

## Lecture 3: Concentration Bounds

Lecturer: Nikolai Matni

Scribes: Zongyu Dai

**Disclaimer:** These notes have not been subjected to the usual scrutiny reserved for formal publications.

The previous lecture focused on identifying the state-space parameters  $(A, B, C, D)$  of an unknown linear-time-invariant (LTI) system using *subspace methods*. Although these methods have been shown to be asymptotically consistent (i.e., up to a similarity transformation, the estimated parameters converge to the true parameters as the number of data points tends to infinity), in this class, we are interested in finite data guarantees. In particular, we wish to quantify the uncertainty in our estimates as a function of the number of data-points, so that it can then be used as part of a robust control synthesis procedure. In this lecture, we introduce technical tools known as *concentration bounds*, and show how they can be used to derive exactly these kinds of bounds for a simplified scalar dynamical system.

## 1 Concentration bounds

### 1.1 Bounded Random Variables

To build some intuition we begin by studying the behavior of almost surely bounded random variables. We assume  $\{x_i\}_{i=1}^N \stackrel{i.i.d.}{\sim} \mathbb{P}^N$ , and let  $x_i \in [a, b]$  with probability 1.

**Theorem 1.1** (McDiarmid's Inequality [1]). *let  $x_i \in \mathcal{X}$  for  $i = 1, 2, \dots, N$  they are independent but not necessarily identical. Then let  $F : \mathcal{X}^n \rightarrow \mathbb{R}$  satisfies for all  $x_1, x_2, \dots, x_N, x_i \in \mathcal{X}$*

$$\sup_{x_1 \dots x_N, x'_i} |F(x_1, \dots, x_N) - F(x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_N)| \leq c_i \quad (1)$$

then we have that

$$\mathbb{P}[F(x_1, \dots, x_N) - \mathbb{E}[F(x_1, \dots, x_N)] \geq t] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right) \quad (2)$$

**Corollary 1** (Hoeffding's inequality for bounded random variables). *Let  $\{X_i\}_{i=1}^N \stackrel{i.i.d.}{\sim} p^N$  be such that  $X_i \in [a, b]$  a.s.. Then*

$$\mathbb{P}\left[\frac{1}{N} \sum_{i=1}^N X_i - \mathbb{E}X_1 \geq t\right] \leq \exp\left(\frac{-2Nt^2}{(b-a)^2}\right). \quad (3)$$

*Proof:* Set  $F(x_1, \dots, x_N) = \frac{1}{N} \sum_{i=1}^N x_i$  and notice that it satisfies the boundedness condition with  $c_i \equiv (b-a)/N$  for all  $i$ .

**Example 1.** *Let  $\{X_i\}_{i=1}^N \stackrel{i.i.d.}{\sim} p^N$  be random vectors in  $\mathcal{X}$ , and let  $\Omega \subseteq \mathcal{X}$  be some set. Let  $\hat{P}_N = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{x \in \Omega}$ , and notice that  $\mathbb{E}\hat{P}_N = \mathbb{P}[x \in \Omega]$ . As  $\mathbb{1}_{x \in \Omega} \in \{0, 1\}$  for all  $x$ , it follows by equation (3) that*

$$\mathbb{P}\left[\hat{P}_N - \mathbb{P}[x \in \Omega] \geq t\right] \leq \exp(-2Nt^2). \quad (4)$$

*We can obtain a similar bound on the probability of the event  $\{\hat{P}_N - \mathbb{P}[x \in \Omega] \leq -t\}$  occurring: it then follows by union bounding over these two events that*

$$\mathbb{P}\left[\left|\hat{P}_N - \mathbb{P}[x \in \Omega]\right| \geq t\right] \leq 2\exp(-2Nt^2). \quad (5)$$

Thus we have seen that in the case of a.s. bounded random variables, concentration of measure does indeed occur. We will now see that similar concentration occurs for random variables drawn from distributions with sufficiently rapidly decaying tails.

## 1.2 Sub-Gaussian Random Variables

We begin by recalling the Chernoff bound, which states that for a random variable  $X$  with mean  $\mathbb{E}X$ , and moment generating function (MGF)  $\mathbb{E}[e^{\lambda(X-\mathbb{E}X)}]$  defined for all  $|\lambda| \leq b$ , it holds that

$$\log \mathbb{P}[X - \mathbb{E}X \geq t] \leq \inf_{\lambda \in [0, b]} \left[ \log \mathbb{E} \left[ e^{\lambda(X-\mathbb{E}X)} \right] - \lambda t \right]. \quad (6)$$

We now turn our attention to Gaussian random variables, and recall that for  $X \sim \mathcal{N}(\mu, \sigma^2)$ , we have that  $\mathbb{E}[e^{\lambda(X-\mu)}] = \exp\left(\frac{\sigma^2 \lambda^2}{2}\right)$  for all  $\lambda \in \mathbb{R}$ . Substituting this into the Chernoff bound (6) and solving for  $\lambda^* = t/\sigma^2$ , we immediately obtain

$$\mathbb{P}[X - \mu \geq t] \leq \exp\left(\frac{-t^2}{2\sigma^2}\right). \quad (7)$$

Recalling that if  $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$  and  $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$  then  $aX_1 + bX_2 \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$ , it follows immediately that for  $X_i \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$ , it holds that

$$\mathbb{P}\left[\frac{1}{N} \sum_{i=1}^N X_i - \mu \geq t\right] \leq \exp\left(\frac{-Nt^2}{2\sigma^2}\right). \quad (8)$$

Once again, a similar bound can be obtained on the probability of event  $\left\{\frac{1}{N} \sum_{i=1}^N X_i - \mu \leq -t\right\}$  occurring: it then follows by union bounding over these two events that

$$\mathbb{P}\left[\left|\frac{1}{N} \sum_{i=1}^N X_i - \mu\right| \geq t\right] \leq 2\exp\left(\frac{-Nt^2}{2\sigma^2}\right) \quad (9)$$

We now generalize these results to random variables with MGFs dominated by that of a Gaussian random variable.

**Definition 1** (Sub-Gaussian Random Variable). *A random variable  $X$  with mean  $\mathbb{E}X$  is sub-Gaussian if there exists a positive number  $\sigma^2$  such that*

$$\mathbb{E}\left[e^{\lambda(X-\mathbb{E}X)}\right] \leq \exp\left(\frac{\lambda^2 \sigma^2}{2}\right) \quad \forall \lambda \in \mathbb{R}. \quad (10)$$

An example of random variables that are sub-Gaussian but not Gaussian are bounded random variables – it can be shown that a random variable  $X$  taking values in  $[a, b]$  almost surely satisfies equation (10) with parameter  $\sigma^2 = (b-a)^2/4$ .

Further, from this definition, it follows immediately that from the Chernoff bound that all sub-Gaussian random variables satisfy the concentration bound (7). One can also check that if  $X_1$  and  $X_2$  are sub-Gaussian with parameters  $\sigma_1^2$  and  $\sigma_2^2$ , then  $X_1 + X_2$  is sub-Gaussian with parameter  $\sigma_1^2 + \sigma_2^2$ , from which we immediately obtain Hoeffding's Inequality.

**Theorem 1.2** (Hoeffding's Inequality). *Let  $\{X_i\}_{i=1}^N$  be iid sub-Gaussian random variables with parameter  $\sigma^2$ . Then*

$$\mathbb{P}\left[\frac{1}{N} \sum_{i=1}^N X_i - \mathbb{E}X_1 \geq t\right] \leq \exp\left(\frac{-Nt^2}{2\sigma^2}\right). \quad (11)$$

**An aside on probability inversion and two sided bounds** Rather than statements about the probability of large deviations, as in bound (11), we are often interested in the probability that a random variable concentrates near its mean. To do so, we employ probability inversion: if we are willing to tolerate a large deviation occurring with probability at most  $\delta$ , one may invert bound (11) by setting  $\delta = \text{RHS of (11)}$  and solving for  $t$ . This allows us to certify that with probability at least  $1 - \delta$  that

$$\frac{1}{N} \sum_{i=1}^N X_i \leq \mathbb{E}X_1 + \sqrt{\frac{2\sigma^2 \log(1/\delta)}{N}}. \quad (12)$$

Applying the same reasoning to the event  $\{X - \mathbb{E}X \leq -t\}$  yields a similar bound, from which it follows, by the union bound, that with probability at least  $1 - 2\delta$  that

$$\left| \frac{1}{N} \sum_{i=1}^N X_i - \mathbb{E}X_1 \right| \leq \sqrt{\frac{2\sigma^2 \log(1/\delta)}{N}}. \quad (13)$$

### 1.3 Sub-Exponential Random Variables

Revisiting the error term defined in (22), we see that we still do not have the requisite tools to perform the desired analysis.

**Example 2** (Products of Gaussians are not Sub-Gaussian). *Motivated by the error term in (22), we compute the MGFs for  $X^2$  and  $XW$ , where  $X, W \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ . Direct computation of the resulting integrals show that*

$$\begin{aligned} \mathbb{E} \left[ e^{\lambda(X^2-1)} \right] &= \frac{e^{-\lambda}}{1-2\lambda} \text{ if } \lambda < 1/2, \\ \mathbb{E} \left[ e^{\lambda(XW)} \right] &= \frac{1}{\sqrt{\pi(1-\lambda^2)}} \text{ if } |\lambda| < 1. \end{aligned} \quad (14)$$

*These random variables are clearly not sub-Gaussian, as their MGFs do not exist for all  $\lambda \in \mathbb{R}$ . However, notice that they can be bounded by the MGF of a Gaussian random variable in a neighborhood of the origin. In particular we have that*

$$\begin{aligned} \mathbb{E} \left[ e^{\lambda(X^2-1)} \right] &= \frac{e^{-\lambda}}{1-2\lambda} \leq \exp\left(\frac{4\lambda^2}{2}\right) \quad \forall |\lambda| < 1/4 \\ \mathbb{E} \left[ e^{\lambda(XW)} \right] &= \frac{1}{\sqrt{\pi(1-\lambda^2)}} \leq \exp\left(\frac{2\lambda^2}{2}\right) \quad \forall |\lambda| < 1/\sqrt{2}. \end{aligned}$$

*The first inequality follows from some calculus, and the second by leveraging that  $-\log(1-x) \leq x(1-x)^{-1}$  for  $0 \leq x < 1$ .*

We now show that MGFs exhibiting behavior as above also concentrate.

**Definition 2.** *A random variable  $X$  with mean  $\mathbb{E}X$  is sub-exponential with parameters  $(\nu^2, \alpha)$  if*

$$\mathbb{E} \left[ e^{\lambda(X-\mathbb{E}X)} \right] \leq \exp\left(\frac{\nu^2 \lambda^2}{2}\right) \quad \forall |\lambda| \leq \frac{1}{\alpha}. \quad (15)$$

Example 2 therefore demonstrated that for  $X, W \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ ,  $X^2$  is sub-exponential with parameters  $(4, 4)$ , and  $XW$  is sub-exponential with parameters  $(2, \sqrt{2})$ .

**Proposition 1** (Sub-exponential tail bound). *if  $X$  is sub-exponential with parameters  $(\nu^2, \alpha)$ , then*

$$\mathbb{P}[X - \mathbb{E}X \geq t] \leq \begin{cases} \exp\left(\frac{-t^2}{2\nu^2}\right) & \text{if } 0 \leq t \leq \frac{\nu^2}{\alpha} \\ \exp\left(\frac{-t}{2\alpha}\right) & \text{if } t > \frac{\nu^2}{\alpha}. \end{cases} \quad (16)$$

**Remark:** Thus we see that for sufficiently small deviations  $0 \leq t \leq \nu^2/\alpha$ , sub-exponential random variables exhibit sub-Gaussian concentration – indeed, informally, one may view sub-Gaussian random variables as the limit of a sub-exponential random variable with  $\alpha \rightarrow 0$ . Finally, we note that we can show that for  $X_1$  and  $X_2$  sub-exponential random variables with parameters  $(\nu_i^2, \alpha_i^2)$ , we have that  $X_1 + X_2$  is a sub-exponential random variable with parameters  $(\nu_1^2 + \nu_2^2, \max\{\alpha_1, \alpha_2\})$ . Now we give the proof for this proposition.

**Proof.** From [2]: by recentering as needed, we may assume without loss of generality that  $\mu = 0$ . We follow the usual Chernoff-type approach: combining it with the definition of a sub-exponential variable yields the upper bound

$$\mathbb{P}[X \geq t] \leq e^{-\lambda t} \mathbb{E}[e^{\lambda t}] \leq \exp\left(-\lambda t + \frac{\lambda^2 \nu^2}{2}\right)$$

valid for all  $\lambda \in [0, \alpha^{-1})$ , and we define  $g(\lambda, t) = -\lambda t + \frac{\lambda^2 \nu^2}{2}$ . In order to complete the proof, it remains to compute, for each fixed  $t \geq 0$ , the quantity  $g^*(t) = \inf_{\lambda \in [0, \alpha^{-1})} g(\lambda, t)$ . Note that the unconstrained minimum of the function  $g(\cdot, t)$  occurs at  $\lambda^* = \frac{t}{\nu^2}$ . If  $0 \leq t \leq \frac{\nu^2}{\alpha}$ , then this unconstrained optimum corresponds to the constrained minimum as well, so that  $g^*(t) = -\frac{t^2}{2\nu^2}$  over this interval.

Otherwise, we may assume that  $t \geq \frac{\nu^2}{\alpha}$ . In this case, since the function  $g(\cdot, t)$  is monotonically decreasing in the interval  $[0, \lambda^*)$ , the constrained minimum is achieved at the boundary point  $\lambda' = \alpha^{-1}$ , and we have

$$g^*(t) = g(\lambda', t) = -\frac{t}{\alpha} + \frac{1}{2\alpha} \frac{\nu^2}{\alpha} \leq -\frac{t}{2\alpha}$$

where the inequality uses the fact that  $\frac{\nu^2}{\alpha} \leq t$ . ■

As shown in Example 2, the sub-exponential property can be verified by explicitly computing or bounding the moment generating function. This direct calculation may be impracticable in many settings, so it is natural to seek alternative approaches. One such method is based on control of the polynomial moments of variable  $X$ . Given a random variable  $X$  with mean  $\mu = \mathbb{E}[X]$  and variance  $\sigma^2 = \mathbb{E}[X^2] - \mu^2$ , we say that Bernstein's condition with parameter  $b$  holds if

$$|\mathbb{E}[(X - \mu)^k]| \leq \frac{1}{2} k! \sigma^2 b^{k-2} \quad \text{for } k = 2, 3, 4, \dots \quad (17)$$

One sufficient condition for Bernstein's condition to hold is that  $X$  be bounded; in particular, if  $|X - \mu| \leq b$ , then it is straightforward to verify that condition 17 holds. Even for bounded variables, our next result will show that the Bernstein condition can be used to obtain tail bounds that may be tighter than the Hoeffding bound. Moreover, Bernstein's condition is also satisfied by various unbounded variables, a property which lends it much border applicability.

When  $X$  satisfies the Bernstein's condition, then it is sub-exponential with parameters determined by  $\sigma^2$  and  $b$ . Indeed, by the power-series expansion of the exponential, we have

$$\begin{aligned} \mathbb{E}[e^{\lambda(X-\mu)}] &= 1 + \frac{\lambda^2 \sigma^2}{2} + \sum_{k=3}^{\infty} \lambda^k \frac{\mathbb{E}[(X-\mu)^k]}{k!} \\ &\leq 1 + \frac{\lambda^2 \sigma^2}{2} + \frac{\lambda^2 \sigma^2}{2} \sum_{k=3}^{\infty} (|\lambda| b)^{k-2} \end{aligned} \quad (18)$$

where the inequality make use of Bernstein's condition. For any  $|\lambda| < \frac{1}{b}$ , we can sum the geometric series so as to obtain

$$\mathbb{E}[e^{\lambda(X-\mu)}] \leq 1 + \frac{\lambda^2 \sigma^2}{2} \frac{1}{1 - b|\lambda|} \leq e^{\frac{\lambda^2 \sigma^2}{2(1-b|\lambda|)}} \quad (19)$$

where the second inequality follows from the bound  $1 + t \leq e^t$ . Consequently, we conclude that

$$\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\frac{\lambda^2 (\sqrt{2}\sigma)^2}{2}} \quad \text{for } |\lambda| < \frac{1}{2b} \quad (20)$$

As a consequence, an application of sub-exponential tail bound leads directly to tail bounds on a random variables satisfying the Bernstein's condition. However, the resulting tail bound can be sharpened slightly, at least in terms of constant factors, by making direct use of the upper bound 19. We summarize in the following:

**Proposition 2** (Bernstein-type bound). *for any random variable satisfying the Bernstein's condition, we have*

$$\mathbb{E} \left[ e^{\lambda(X - \mathbb{E}X)} \right] \leq e^{\frac{\lambda^2 \sigma^2}{1 - b|\lambda|}} \quad \text{for } |\lambda| < \frac{1}{b}$$

and, moreover, the concentration inequality

$$\mathbb{P} [|X - \mu| \geq t] \leq 2e^{-\frac{t^2}{2(\sigma^2 + bt)}} \quad \text{for } t \geq 0$$

We have proved the first inequality in the discussion preceding this proposition. Using this bound on the moment generating function, the tail bound follows by setting  $\lambda = \frac{t}{\sigma^2 + bt} \in [0, \frac{1}{b})$  in the Chernoff bound, and then simplifying the resulting expression.

## 2 Scalar dynamical system

Consider the scalar dynamical system

$$x_{t+1} = ax_t + u_t + w_t, \quad (21)$$

for  $w_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_w^2)$ , and  $a \in \mathbb{R}$  an unknown parameter. Our goal is to estimate  $a$ , and to do so we inject excitatory Gaussian noise via  $u_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_u^2)$ . We run  $N$  experiments over a horizon of  $T + 1$  time-steps, and then solve for our estimate  $\hat{a}$  via the least-squares problem

$$\begin{aligned} \hat{a} &= \arg \min_a \sum_{i=1}^N (x_{T+1}^{(i)} - ax_T^{(i)} - u_T^{(i)})^2 \\ &= a + \frac{\sum_{i=1}^N x_T^{(i)} w_T^{(i)}}{\sum_{i=1}^N (x_T^{(i)})^2} =: a + e_N. \end{aligned} \quad (22)$$

Notice that we are using only the last two data-points from each trial – this simplifies the analysis of the error term  $e_N$  greatly as each of the summands in the numerator and denominator are now i.i.d. random variables. Our goal is to provide high-probability bounds on this error term in this lecture.

First, we observe that

$$x_T^{(i)} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, (\sigma_w^2 + \sigma_u^2) \sum_{t=0}^T a^{2t}). \quad (23)$$

In what follows, we let  $\sigma_x^2 := (\sigma_w^2 + \sigma_u^2) \sum_{t=0}^T a^{2t}$ , which we recognize as the (variance weighted) finite-time controllability Gramian of the scalar system (21).

**Theorem 2.1.** *Consider the least squares estimator (22). Fix a failure probability  $\delta \in (0, 1]$ , and assume that  $N \geq 32 \log(2/\delta)$ . Then with probability at least  $1 - \delta$ , we have that*

$$|e_N| \leq 4 \frac{\sigma_w}{\sigma_x} \sqrt{\frac{\log(4/\delta)}{N}}. \quad (24)$$

This theorem follows immediately by invoking the next two propositions with failure probability  $\delta/2$  and union bounding.

**Proposition 2.2.** *Fix  $\delta \in (0, 1]$ , and let  $N \geq 32 \log(1/\delta)$ . Then with probability at least  $1 - \delta$*

$$\sum_{i=1}^N (x_T^{(i)})^2 \geq \sigma_x^2 \frac{N}{2}. \quad (25)$$

**Proof.** From (23), we have that  $x_T/\sigma_x \sim \mathcal{N}(0, 1)$ . Thus from Example 2,  $x_T^2/\sigma_x^2$  is sub-exponential with parameters (4, 4), and  $\sum_{i=1}^N \left(x_T^{(i)}/\sigma_x\right)^2$  is sub-exponential with parameters (4N, 4). From the tail bound (16), we see that

$$\mathbb{P} \left[ \sigma_x^2 \sum_{i=1}^N \left( \frac{x_T^{(i)}}{\sigma_x} \right)^2 - N\sigma_x^2 \leq -t \right] = \mathbb{P} \left[ \sum_{i=1}^N \left( \frac{x_T^{(i)}}{\sigma_x} \right)^2 - N \leq \frac{-t}{\sigma_x^2} \right] \leq \exp \left( \frac{-t^2}{8N\sigma_x^4} \right),$$

for all  $t \leq N\sigma_x^2$ . Inverting this bound to solve for a failure probability of  $\delta$ , we see that  $t = \sigma_x^2 \sqrt{8N \log(1/\delta)} \leq N\sigma_x^2$ , where the inequality follows from our assumed lower bound on  $N$ . We therefore have, with probability at least  $1 - \delta$ , that

$$\sum_{i=1}^N (x_T^{(i)})^2 \geq \sigma_x^2 (N - \sqrt{8N \log(1/\delta)}) \geq \sigma_x^2 \frac{N}{2},$$

where the final inequality follows from the assumed lower bound on  $N$ . ■

**Proposition 3.** Fix  $\delta \in (0, 1]$ , and let  $N \geq \frac{1}{2} \log(2/\delta)$ . Then with probability at least  $1 - \delta$

$$\left| \sum_{i=1}^N x_T^{(i)} w_T^{(i)} \right| \leq 2\sigma_x \sigma_w \sqrt{N \log(2/\delta)}. \quad (26)$$

**Proof.** By a similar argument as the previous proof, we have that  $\sum_{i=1}^N x_T^{(i)} w_T^{(i)}/\sigma_x \sigma_w$  is sub-exponential with parameters (4N,  $\sqrt{2}$ ), from which it follows that

$$\mathbb{P} \left[ \left| \sum_{i=1}^N x_T^{(i)} w_T^{(i)} \right| \geq t \right] = \mathbb{P} \left[ \left| \sum_{i=1}^N \frac{x_T^{(i)} w_T^{(i)}}{\sigma_x \sigma_w} \right| \geq \frac{t}{\sigma_x \sigma_w} \right] \leq 2 \exp \left( \frac{-t^2}{4N\sigma_x^2 \sigma_w^2} \right),$$

if  $t \leq 2\sqrt{2}N\sigma_x \sigma_w$ . Inverting with probability failure  $\delta$ , we obtain  $t = 2\sigma_x \sigma_w \sqrt{N \log(2/\delta)} \leq 2\sqrt{2}N\sigma_x \sigma_w$ , where the final inequality holds by the assumed lower bound on  $N$ . Thus, with probability at least  $1 - \delta$ , (26) holds. ■

### 3 Numerical experiment

We use scalar linear system  $x_{t+1} = ax_t + u_t + w_t$  to illustrate our bound here. This system has been analyzed in section 2. We set  $a = 2, T = 50, \delta = 0.1, N = 100 \sim 2100$  and all these experiment numbers satisfy the inequality  $N \geq 32 \log(\frac{2}{\delta})$  illustrated in Theorem 3. In both figures, the red line represents true error  $e_N$ , the blue line represents the error bound we get in Theorem 3. In figure 1, we set  $\sigma_w = \sigma_u = 1$ . In figure 2, we set  $\sigma_w = 1, \sigma_u = 2$ .

## References

- [1] Nikolai Matni and Stephen Tu. A tutorial on concentration bounds for system identification. *arXiv preprint arXiv:1906.11395*, 2019.
- [2] Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

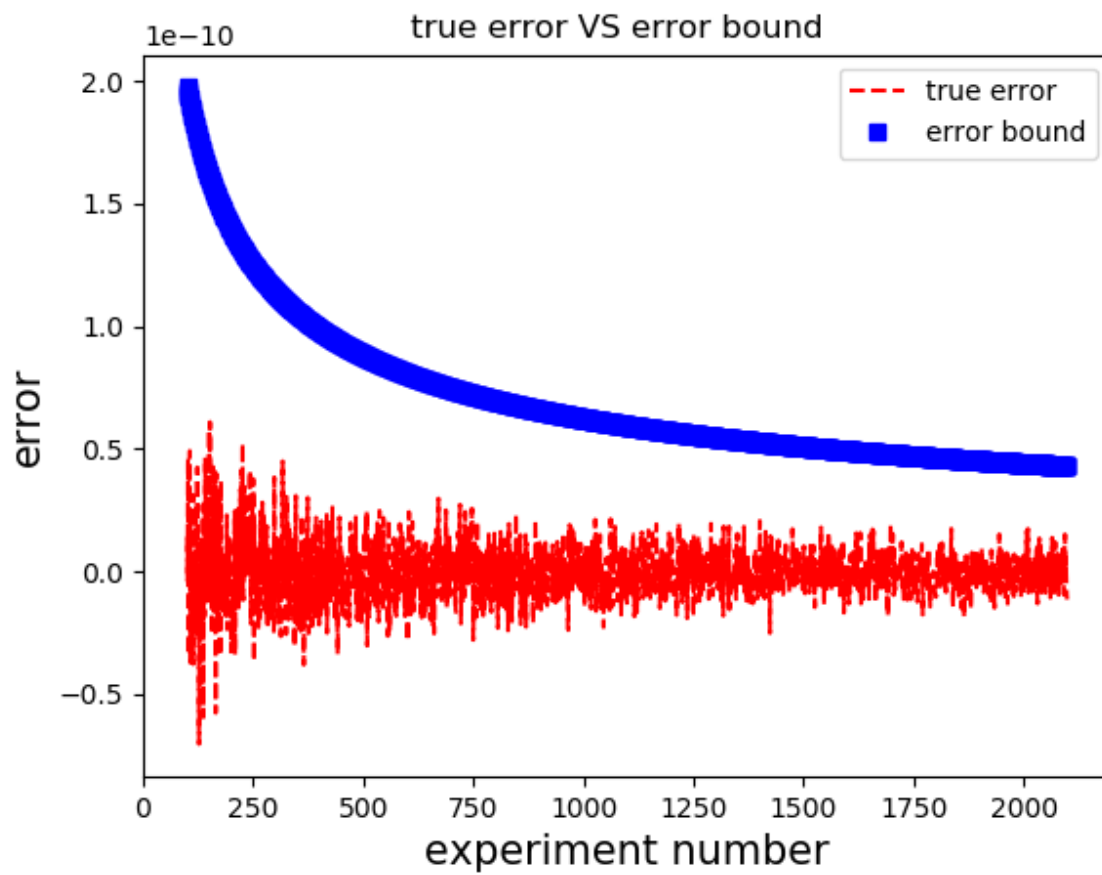
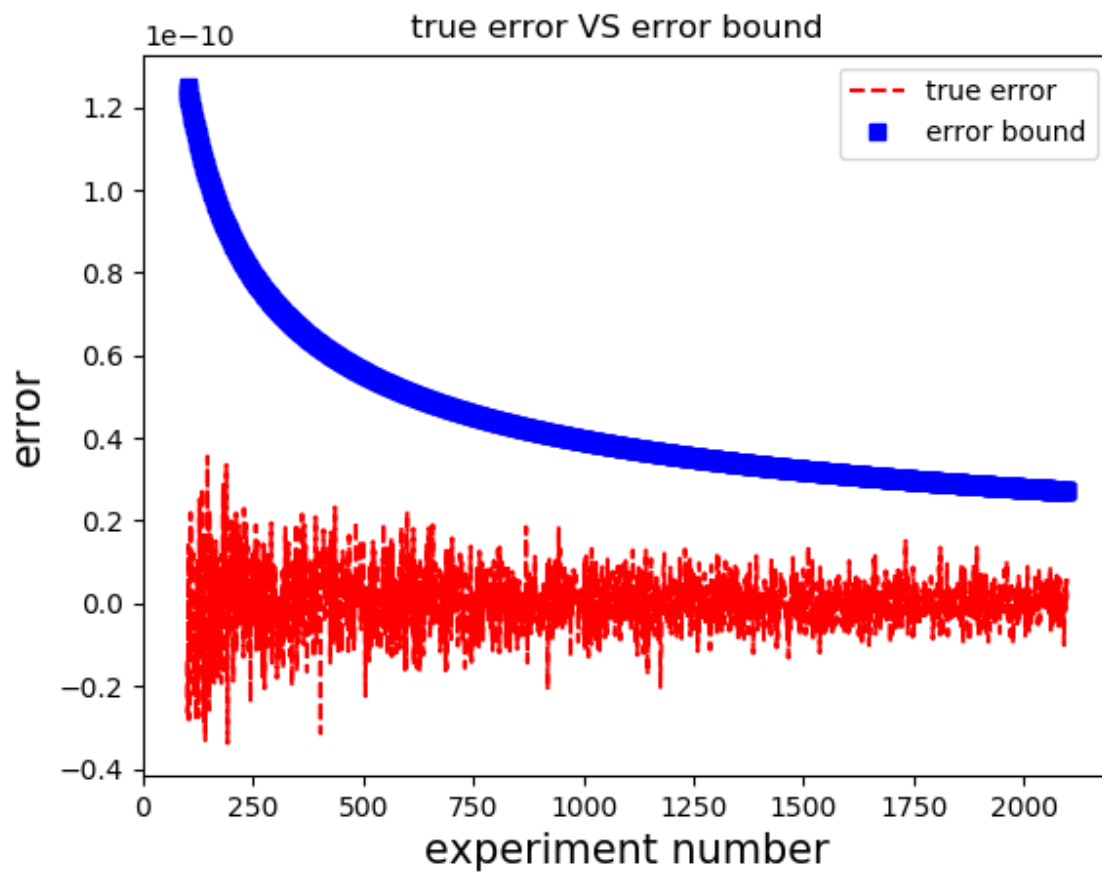


Figure 1:  $\delta_u = 1$   $\delta_w = 1$   $N = 100 \sim 2100$ , a single experiment has a horizon of 50+1 time-steps, as experiment number increases, we see error bound and true error both converge to 0

Figure 2:  $\delta_u = 2$   $\delta_w = 1$   $N = 100 \sim 2100$