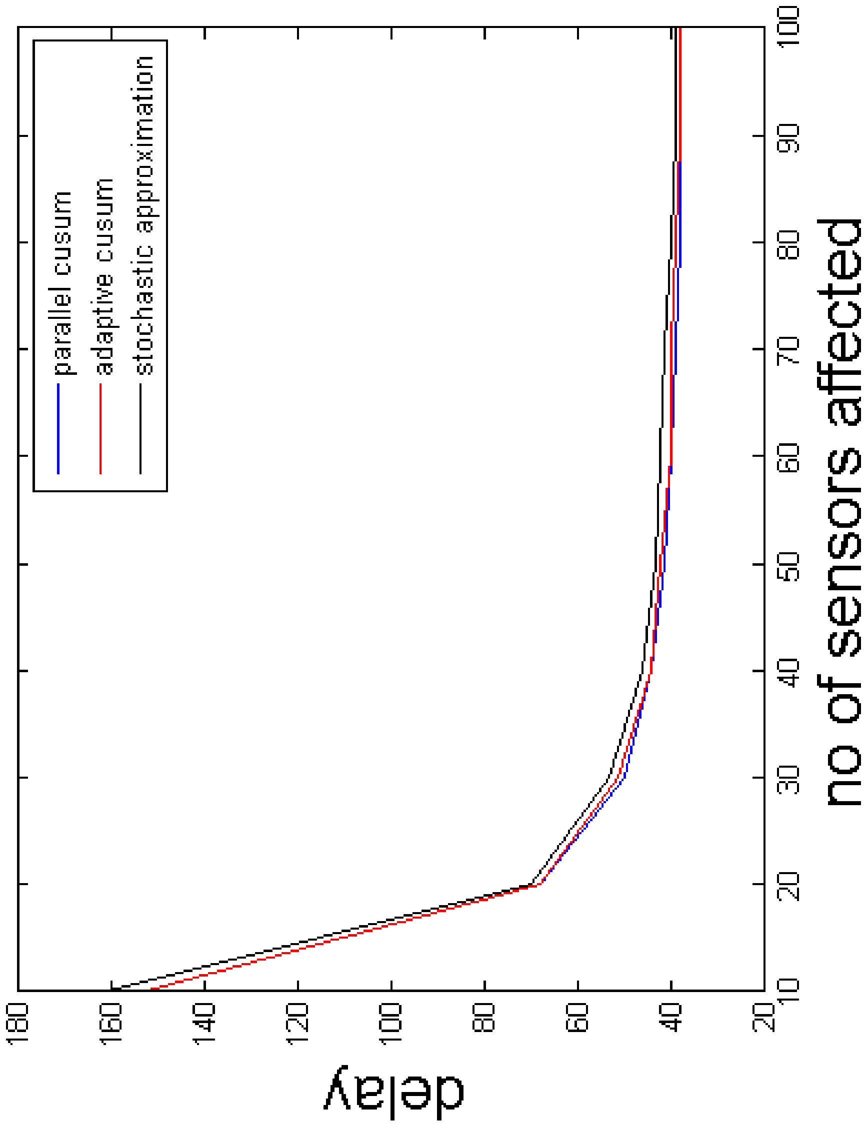
The parameter tracking of the Stochastic Approximation Cusum algorithm as explained in subsection 4.1.1 depends on two parameters:

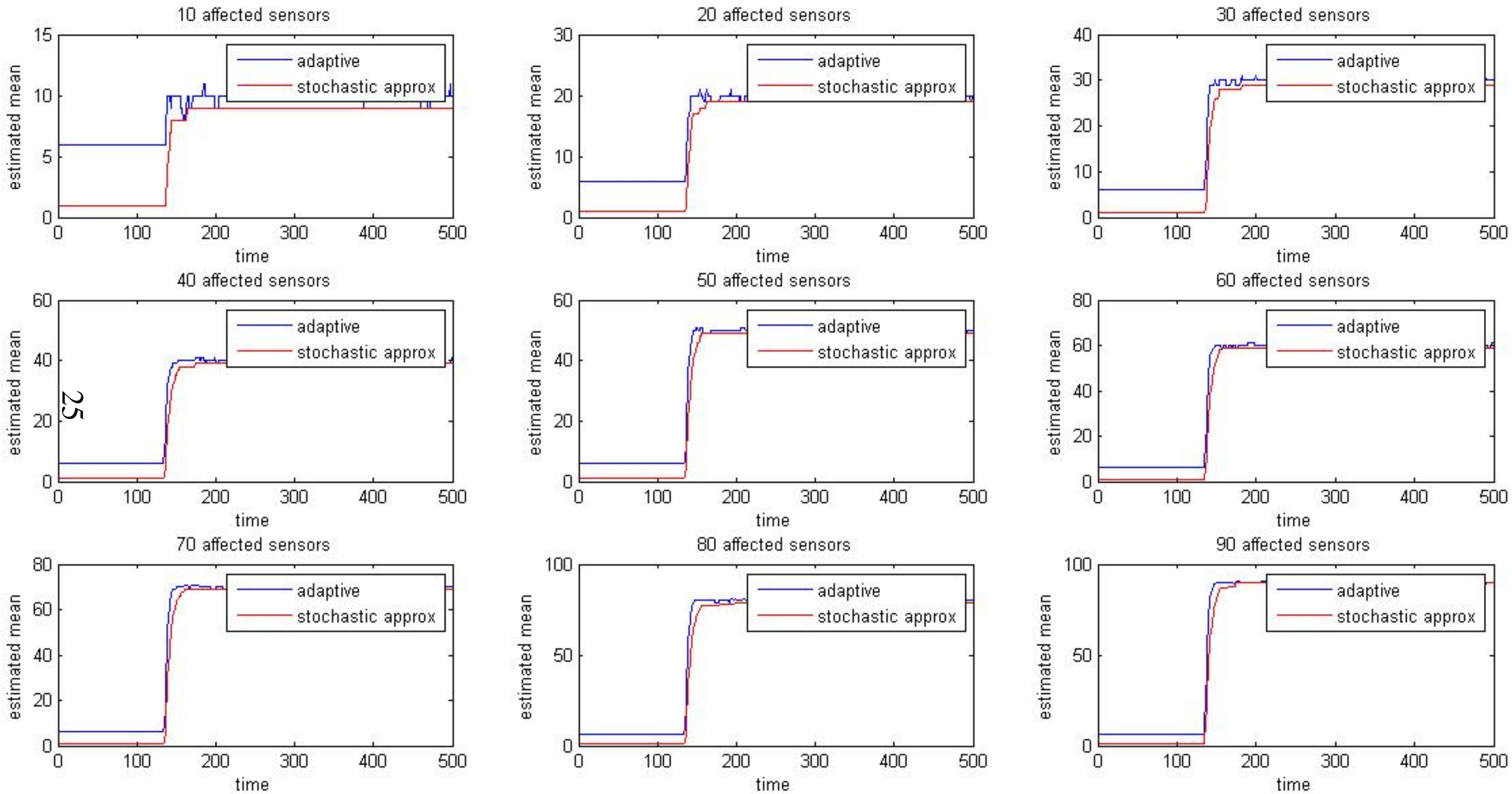
1. Series a_n
2. Series c_n

It is not possible to choose the parameters of the respective algorithms appropriately to improve performance without any prior knowledge of the number of sensors that will get affected.

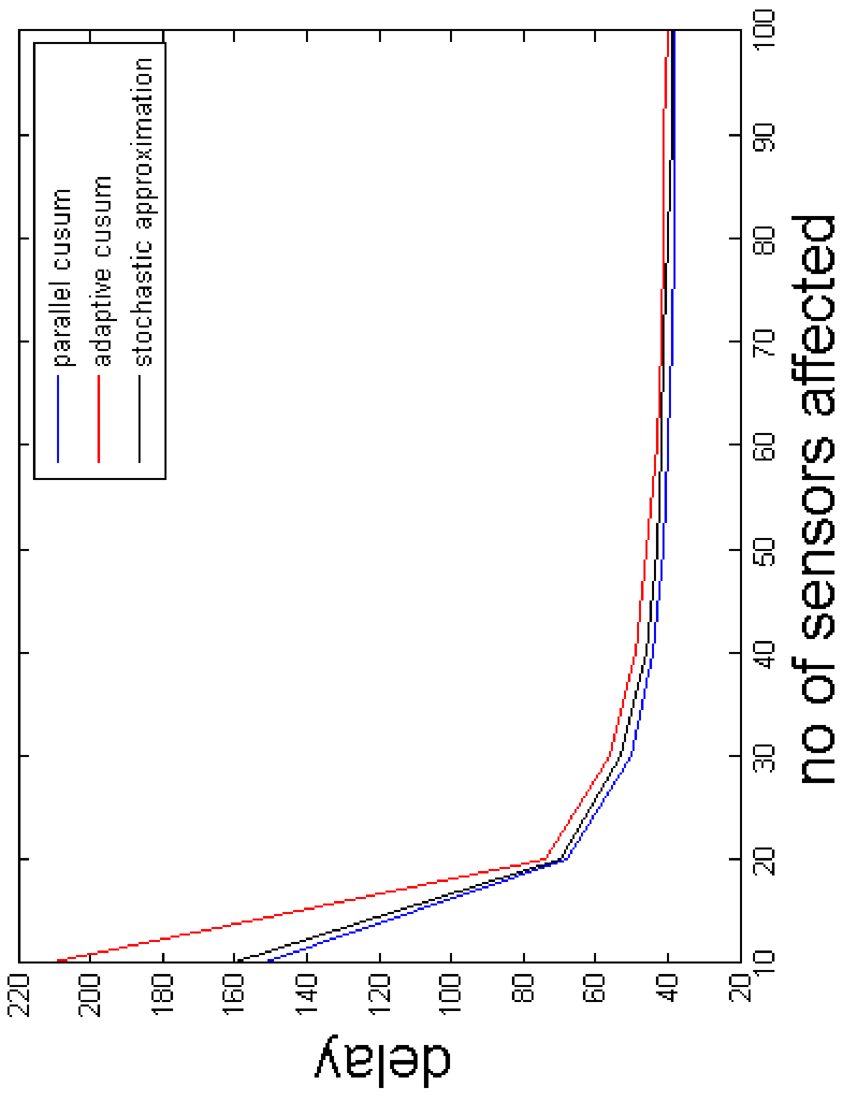
In the case of our problem formulation, where there is no prior information about the change point or the number of affected sensors, depending on the parameters chosen, either algorithm can be shown to outperform each other. This is illustrated below.

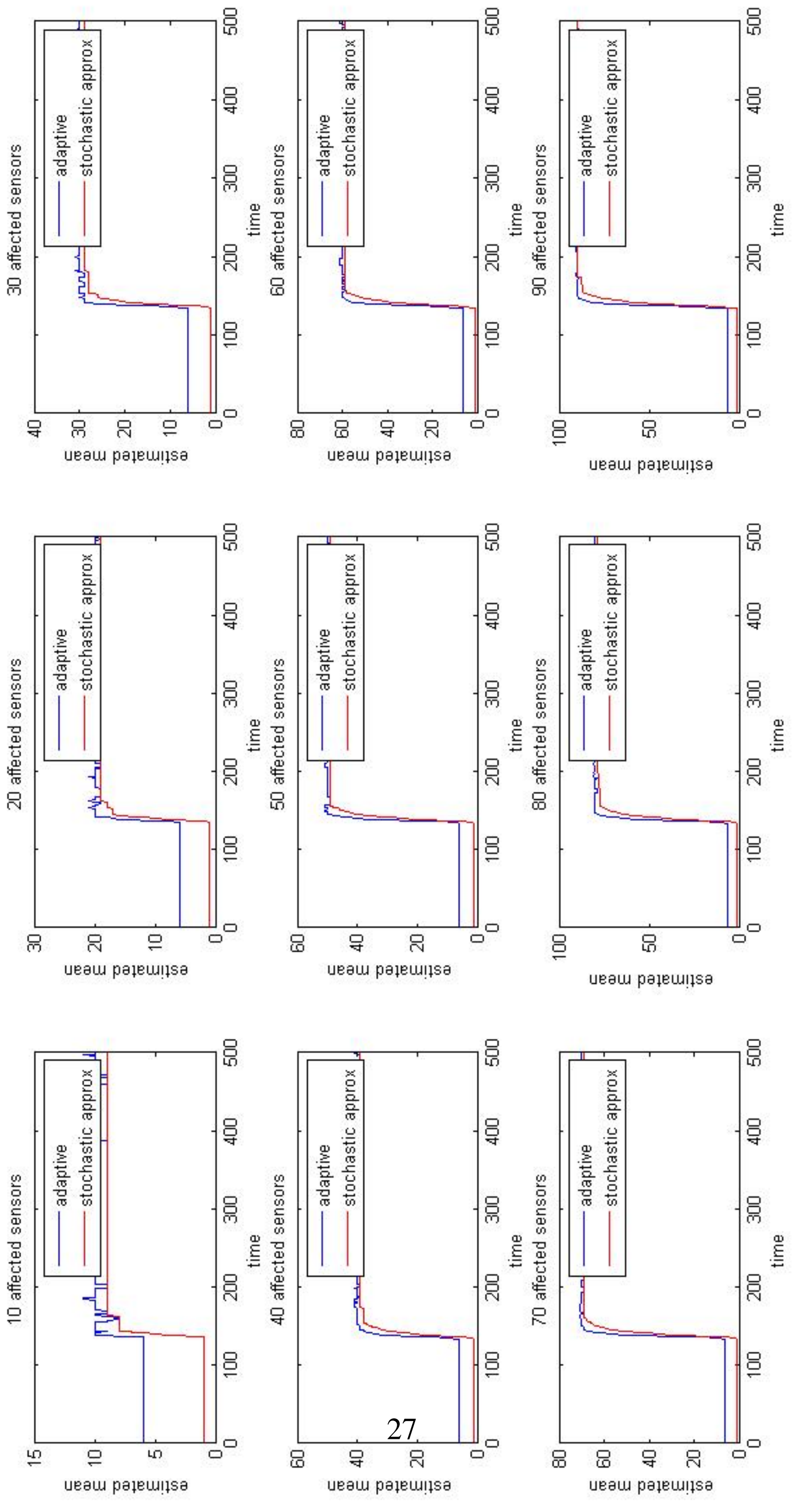
$\epsilon = 10, \xi = 0.03, a_n = n^-1, c_n = n^{-1/3}$




10 affected sensors
20 affected sensors
30 affected sensors
40 affected sensors
50 affected sensors
60 affected sensors
70 affected sensors
80 affected sensors
90 affected sensors
estimated mean
time
adaptive
stochastic approx

$\epsilon = 2, \xi = 0.03, a_n = n^-1, c_n = n^{-1/3}$




10 affected sensors
20 affected sensors
30 affected sensors
40 affected sensors
50 affected sensors
60 affected sensors
70 affected sensors
80 affected sensors
90 affected sensors
adaptive
stochastic approx
estimated mean
time

4.5 Further Work

In this study, we consider only the exponential family of distributions where there is change in only one parameter at the change point.

This choice of exponential family of distributions ensures that the $F(m)$ function (define in equation 3.1) is symmetric about its maxima.

A possible area of expansion would be to explore changes in distribution that ensure that the resulting $F(m)$, is a strictly concave function but asymmetric about its maxima.

In such a situation, the parameter tracking part of the Adaptive Cusum algorithm will not converge to the maximising argument. The Stochastic Approximation algorithm will still converge to the maxima and will hence outperform the Adaptive Cusum algorithm.

CHAPTER 5

Conclusion

Quickest change detection algorithms for sensor networks with an unknown number of affected sensors at an unknown time instant are studied. Lorden's formulation for the quickest change detection problem is the one used as the base for the analysis. Existing optimal schemes like the Parallel Cusum Algorithm and suboptimal schemes like the Adaptive Cusum algorithm are discussed. An alternate suboptimal scheme based on Stochastic Approximation is suggested. The performance in terms of the detection delay at the fusion center is studied and simulated for each of the schemes. The convergence of the newly suggested suboptimal scheme is also theoretically proven. A comparison is made between the two suboptimal schemes and a further possible area of work on the topic is suggested.

REFERENCES

1. **Basseville, M.** and **I. V. Nikiforov**, *Detection of Abrupt Changes : Theory and Practice*. Prentice Hall Inc., .

2. **Chengzhi Li, H. L., Huaiyu Dai** (2009). Adaptive quickest change detection with unknown parameter. *IEEE*, **24**(6).

3. **Kiefer, J.** and **J. Wolfowitz** (1952). Stochastic estimation of the maximum of a regression function. *Annals of Mathematical Statistics*.

4. **Lorden, G.** (1971). Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics*, **42**(6), 1897–1908.

5. **Mei, J.** (2003). Asymptotically optimal methods for sequential detection. *PhD dissertation, California Institute of Technology*.

6. **T.Bannerjee, V. K., V. Sharma** and **A. Jayaprakasam** (2011). Generalized analysis of a distributed energy efficient algorithm for change detection. *Wireless Communications,IEEE Transactions on*, **10**, 91–101.

7. **V.Nikiforov, I.** (2000). A suboptimal quadratic change detection scheme. *IEEE Transactions on Information Theory*, **46**(6), 1864–1878.