Online Heavy-tailed Change-point detection (Supplementary Materials)

Abishek Sankararaman[*1]   **Balakrishnan (Murali) Narayanaswamy**[1]

[1]AWS AI Labs

A CHANGE-POINT LOCALIZATION

Algorithm 1 Online `Clipped-SGD` Change Point Detection and Localization

Input: $(\eta_t)_{t\geq 1}$, $\lambda > 0$, $\theta_0 \in \Theta$, $\delta \in (0,1)$ FPR guarantee
$r \leftarrow 1$
$\widehat{\theta}_{t,t-1} \leftarrow \theta_0$, for all $t \geq 1$.
Set $\tau_c^{(0)} \leftarrow 0$
Set `Num-change-points` $\leftarrow 0$
for each time $t = 1, 2, \cdots,$ **do**
    Receive sample X_t
    $\widehat{\theta}_{s,t} \leftarrow \prod_\theta(\widehat{\theta}_{s,t-1} - \eta_{t-s}\text{clip}(X_t - \widehat{\theta}_{s,t-1}, \lambda))$, for every $r \leq s \leq t$.
    if $\exists s \in (r,t)$ such that $\|\widehat{\theta}_{r:s} - \widehat{\theta}_{s+1:t}\|_2^2 > \mathcal{B}\left(s-r, \frac{\delta}{2(t-r)(t-r+1)}\right) + \mathcal{B}\left(t-s-1, \frac{\delta}{2(t-r)(t-r+1)}\right)$ {$B(\cdot,\cdot)$ is defined in Equation (5)} **then**
        Set **Restart**$_t \leftarrow 1$ {Change point detected}
        Set `Num-change-points` $\leftarrow$`Num-change-points` $+1$ {Increment number of change-points detected}
        Output time interval $[\inf\{s \in (r,t) \text{ s.t. } \mathfrak{B}(r,s,t,\delta) = 1\}, \sup\{s \in (r,t) \text{ s.t. } \mathfrak{B}(r,s,t,\delta) = 1\}]$ as the location of the change-point {$\mathfrak{B}()$ defined in Equation (8)}
        $r \leftarrow t+1$
    else
        Set **Restart**$_t \leftarrow 0$
    end if
end for

B PROOF FOR ROBUST ESTIMATION IN THEOREM 3.1

We follow the same proof architecture as that of Proof of [Tsai et al., 2022].

Fix a time $t \in \mathbb{N}$. We define a sequence of random variable $(\psi_t)_{t\geq 1}$ as follows.

$$\psi_t := \text{clip}((X_t - \widehat{\theta}_{t-1}), \lambda) - (\theta^* - \widehat{\theta}_{t-1}),$$

Consider any time t. We have

$$\|\theta_t - \theta^*\|_2^2 = \|\prod_\Theta(\widehat{\theta}_{t-1} - \eta_t\text{clip}(X_t - \widehat{\theta}_{t-1}, \lambda)) - \theta^*\|_2^2, \tag{1}$$

$$\stackrel{(a)}{\leq} \|\widehat{\theta}_{t-1} - \eta_t\text{clip}(X_t - \widehat{\theta}_{t-1}, \lambda) - \theta^*\|_2^2, \tag{2}$$

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

$$
\begin{aligned}
&= \|\widehat{\theta}_{t-1} - \eta_t(\psi_t + (\theta^* - \widehat{\theta}_{t-1})) - \theta^*\|_2^2, \\
&= \|\widehat{\theta}_{t-1} - \theta^*\|_2^2 + \eta_t^2\|\psi_t + (\theta^* - \widehat{\theta}_{t-1})\|_2^2 - 2\eta_t\langle\widehat{\theta}_{t-1} - \theta^*, \psi_t + (\theta^* - \widehat{\theta}_{t-1})\rangle, \\
&\overset{(b)}{\leq} \|\widehat{\theta}_{t-1} - \theta^*\|_2^2 + 2\eta_t^2\|\psi_t\|_2^2 + 2\eta_t^2\|(\theta^* - \widehat{\theta}_{t-1})\|_2^2 - 2\eta_t\langle\widehat{\theta}_{t-1} - \theta^*, \psi_t + (\theta^* - \widehat{\theta}_{t-1})\rangle, \qquad (3)
\end{aligned}
$$

Step (a) follows since Θ is a convex set, $\|\mathcal{P}_\Theta(\widehat{\theta}_t) - \theta^*\| \leq \|\widehat{\theta}_t - \theta^*\|$, since $\theta^* \in \Theta$. In step (b), we use the fact that $\|a+b\|_2^2 \leq 2\|a\|_2^2 + 2\|b\|_2^2$, for all $a, b \in \mathbb{R}^d$. Substituting Equation (35) into (3), we get that

$$
\begin{aligned}
\|\theta^* - \theta_t\|_2^2 &\leq \|\widehat{\theta}_{t-1} - \theta^*\|_2^2 + 2\eta_t^2\|\psi_t\|_2^2 - 2\eta_t\langle\widehat{\theta}_{t-1} - \theta_t^*, \psi_t\rangle \\
&\quad + 2\eta_t^2\left((M+m)\langle(\theta^* - \widehat{\theta}_{t-1}), \widehat{\theta}_{t-1} - \theta_t^*\rangle - mM\|\widehat{\theta}_{t-1} - \theta^*\|_2^2\right) - 2\eta_t\langle(\theta^* - \widehat{\theta}_{t-1}), \widehat{\theta}_{t-1} - \theta_t^*\rangle.
\end{aligned}
$$

Re-arranging the equation above yields

$$
\begin{aligned}
\|\theta^* - \theta_t\|_2^2 &\leq (1 - 2\eta_t^2 mM)\|\widehat{\theta}_{t-1} - \theta^*\|_2^2 + 2\eta_t^2\|\psi_t\|_2^2 - 2\eta_t\langle\widehat{\theta}_{t-1} - \theta^*, \psi_t\rangle \\
&\quad - 2\eta_t(1 - \eta_t\left((M+m)\right)\langle(\theta^* - \widehat{\theta}_{t-1}), \widehat{\theta}_{t-1} - \theta^*\rangle.
\end{aligned}
$$

Further substituting Equation (34) into the display above yields that

$$
\begin{aligned}
\|\theta^* - \widehat{\theta}_t\|_2^2 &\leq (1 - 2\eta_t m + 2\eta_t^2 m^2)\|\widehat{\theta}_{t-1} - \theta^*\|_2^2 + 2\eta_t^2\|\psi_t\|_2^2 - 2\eta_t\langle\widehat{\theta}_{t-1} - \theta^*, \psi_t\rangle, \\
&\leq (1 - \eta_t m)\|\widehat{\theta}_{t-1} - \theta^*\|_2^2 + 2\eta_t^2\|\psi_t\|_2^2 - 2\eta_t\langle\widehat{\theta}_{t-1} - \theta^*, \psi_t\rangle,
\end{aligned}
$$

where the inequality comes from the fact that if $\eta_t m < 1 \implies 2\eta_t m - 2\eta_t^2 m^2 > \eta m$.

$$
\|\theta^* - \widehat{\theta}_t\|_2^2 \leq (1 - \eta_t m)\|\widehat{\theta}_{t-1} - \theta^*\|_2^2 + 2\eta_t^2\|\psi_t\|_2^2 - 2\eta_t\langle\widehat{\theta}_{t-1} - \theta^*, \psi_t\rangle. \qquad (4)
$$

Unrolling the recursion yields,

$$
\begin{aligned}
\|\theta^* - \widehat{\theta}_t\|_2^2 \leq \prod_{u=1}^{t}(1 - \eta_u m)\|\theta_1 - \theta^*\|_2^2 &+ 2\eta_t^2\sum_{s=1}^{t-1}\prod_{u=1}^{s}(1 - \eta_{t-u+1}m)\|\psi_{t-s+1}\|_2^2 \\
&-2\eta_t\sum_{s=1}^{t-1}\prod_{u=1}^{s}(1 - \eta_{t-u+1}m)\langle\theta_{t-s} - \theta^*, \psi_{t-s+1}\rangle.
\end{aligned}
$$

Using the fact that $\prod_{u=1}^{s}(1 - \eta_{t-u+1}m) = \frac{(t-s+\gamma-3)(t-s+\gamma-2)}{(t+\gamma)(t+\gamma-1)}$, we get that

$$
\|\theta^* - \widehat{\theta}_t\|_2^2 \leq \frac{(\gamma-2)(\gamma-1)\|\theta_1 - \theta^*\|_2^2}{(t+\gamma)(t+\gamma-1)} \qquad (5)
$$

$$
-2\eta_t\sum_{s=1}^{t-1}\frac{(t-s+\gamma-3)(t-s+\gamma-2)\langle\theta_{t-s} - \theta^*, \psi_{t-s+1}\rangle}{(t+\gamma)(t+\gamma-1)}. \qquad (6)
$$

Denote by $\psi_t := \psi_t^{(b)} + \psi_t^{(v)}$, where $\psi_t^{(b)} := \mathbb{E}_{Z_t}[\psi_t|\mathcal{F}_{t-1}]$ and $\psi_t^{(v)} := \psi_t - \psi_t^{(b)}$. Using this in the display above and using that fact that $\|a+b\|_2^2 \leq 2\|a\|_2^2 + 2\|b\|_2^2$, we get

$$
\begin{aligned}
\|\theta_t^* - \theta\|_2^2 \leq \frac{(\gamma-1)(\gamma-2)\|\theta_1 - \theta^*\|_2^2}{(t+\gamma)(t+\gamma-1)} &+ 4\eta_t^2\sum_{s=1}^{t-1}\frac{(t-s+\gamma-3)(t-s+\gamma-2)\|\psi_{t-s+1}\|_2^2}{(t+\gamma)(t+\gamma-1)} + \\
&- 2\eta_t\sum_{s=1}^{t-1}\frac{(t-s+\gamma-3)(t-s+\gamma-2)\langle\theta_{t-s} - \theta^*, \psi_{t-s+1}^{(b)}\rangle}{(t+\gamma)(t+\gamma-1)} \\
&- 2\eta_t\sum_{s=1}^{t-1}\frac{(t-s+\gamma-3)(t-s+\gamma-2)\langle\theta_{t-s} - \theta^*, \psi_{t-s+1}^{(v)}\rangle}{(t+\gamma)(t+\gamma-1)}.
\end{aligned}
$$

Further simplifying by adding and subtracting $\mathbb{E}_{Z_t}[\|\psi_t^{(v)}\|_2^2|\mathcal{F}_{t-1}]$ to be above display, we get

$$\|\theta^* - \widehat{\theta}_t\|_2^2 \leq \frac{(\gamma-1)(\gamma-2)\|\theta_1 - \theta^*\|_2^2}{(t+\gamma)(t+\gamma-1)} + 4\eta_t^2 \sum_{s=1}^{t-1} \frac{(t-s+\gamma-3)(t-s+\gamma-2)\|\psi_{t-s+1}^{(b)}\|_2^2}{(t+\gamma)(t+\gamma-1)} \tag{7}$$

$$+ 4\eta_t^2 \sum_{s=1}^{t-1} \frac{(t-s+\gamma-3)(t-s+\gamma-2)\mathbb{E}_{Z_{t-s+1}}[\|\psi_{t-s+1}^{(v)}\|_2^2|\mathcal{F}_{t-s}]}{(t+\gamma)(t+\gamma-1)}$$

$$+ 4\eta_t^2 \sum_{s=1}^{t-1} \frac{(t-s+\gamma-3)(t-s+\gamma-2)(\|\psi_{t-s+1}^{(v)}\|_2^2 - \mathbb{E}_{Z_{t-s+1}}[\|\psi_{t-s+1}^{(v)}\|_2^2|\mathcal{F}_{t-s}])}{(t+\gamma)(t+\gamma-1)} \tag{8}$$

$$- 2\eta_t \sum_{s=1}^{t-1} \frac{(t-s+\gamma-3)(t-s+\gamma-2)\langle \theta_{t-s} - \theta^*, \psi_{t-s+1}^{(b)}\rangle}{(t+\gamma)(t+\gamma-1)}$$

$$- 2\eta \sum_{s=1}^{t-1} \frac{(t-s+\gamma-3)(t-s+\gamma-2)\langle \theta_{t-s} - \theta^*, \psi_{t-s+1}^{(v)}\rangle}{(t+\gamma)(t+\gamma-1)}. \tag{9}$$

Lemma B.1 (Lemma F.5 [Gorbunov et al., 2020])**.** *If $\lambda \geq 2G$, the following inequalities hold almost-surely for all times t.*

$$\|\psi_t^{(v)}\| \leq 2\lambda \mathbf{1}_{\sigma>0} \tag{10}$$

$$\|\psi_t^{(b)}\|_2 \leq \frac{4\sigma^2}{\lambda} \tag{11}$$

$$\mathbb{E}_{Z_t}[\|\psi_t^{(v)}\|_2^2|\mathcal{F}_{t-1}] \leq 10\sigma^2 \tag{12}$$

Simplifying Equation (9) using bounds in Lemma B.1, along with the fact that for all $1 \leq s \leq t$ and $\gamma \geq 1$, $\frac{(t-s+\gamma-3)(t-s+\gamma-2)}{(t+\gamma)(t+\gamma-1)} \leq \frac{t-s+\gamma}{t+\gamma}$ we get

$$\|\theta^* - \widehat{\theta}_t\|_2^2 \leq \frac{(\gamma-1)(\gamma-2)\|\theta_1 - \theta^*\|_2^2}{(t+\gamma)(t+\gamma-1)} + \frac{16\eta_t^2\sigma^2}{\lambda} \sum_{s=1}^{t-1} \frac{t-s+\gamma}{t+\gamma} + 4\eta_t^2\sigma^2 \sum_{s=1}^{t-1} \frac{t-s+\gamma}{t+\gamma}$$

$$+ 4\eta_t^2 \sum_{s=1}^{t-1} \frac{(t-s+\gamma)(\|\psi_{t-s+1}^{(v)}\|_2^2 - \mathbb{E}_{Z_{t-s+1}}[\|\psi_{t-s+1}^{(v)}\|_2^2|\mathcal{F}_{t-s+1}])}{t+\gamma}$$

$$+ 2\eta_t \sum_{s=1}^{t-1} \frac{(t-s+\gamma)\|\theta_{t-s} - \theta^*\|\|\psi_{t-s+1}^{(b)}\|}{t+\gamma} + -2\eta_t \sum_{s=1}^{t-1} \frac{(t-s+\gamma)\langle \theta_{t-s} - \theta^*, \psi_{t-s+1}^{(v)}\rangle}{t+\gamma}. \tag{13}$$

Further applying the bound that $\|\psi_t^{(b)}\| \leq \frac{4\sigma^2}{\lambda}$

$$\|\theta^* - \widehat{\theta}_t\|_2^2 \leq \frac{(\gamma-1)(\gamma-2)\|\theta_1 - \theta^*\|_2^2}{(t+\gamma)(t+\gamma-1)} + \underbrace{\left(\frac{16\eta_t^2\sigma^2}{\lambda} + 4\eta_t^2\sigma^2\right) \sum_{s=1}^{t-1} \frac{t-s+1}{t+\gamma}}_{\text{Term 1}}$$

$$+ \underbrace{4\eta_t^2 \sum_{s=1}^{t-1} \frac{(t-s+\gamma)(\|\psi_{t-s+1}^{(v)}\|_2^2 - \mathbb{E}_{Z_{t-s+1}}[\|\psi_{t-s+1}^{(v)}\|_2^2|\mathcal{F}_{t-s+1}])}{t+\gamma}}_{\text{Term 2}}$$

$$+ \underbrace{\frac{8\sigma^2\eta_t}{\lambda} \sum_{s=1}^{t-1} \frac{(t-s+\gamma)\|\theta_{t-s} - \theta^*\|}{t+\gamma}}_{\text{Term 3}} \underbrace{-2\eta_t \sum_{s=1}^{t-1} \frac{(t-s+\gamma)\langle \theta_{t-s} - \theta^*, \psi_{t-s+1}^{(v)}\rangle}{t+\gamma}}_{\text{Term 4}}. \tag{14}$$

B.1 PROBABILISTIC ANALYSIS

Definitions

For every $t \geq 1$, denote by the constant

$$C_t = \max\left(\frac{1024\sigma^4}{G^2 m^2 \lambda^2}, \frac{8\lambda\sqrt{\ln\left(\frac{2t^3}{\delta}\right)}}{\gamma^2 G}\right). \tag{15}$$

Denote by the deterministic constant $\xi_u^{(t)}$ for $u = 1, \cdots, t$ as

$$\left(\xi_u^{(t)}\right)^2 := C_t\Bigg[\left(\frac{16\sigma^2}{\lambda} + 4\sigma^2\right)\frac{1}{2m^2(u+1)} + \frac{96\lambda^2 \ln\left(\frac{2t^3}{\delta}\right)\sigma(\sigma+1)}{m(u+\gamma)\sqrt{u+1}}\Bigg]. \tag{16}$$

From the definition, the following in-equalities hold.

Proposition B.2. *For all times $u \in \{1, \cdots, t\}$,*

$$\sum_{s=1}^{u-1}(u-s+\gamma)\xi_s^{(t)} \leq 2(u+\gamma)\sqrt{u+1}\xi_u^{(t)}, \tag{17}$$

$$\sum_{s=1}^{u-1}(\xi_s^{(t)})^2 \leq 2(u+1)\ln(u+1)(\xi_u^{(t)})^2 \tag{18}$$

Proof. This follows from the following fact.

Proposition B.3. *For all $u \in \mathbb{N}$ and $\gamma \geq 0$, we have*

$$\sum_{s=1}^{u-1}\frac{u-s+\gamma}{\sqrt{u+1}} \leq 2(u+\gamma)\sqrt{u+1}.$$

□

For each time $u \in \{1, \cdots, t\}$, denote by the random variable $\nu_u^{(t)}$ by

$$\nu_u^{(t)} := \begin{cases} \theta_u - \theta^* & \text{if } \|\theta_u - \theta^*\|^2 \leq (\xi_u^{(t)})^2 + \frac{C_t\gamma^2 G^2}{(u+1)} \\ 0 & \text{if otherwise} \end{cases}$$

For every $u \in \{1, \cdots, t\}$, denote by the event $\mathcal{E}_{u;1}^{(t)}$ to be the one in which the following inequality holds for all $u \in \{1, \cdots, t\}$.

$$\mathcal{E}_{u;1}^{(t)} := \Bigg\{4\eta_t^2\sum_{s=1}^{u-1}\frac{(u-s+\gamma)(\|\psi_{u-s+1}^{(v)}\|_2^2 - \mathbb{E}_{Z_{u-s+1}}[\|\psi_{u-s+1}^{(v)}\|_2^2|\mathcal{F}_{u-s+1}])}{t+\gamma} \leq \frac{96\lambda^2\ln\left(\frac{2t^2(t+1)}{\delta}\right)\sigma(\sigma+1)}{m(u+\gamma)\sqrt{u+1}}\Bigg\}. \tag{19}$$

and $\mathcal{E}_{u;2}^{(t)}$ as

$$\mathcal{E}_{u;2}^{(t)} := \Bigg\{-2\eta_u\sum_{s=1}^{u-1}\frac{(u-s+\gamma)\langle\upsilon_{u-s}, \psi_{u-s+1}^{(v)}\rangle}{t+\gamma} \leq \frac{\xi_u^{(t)}\ln\left(\frac{2t^2(t+1)}{\delta}\right)}{10\sqrt{u+1}} + \frac{C_u\gamma^2G^2}{4(u+1)}.\Bigg\} \tag{20}$$

Denote by the event $\mathcal{E}^{(t)}$ as

$$\mathcal{E}^{(t)} := \bigcap_{u=1}^{t}\left(\mathcal{E}_{u;1}^{(t)} \cap \mathcal{E}_{u;2}^{(t)}\right). \tag{21}$$

Lemma B.4. *For all $t \geq 1$,*

$$\mathbb{P}[\mathcal{E}^{(t)}] \geq 1 - \frac{\delta}{t(t+1)}.$$

We now prove by induction hypothesis that

Lemma B.5. *For every t, under the event $\mathcal{E}^{(t)}$, the following holds.*

$$\|\widehat{\theta}_u - \theta^*\|_2^2 \leq \frac{C_t \gamma^2 G^2}{(u+1)^2} + (\xi_u^{(t)})^2, \tag{22}$$

for all $u \in \{1, \cdots, t\}$.

Proof. Proof of Lemma B.1. We will prove this lemma by induction on u by analyzing Equation (14). The base-case of $u = 1$ holds trivially with probability 1 since $C_t > 1$, $\forall t \geq 1$ and $\gamma > 2$.

Now, assume that on the event $\mathcal{E}^{(t)}$, the induction hypothesis in Equation (22) holds for all times $1, \cdots, u-1$. We prove this by expanding Equation (14) and bounding each of the terms.

Term 1

It is easy to verify that

$$\left(\frac{16\eta_u^2\sigma^2}{\lambda} + 4\eta_u^2\sigma^2\right)\sum_{s=1}^{u-1}\frac{u-s+\gamma}{u+\gamma} \leq \left(\frac{16\sigma^2}{\lambda} + 4\sigma^2\right)\frac{u}{2m^2(u+\gamma)^2},$$
$$\leq \frac{\left(\frac{16\sigma^2}{\lambda} + 4\sigma^2\right)}{2m^2(u+1)}.$$

The last inequality follows since $\gamma^2 > 1$.

Term 2

First notice that

$$4\eta_u^2 \sum_{s=1}^{u-1} \frac{(u-s+\gamma)(\|\psi_{u-s+1}^{(v)}\|_2^2 - \mathbb{E}_{Z_{u-s+1}}[\|\psi_{u-s+1}^{(v)}\|_2^2|\mathcal{F}_{u-s+1}])}{t+\gamma} \leq$$
$$\frac{4\eta_u}{u+\gamma}\sum_{s=1}^{u-1}(\|\psi_{u-s+1}^{(v)}\|_2^2 - \mathbb{E}_{Z_{u-s+1}}[\|\psi_{u-s+1}^{(v)}\|_2^2|\mathcal{F}_{u-s+1}])$$

From the definition of event $\mathcal{E}^{(t)}$ in Equation (21), we get that

$$\text{Term 2} \leq \frac{96\lambda^2 \ln\left(\frac{2t^2(t+1)}{\delta}\right)\sigma(\sigma+1)}{m(u+\gamma)\sqrt{u+1}}.$$

Term 3

$$\frac{8\sigma^2\eta_u}{\lambda}\sum_{s=1}^{u-1}\frac{(u-s+\gamma)\|\theta_{u-s}-\theta^*\|}{u+\gamma} \leq \frac{8\sigma^2}{m\lambda(u+\gamma)^2}\sum_{s=1}^{u-1}\left((u-s+\gamma)\xi_{u-s}^{(t)} + \sqrt{C_t}\gamma G\frac{(u-s+\gamma)}{(u-s+1)}\right),$$

$$
\begin{aligned}
&\overset{(18)}{\leq} \frac{16\sigma^2\sqrt{(u+1)}\xi_u^{(t)}}{m(u+\gamma)} + \frac{8\sqrt{C_t}\sigma^2\gamma^2 Gu}{m\lambda(u+\gamma)^2}, \\
&\leq \frac{16\sigma^2\sqrt{(u+1)}\xi_u^{(t)}}{m(u+\gamma)} + \frac{8\sqrt{C_t}\sigma^2\gamma^2 G}{m\lambda(u+\gamma)}, \\
&\overset{(a)}{\leq} \frac{\xi_u^{(t)}}{10\sqrt{u+1}} + \frac{C_t\gamma^2 G^2}{4(u+1)}.
\end{aligned}
$$

The last inequality follows since $\gamma \geq \frac{320\sigma^2}{m} + 1 \implies \frac{8\sigma^2(u+1)^{1/2}\log(u+1)}{m(u+\gamma)} \leq \frac{1}{10\sqrt{u+1}}$, for all $u \leq t$ and the fact that $C_t \geq \frac{1024\sigma^4}{G^2m^2\lambda^2}$.

Term 4

The definition of event $\mathcal{E}^{(t)}$ in Equation (21) gives that Term $4 \leq \frac{\xi_u^{(t)}\ln\left(\frac{2t^2(t+1)}{\delta}\right)}{10\sqrt{u+1}} + \frac{C_t\gamma^2G^2}{4(u+1)}$

Now, adding in the bounds together into Equation (14),

$$
\|\widehat{\theta}_u - \theta^*\|_2^2 \leq \frac{\gamma^2G^2}{u+1} + \frac{\left(\frac{16\sigma^2}{\lambda} + 4\sigma^2\right)}{2m^2(u+1)} + \frac{\xi_u^{(t)}}{10\sqrt{u+1}} + \frac{1600\lambda^2\ln\left(\frac{2t^2(t+1)}{\delta}\right)\sigma(\sigma+1)}{m(u+\gamma)\sqrt{u+1}} + \frac{\xi_u^{(t)}\ln\left(\frac{2t^2(t+1)}{\delta}\right)}{10\sqrt{u+1}} + \frac{C_t\gamma^2G^2}{2(u+1)}.
$$

Now using the fact that $\frac{\xi_u^{(t)}\ln\left(\frac{2t^3}{\delta}\right)}{\sqrt{u+1}} \leq (\xi_u^{(t)})^2$, we get that

$$
\|\widehat{\theta}_u - \theta^*\|_2^2 \leq \left(1 + \frac{C_t}{2}\right)\frac{\gamma^2G^2}{u+1} + \frac{\left(\frac{16\sigma^2}{\lambda} + 4\sigma^2\right)}{2m^2(u+1)} + \frac{(\xi_u^{(t)})^2}{5} + \frac{96\lambda^2\ln\left(\frac{2t^2(t+1)}{\delta}\right)\sigma(\sigma+1)}{m(u+\gamma)\sqrt{u+1}}.
$$

Substituting the definition of $\xi_u^{(t)}$ from Equation (16), we get that

$$
\begin{aligned}
\|\widehat{\theta}_u - \theta^*\|_2^2 &\leq \left(1 + \frac{C_t}{2}\right)\left[\frac{\gamma^2G^2}{u+1} + \frac{\left(\frac{16\sigma^2}{\lambda} + 4\sigma^2\right)}{2m^2(u+1)} + \frac{96\lambda^2\ln\left(\frac{2t^2(t+1)}{\delta}\right)\sigma(\sigma+1)}{m(u+\gamma)\sqrt{u+1}}\right], \\
&\leq (\xi_u^{(t)})^2 + \frac{C_t\gamma^2G^2}{u+1}.
\end{aligned}
$$

The last inequality follows since $C_t = \max\left(\frac{1024\sigma^4}{G^2m^2\lambda^2}, \frac{8\lambda\sqrt{\ln\left(\frac{2t^3}{\delta}\right)}}{\gamma^2G}\right) \implies C_t \geq 2.$

□

□

B.2 PROOF OF LEMMA B.4

We first reproduce an useful result.

Lemma B.6 (Freedman's inequality[Victor, 1999]). *Suppose $Y_1, \cdots, Y_T$ is a bounded martingale with respect to a filtration $(\mathcal{F}_t)_{t=0}^T$ with $\mathbb{E}[Y_t|\mathcal{F}_{t-1}] = 0$ and $\mathbb{P}[|Y_t| \leq B] = 1$ for all $t \in \{1, \cdots, T\}$. Denote by $V_s := \sum_{n=1}^s Var(Y_n|\mathcal{F}_{n-1})$ be the sum of conditional variances. Then, for every $a, v > 0$,*

$$
\mathbb{P}\left(\exists n \in [1, T] \text{ such that } \sum_{t=1}^n Y_t \geq a \text{ and } V_n \leq v\right) \leq \exp\left(\frac{-a^2}{2(v + Ba)}\right). \tag{23}
$$

Re-arranging the above inequality, we see that if

$$a \geq B \ln\left(\frac{2T}{\delta}\right) + \sqrt{\left(B \ln\left(\frac{2T}{\delta}\right)\right)^2 + 2v \ln\left(\frac{2T}{\delta}\right)}, \tag{24}$$

then the RHS of Equation (23) is bounded above by $\frac{\delta}{2}$.

Proof of Lemma B.4. **Proof of Equation (19)**

Fix a $u \in \{1, \cdots, t\}$. For $s \in \{1, \cdots, u-1\}$, denote by the random variable $Y_s^{(u)} := \frac{(u-s+\gamma)}{u+\gamma}(\|\psi_{u-s+1}^{(v)}\|_2^2 - \mathbb{E}_{Z_{u-s+1}}[\|\psi_{u-s+1}^{(v)}\|_2^2|\mathcal{F}_{u-s}])$. Thus,

$$4\eta_u^2 \sum_{s=1}^{u-1} \frac{(u-s+\gamma)(\|\psi_{u-s+1}^{(v)}\|_2^2 - \mathbb{E}_{Z_{u-s+1}}[\|\psi_{u-s+1}^{(v)}\|_2^2|\mathcal{F}_{u-s+1}])}{u+\gamma} \leq 4\eta_u^2 \sum_{s=1}^{u-1} Y_s^{(u)}.$$

Observe that the sequence $(Y_s^{(u)})_{s=1}^{u-1}$ is a martingale difference sequence with respect to the filtration $(\mathcal{G}_s)_{s=1}^{t-1}$, where $\mathcal{G}_s := \mathcal{F}_{u-s}$. Furthermore, observe that with probability 1, $|Y_s^{(u)}| \leq 4\lambda^2 \mathbf{1}_{\sigma>0} + 4\lambda^2 \mathbf{1}_{\sigma>0} \leq 8\lambda^2 \mathbf{1}_{\sigma>0}$. We can bound the conditional variance as

$$\begin{aligned}
\sum_{s=1}^{u-1} \mathrm{Var}(Y_s^{(u)}|\mathcal{G}_s) &\leq \sum_{s=1}^{u-1} \left(\frac{(u-s+\gamma)}{u+\gamma}\right)^2 \mathbb{E}_{Z_{u-s}}[(\|\psi_{u-s+1}^{(v)}\|_2^2 - \mathbb{E}_{Z_{u-s+1}}[\|\psi_{u-s+1}^{(v)}\|_2^2|\mathcal{F}_{u-s}])^2|\mathcal{F}_{u-s}], \\
&\overset{10}{\leq} 8\lambda^2 \sum_{s=1}^{u-1} \mathbb{E}_{Z_{u-s}}[||\|\psi_{u-s+1}^{(v)}\|_2^2 - \mathbb{E}_{Z_{u-s+1}}[\|\psi_{u-s+1}^{(v)}\|_2^2|\mathcal{F}_{u-s}]||\mathcal{F}_{u-s}], \\
&\leq 8\lambda^2 \sum_{s=1}^{u-1} 2\mathbb{E}_{Z_{u-s}}[||\psi_{u-s+1}^{(v)}\|_2^2|\mathcal{F}_{u-s}], \\
&\overset{12}{\leq} 160\lambda^2\sigma^2(u-1).
\end{aligned}$$

Now, putting $B := 8\lambda^2$ and $v = 160\lambda^2\sigma^2 u$, we get from Equation (24) that with probability at-least $1 - \delta/(2t^2(t+1))$,

$$\begin{aligned}
\sum_{s=1}^{u-1} Y_s^{(u)} &\leq 8\lambda^2 \ln\left(\frac{2t^2(t+1)}{\delta}\right) \mathbf{1}_{\sigma>0} + \sqrt{\left(8\lambda^2 \ln\left(\frac{2t^2(t+1)}{\delta}\right) \mathbf{1}_{\sigma>0}\right)^2 + 160\lambda^2\sigma^2 u \ln\left(\frac{2t^2(t+1)}{\delta}\right)}, \\
&\overset{(a)}{\leq} 32\lambda^2 \ln\left(\frac{2t^2(t+1)}{\delta}\right) \sigma(\sigma+1)\sqrt{u+1}.
\end{aligned}$$

Step (a) follows from the fact that $\lambda \geq 1$. Thus, we have with probability at-least $1 - \frac{\delta}{2t^2(t+1)}$,

$$\begin{aligned}
4\eta_u^2 \sum_{s=1}^{u-1} \frac{(u-s+\gamma)(\|\psi_{u-s+1}^{(v)}\|_2^2 - \mathbb{E}_{Z_{u-s+1}}[\|\psi_{u-s+1}^{(v)}\|_2^2|\mathcal{F}_{u-s+1}])}{u+\gamma} &\leq 96\eta_u^2\lambda^2 \ln\left(\frac{2t^2(t+1)}{\delta}\right) \sigma(\sigma+1)\sqrt{u+1}, \\
&\leq \frac{96\lambda^2 \ln\left(\frac{2t^2(t+1)}{\delta}\right) \sigma(\sigma+1)\sqrt{u+1}}{m^2(u+\gamma)^2}, \\
&\leq \frac{96\lambda^2 \ln\left(\frac{2t^2(t+1)}{\delta}\right) \sigma(\sigma+1)}{m^2(u+\gamma)\sqrt{u+1}}.
\end{aligned}$$

Now taking an union bound over all $u \in \{1, \cdots, t\}$ yields that with probability at-least $1 - \frac{\delta}{2t(t+1)}$, for all time $u \in \{1, \cdots, t\}$,

$$4\eta_u^2 \sum_{s=1}^{u-1} \frac{(t-s+\gamma)(\|\psi_{u-s+1}^{(v)}\|_2^2 - \mathbb{E}_{Z_{u-s+1}}[\|\psi_{u-s+1}^{(v)}\|_2^2|\mathcal{F}_{u-s+1}])}{t+\gamma} \leq \frac{96\lambda^2 \ln\left(\frac{2t^2(t+1)}{\delta}\right) \sigma(\sigma+1)}{m(u+\gamma)\sqrt{u+1}}$$

Proof of Equation (20)

$$-2\eta_u \sum_{s=1}^{u-1} \frac{(u-s+\gamma)\langle \upsilon_{u-s}, \psi_{u-s+1}^{(v)}\rangle}{u+\gamma} \leq \frac{2}{m(u+\gamma)^2} \sum_{s=1}^{u-1} \langle \theta_{u-s} - \theta^*, \psi_{u-s+1}^{(v)}\rangle$$

Fix a $u \in \{1, \cdots, t\}$ and denote by $Y_s^{(u)} := (u-s+\gamma)\langle \theta_{u-s} - \theta^*, \psi_{u-s+1}^{(v)}\rangle$. Since θ_{u-s} is measurable with respect to the sigma-algebra generated by $\mathcal{F}_{u-s}$, the conditional expectation $\mathbb{E}[Y_s^{(u)}|\mathcal{F}_{u-s}] = 0$. Thus, $(Y_s^{(u)})_{s=1}^{u-1}$ is a martingale difference sequence with respect to the filtration $(\mathcal{F}_{u-s})_{s=1}^{u-1}$. Furthermore, we have from Equation (10) that $|Y_s^{(u)}| \leq 2(u-s+\gamma)\left(\xi_{u-s}^{(t)} + \frac{\gamma R_1}{(u+\gamma-1)}\right)\lambda \leq 2\lambda(u+\gamma)\xi_t^{(t)} + 2\lambda\gamma G$. We can now bound the sum of conditional variances as

$$\begin{aligned}\sum_{s=1}^{u-1} \text{Var}(Y_s^{(u)}|\mathcal{F}_{u-s}) &\leq \sum_{s=1}^{u-1} 4(u-s+\gamma)^2(\xi_{u-s}^{(t)})^2\lambda^2\sigma^2 + 4\lambda^2 G^2, \\ &\overset{(18)}{\leq} 12\lambda^2\sigma^2(u+\gamma)^2(u+1)\log(u+1)(\xi_u^{(t)})^2 + 4\lambda^2\gamma^2 G^2 u.\end{aligned}$$

Step (a) follows since $\eta m < 1$. Now applying the bound in Equation (24) with $B := 2\lambda(u+\gamma)\xi_u^{(t)} + 2\lambda G$ and $v = 12\lambda^2\sigma^2(u+\gamma)^2(u+1)\log(u+1)(\xi_u^{(t)})^2 + 4\lambda^2\gamma^2G^2u$, we get that with probability at-least $1 - \delta/(2t^2(t+1))$,

$$\begin{aligned}\sum_{s=1}^{u-1}(u-s+\gamma)\langle \upsilon_{u-s}, \psi_{u-s+1}^{(v)}\rangle &\leq 2\lambda\left((u+\gamma)\xi_u^{(t)} + R_1\right)\ln\left(\frac{2t^2(t+1)}{\delta}\right) + \Bigg[\left(2\lambda\left((u+\gamma)\xi_u^{(t)} + G\right)\ln\left(\frac{2t^2(t+1)}{\delta}\right)\right)^2 \\ &\quad + \left(\lambda^2\sigma^2(u+\gamma)^2(u+1)\log(u+1)(\xi_u^{(t)})^2 + 4\lambda^2\gamma^2G^2(u+1)\right)\ln\left(\frac{2t^2(t+1)}{\delta}\right)\Bigg]^{\frac{1}{2}}, \\ &\leq 6(u+\gamma)\sqrt{u+1}\log(u+1)(\xi_u^{(t)})\lambda\sigma(\sigma+1)\ln\left(\frac{2t^2(t+1)}{\delta}\right) + 2\lambda\gamma G\sqrt{(u+1)\ln\left(\frac{2t^2(t+1)}{\delta}\right)}.\end{aligned}$$

Thus,

$$\begin{aligned}-2\eta_u \sum_{s=1}^{u-1} \frac{(u-s+\gamma)\langle \upsilon_{u-s}, \psi_{u-s+1}^{(v)}\rangle}{u+\gamma} &\leq \frac{12\sqrt{u+1}\log(u+1)(\xi_u^{(t)})\lambda\sigma(\sigma+1)\ln\left(\frac{2t^2(t+1)}{\delta}\right)}{(u+\gamma)} + \frac{C_t\gamma G}{10(u+1)}, \\ &\leq \frac{\xi_u^{(t)}\ln\left(\frac{2t^2(t+1)}{\delta}\right)}{10\sqrt{u+1}} + \frac{C_t G}{10(u+1)}.\end{aligned}$$

The first inequality follows since $C_t \geq \frac{8\lambda\sqrt{\ln\left(\frac{2t^3}{\delta}\right)}}{\gamma^2 G}$. The last inequality follows since for all times $u \leq t$, we have

$$\frac{12\sqrt{u+1}\log(u+1)\lambda\sigma(\sigma+1)\ln\left(\frac{2t^2(t+1)}{\delta}\right)}{(u+\gamma)} \leq \frac{\ln\left(\frac{2t^2(t+1)}{\delta}\right)}{10}$$

as a consequence of $\gamma \geq 120\lambda\sigma(\sigma+1)$.

□

C PROOFS FROM SECTION 4.2

C.1 PROOF OF THEOREM 4.1

We bound this probability using the result of 3.1 and a simple union bound argument. For any process $\mathfrak{M}$, observe that

$$\mathbb{P}[\exists t \in [r+1, \tau_c^{(r)}) \text{ s.t.} \mathcal{A}_t = 1 | \mathcal{A}_r = 1] = \mathbb{P}[\cup_{t=r+1}^{\tau_c-1} \mathcal{A}_t = 1 | \mathcal{A}_r = 1]$$

$$\leq \sum_{t=r+1}^{\tau_c - 1} \mathbb{P}[\mathcal{A}_t = 1 | \mathcal{A}_r = 1]. \tag{25}$$

We now examine the above Equation to bound it. For any fixed $t \in (r, \tau_c^{(r)})$

$$\begin{aligned}
\mathbb{P}[\mathcal{A}_t = 1 | \mathcal{A}_r = 1] &= \mathbb{P}\left[\bigcup_{s=r+1}^{t-1} \|\widehat{\theta}_{r:s} - \widehat{\theta}_{s+1:t}\| \geq \mathcal{B}\left(s-r, \frac{\delta}{2t(t+1)}\right) + \mathcal{B}\left(t-s-1, \frac{\delta}{2t(t+1)}\right)\right], \\
&\leq \sum_{s=r+1}^{t-1}\left(\mathbb{P}\left[\|\widehat{\theta}_{r:s} - \theta_{c-1}\| \geq \mathcal{B}\left(s-r, \frac{\delta}{2t(t+1)}\right)\right] + \mathbb{P}\left[\|\widehat{\theta}_{s+1:t} - \theta_{c-1}\| \geq \mathcal{B}\left(t-s-1, \frac{\delta}{2t(t+1)}\right)\right]\right), \\
&\overset{(a)}{\leq} \sum_{s=r+1}^{t-1}\left(\frac{\delta}{2t(t+1)(s-r)(s-r+1)} + \frac{\delta}{2t(t+1)(t-s-1)(t-s)}\right), \\
&= \frac{\delta}{2t(t+1)}\left(\sum_{s=r+1}^{t-1}\frac{1}{(s-r)(s-r+1)} + \sum_{s=r+1}^{t-1}\frac{1}{(t-s-1)(t-s)}\right), \\
&\leq \frac{\delta}{2t(t+1)}\left(\sum_{s=1}^{t-1-r}\frac{1}{s(s+1)} + \sum_{s=1}^{t-1-r}\frac{1}{s(s+1)}\right), \\
&\overset{(b)}{\leq} \frac{\delta}{t(t+1)}.
\end{aligned} \tag{26}$$

Since for all $t < \tau_c^{(r)}$, the mean of the random variables $X_{r+1}, \cdots, X_t$ are identical and equal to θ_{c-1} (see notation in Section 2), Theorem 3.1 gives rise to inequality (a). Step (b) follows from the fact that $\sum_{s \geq 1} \frac{1}{s(s+1)} = 1$. Now substituting the bound from Equation (26) into Equation (25), we get that

$$\begin{aligned}
\mathbb{P}[\exists t \in [r+1, \tau_c^{(r)}) \text{ s.t. } \mathcal{A}_t = 1 | \mathcal{A}_r = 1] &\leq \sum_{t=r+1}^{\tau_c - 1} \frac{\delta}{t(t+1)}, \\
&\leq \sum_{t \geq 1} \frac{\delta}{t(t+1)}, \\
&= \delta.
\end{aligned}$$

Since the above bound holds for all r and process $\mathfrak{M}$, we have

$$\sup_{\mathfrak{M}, r} \mathbb{P}[\exists t \in [r+1, \tau_c^{(r)}) \text{ s.t.} \mathcal{A}_t = 1 | \mathcal{A}_r = 1] \leq \delta.$$

C.2 PROOF OF LEMMA 4.2

Recall from the definition that the rth detection is false if

$$\chi_r^{(A)} = \mathbf{1}(\nexists c \text{ s.t. } \tau_c \in (t_{r-1}^{(A)}, t_r^{(A)}]).$$

We will show that $\mathbb{E}[\chi_r^{(A)}] \leq \delta$. This will then conclude the proof of the lemma.

$$\begin{aligned}
\mathbb{E}[\chi_r^{(A)}] &= \mathbb{P}[\nexists c \text{ s.t. } \tau_c \in (t_{r-1}^{(A)}, t_r^{(A)}]], \\
&= \mathbb{E}\left[\mathbb{P}[\nexists c \text{ s.t. } \tau_c^{(s)} \in (s, t_r^{(A)}]] \middle| t_{r-1}^{(A)} = s\right], \\
&\leq \mathbb{E}\left[\mathbb{P}[\cup_{t=s+1}^{\infty} \tau_c^{(s)} = t, t_r^{(A)} < t] \middle| t_{r-1}^{(A)} = s\right],
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathbb{E}\left[\mathbb{P}[\exists t \in [s+1, \tau_c^{(s)}), \mathcal{A}_t = 1] \middle| t_{r-1}^{(A)} = s\right], \\
&\overset{(a)}{\leq} \mathbb{E}\left[\mathbb{P}[\exists t \in [s+1, \tau_c^{(s)}), \mathcal{A}_t = 1 | \mathcal{A}_s = 1] \middle| t_{r-1}^{(A)} = s\right], \\
&\overset{(b)}{\leq} \delta.
\end{aligned}
$$

Inequality (a) follows from the fact that on the event $t_{r-1}^{(\mathcal{A})} = s$, $\mathcal{A}_s = 1$. Inequality (b) follows from Theorem 4.1.

D PROOF OF LEMMA 4.3

The proof follows from a straightforward application of Theorem 3.1 as follows. Let $n \in \mathbb{N}, \Delta > 0$ and $\delta' \in (0,1)$ be arbitrary.

$$
\begin{aligned}
\mathbb{P}[\mathcal{D}(n, \Delta, \delta') \geq d] &= \mathbb{P}[\cap_{s=1}^{n+d} \mathcal{A}(X_{1:s}) = 0], \\
&= \mathbb{P}\left[\bigcap_{s=1}^{n+d} \|\widehat{\theta}_{1:s} - \widehat{\theta}_{s+1:n+d}\|_2^2 \leq \mathcal{B}\left(s, \frac{\delta}{2(n+d)(n+d+1)}\right) + \mathcal{B}\left(n+d-s-1, \frac{\delta}{2(n+d)(n+d+1)}\right)\right], \\
&\leq \mathbb{P}\left[\|\widehat{\theta}_{1:n-1} - \widehat{\theta}_{n:n+d}\|_2^2 \leq \mathcal{B}\left(n-1, \frac{\delta}{2(n+d)(n+d+1)}\right) + \mathcal{B}\left(d, \frac{\delta}{2(n+d)(n+d+1)}\right)\right].
\end{aligned} \tag{27}
$$

From triangle-inequality, we know that

$$
\begin{aligned}
\|\widehat{\theta}_{1:n-1} - \widehat{\theta}_{n:n+d}\|_2^2 &\geq \|\theta_1 - \theta_2\|_2^2 - \|\widehat{\theta}_{1:n-1} - \theta_1\|_2^2 - \|\widehat{\theta}_{n:n+d} - \theta_2\|_2^2, \\
&= \Delta^2 - \|\widehat{\theta}_{1:n-1} - \theta_1\|_2^2 - \|\widehat{\theta}_{n:n+d} - \theta_2\|_2^2.
\end{aligned} \tag{28}
$$

Thus, substituting Equation (28 into Equation (27), we get that

$$
\mathbb{P}[\mathcal{D}(n, \Delta, \delta') \geq d] \leq \mathbb{P}\Big[\Delta^2 - \|\widehat{\theta}_{1:n-1} - \theta_1\|_2^2 - \|\widehat{\theta}_{n:n+d} - \theta_2\|_2^2 \leq \mathcal{B}\left(n-1, \frac{\delta}{2(n+d)(n+d+1)}\right) + \mathcal{B}\left(d, \frac{\delta}{2(n+d)(n+d+1)}\right)\Big].
$$

Denote by the events $\mathcal{E}_i$ for $i \in \{1, 2\}$ as

$$
\begin{aligned}
\mathcal{E}_1 &:= \left\{\|\widehat{\theta}_{1:n-1} - \theta_1\|_2^2 > \mathcal{B}\left(n-1, \frac{\delta'}{2}\right)\right\}, \\
\mathcal{E}_2 &:= \left\{\|\widehat{\theta}_{n:n+d} - \theta_2\|_2^2 > \mathcal{B}\left(d, \frac{\delta'}{2}\right)\right\},
\end{aligned}
$$

Denote by $\mathcal{E} := \mathcal{E}_1 \cup \mathcal{E}_2$. Theorem 3.1 gives that $\mathbb{P}[\mathcal{E}_1] \leq \frac{\delta'}{2(n(n+1))} \leq \frac{\delta'}{2}$ and $\mathbb{P}[\mathcal{E}_2] \leq \frac{\delta'}{2d(d+1)} \leq \frac{\delta'}{2}$. Thus, an union bound gives that $\mathbb{P}[\mathcal{E}] \leq \delta'$. Let $d' \in \mathcal{G}$ be arbitrary, where

$$
\mathcal{G} := \left\{d \in \mathbb{N} : \Delta^2 \geq \mathcal{B}\left(n-1, \frac{\delta'}{2}\right) + \mathcal{B}\left(d, \frac{\delta'}{2}\right) + \mathcal{B}\left(n, \frac{\delta}{2(n+d+1)(n+d)}\right) + \mathcal{B}\left(d, \frac{\delta}{2(n+d+1)(n+d)}\right)\right\} \tag{29}
$$

Claim : If the event $\mathcal{E}^c$ holds, then $\mathcal{D}(n, \Delta, \delta) \leq d$ for all $d \in \mathcal{G}$.

Suppose $d \in \mathcal{G}$ and event $\mathcal{E}^c$ holds. Then, we know by triangle inequality in Equation (28) that

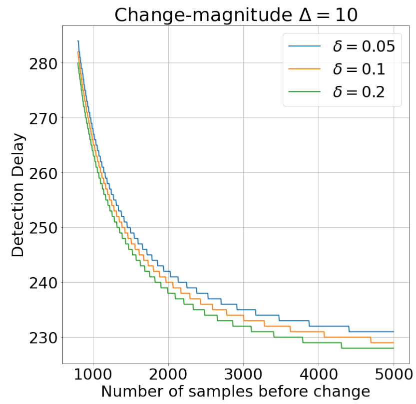


Figure 1: Plot of $\mathcal{D}(n, \Delta, \delta')$ in Lemma 4.3 for fixed $\Delta = 10, \delta = 0.1$.

$$\begin{aligned}
\|\widehat{\theta}_{1:n-1} - \widehat{\theta}_{n:n+d}\|_2^2 &\geq \|\theta_1 - \theta_2\|_2^2 - \|\widehat{\theta}_{1:n-1} - \theta_1\|_2^2 - \|\widehat{\theta}_{n:n+d} - \theta_2\|_2^2, \\
&= \Delta^2 - \|\widehat{\theta}_{1:n-1} - \theta_1\|_2^2 - \|\widehat{\theta}_{n:n+d} - \theta_2\|_2^2, && (30) \\
&\overset{(a)}{\geq} \Delta^2 - \mathcal{B}\left(n-1, \frac{\delta'}{2}\right) - \mathcal{B}\left(d, \frac{\delta'}{2}\right), && (31) \\
&\overset{(b)}{\geq} \mathcal{B}\left(n, \frac{\delta}{2(n+d+1)(n+d)}\right) + \mathcal{B}\left(d, \frac{\delta}{2(n+d+1)(n+d)}\right). && (32)
\end{aligned}$$

Step (a) follows from the definition of event $\mathcal{E}$ and on the assumption of the claim that event $\mathcal{E}^c$ holds. Step (b) follows from the fact that $d \in \mathcal{G}$ is arbitrary (cf. Equation (29). The last step says from Line 8 of Algorithm 1 that if no detection has been made till time $n + d$, then under the event $\mathcal{E}^c$, time step d is a detection time. Since event $\mathcal{E}^c$ holds with probability at-least $1 - \delta'$, this concludes the proof.

D.1 USEFUL CONVEXITY BASED INEQUALITIES

Let $f : \Theta \to \mathbb{R}$ be a strongly convex function with strong convexity parameters $0 < m \leq M < \infty$. Denote by $\theta^* := \arg\min_{\theta \in \Theta} f(\theta)$. Since $f(\cdot)$ is convex and Θ is convex and compact, the existence and uniqueness of θ^* is guaranteed. Strong convexity gives that for any $\widehat{\theta}_{t-1} \in \Theta$,

$$f(\theta^*) \geq f(\widehat{\theta}_{t-1}) + \langle \nabla f(\widehat{\theta}_{t-1}), \theta^* - \widehat{\theta}_{t-1} \rangle + \frac{m}{2}\|\theta^* - \widehat{\theta}_{t-1}\|_2^2. \quad (33)$$

Further since $\theta^* = \arg\min_{\theta \in \Theta} f(\theta)$., we have that

$$f(\widehat{\theta}_{t-1}) - f(\theta_t^*) \geq \frac{m}{2}\|\widehat{\theta}_{t-1} - \theta^*\|_2^2.$$

Putting these two together, we see that

$$\langle \nabla f(\widehat{\theta}_{t-1}), \widehat{\theta}_{t-1} - \theta^* \rangle \geq m\|\widehat{\theta}_{t-1} - \theta^*\|_2^2. \quad (34)$$

Also, We further use the following lemma.

Lemma D.1 (Lemma 3.11 from [Bubeck, 2015])**.** *Let $g : \mathbb{R}^d \to \mathbb{R}$ be a M smooth and m strongly convex function. Then for all $x, y \in \mathbb{R}^d$,*

$$\langle \nabla g(x) - \nabla g(y), x - y \rangle \geq \frac{mM}{M+m}\|x - y\|_2^2 + \frac{1}{M+m}\|\nabla g(x) - \nabla g(y)\|_2^2.$$

By substituting $x = \widehat{\theta}_{t-1}$, $y = \theta_t^*$ and $g(\cdot) = f(\cdot)$ and by leveraging the fact that $\nabla f(\theta^*) = 0$, we get the inequality that

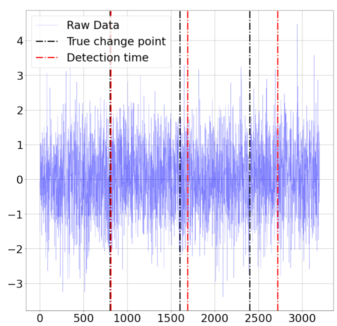
$$\langle \nabla f(\widehat{\theta}_{t-1}), \widehat{\theta}_{t-1} - \theta^* \rangle \geq \frac{mM}{m+M}\|\widehat{\theta}_{t-1} - \theta^*\|_2^2 + \frac{1}{M+m}\|\nabla f(\widehat{\theta}_{t-1})\|_2^2.$$

Re-arranging, we see that

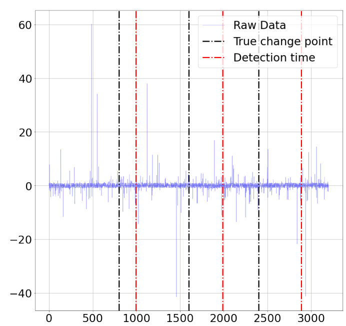
$$\|\nabla f(\widehat{\theta}_{t-1})\|_2^2 \leq (M+m)\langle \nabla f(\widehat{\theta}_{t-1}), \widehat{\theta}_{t-1} - \theta^* \rangle - mM\|\widehat{\theta}_{t-1} - \theta^*\|_2^2. \tag{35}$$

E ADDITIONAL SIMULATIONS

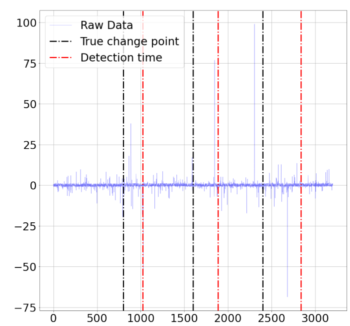
In Figure 2, we plot a sample path of observed data and mark out the true change-points and the detected time-instants by Algorithm 1. The plots indicate that although visually identifying the change in the means is hard, our change-point detection algorithm is able to consistently across variety of distribution families.



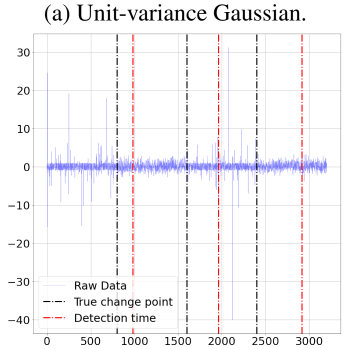
(a) Unit-variance Gaussian.



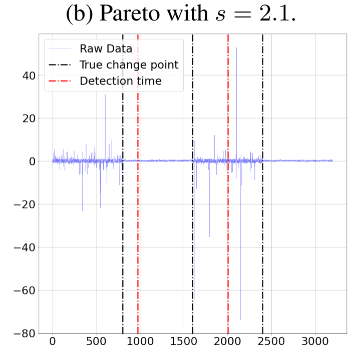
(b) Pareto with $s = 2.1$.



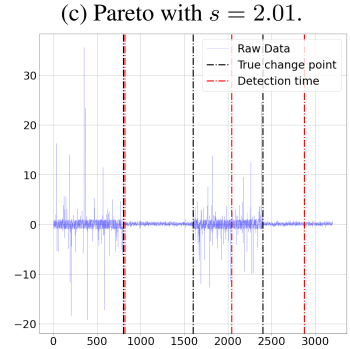
(c) Pareto with $s = 2.01$.



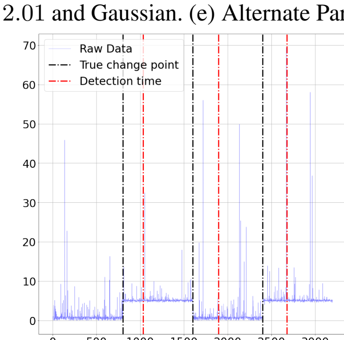
(d) Alternate Pareto $s = 2.01$ and Gaussian.



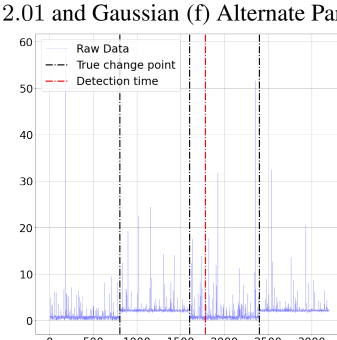
(e) Alternate Pareto $s = 2.01$ and Gaussian



(f) Alternate Pareto $s = 2.01$ and Gaussian



(g) Pareto $s = 2.01, d = 15, \Delta = 5$



(h) Pareto $s = 2.01, d = 15, \Delta = 2$

Figure 2: In all plots, we choose the change-point gap to be $\Delta = 0.1$ and $\delta = 0.05$ except (g) and (h) where $\Delta = 5$ and 2 respectively. In plots (g) and (h), we plot the norm of the observed random vector and thus the Y-axis is non-negative. We see missed detection in Figures (e) and (h) with the last change-point on the right being missed. We do not observe False-positives in these plots.