ACTIVE SEQUENTIAL HYPOTHESIS TESTING WITH EXTENSION TO ACTIVE REGRESSION AND MULTI-ARMED BANDITS

by

Subhojyoti Mukherjee

A preliminary report submitted in partial fulfillment of
the requirements for the degree of

Master of Research (Project)

(Electrical Engineering)

at the

UNIVERSITY OF WISCONSIN–MADISON

2021

To all my friends, family, collaborators and teachers who helped me reach here.

ACKNOWLEDGMENTS

I am grateful to my thesis advisor Dr. Robert Nowak from Electrical and Computer Engineering Department at University of Wisconsin-Madison. Without his constant guidance, motivation and perseverance with me, none of these works would have been completed. His courses of Mathematical Machine Learning and Non-parametric Methods in Data Science helped me a lot in my research. Not only these courses carried a wealth of information on the current state of literature in the respective fields but they used to be hugely competitive which used to constantly motivate me to strive for excellence. One of his greatest influences on me is to inculcate the constant urge to collaborate with as many interested people as possible, both inside and outside of my immediate field of research. This became a huge contributing factor in the research that I conducted for this thesis.

I am also grateful to my co-author Dr. Ardhendu Tripathy who is an Assistant Professor at the Department of Computer Science in Missouri University of Science & Technology. I fondly remember that how intensely he used to scrutinize my drafts for errors and incoherent ideas. I am thankful towards a host of my colleagues in the Nowak Lab at Wisconsin Institute of Discovery where I spent a significant time while conducting research at UW-Madison. Thanks Blake Mason, Yinglun Zhu, Rahul Parhi, Scott Sievert, and Julian-Katz Samuels for creating such an environment in the lab, conducive to research. Last but not the least, I must thank Dr. A.P. Vijay Rengarajan who was doing his post-doc at Carnegie Mellon University. He was always there, listening to all my concerns on a host of issues that I faced in my research and personal life. Vijay without you life will not be same.

DISCARD THIS PAGE

TABLE OF CONTENTS

Page

Appendix

Page

DISCARD THIS PAGE

LIST OF TABLES

DISCARD THIS PAGE

LIST OF FIGURES

DISCARD THIS PAGE

NOMENCLATURE

n	Total number of actions
J	Total number of hypotheses
$\boldsymbol{\Theta}$	Parameter Space
$\mu_i(\boldsymbol{\theta})$	Mean of action i under hypothesis $\boldsymbol{\theta}$
π	Policy
δ	Probability of error of δ-PAC policy
τ_δ	Stopping time
$\beta(J, \delta)$	$\log(J/\delta)$
τ_δ	Stopping time
$\alpha(J)$	$\ell \log(J), \ell > 0$
Y^t	Vector of rewards observed till round t
I^t	Vector of actions sampled till round t

I_s	Action sampled at round s
$Z_i(t)$	Number of time action i is sampled till round t
$\mathbf{p}_{\boldsymbol{\theta}}$	p.m.f. to verify hypothesis $\boldsymbol{\theta}$ (Solution to Chernoff optimization in (1.3))
$\mathrm{KL}(.\|\|.)$	Kullback-Leibler divergence
$\widehat{\boldsymbol{\theta}}(s)$	Most likely hypothesis at round s
$\tilde{\boldsymbol{\theta}}(s)$	Second most likely hypothesis at round s
$L_{Y^t}(\boldsymbol{\theta})$	Sum of squared errors till round t under hypothesis $\boldsymbol{\theta}$
$T_{Y^t}(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$	$L_{Y^t}(\boldsymbol{\theta}) - L_{Y^t}(\boldsymbol{\theta}^*)$
$\xi^{\delta}(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$	Event that $\{L_{Y^{\tau_\delta}}(\boldsymbol{\theta}') - L_{Y^{\tau_\delta}}(\boldsymbol{\theta}) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}\}$
$(1+c)M$	Critical number of samples $(1+c)\left({\log J}/{D_1} + {\log(J/\delta)}/{D_0}\right), c > 0$
$\mathcal{I}$	$\{i \in [n] : i = \arg\max_{i' \in [n]} (\mu_{i'}(\boldsymbol{\theta}) - \mu_{i'}(\boldsymbol{\theta}'))^2$ for some $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}\}$

ACTIVE SEQUENTIAL HYPOTHESIS TESTING WITH EXTENSION TO ACTIVE REGRESSION AND MULTI-ARMED BANDITS

Subhojyoti Mukherjee

Under the supervision of Professor Robert Nowak

At the University of Wisconsin-Madison

Pool-based active learning promises more efficient use of noisy oracle responses by querying informative data points. In this setting the learner has access to a set of data points or actions and it must choose the most informative action in a sequential manner. In the end it must declare an estimate of its quantity of interest and suffer a penalty for wrong declaration. This is called the general active testing setting which shows up in many real-world scenario such as placing ads on the internet, designing medical trials, managing industrial workers, etc.

In this thesis, we build a connection between active sequential hypothesis testing, a framework initiated by Chernoff (1959), and active regression. In the first part of the thesis we study the active sequential hypothesis testing setting. For active testing, using techniques introduced by Naghshvar and Javidi (2013), we obtain a novel sample complexity bound for Chernoff's original procedure, uncovering non-asymptotic terms that reduce in significance as the allowed error probability $\delta \to 0$.

In the second part of the thesis we study the active regression setting. Initially proposed for testing among finitely many hypotheses, we obtain the analogue of Chernoff sampling for the case when the hypotheses belong to a compact space. This allows us to use it in estimating the parameter in active regression. Using techniques introduced by Chaudhuri et al. (2015), we show that the risk of our procedure converges to the optimal value. Unlike existing methods that typically change the sampling proportion in stages, our procedure is fully adaptive. We demonstrate its potential by actively learning neural network models and in real-data linear and non-linear regression problems, where our general-purpose approach outperforms state-of-the-art methods.

Finally, in the third part of the thesis we study the best-arm identification in multi-armed bandit setting using the ideas from active regression. The extension to active regression allows us to directly apply it to structured bandit problems, where the unknown parameter specifying the arm

means is often assumed to be an element of Euclidean space. Empirically, we demonstrate the potential of our proposed approach in linear and non-linear bandit settings, where we observe that our general-purpose approach compares favorably to state-of-the-art bandit methods.

ABSTRACT

Pool-based active learning promises more efficient use of noisy oracle responses by querying informative data points. In this setting the learner has access to a set of data points or actions and it must choose the most informative action in a sequential manner. In the end it must declare an estimate of its quantity of interest and suffer a penalty for wrong declaration. This is called the general active testing setting which shows up in many real-world scenario such as placing ads on the internet, designing medical trials, managing industrial workers, etc.

In this thesis, we build a connection between active sequential hypothesis testing, a framework initiated by Chernoff (1959), and active regression. In the first part of the thesis we study the active sequential hypothesis testing setting. For active testing, using techniques introduced by Naghshvar and Javidi (2013), we obtain a novel sample complexity bound for Chernoff's original procedure, uncovering non-asymptotic terms that reduce in significance as the allowed error probability $\delta \to 0$.

In the second part of the thesis we study the active regression setting. Initially proposed for testing among finitely many hypotheses, we obtain the analogue of Chernoff sampling for the case when the hypotheses belong to a compact space. This allows us to use it in estimating the parameter in active regression. Using techniques introduced by Chaudhuri et al. (2015), we show that the risk of our procedure converges to the optimal value. Unlike existing methods that typically change the sampling proportion in stages, our procedure is fully adaptive. We demonstrate its potential by actively learning neural network models and in real-data linear and non-linear regression problems, where our general-purpose approach outperforms state-of-the-art methods.

Finally, in the third part of the thesis we study the best-arm identification in multi-armed bandit setting using the ideas from active regression. The extension to active regression allows us to directly apply it to structured bandit problems, where the unknown parameter specifying the arm

means is often assumed to be an element of Euclidean space. Empirically, we demonstrate the potential of our proposed approach in linear and non-linear bandit settings, where we observe that our general-purpose approach compares favorably to state-of-the-art bandit methods.

Chapter 1

Active Sequential Hypothesis Testing

1.1 Introduction

Consider a learner who wants to dynamically collect observations so as to improve their information about an underlying phenomena of interest. The learner needs to collect the observations quickly as well as account for the penalty of wrong declaration. This is a sequential learning problem where the learner relies on his current information state to adaptively select the most "informative" action from the available action set. In this thesis we build a connection to the sequential design of experiments originally proposed by Chernoff in Chernoff (1959) and develop new approaches in active regression. Let us first establish some basic notation and a problem statement. Consider a sequential learning problem in which the learner may select one of n possible measurement actions at each step. A sample resulting from measurement action i is a realization of a sub-Gaussian random variable with mean $\mu_i(\boldsymbol{\theta}^*)$, where the mean $\mu_i(\boldsymbol{\theta})$ is a known function parameterized by $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and ϵ is a zero mean random variable. The specific $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$ that governs the observations is not known. Each measurement action may be performed multiple times, resulting in i.i.d. observations of this form, and all measurements (from different actions) are also statistically independent. This thesis considers the problem of sequentially and adaptively choosing measurement actions to accomplish one or more of the following goals:

Active Testing: $\boldsymbol{\Theta}$ is finite and the goal is to correctly determine the true hypothesis $\boldsymbol{\theta}^*$.

Active Regression: $\boldsymbol{\Theta}$ is a compact (uncountable) space and the goal is to accurately estimate $\boldsymbol{\theta}^*$.

Structured Bandit: The goal is to identify measurement action(s) that achieve $\max_i \mu_i(\boldsymbol{\theta}^*)$.

Note that knowledge of $\boldsymbol{\theta}^*$ is sufficient, but may not be necessary, to determine the best action(s). In the parlance of multi-armed bandits, the actions are "arms" and the structure is determined by the functions $\mu_i(\cdot)$ for every $i \in [n] := \{1, \ldots, n\}$. For example, in the linear bandit setting arm i is associated with a feature vector $\mathbf{x}_i$ and the function $\mu_i(\boldsymbol{\theta}) := \boldsymbol{\theta}^T \mathbf{x}_i$. In general, μ_i may be a nonlinear function and the value of $\mu_i(\boldsymbol{\theta}^*)$ is the "reward" of action/arm i. This thesis also studies the "best-arm" problem, maximizing the cumulative rewards (i.e., minimizing the cumulative regret) over time is another common goal in structured bandit problems, but not a focus of this thesis.

We now introduce the active testing setting. A sequential policy π tasked to find $\boldsymbol{\theta}^*$ interacts with the environment in an iterative fashion. In this thesis we use the term policy and algorithm interchangeably. At time t, the policy samples arm I_t and receives a random reward Y_t that follows the distribution $\nu_{I_t}(\cdot;\boldsymbol{\theta}^*)$, where $\boldsymbol{\theta}^*$ is the true value of the unknown parameter that belongs to a set $\boldsymbol{\Theta}$. Let $\mathcal{F}_t := \sigma(I_1, Y_1, I_2, Y_2, \ldots, I_t, Y_t)$ denote the sigma-algebra generated by the sequence of actions and observations till time t. Policy π is said to be δ-PAC if: (1) at each t the sampling rule I_t is $\mathcal{F}_{t-1}$ measurable, (2) it has a finite stopping time τ_δ with respect to $\mathcal{F}_t$, and (3) its final prediction $\widehat{\boldsymbol{\theta}}(\tau_\delta)$ is based on $\mathcal{F}_{\tau_\delta}$ and satisfies $\mathbb{P}(\widehat{\boldsymbol{\theta}}(\tau_\delta) \neq \boldsymbol{\theta}^*) \leq \delta$. Based on the observations $Y^t := (Y_1, Y_2, \ldots, Y_t)$, we define for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ the sum of squared errors as follows:

$$L_{Y^t}(\boldsymbol{\theta}) := \sum_{s=1}^{t} (Y_s - \mu_{I_s}(\boldsymbol{\theta}))^2. \tag{1.1}$$

Next we discuss the main contributions of this thesis before delving into the individual parts of the thesis.

1.2 Main Contributions

The main contributions of the thesis are as follows:

1. In the first part of the thesis we revisit Chernoff's method of sequential design of experiments, which is equivalent to the situation where the parameter space $\boldsymbol{\Theta}$ is a finite set. Chernoff proposed an algorithm that provably minimizes the number of samples used to identify $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$ in the asymptotic high-confidence setting. At each step, Chernoff's algorithm solves an optimization problem to obtain a sampling distribution over the set of actions and draws the next sample accordingly. We derive finite-sample complexity bounds for Chernoff's algorithm, exposing a new term that may dominate in low/medium confidence regimes. We demonstrate the importance of the new term in the context of active testing and discuss it in relation to recent work by Naghshvar and Javidi (2013). We also generalize the Chernoff's policy to handle sub-Gaussian noise distributions. Consequently, we replace the maximum likelihood criterion with the minimum sum-of-squared errors criterion (which depends only on the mean functions and not on the probability distribution of the noise).

2. In the second part of the thesis we extend Chernoff's method to handle smoothly parameterized functions $\mu_i(\boldsymbol{\theta})$ where $\boldsymbol{\Theta} \subseteq \mathbb{R}^d$. A brute-force approach could involve using a finite, discrete covering of $\boldsymbol{\Theta}$, but this is impractical. Instead, we prove that an optimal sampling distribution (according to the Chernoff method) is generally sparse and may be obtained by solving a relatively simple eigenvalue optimization problem closely related to the notion of E-optimality in experimental design (Dette and Studden, 1993). We provide convergence guarantee for the smoothly parameterized setting which requires a new metric of comparison. We demonstrate that our fully online GC procedure for the smoothly parameterized setting outperforms state-of-the-art algorithms in several benchmark real-life dataset and in neural networks experiments.

3. Finally, in the third part we investigate the best-arm problem in structured bandits using Generalized Chernoff sampling. The actions partition $\boldsymbol{\Theta}$ such that the cell associated with action j consists of all $\boldsymbol{\theta}$ such that $\mu_j(\boldsymbol{\theta}) = \max_i \mu_i(\boldsymbol{\theta})$ (the maximum is assumed to be unique by assumption). This partitioning leads to a natural stopping condition for Chernoff's algorithm when the goal is only to identify the best action. For both linear and non-linear reward functions, our approach performs favorably compared to existing methods and other baselines.

Within Active Testing, Chernoff (1959) had given a sampling procedure that is asymptotically optimal, under less stringent assumptions Naghshvar and Javidi (2013) obtained non-asymptotic terms in their upper bound. We show that Chernoff's original procedure also has a similar non-asymptotic term as that in Naghshvar and Javidi (2013). For Active Regression, we bound a time-dependent loss. As $t \to \infty$, the loss $P_t \to L_{\mathbf{p}_{\boldsymbol{\theta}^*}}$, where $\mathbf{p}_{\boldsymbol{\theta}^*}$ is the optimal verification proportion.

1.3 Related Works in Active Testing

Here is the core idea of Chernoff (1959). Consider a partition of $\boldsymbol{\Theta} = \boldsymbol{\Theta}_1 \cup \boldsymbol{\Theta}_2$ and a hypothesis test $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}_1$ or $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}_2$. Assume we have flexibility in selecting among a number of measurement actions to gather the data for testing, and that the distribution of each action under each hypothesis $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is known. A maximum-likelihood estimate $\widehat{\boldsymbol{\theta}}$ is found using past measurements, and suppose

$\widehat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}_1$. The next action is chosen according to a probability mass function over actions obtained by

$$\arg\max_{\mathbf{p}} \inf_{\boldsymbol{\theta}' \in \boldsymbol{\Theta}_2} \sum_{i=1}^{n} p(i)\, \mathrm{KL}(\nu_i(x; \widehat{\boldsymbol{\theta}}) \| \nu_i(x; \boldsymbol{\theta}')), \tag{1.2}$$

where $\mathbf{p} = (p(1), \ldots, p(n))$, $\nu_i(\cdot; \boldsymbol{\theta})$ denotes the probability distribution of the value of action i if $\boldsymbol{\theta}$ were the true hypothesis, and KL denotes the Kullback-Leibler divergence. Albert (1961) showed that a similar sampling rule is an asymptotically optimal strategy when $\boldsymbol{\Theta}$ contains infinitely many hypotheses. The optimization (1.2) is similar to those appearing in sample complexity lower bounds for multi-armed bandit problems, e.g., Garivier and Kaufmann (2016); Combes et al. (2017); Degenne et al. (2020a). Roughly speaking, in that line of work the $\inf$ is taken over all $\boldsymbol{\theta}'$ with an optimal action that differs from that of a reference $\boldsymbol{\theta}$.

Within the active testing literature, Naghshvar and Javidi (2013) identified that while Chernoff's procedure is asymptotically optimal as in the number of samples, it does not consider any notion of optimality when the number of hypotheses tends to infinity. Naghshvar and Javidi (2013) proposed a two-stage Bayesian policy TP which conducts forced exploration in the first stage, builds a posterior belief over the hypotheses and once the probability of one hypothesis crosses a threshold it samples according to the optimal sampling proportion. As opposed to this we show that the fully online frequentist algorithm GC with no such separation of stages empirically outperforms TP as well as enjoys both moderate and optimal asymptotic guarantees. Nitinawarat et al. (2013) have shown that Chernoff's procedure is asymptotically optimal in a stronger sense relevant for testing among multiple hypotheses and proposed an algorithm that has no moderate confidence guarantee. In a different setting Vaidhiyan and Sundaresan (2017) have modified Chernoff's procedure to quickly identify an odd Poisson point process having a different rate of arrival than others.

1.4 Verification Proportion

Suppose we are asked to verify (i.e. output "yes"/"no") in a δ-PAC manner if a given $\boldsymbol{\theta}$ is the true hypothesis. What is an optimal sampling rule that achieves this using the fewest expected number of samples? The answer was given by Chernoff (1959) and we re-derive it in the Gaussian case.

We first introduce the following assumption for bounded rewards required to prove the verification Theorem 1.

Assumption 1. *The reward from any arm under any hypothesis has bounded range, i.e.,* $Y_s \in [-\sqrt{\eta}/2, \sqrt{\eta}/2]$ *almost surely at every round* s *for some fixed* $\eta > 0$.

Next we prove the verification theorem to verify if a given $\boldsymbol{\theta}$ is the true hypothesis

Theorem 1. ***(Verification Proportion)*** *Let* $\nu_i(Y;\boldsymbol{\theta})$ *be Gaussian with mean* $\mu_i(\boldsymbol{\theta})$ *and variance* $1/2$. *For a* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, *a* δ*-PAC policy* π *that verifies if* $\boldsymbol{\theta}$ *is the true hypothesis using the fewest expected number of samples must follow at all times* t *the randomized sampling rule* $\mathbb{P}(I_t = i) = p_{\boldsymbol{\theta}}(i) \, \forall i \in [n]$, *where*

$$\mathbf{p}_{\boldsymbol{\theta}} := \arg\max_{\mathbf{p}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}} \sum_{i=1}^{n} p(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}))^2. \tag{1.3}$$

Here $\mathbf{p}$ *is the optimization variable restricted to be a probability mass function, its* i*-th component is denoted as* $p(i)$.

Let $L_{Y^t}(\boldsymbol{\theta})$ denote the sum of squared errors of hypothesis $\boldsymbol{\theta}$ based on the reward vector Y^t. The optimal hypothesis test between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ is a likelihood ratio test. If the test statistic $T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}) = L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta})$ crosses a threshold $\beta(J, \delta) > 0$ for all hypotheses $\boldsymbol{\theta}' \in \boldsymbol{\Theta} \setminus \{\boldsymbol{\theta}\}$ then we stop and declare $\boldsymbol{\theta}$. $T_{Y^t}(\cdot, \boldsymbol{\theta})$ is a Gaussian random variable and if $\boldsymbol{\theta}$ is true, sampling sources according to the p.m.f. in (1.3) allows π to perform all hypotheses tests of $\boldsymbol{\theta}$ against $\boldsymbol{\theta}'$ in the fewest expected number of samples. The full proof is given below.

We can solve the optimization problem in (1.3) as a linear program to verify hypothesis $\boldsymbol{\theta}$ as follows:

$$\max_{\mathbf{p}} \ell \text{ **s.t.** } \sum_{i=1}^{n} p(i)(\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}'))^2 \geq \ell \, \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}, \tag{1.4}$$

where $\mathbf{p}$ is a p.m.f. and the optimization variable that satisfies the constraint $\sum_{i=1}^{n} p(i) = 1$.

1.4.1 Proof of Main Theorem

Proof. Denote the sequence of sampling actions till time τ as $I^\tau := (I_1, I_2, \ldots, I_\tau)$ and sequence of rewards as $Y^\tau := (Y_1, Y_2, \ldots, Y_\tau)$. Given I^τ and Y^τ, the log-likelihood of a hypothesis $\boldsymbol{\theta}'$ is as follows.

$$\begin{aligned}\log \mathbb{P}(Y^\tau | I^\tau, \boldsymbol{\theta}') &= \log \prod_{t=1}^{\tau} \mathbb{P}(Y_t | I_t, \boldsymbol{\theta}') = \sum_{t=1}^{\tau} \log(\exp(-(Y_t - \mu_{I_t}(\boldsymbol{\theta}'))^2)/2\sqrt{\pi}) \\ &= -\tau \log(2\sqrt{\pi}) - \sum_{t=1}^{\tau} Y_t^2 + \sum_{t=1}^{\tau} (2\mu_{I_t}(\boldsymbol{\theta}')Y_t - \mu_{I_t}(\boldsymbol{\theta}')^2).\end{aligned}$$

We are asked to verify if $\boldsymbol{\theta}$ is the true hypothesis. Assuming all hypotheses are equally probable a priori, the optimal test to disambiguate between hypothesis $\boldsymbol{\theta}$ and $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ is the likelihood ratio test. Equivalently, we can consider the ratio of the log-likelihoods as our test statistic $T_{Y^\tau}(\boldsymbol{\theta}, \boldsymbol{\theta}')$ and decide based on the rule (where τ is determined later)

$$T_{Y^\tau}(\boldsymbol{\theta}, \boldsymbol{\theta}') := \log\left(\frac{\mathbb{P}(Y^\tau | I^\tau, \boldsymbol{\theta})}{\mathbb{P}(Y^\tau | I^\tau, \boldsymbol{\theta}')}\right) = \sum_{t=1}^{\tau} \left(2Y_t(\mu_{I_t}(\boldsymbol{\theta}) - \mu_{I_t}(\boldsymbol{\theta}')) - (\mu_{I_t}(\boldsymbol{\theta})^2 - \mu_{I_t}(\boldsymbol{\theta}')^2)\right) \underset{\boldsymbol{\theta}'}{\overset{\boldsymbol{\theta}}{\gtrless}} 0. \tag{1.5}$$

If $\boldsymbol{\theta}$ is the true hypothesis, $Y_t \sim \mathcal{N}(\mu_{I_t}(\boldsymbol{\theta}), 1/2)$. $T_{Y^\tau}(\boldsymbol{\theta}, \boldsymbol{\theta}')$ is a Gaussian random variable with $\mathbb{E}[T_{Y^\tau}(\boldsymbol{\theta}, \boldsymbol{\theta}')|I^\tau, \boldsymbol{\theta}] = \sum_{t=1}^{\tau} 2(\mu_{I_t}(\boldsymbol{\theta}) - \mu_{I_t}(\boldsymbol{\theta}'))^2 = \mathrm{Var}[T_{Y^\tau}(\boldsymbol{\theta}, \boldsymbol{\theta}')|I^\tau, \boldsymbol{\theta}]$. Using a bound for the Gaussian Q-function, the probability of making an error by declaring $\boldsymbol{\theta}'$ when $\boldsymbol{\theta}$ is true is

$$\mathbb{P}(T_{Y^\tau}(\boldsymbol{\theta}, \boldsymbol{\theta}') < 0 | I^\tau, \boldsymbol{\theta}) \leq \exp\left(\frac{-\sum_{t=1}^{\tau} 2(\mu_{I_t}(\boldsymbol{\theta}) - \mu_{I_t}(\boldsymbol{\theta}'))^2}{2}\right). \tag{1.6}$$

We have to minimize the probability of error for all $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$. In order to guarantee a desired error probability δ, one should choose I^τ to ensure

$$\sum_{t=1}^{\tau} (\mu_{I_t}(\boldsymbol{\theta}) - \mu_{I_t}(\boldsymbol{\theta}'))^2 \geq \beta(J, \delta) \quad \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}, \tag{1.7}$$

where $\beta(J, \delta) = \log(J/\delta)$ and the above inequality (1.7) ensures that in (1.6) the $\mathbb{P}(T_{Y^\tau}(\boldsymbol{\theta}, \boldsymbol{\theta}') < 0 | I^\tau, \boldsymbol{\theta}) \leq \delta/J$ for all $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$. If we follow the rule in (1.5) then we will conclude that $\boldsymbol{\theta}$ is the true hypothesis with high probability.

If $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ is the true hypothesis instead, then one can evaluate that $\mathrm{Var}[T_{Y^\tau}(\boldsymbol{\theta}', \boldsymbol{\theta})|I^\tau, \boldsymbol{\theta}'] = \sum_{t=1}^{\tau} 2(\mu_{I_t}(\boldsymbol{\theta}) - \mu_{I_t}(\boldsymbol{\theta}'))^2 = -\mathbb{E}[T_{Y^\tau}(\boldsymbol{\theta}', \boldsymbol{\theta})|I^\tau, \boldsymbol{\theta}']$, yielding

$$\mathbb{P}(T_{Y^\tau}(\boldsymbol{\theta}, \boldsymbol{\theta}') < 0|I^\tau, \boldsymbol{\theta}') = 1 - \mathbb{P}(T_{Y^\tau}(\boldsymbol{\theta}, \boldsymbol{\theta}') \geq 0|I^\tau, k) \geq 1 - \exp\left(\frac{-\sum_{t=1}^{\tau} 2(\mu_{I_t}(\boldsymbol{\theta}) - \mu_{I_t}(\boldsymbol{\theta}'))^2}{2}\right).$$

Again using the inequality (1.7) and $\beta(J, \delta) = \log(\delta/J)$ in the above inequality ensures that $\mathbb{P}(T_{Y^\tau}(\boldsymbol{\theta}, \boldsymbol{\theta}') < 0|I^\tau, \boldsymbol{\theta}') \geq 1 - \delta/J$. If we follow the rule in (1.5) comparing $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}$ then we will conclude that $\boldsymbol{\theta}$ is not the true hypothesis with high probability.

If I_t is chosen such that $\mathbb{P}(I_t = i) = p_{\boldsymbol{\theta}}(i)$ at all times t, then by Wald's lemma, the expected value of the LHS in (1.7) is $\mathbb{E}[\tau] \sum_{i=1}^{n} p(i)(\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}'))^2$. Thus the expected value of the LHS in (1.7) meets its required condition with the smallest value of $\mathbb{E}[\tau]$, if $\mathbf{p}_{\boldsymbol{\theta}}$ is chosen as follows:

$$\mathbf{p}_{\boldsymbol{\theta}} := \arg\max_{\mathbf{p}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}} \sum_{i=1}^{n} \mathbf{p}(i)(\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}'))^2. \tag{1.8}$$

The claim of the theorem follows. □

Once we have established the claim of Theorem 1 we now show in the next chapter how to extend this idea to come up with the GC policy.

Chapter 2

Generalized Chernoff Sampling

In this chapter we introduce the GC policy. We first show how the main idea of GC follows directly from the verification proportion proved in Theorem 1. Then we prove a sample complexity bound for GC policy that holds for all $\delta > 0$.

2.1 GC Sampling

Inspired by the sampling proportion in Theorem 1, we use the same sampling strategy even though the distributions $\{\nu_i(Y;\boldsymbol{\theta})\}_{i=1}^n$ are only assumed to be sub-Gaussian (Algorithm 1). Our estimate of the most likely hypothesis (breaking ties at random) given the data is $\widehat{\boldsymbol{\theta}}(t)$ given by

$$\widehat{\boldsymbol{\theta}}(t) := \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_{Y^t}(\boldsymbol{\theta}). \tag{2.1}$$

The arm sampled at the next time $t+1$ is chosen by the randomized rule $\mathbb{P}(I_{t+1} = i) = p_{\widehat{\boldsymbol{\theta}}(t)}(i), \forall i \in [n]$. By the sub-Gaussian assumption, we can apply similar concentration bounds as used in the proof of Theorem 1. We stop sampling at τ_δ if the sum of squared errors for all competing hypothesis is greater than that of $\widehat{\boldsymbol{\theta}}(\tau_\delta)$ by a threshold

$$\beta(J,\delta) := \log(J/\delta).$$

The pseudo-code of Algorithm 1 is given below.

Algorithm 1 Generalized Chernoff Sampling (GC)

1: **Input:** Confidence parameter δ, $\beta(J,\delta)$.
2: Sample $I_1 \in [n]$ randomly, observe Y_1 and find $\widehat{\boldsymbol{\theta}}(1)$.
3: **for** $t = 2, 3, \ldots$ **do**
4: Sample $I_t \sim \mathbf{p}_{\widehat{\boldsymbol{\theta}}(t-1)}$ from (1.3) and observe Y_t.
5: Calculate $L_{Y^t}(\boldsymbol{\theta})$ from (1.1) $\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$, find $\widehat{\boldsymbol{\theta}}(t)$.
6: **if** $L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\widehat{\boldsymbol{\theta}}(t)) > \beta(J,\delta) \forall \boldsymbol{\theta}' \neq \widehat{\boldsymbol{\theta}}(t)$ **then**
7: **Return** $\widehat{\boldsymbol{\theta}}(t)$ as the true hypothesis.

In Chernoff (1959) they provide a sample complexity upper bound only for the asymptotic regime when $\delta \to 0$. We give a non-asymptotic fixed confidence sample complexity upper bound in

Theorem 2. The following assumption, which was also made by Chernoff (1959), is used to prove Theorem 2.

Assumption 2. *The mean of any arm under $\boldsymbol{\theta}^*$ is different than its mean under any other hypothesis, i.e.,* $\min_{i\in[n]} \min_{\boldsymbol{\theta}\neq\boldsymbol{\theta}^*} |\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}^*)| > 0$.

Note that Assumption 2 is used to obtain theoretical guarantees and the value of η_0 is not needed to run the algorithm. In subsequent works by Nitinawarat et al. (2013) and Naghshvar and Javidi (2013), it was shown that the above assumption can be relaxed if the algorithm is modified. We make the assumption since we give a non-asymptotic sample complexity bound for the original algorithm.

Definition 1. *Define the smallest squared difference of means between any two arms under any pair of hypotheses $\boldsymbol{\theta}, \boldsymbol{\theta}'$ as $\eta_0 := \min_{i\in[n]} \min_{\boldsymbol{\theta}\neq\boldsymbol{\theta}'} (\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}'))^2$. By Assumption 2 we have $\eta_0 > 0$.*

Theorem 2. ***(GC Sample Complexity)*** *Let τ_δ denote the stopping time of GC in Algorithm 1. Let D_0 be the objective value of the* $\max\min$ *optimization in* (1.3) *when $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, i.e.,*

$$D_0 := \max_{\mathbf{p}} \min_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*} \sum_{i=1}^{n} p(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2, D_1 := \min_{\{\mathbf{p}_{\boldsymbol{\theta}}:\boldsymbol{\theta}\in\boldsymbol{\Theta}\}} \min_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*} \sum_{i=1}^{n} p_{\boldsymbol{\theta}}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2,$$

where $\mathbf{p}_{\boldsymbol{\theta}}$ is the solution of (1.3) *and from Assumption 2 we get that $D_1 > 0$. The sample complexity of the δ-PAC GC has the following upper bound, where $J := |\boldsymbol{\Theta}|$, $C = O((\eta/\eta_0)^2)$ is a constant:*

$$\mathbb{E}[\tau_\delta] \leq O\left(\frac{\eta \log(C) \log J}{D_1} + \frac{\log(J/\delta)}{D_0} + JC^{\frac{1}{\eta}}\delta^{\frac{D_0}{\eta^2}}\right).$$

In any run of GC , we can think of two implicit stages that are determined by $\{L_{Y^t}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$. The first stage can be called *exploration*, when the most likely hypothesis is not necessarily $\boldsymbol{\theta}^*$. This stage ends at the last time when $\widehat{\boldsymbol{\theta}}(t) \neq \boldsymbol{\theta}^*$. The second stage can be called *exploitation*, when the most likely hypothesis is indeed $\boldsymbol{\theta}^*$, and we are collecting samples in the most efficient way to verify that $\widehat{\boldsymbol{\theta}}(t) = \boldsymbol{\theta}^*$. In Theorem 2 the $C(\eta_0, \eta)\log(J)/D_1$ term bounds the number of samples taken in the exploration phase. Chernoff (1959) does not give any guarantees for this non-asymptotic regime for some finite δ. Our proof technique uses the Assumption 2 that $D_1 > 0$, along with

sub-Gaussian concentration inequality (Lemma 1), to obtain the upper bound. The $\log(J/\delta)/D_0$ term in Theorem 2 counts the number of samples in the exploitation phase, which is after the last time any $\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*$ was the most likely hypothesis. Note that $\log(J/\delta)/D_0$ is the dominating term when $\delta \to 0$, and it matches the asymptotic sample complexity expression of Chernoff (1959). Finally, the $JC^{1/\eta}\delta^{D_0/\eta^2}$ term counts the contribution to the expected sample complexity in the error event, when $\hat{\boldsymbol{\theta}}(\tau_\delta) \neq \boldsymbol{\theta}^*$. This term becomes negligible as $\delta \to 0$. Naghshvar and Javidi (2013) also derive a moderate confidence bound for their policy called TP but suffer from a worse non-asymptotic term $\log(J/\delta)/D_{\text{NJ}}$ where $D_{\text{NJ}} < D_1$. TP do not require require the assumption that $D_1 > 0$ as it conducts forced exploration in the first stage till some hypothesis becomes more likely than all the others and then follows the Chernoff proportion to verify that hypothesis. Note that our method GC is fully adaptive and do not require such explicit stages. The TP policy is asymptotically optimal but performs poorly in some instances (see Example 1 and 3 group setting experiment) due to the fixed exploration in the first stage. Next we prove the result of Theorem 2 below.

Proof. **Step 1 (Definitions):** Define $L_{Y^t}(\boldsymbol{\theta})$ as the total sum of squared errors of hypothesis $\boldsymbol{\theta}$ till round t. Let, $T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) := L_{Y^t}(\boldsymbol{\theta}^*) - L_{Y^t}(\boldsymbol{\theta})$ be the difference of squared errors between $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}$. Note that the p.m.f. $\mathbf{p}_{\boldsymbol{\theta}}$ is the Chernoff verification proportion for verifying hypothesis $\boldsymbol{\theta}$.

Step 2 (Define stopping time and partition): We define the stopping time τ_δ for the policy π as follows:

$$\tau_\delta := \min\{t : \exists \boldsymbol{\theta} \in \boldsymbol{\Theta}, L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}\} \tag{2.2}$$

where, $\beta(J, \delta)$ is the threshold function defined in (7.7).

Step 3 (Define bad event): We define the bad event $\xi^\delta(\boldsymbol{\theta})$ for the sub-optimal hypothesis $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$ as follows:

$$\xi^\delta(\boldsymbol{\theta}) = \{\widehat{\boldsymbol{\theta}}(t) = \boldsymbol{\theta}, L_{Y^{\tau_\delta}}(\boldsymbol{\theta}') - L_{Y^{\tau_\delta}}(\boldsymbol{\theta}) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}\}. \tag{2.3}$$

The event $\xi^\delta(\boldsymbol{\theta})$ denotes that a sub-optimal hypothesis $\boldsymbol{\theta}$ is declared the optimal hypothesis when it has a smaller sum of squared errors than any other hypothesis $\boldsymbol{\theta}'$ at τ_δ.

Step 4 (Decomposition of bad event): In this step we decompose the bad event to show that only comparing $\boldsymbol{\theta}$ against $\boldsymbol{\theta}^*$ is enough to guarantee a δ-PAC policy. First we decompose the bad event $\xi^{\delta}(\boldsymbol{\theta})$ as follows:

$$\begin{aligned}\xi^{\delta}(\boldsymbol{\theta}) =& \{\widehat{\boldsymbol{\theta}}(\tau_\delta) = \boldsymbol{\theta}, L_{Y^{\tau_\delta}}(\boldsymbol{\theta}') - L_{Y^{\tau_\delta}}(\boldsymbol{\theta}) > \beta(J,\delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}\} \\ \overset{(a)}{=}& \{\underbrace{\widehat{\boldsymbol{\theta}}(\tau_\delta) = \boldsymbol{\theta}, L_{Y^{\tau_\delta}}(\boldsymbol{\theta}') - L_{Y^{\tau_\delta}}(\boldsymbol{\theta}) > \beta(J,\delta), \forall \boldsymbol{\theta}' \in \boldsymbol{\Theta} \setminus \{\boldsymbol{\theta}^*\}}_{\textbf{part A}}\} \\ & \bigcap \{\underbrace{\widehat{\boldsymbol{\theta}}(\tau_\delta) = \boldsymbol{\theta}, L_{Y^{\tau_\delta}}(\boldsymbol{\theta}^*) - L_{Y^{\tau_\delta}}(\boldsymbol{\theta}) > \beta(J,\delta)}_{\textbf{part B}}\} \\ \subseteq& \{\widehat{\boldsymbol{\theta}}(\tau_\delta) = \boldsymbol{\theta}, L_{Y^{\tau_\delta}}(\boldsymbol{\theta}^*) - L_{Y^{\tau_\delta}}(\boldsymbol{\theta}) > \beta(J,\delta)\} \end{aligned} \tag{2.4}$$

where, (a) follows by decomposing the event in two parts containing $\boldsymbol{\theta} \in \boldsymbol{\Theta} \setminus \{\boldsymbol{\theta}^*\}$ and $\{\boldsymbol{\theta}^*\}$, (b) follows by noting that the intersection of events holds by taking into account only the event in part B.

Step 5 (Proof of correctness): In this step we want to show that based on the stopping time definition and the bad event $\xi^{\delta}(\boldsymbol{\theta})$ the GC stops and outputs the correct hypothesis $\boldsymbol{\theta}^*$ with $1-\delta$ probability. As shown in Step 4, we can define the error event $\xi^{\delta}(\boldsymbol{\theta})$ as follows:

$$\xi^{\delta}(\boldsymbol{\theta}) \subseteq \{\widehat{\boldsymbol{\theta}}(\tau_\delta) = \boldsymbol{\theta}, L_{Y^{\tau_\delta}}(\boldsymbol{\theta}^*) - L_{Y^{\tau_\delta}}(\boldsymbol{\theta}) > \beta(J,\delta)\}$$

Define $\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} = \min\{t : T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > \beta(J,\delta)\}$. Then we can show for the stopping time τ_δ, the round $\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}}$ from Lemma 7.10 and the threshold $\beta(J,\delta) := \log\left(\frac{(1+\eta^2/2\eta_0^2)\,J}{\delta}\right)$ we have

$$\begin{aligned}\mathbb{P}\left(\tau_\delta < \infty, \widehat{\boldsymbol{\theta}}(\tau_\delta) \neq \boldsymbol{\theta}^*\right) &\leq \mathbb{P}\left(\exists \boldsymbol{\theta} \in \boldsymbol{\Theta} \setminus \{\boldsymbol{\theta}^*\}, \exists t \in \mathbb{N} : T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > \beta(J,\delta)\right) \\ &\leq \sum_{\boldsymbol{\theta} \neq \boldsymbol{\theta}^*} \mathbb{P}\left(L_{Y^{\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}}}}(\boldsymbol{\theta}^*) - L_{Y^{\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}}}}(\boldsymbol{\theta}) > \beta(J,\delta), \tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} < \infty\right) \overset{(a)}{\leq} \sum_{\boldsymbol{\theta} \neq \boldsymbol{\theta}^*} \frac{\delta}{J} \leq \delta\end{aligned}$$

where, (a) follows from Lemma 7.10.

Step 6 (Sample complexity analysis): In this step we bound the total sample complexity satisfying the δ-PAC criteria. We define the stopping time τ_δ as follows:

$$\tau_\delta := \min\left\{t : L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\widehat{\boldsymbol{\theta}}(t)) > \beta(J,\delta), \forall \boldsymbol{\theta}' \neq \widehat{\boldsymbol{\theta}}(t)\right\}$$

We further define the stopping time $\tau_{\boldsymbol{\theta}^*}$ for the hypothesis $\boldsymbol{\theta}^*$ as follows:

$$\tau_{\boldsymbol{\theta}^*} := \min\left\{t : L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*) > \beta(J,\delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*\right\}. \tag{2.5}$$

We also define the critical number of samples as $(1+c)M$ where M is defined as follows:

$$M := \left(\frac{\alpha(J)}{D_1} + \frac{C'(\eta_0,\eta) + \log(J/\delta)}{D_0}\right) \tag{2.6}$$

where, $C'(\eta_0,\eta) = \log\left(1 + {}^{\eta^2}/_{\eta_0^2}\right)$. We define the term D_1 as follows:

$$D_1 := \min_{\{\mathbf{p}_{\boldsymbol{\theta}}:\boldsymbol{\theta}\in\boldsymbol{\Theta}\}} \min_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*} \sum_{i=1}^{n} p_{\boldsymbol{\theta}}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2 \tag{2.7}$$

and the term D_0 as follows:

$$D_0 := \min_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*} \sum_{i=1}^{n} p_{\boldsymbol{\theta}^*}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2 \tag{2.8}$$

It then follows that

$$\begin{aligned}
\mathbb{E}[\tau_\delta] \leq \mathbb{E}[\tau_{\boldsymbol{\theta}^*}] &= \sum_{t=0}^{\infty} \mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) \overset{(a)}{\leq} 1 + (1+c)M + \sum_{t:t>(1+c)M} \mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) \\
&\overset{(b)}{\leq} 1 + (1+c)\left(\frac{\alpha(J)}{D_1} + \frac{C'(\eta_0,\eta) + 4\log(J/\delta)}{D_0}\right) + J \sum_{t:t>\left(\frac{\alpha(J)}{D_1}+\frac{\log(J/\delta)}{D_0}\right)(1+c)} C_1 \exp(-C_2 t) \\
&\overset{(c)}{\leq} 1 + (1+c)\left(\frac{\alpha(J)}{D_1} + \frac{C'(\eta_0,\eta) + 4\log(J/\delta)}{D_0}\right) + JC_1 \frac{\exp\left(-C_2(1+c)\left(\frac{\alpha(J)}{D_1} + \frac{4\log(J/\delta)}{D_0}\right)\right)}{1-\exp\left(-C_2\right)} \\
&\overset{(d)}{\leq} 1 + (1+c)\left(\frac{\alpha(J)}{D_1} + \frac{C'(\eta_0,\eta) + 4\log(J/\delta)}{D_0}\right) \\
&\quad + JC_1 \frac{\exp\left(-C_2(1+c)\left(\frac{\alpha(J)}{D_0} + \frac{4\log(J)}{D_0} + \frac{4\log(1/\delta)}{D_0}\right)\right)}{1-\exp\left(-C_2\right)} \\
&\overset{(e)}{\leq} 1 + (1+c)\left(\frac{\alpha(J)}{D_1} + \frac{C'(\eta_0,\eta) + \log(J/\delta)}{D_0}\right) \\
&\quad + JC_1 \frac{\exp\left(-C_2\left(\frac{\alpha(J)+4\log J}{D_0}\right)\exp\left(-4C_2\frac{\log(1/\delta)}{D_0}\right)\right)}{1-\exp\left(-C_2\right)}
\end{aligned}$$

$$
\begin{aligned}
&\overset{(f)}{=} 1+(1+c)\left(\frac{\ell\log J}{D_1}+\frac{C'(\eta_0,\eta)+\log(J/\delta)}{D_0}\right)\\
&\quad + JC_1\frac{\exp\left(-C_2\left(\frac{(\ell+4)\log J}{D_0}\right)\exp\left(-4C_2\frac{\log(1/\delta)}{D_0}\right)\right)}{1-\exp\left(-C_2\right)}\\
&\leq 1+(1+c)\left(\frac{\ell\log J}{D_1}+\frac{C'(\eta_0,\eta)+\log(J/\delta)}{D_0}\right)+JC_1\left(\frac{1}{J}\right)^{C_2(\ell+4)/D_0}\delta^{4C_2/D_0}\left(1+\max\{1,\frac{1}{C_2}\}\right)\\
&\leq 1+(1+c)\left(\frac{\ell\log J}{D_1}+\frac{C'(\eta_0,\eta)+\log(J/\delta)}{D_0}\right)+C_1(2+\frac{1}{C_2})J^{1-\frac{C_2(\ell+4)}{D_0}}\delta^{\frac{4C_2}{D_0}}\\
&\overset{(g)}{\leq} 1+2\left(\frac{\ell\log J}{D_1}+\frac{C'(\eta_0,\eta)+\log(J/\delta)}{D_0}\right)+\left(44+\frac{\eta^2}{D_1^2}\right)\left(2+\frac{\eta^2}{D_1^2}\right)J^{1-\frac{D_1^2(\ell+4)}{2\eta^2D_0}}\delta^{\frac{D_1^2}{\eta^2D_0}}\\
&\leq 1+2\left(\frac{\ell\log J}{D_1}+\frac{C'(\eta_0,\eta)+\log(J/\delta)}{D_0}\right)+\left(44+\frac{\eta^2}{D_1^2}\right)^2J^{1-\frac{D_1^2(\ell+4)}{2\eta^2D_0}}\delta^{\frac{D_0}{\eta^2}}\\
&\overset{(h)}{\leq} 1+\underbrace{2\left(\frac{\ell\log J}{D_1}+\frac{C'(\eta_0,\eta)+\log(J/\delta)}{D_0}\right)}_{\textbf{Term A}}+\underbrace{\left(44+\frac{\eta^2}{\eta_0^2}\right)^2J^{1-\frac{\eta_0^2(\ell+4)}{2\eta^3}}}_{\textbf{Term B}}\underbrace{\delta^{\frac{D_0}{\eta^2}}}_{\textbf{Term C}} \qquad (2.9)
\end{aligned}
$$

where, (a) follows from definition of M in (2.6), (b) follows from Lemma 7.5, $C_1 := 23 + 11\max\left\{1, \frac{\eta^2}{2D_1^2}\right\}$, $C_2 := \frac{2D_1^2\min\{(c/2-1)^2, c\}}{\eta^2}$, (c) follows by applying the geometric progression formula, (d) follows as $D_1 \leq D_0$. The inequality (e) follows as $c > 0$, (f) follows by setting $\alpha(J) = \ell\log J$ for some constant $\ell > 1$, (g) follows by setting $c = \frac{1}{2}$ in C_1 and C_2, and (h) follows as $D_1 \geq \eta_0$, and $D_0 \leq \eta$.

Now, note that in (2.9) the term C ≤ 1 as $\delta \in (0,1)$. Now for the Term B we need to find an ℓ such that Term B $\leq J^{1-\frac{\eta_0^2(\ell+4)}{2\eta^3}+\frac{\ell}{\eta\log J}}$. Hence,

$$
\begin{aligned}
&\left(44+\frac{\eta^2}{\eta_0^2}\right)^2J^{1-\frac{\eta_0^2(\ell+4)}{2\eta^3}} \leq J^{1-\frac{\eta_0^2(\ell+4)}{2\eta^3}+\frac{\ell}{\eta\log J}}\\
&\implies \left(44+\frac{\eta^2}{\eta_0^2}\right)^2 \leq J^{\frac{\ell}{\eta\log J}} \implies \eta\log\left(44+\frac{\eta^2}{\eta_0^2}\right) \leq \ell
\end{aligned}
$$

So for a constant $\ell > 1$ such that if ℓ satisfies the following condition

$$\ell = \eta \log\left(44 + \frac{\eta^2}{\eta_0^2}\right) > \log\left(1 + \frac{\eta^2}{\eta_0^2}\right) \tag{2.10}$$

then we have that Term B $\leq J^{1-\frac{\eta_0^2(\ell+4)}{2\eta^3}+\frac{\ell}{\eta \log J}}$. Plugging this in (2.9) we get that the expected sample complexity is upper bounded by

$$\begin{aligned}
\mathbb{E}[\tau_\delta] &\leq 1 + 2\left(\frac{\eta \log\left(44 + \frac{\eta^2}{\eta_0^2}\right)\log J}{D_1} + \frac{\log\left(1 + \frac{\eta^2}{\eta_0^2}\right) + \log(J/\delta)}{D_0}\right) + J^{1-\frac{\eta_0^2(\ell+4)}{2\eta^3}+\frac{\ell}{\eta \log J}}\frac{D_0}{\delta\,\eta^2} \\
&\overset{(a)}{\leq} 1 + 2\left(\frac{\eta \log\left(44 + \frac{8\eta^2}{\eta_0^2}\right)\log J}{D_1} + \frac{\log\left(\left(44 + \frac{\eta^2}{\eta_0^2}\right)\frac{J}{\delta}\right)}{D_0}\right) + J^{1+\frac{\log\left(44 + \frac{\eta^2}{\eta_0^2}\right)}{\eta \log J}}\frac{D_0}{\delta\,\eta^2} \\
&\overset{(b)}{\leq} 1 + 2\left(\frac{\eta \log(C(\eta_0,\eta))\log J}{D_1} + \frac{\log(C(\eta_0,\eta)J/\delta)}{D_0}\right) + J^{1+\frac{\log(C(\eta_0,\eta))}{\eta \log J}}\frac{D_0}{\delta\,\eta^2} \\
&= 1 + 2\left(\frac{\eta \log(C(\eta_0,\eta))\log J}{D_1} + \frac{\log(C(\eta_0,\eta)J/\delta)}{D_0}\right) + J \cdot J^{\frac{\log(C(\eta_0,\eta))^{1/\eta}}{\log J}}\frac{D_0}{\delta\,\eta^2} \\
&= O\left(\frac{\eta \log(C(\eta_0,\eta))\log J}{D_1} + \frac{\log(J/\delta)}{D_0} + J(C(\eta_0,\eta))^{1/\eta}\delta^{D_0/\eta^2}\right)
\end{aligned}$$

where, (a) follows as $2\log\left(44 + \frac{\eta^2}{\eta_0^2}\right) \geq \log\left(1 + \frac{\eta^2}{2\eta_0^2}\right)$, and in (b) we substitute $C(\eta_0,\eta) = \left(44 + \frac{\eta^2}{\eta_0^2}\right)$. The claim of the Theorem follows. □

2.2 Minimax lower bound

While Chernoff (1959) had shown the policy to be optimal as $\delta \to 0$, we demonstrate an environment where GC has optimal sample complexity for any fixed value of $\delta > 0$. Let $\Gamma = \sqrt{\eta}/2$.

The following table depicts the values for $\mu_1(\cdot), \mu_2(\cdot), \ldots, \mu_n(\cdot)$ under J different hypotheses:

$$\begin{array}{cccccc} \boldsymbol{\theta} & = & \boldsymbol{\theta}^* & \boldsymbol{\theta}_2 & \boldsymbol{\theta}_3 & \ldots & \boldsymbol{\theta}_J \\ \hline \mu_1(\boldsymbol{\theta}) & = & \Gamma & \Gamma - \frac{\Gamma}{J} & \Gamma - \frac{2\Gamma}{J} & \ldots & \Gamma - \frac{(J-1)\Gamma}{J} \\ \mu_2(\boldsymbol{\theta}) & = & \epsilon_{21} & \epsilon_{22} & \epsilon_{23} & \ldots & \epsilon_{2J} \\ & \vdots & & & \vdots & & \\ \mu_n(\boldsymbol{\theta}) & = & \epsilon_{n1} & \epsilon_{n2} & \epsilon_{n3} & \ldots & \epsilon_{nJ} \end{array} \tag{2.11}$$

In the above, each ϵ_{ij} is distinct and satisfies $\epsilon_{ij} < \Gamma/4J$. $\mu_1(\cdot)$ is such that the difference of means across any pair of hypotheses is at least Γ/J. Theorem 4 is proved below by a change of measure argument (see Theorem 1 in Huang et al. (2017)). Note that arm 1 is better than all others in discriminating between any pair of hypotheses, and any policy to identify $\boldsymbol{\theta}^*$ cannot do better than allocating all its samples to arm 1.

Theorem 4. ***(Lower Bound)*** *In the environment in* (2.11)*, any δ-PAC policy π that identifies $\boldsymbol{\theta}^*$ satisfies $\mathbb{E}[\tau_\delta] \geq \Omega\left(J^2\Gamma^{-2}\log(1/\delta)\right)$. Applying Theorem 2 to the same environment, the sample complexity of GC is $O\left(J^2\Gamma^{-2}\log(J/\delta)\right)$ which matches the lower bound upto constants.*

Proof. The proof follows the standard change of measure argument. We follow the proof technique in Theorem 1 of Huang et al. (2017). Let, Λ_1 be the set of alternate bandit models having a different optimal hypothesis than $\boldsymbol{\theta}^* = \boldsymbol{\theta}_1$ such that all bandit models having different optimal hypothesis than $\boldsymbol{\theta}_1$ such as $B_2, B_3, \ldots B_J$ are in Λ_1. Let τ_δ be a stopping time for any δ-PAC policy π. Let $Z_i(t)$ denote the number of times the action i has been sampled till round t. Let $\widehat{\boldsymbol{\theta}}(t)$ be the predicted optimal hypothesis at round τ_δ. We first consider the bandit model B_1. Define the event $\xi = \{\widehat{\boldsymbol{\theta}}(t) \neq \boldsymbol{\theta}^*\}$ as the error event in bandit model B_1. Let the event $\xi' = \{\widehat{\boldsymbol{\theta}}(t) \neq \boldsymbol{\theta}'^*\}$ be the corresponding error event in bandit model B_2. Note that $\xi^C \subset \xi'$. Now since π is δ-PAC policy we have $\mathbb{P}_{B_1,\pi}(\xi) \leq \delta$ and $\mathbb{P}_{B_2,\pi}(\xi^C) \leq \delta$. Hence we can show that,

$$
\begin{aligned}
2\delta \geq \mathbb{P}_{B_1,\pi}(\xi) + \mathbb{P}_{B_2,\pi}(\xi^c) &\overset{(a)}{\geq} \frac{1}{2}\exp\left(-\mathrm{KL}\left(P_{B_1,\pi}||P_{B_2,\pi}\right)\right) \\
\mathrm{KL}\left(P_{B_1,\pi}||P_{B_2,\pi}\right) &\geq \log\left(\frac{1}{4\delta}\right) \\
\sum_{i=1}^{n} \mathbb{E}_{B_1,\pi}[Z_i(\tau_\delta)] \cdot \left(\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta}'^*)\right)^2 &\overset{(b)}{\geq} \log\left(\frac{1}{4\delta}\right) \\
\left(\Gamma - \Gamma + \frac{\Gamma}{J}\right)^2 \mathbb{E}_{B_1,\pi}[Z_1(\tau_\delta)] + \sum_{i=2}^{n} (\epsilon_{i1} - \epsilon_{i2})^2 \mathbb{E}_{B_1,\pi}[Z_i(\tau_\delta)] &\overset{(c)}{\geq} \log\left(\frac{1}{4\delta}\right) \\
\left(\frac{1}{J}\right)^2 \Gamma^2 \mathbb{E}_{B_1,\pi}[Z_1(\tau_\delta)] + \sum_{i=2}^{n} (\epsilon_{i1} - \epsilon_{i2})^2 \mathbb{E}_{B_1,\pi}[Z_i(\tau_\delta)] &\geq \log\left(\frac{1}{4\delta}\right) \\
\left(\frac{1}{J}\right)^2 \Gamma^2 \mathbb{E}_{B_1,\pi}[Z_1(\tau_\delta)] + \sum_{i=2}^{n} \frac{\Gamma^2}{4J^2} \mathbb{E}_{B_1,\pi}[Z_i(\tau_\delta)] &\overset{(d)}{\geq} \log\left(\frac{1}{4\delta}\right) \qquad (2.12)
\end{aligned}
$$

where, (a) follows from Lemma 7.2, (b) follows from Lemma 7.1, (c) follows from the construction of the bandit environments, and (d) follows as $(\epsilon_{ij} - \epsilon_{ij'})^2 \leq \frac{\Gamma^2}{4J^2}$ for any i-th action and j-th hypothesis pair.

Now, we consider the alternate bandit model B_3. Again define the event $\xi = \{\widehat{\boldsymbol{\theta}}(t) \neq \boldsymbol{\theta}^*\}$ as the error event in bandit model B_1 and the event $\xi' = \{\widehat{\boldsymbol{\theta}}(t) \neq \boldsymbol{\theta}''^*\}$ be the corresponding error event in bandit model B_3. Note that $\xi^C \subset \xi'$. Now since π is δ-PAC policy we have $\mathbb{P}_{B_1,\pi}(\xi) \leq \delta$ and $\mathbb{P}_{B_3,\pi}(\xi^C) \leq \delta$. Following the same way as before we can show that,

$$
\left(\frac{2}{J}\right)^2 \Gamma^2 \mathbb{E}_{B_1,\pi}[Z_1(\tau_\delta)] + \sum_{i=2}^{n} \frac{\Gamma^2}{4J^2} \mathbb{E}_{B_1,\pi}[Z_i(\tau_\delta)] \overset{(d)}{\geq} \log\left(\frac{1}{4\delta}\right). \qquad (2.13)
$$

Similarly we get the equations for all the other $(J-2)$ alternate models in Λ_1. Now consider an optimization problem

$$\begin{aligned}
&\min_{x_i : i\in[n]} \sum x_i \quad s.t. \quad \left(\frac{1}{J}\right)^2 \Gamma^2 x_1 + \frac{\Gamma^2}{4J^2}\sum_{i=2}^{n} x_i \geq \log(1/4\delta) \\
&\left(\frac{2}{J}\right)^2 \Gamma^2 x_1 + \frac{\Gamma^2}{4J^2}\sum_{i=2}^{n} x_i \geq \log(1/4\delta) \\
&\vdots \\
&\left(\frac{J-1}{J}\right)^2 \Gamma^2 x_1 + \frac{\Gamma^2}{4J^2}\sum_{i=2}^{n} x_i \geq \log(1/4\delta) \\
&x_i \geq 0, \forall i \in [n]
\end{aligned}$$

where the optimization variables are x_i. It can be seen that the optimum objective value is $J^2\Gamma^{-2}\log(1/4\delta)$. Interpreting $x_i = \mathbb{E}_{B_1,\pi}[Z_i(\tau_\delta)]$ for all i, we get that $\mathbb{E}_{B_1,\pi}[\tau_\delta] = \sum_i x_i$ which gives us the required lower bound. Let $\mathbf{p}_{\boldsymbol{\theta}^*}$ be the sampling p.m.f for the environment B_1 for verifying $\boldsymbol{\theta}^*$. We also know that the Chernoff verification in (1.3) has a nice linear programming formulation as stated in (1.4). Using that for B_1 we can show that,

$$\begin{aligned}
&\max \ell \quad s.t. \quad \mathbf{p}_{\boldsymbol{\theta}^*}(1)\Gamma^2\left(\frac{1}{J}\right)^2 + \sum_{i=2}^{n} \mathbf{p}_{\boldsymbol{\theta}^*}(i)(\epsilon_{i\boldsymbol{\theta}^*} - \epsilon_{i\boldsymbol{\theta}'})^2 \geq \ell \\
&\mathbf{p}_{\boldsymbol{\theta}^*}(1)\Gamma^2\left(\frac{2}{J}\right)^2 + \sum_{i=2}^{n} \mathbf{p}_{\boldsymbol{\theta}^*}(i)(\epsilon_{i\boldsymbol{\theta}^*} - \epsilon_{i\boldsymbol{\theta}''})^2 \geq \ell \\
&\vdots \\
&\mathbf{p}_{\boldsymbol{\theta}^*}(1)\Gamma^2\left(\frac{J-1}{J}\right)^2 + \sum_{i=2}^{n} \mathbf{p}_{\boldsymbol{\theta}^*}(i)(\epsilon_{i\boldsymbol{\theta}^*} - \epsilon_{i\boldsymbol{\theta}'''})^2 \geq \ell \\
&\sum_{i=1}^{n} \mathbf{p}_{\boldsymbol{\theta}^*}(i) = 1.
\end{aligned}$$

We can relax this further by noting that $(\epsilon_{i\boldsymbol{\theta}^*} - \epsilon_{i\boldsymbol{\theta}'})^2 \leq \frac{\Gamma^2}{4J^2}$ for all $i \in [n]$ and $\boldsymbol{\theta}^*, \boldsymbol{\theta}' \in \boldsymbol{\Theta}$. Hence,

$$\begin{aligned}
&\max \ell \quad s.t. \quad \mathbf{p}_{\boldsymbol{\theta}^*}(1)\Gamma^2 \left(\frac{1}{J}\right)^2 + \sum_{i=2}^{n} \mathbf{p}_{\boldsymbol{\theta}^*}(i)\frac{\Gamma^2}{4J^2} \geq \ell \\
&\mathbf{p}_{\boldsymbol{\theta}^*}(1)\Gamma^2 \left(\frac{2}{J}\right)^2 + \sum_{i=2}^{n} \mathbf{p}_{\boldsymbol{\theta}^*}(i)\frac{\Gamma^2}{4J^2} \geq \ell \\
&\vdots \\
&\mathbf{p}_{\boldsymbol{\theta}^*}(1)\Gamma^2 \left(\frac{J-1}{J}\right)^2 + \sum_{i=2}^{n} \mathbf{p}_{\boldsymbol{\theta}^*}(i)\frac{\Gamma^2}{4J^2} \geq \ell \\
&\sum_{i=1}^{n} \mathbf{p}_{\boldsymbol{\theta}^*}(i) = 1.
\end{aligned}$$

Solving this above optimization gives that $\mathbf{p}_{\boldsymbol{\theta}^*}(1) = 1$, and $\mathbf{p}_{\boldsymbol{\theta}^*}(2) = \mathbf{p}_{\boldsymbol{\theta}^*}(3) = \ldots = \mathbf{p}_{\boldsymbol{\theta}^*}(n) = 0$. Similarly, for verifying any hypothesis $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ we can show that the verification proportion is given by $\mathbf{p}_{\boldsymbol{\theta}} = (1, \underbrace{0, 0, \ldots, 0}_{\text{(J-1) zeros}})$. This also shows that for example in eq. (2.11),

$$D_0 = \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^{n} p_{\boldsymbol{\theta}^*}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2 = p_{\boldsymbol{\theta}^*}(1)\Gamma^2 \left(\frac{1}{J}\right)^2 + \sum_{i=2}^{n} p_{\boldsymbol{\theta}^*}(i)(\epsilon_{i\boldsymbol{\theta}^*} - \epsilon_{i\boldsymbol{\theta}'})^2 = \frac{\Gamma^2}{J^2}$$

$$D_1 = \min_{\{\mathbf{p}_{\boldsymbol{\theta}}:\boldsymbol{\theta}\in\boldsymbol{\Theta}\}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^{n} p_{\boldsymbol{\theta}}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2 \overset{(a)}{\geq} p_{\boldsymbol{\theta}}(1)\Gamma^2 \left(\frac{1}{J}\right)^2 + \sum_{i=2}^{n} p_{\boldsymbol{\theta}}(i)(\epsilon_{i\boldsymbol{\theta}} - \epsilon_{i\boldsymbol{\theta}'})^2 = \frac{\Gamma^2}{J^2}$$

where, (a) follows as the verification of any hypothesis $\boldsymbol{\theta}$ is a one hot vector $p_{\boldsymbol{\theta}} = (1, \underbrace{0, 0, \ldots, 0}_{\text{(J-1) zeros}})$. Note that $\eta/4 = \Gamma^2$. Plugging this in Theorem 2 gives us that the upper bound of GC as

$$\begin{aligned}
\mathbb{E}[\tau_\delta] &\leq O\left(\frac{J^2\Gamma^2 \log(\Gamma^4/\eta_0^4)\Gamma \log J}{\Gamma^2} + \frac{J^2 \log(J/\delta)}{\Gamma^2} + J \log(\Gamma^4/\eta_0^4)^{1/\Gamma^2} \delta^{\Gamma^2/(J^2\Gamma^4)}\right) \\
&\leq O\left(J^2 \log(\Gamma^4/\eta_0^4)\Gamma \log J + \frac{J^2 \log(J/\delta)}{\Gamma^2}\right) \leq O\left(\frac{J^2 \log(J/\delta)}{\Gamma^2}\right).
\end{aligned}$$

The claim of the theorem follows. □

2.3 A Non-asymptotic Example

We now introduce an example where we show that the non-asymptotic term introduced in Theorem 2 can be significant.

Example 1. (Non-asymptotes matter) Consider an environment with two arms and $\boldsymbol{\Theta} = \{\boldsymbol{\theta}^*, \boldsymbol{\theta}', \boldsymbol{\theta}''\}$. The following table describes the values of $\mu_1(\cdot), \mu_2(\cdot)$ under these three hypotheses.

$\boldsymbol{\theta}$	$=$	$\boldsymbol{\theta}^*$	$\boldsymbol{\theta}'$	$\boldsymbol{\theta}''$
$\mu_1(\boldsymbol{\theta})$	$=$	1	0.001	0
$\mu_2(\boldsymbol{\theta})$	$=$	1	1.002	0.998

For a choice of $\delta = 0.1$, we can evaluate that $\log(J)/D_1 \approx 3 \times 10^5$ and $\log(J/\delta)/D_0 \approx 3.4$. While Theorem 2 is only an upper bound, empirically we do see that the non-asymptotic term dominates the sample complexity. The Figure 3.1a shows a box plot of the stopping times of four algorithms over 100 independent trials on the above environment. A box plot depicts a set of numerical data using their quartiles (Tukey, 1977). A uniform sampling baseline (Unif) performs much better than GC , as it samples arm 1 half the time in expectation, and arm 1 is the best choice to distinguish $\boldsymbol{\theta}^*$ from both $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}''$. TP also performs poorly compared to Unif and GC in this setting. The non-asymptotic term of TP scales as $\log(J)/D_{\text{NJ}} \approx 4 \times 10^6$.

2.4 Theoretical Comparison with Other Works

In this section we compare theoretically our work against other existing works in active testing. We compare our work against Chernoff (1959), Albert (1961) which extended Chernoff (1959) to continuous hypotheses space partitioned into two disjoint space, and the recent active testing algorithms of Naghshvar and Javidi (2013); Nitinawarat et al. (2013). We first recall the following problem complexity parameters as follows:

$$D_0 := \max_{\mathbf{p}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^{n} p(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2, \qquad D_1 := \min_{\{\mathbf{p}_{\boldsymbol{\theta}}: \boldsymbol{\theta} \in \boldsymbol{\Theta}\}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^{n} p_{\boldsymbol{\theta}}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$$

$$D_{\text{NJ}} := \left(\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}} \sum_{i=1}^{n} p_{\boldsymbol{\theta}'}(i)(\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}'))^2 \right)^2 \max_{\mathbf{p}} \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}} \sum_{i=1}^{n} p(i)(\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}'))^2 \tag{2.14}$$

$$D_e := \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^{n} \frac{1}{n} (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2 \tag{2.15}$$

where, the quantity D_{NJ} is defined in Naghshvar and Javidi (2013) and is an artefact of the forced exploration conducted by their algorithm. Note that $D_{NJ} < D_1$ by definition. Similarly, $D_e > D_1$

and $D_e > 0$ by definition. Note that due to the factor D_{NJ}, TP can perform worse than GC in certain instances (see Active testing experiment in Figure 3.1a, and Figure 3.1b). We summarize out result in context of other existing results in the following table:

Sample Complexity Bound	**Comments**
$\mathbb{E}[\tau_\delta] \leq \frac{C \log J/\delta}{D_0} + o\left(\log \frac{1}{\delta}\right)$	Upper bound in Chernoff (1959).Optimal for $\delta \to 0$.
$\liminf_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \geq \frac{1}{D_0}$	Lower Bound in (Chernoff, 1959). Optimal for $\delta \to 0$.
$\mathbb{E}[\tau_\delta] \leq \frac{C \log J/\delta}{D_0} + o\left(\log \frac{1}{\delta}\right)$	Upper bound in Albert (1961). Optimal for $\delta \to 0$. Extension to compound hypotheses.
$\mathbb{E}[\tau_\delta] \leq \frac{C \log J/\delta}{D_0} + o\left(\log \frac{1}{\delta}\right)$	Upper bound in Nitinawarat et al. (2013). Bound valid for discrete hypotheses. It is under weaker assumptions than Chernoff (1959).
$\mathbb{E}[\tau_\delta] \leq O\left(\frac{\log(J)}{D_{\text{NJ}}} + \frac{\log(J/\delta)}{D_0}\right)$	Upper bound of TP in Naghshvar and Javidi (2013). Valid for any $\delta \in (0, 1]$. Asymptotically optimal for $\delta \to 0$.
$\mathbb{E}[\tau_\delta] \leq O\left(\frac{\log(J)}{D_1} + \frac{\log(J/\delta)}{D_0}\right)$	GC (Ours). Valid for any $\delta \in (0, 1]$. Asymptotically optimal for $\delta \to 0$.
$\mathbb{E}[\tau_\delta] \leq O\left(\frac{\log(J)}{D_e} + \frac{\log(J/\delta)}{D_0}\right)$	GCE (Ours). Valid for any $\delta \in (0, 1]$. Asymptotically optimal for $\delta \to 0$. Does not require Assumption 2.

Table 2.1: Active Testing comparison. $D_{\text{NJ}} < D_1 < D_0$, and $D_e < D_0$.

In this chapter we introduced the GC policy and established its moderate confidence sample complexity bounds. In the next chapter we introduce the T2 sampling scheme and other important variations of the GC policy.

Chapter 3

Variations of Chernoff Sampling

In this chapter we introduce several variations of GC sampling. Note that evaluating the GC policy at every round can be costly. To minimize this computational load we introduce the two policies T2 and Batch-GC in this chapter. Further we show that employing a little bit exploration can lead to a drastic improvement in GC policy performance and so we introduce the exploring GCE policy in this chapter.

3.1 T2 Sampling

Instead of sampling according to the optimal verification proportion in line 4 of Algorithm 1, we can instead use the following heuristic argument. At each time, consider the current most likely $\widehat{\boldsymbol{\theta}}(t)$ and its "closest" competing hypothesis defined as $\tilde{\boldsymbol{\theta}}(t) := \arg\min_{\boldsymbol{\theta}\neq\widehat{\boldsymbol{\theta}}(t)} L_{Y^t}(\boldsymbol{\theta})$. The heuristic called Top-2 sampling (abbreviated as (T2)) suggests to sample an arm that best discriminates between them, i.e.,

$$I_{t+1} := \arg\max_{i\in[n]} (\mu_i(\widehat{\boldsymbol{\theta}}(t)) - \mu_i(\tilde{\boldsymbol{\theta}}(t)))^2. \tag{3.1}$$

This strategy requires lesser computation as we don't need to compute $\boldsymbol{p}_{\widehat{\boldsymbol{\theta}}(t)}$ which could be useful when J or n is very large. It was proposed by Chernoff (1959) (without any sample complexity proof), where he showed using an example that it is not optimal. The proof technique of Theorem 2 can be reused to obtain a novel moderate confidence sample complexity expression for T2 as follows:

Theorem 3. ***(T2 Sample Complexity)*** *Let τ_δ denote the stopping time of T2 following the sampling strategy of* (3.1)*. Consider the set $\mathcal{I}(\boldsymbol{\theta},\boldsymbol{\theta}') \subset [n]$ of arms that could be sampled following* (3.1) *when $\widehat{\boldsymbol{\theta}}(t) = \boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}(t) = \boldsymbol{\theta}'$, and let $\mathbf{u}_{\boldsymbol{\theta},\boldsymbol{\theta}'}$ denote a uniform p.m.f. supported on $\mathcal{I}(\boldsymbol{\theta},\boldsymbol{\theta}')$. Define*

$$D_0' := \min_{\boldsymbol{\theta},\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*} \sum_{i=1}^{n} u_{\boldsymbol{\theta}^*\boldsymbol{\theta}}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2, \quad D_1' := \min_{\boldsymbol{\theta}\neq\boldsymbol{\theta}',\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*} \sum_{i=1}^{n} u_{\boldsymbol{\theta}\boldsymbol{\theta}'}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2,$$

where we assume that $D_1' > 0$. The sample complexity of T2 has the following upper bound:

$$\mathbb{E}[\tau_\delta] \leq O\left(\frac{\eta \log(C(\eta_0,\eta)) \log J}{D_1'} + \frac{\log(J/\delta)}{D_0'} + JC(\eta_0,\eta)^{1/\eta}\delta^{D_0'/\eta^2}\right).$$

The bound has a similar form as in Theorem 2 with three terms. The first term does not scale with error probability δ, while the second term scales with $\log(J/\delta)$. The denominators of the two terms are different from D_0 and D_1 due to the different sampling rule of T2 and hence T2 is not asymptotically optimal. We provide the main proof below.

Proof. **Step 1 (Definitions):** Let the action $i_{\boldsymbol{\theta}\boldsymbol{\theta}'} := \arg\max_{i\in[n]}(\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}'))^2$. Let $\widehat{\boldsymbol{\theta}}(t)$ denote the most likely hypothesis at round s and $\tilde{\boldsymbol{\theta}}(t)$ be the second most likely hypothesis at round s. Note that T2 only samples the action $i_{\boldsymbol{\theta}\boldsymbol{\theta}'}$ at round s when $\widehat{\boldsymbol{\theta}}(s) = \boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}(s) = \boldsymbol{\theta}'$. Again, let $L_{Y^t}(\boldsymbol{\theta})$ denote the total sum of squared errors of hypothesis $\boldsymbol{\theta}$ till round t. We further define the set $\mathcal{I} := \{i \in [n] : i = \arg\max_{i'\in[n]}(\mu_{i'}(\boldsymbol{\theta}) - \mu_{i'}(\boldsymbol{\theta}'))^2 \text{ for some } \boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}\}$.

Step 2 (Define stopping time): We define the stopping time τ_δ for the policy T2 as follows:

$$\tau_\delta := \min\{t : \exists \boldsymbol{\theta} \in \boldsymbol{\Theta}, L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}) > \beta(J,\delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}\} \tag{3.2}$$

where, $\beta(J,\delta)$ is the threshold function defined in (7.7).

Step 3 (Define bad event): We define the bad event $\xi^\delta(\boldsymbol{\theta})$ for the sub-optimal hypothesis $\boldsymbol{\theta}$ as follows:

$$\xi^\delta(\boldsymbol{\theta}) := \{L_{Y^{\tau_\delta}}(\boldsymbol{\theta}') - L_{Y^{\tau_\delta}}(\boldsymbol{\theta}) > \beta(J,\delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}\}. \tag{3.3}$$

The event $\xi^\delta(\boldsymbol{\theta})$ denotes that a sub-optimal hypothesis $\boldsymbol{\theta}$ has been declared optimal at stopping time $\boldsymbol{\theta}$ and its sum of squared error is smaller than any other hypothesis $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ at τ_δ.

Step 4 (Decomposition of bad event): Decomposing the bad event follows the same approach in Theorem 2. A crucial thing to note is that the stopping time only depends on the threshold function $\beta(J,\delta)$ and not on the sampling rule. Again we can decompose the bad event to show that only comparing $\boldsymbol{\theta}$ against $\boldsymbol{\theta}^*$ is enough to guarantee a δ-PAC policy. Finally following (2.4) we can decompose the bad event $\xi^\delta(\boldsymbol{\theta})$ as follows:

$$\xi^\delta(\boldsymbol{\theta}) \subseteq \{\widehat{\boldsymbol{\theta}}(\tau_\delta) = \boldsymbol{\theta}, L_{Y^{\tau_\delta}}(\boldsymbol{\theta}^*) - L_{Y^{\tau_\delta}}(\boldsymbol{\theta}) > \beta(J,\delta)\} \tag{3.4}$$

such that we compare the sub-optimal hypothesis $\boldsymbol{\theta}$ only with optimal hypothesis $\boldsymbol{\theta}^*$.

Step 5 (Control bad event): The control of the bad event follows the same approach in Theorem 2. We want to show that based on the stopping time definition and the bad event $\xi^{\delta}(\boldsymbol{\theta})$ the T2 stops and outputs the correct hypothesis $\boldsymbol{\theta}^*$ with $1-\delta$ probability. As shown in Step 4, we can define the error event $\xi^{\delta}(\boldsymbol{\theta})$ as follows:

$$\xi^{\delta}(\boldsymbol{\theta}) \subseteq \{\widehat{\boldsymbol{\theta}}(\tau_\delta) = \boldsymbol{\theta}, L_{Y^{\tau_\delta}}(\boldsymbol{\theta}^*) - L_{Y^{\tau_\delta}}(\boldsymbol{\theta}) > \beta(J,\delta)\}$$

Again define $\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} := \min\{t : T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > \beta(J,\delta)\}$. Then following the same steps as in Step 5 of Theorem 2 we can show that

$$\begin{aligned}\mathbb{P}\left(\tau_\delta < \infty, \widehat{\boldsymbol{\theta}}(\tau_\delta) \neq \boldsymbol{\theta}^*\right) &\leq \mathbb{P}\left(\exists \boldsymbol{\theta} \in \boldsymbol{\Theta} \setminus \{\boldsymbol{\theta}^*\}, \exists t \in \mathbb{N} : T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > \beta(J,\delta)\right) \\ &\leq \sum_{\boldsymbol{\theta} \neq \boldsymbol{\theta}^*} \mathbb{P}\left(L_{Y^{\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}}}}(\boldsymbol{\theta}^*) - L_{Y^{\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}}}}(\boldsymbol{\theta}) > \beta(J,\delta), \tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} < \infty\right) \overset{(a)}{\leq} \sum_{\boldsymbol{\theta} \neq \boldsymbol{\theta}^*} \frac{\delta}{J} \leq \delta\end{aligned}$$

where (a) follows follows from Lemma 7.10 and the definition of $\beta(J,\delta)$.

Step 6 (Sample complexity analysis): In this step we bound the total sample complexity of T2 satisfying the δ-PAC criteria. Recall that the set $\mathcal{I} := \{i \in [n] : i = \arg\max_{i' \in [n]}(\mu_{i'}(\boldsymbol{\theta}) - \mu_{i'}(\boldsymbol{\theta}'))^2$ for some $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}\}$. Note that T2 does not sample by the Chernoff p.m.f. $\mathbf{p}_{\boldsymbol{\theta}}$. Rather it samples by the p.m.f.

$$u_{\boldsymbol{\theta}\boldsymbol{\theta}'} := \frac{1}{|\mathcal{I}(\boldsymbol{\theta}\boldsymbol{\theta}')|} \tag{3.5}$$

where, $\mathcal{I}(\boldsymbol{\theta}\boldsymbol{\theta}') := \{i \in \mathcal{I} : i = \arg\max_{i' \in [n]}(\mu_{i'}(\boldsymbol{\theta}) - \mu_{i'}(\boldsymbol{\theta}'))^2$ for $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}\}$. Hence $u_{\boldsymbol{\theta}\boldsymbol{\theta}'}$ is a uniform random p.m.f between all the maximum mean squared difference actions between hypotheses $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ which are $\widehat{\boldsymbol{\theta}}(t)$ and $\tilde{\boldsymbol{\theta}}(t)$ respectively for some rounds $s \in [\tau_\delta]$. The rest of the analysis follows the same steps as in Step 6 of Theorem 2 as the proof does not rely on any specific type of sampling proportion. We define the stopping time τ_δ as follows:

$$\tau_\delta = \min\left\{t : L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\widehat{\boldsymbol{\theta}}(t)) \geq \beta(J,\delta), \forall \boldsymbol{\theta}' \neq \widehat{\boldsymbol{\theta}}(t)\right\}.$$

We further define the stopping time $\tau_{\boldsymbol{\theta}^*}$ for the hypothesis $\boldsymbol{\theta}^*$ as follows:

$$\tau_{\boldsymbol{\theta}^*} := \min\left\{t : L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*) \geq \beta, \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*\right\} \tag{3.6}$$

We also define the critical number of samples as $(1+c)M$ where M is defined as follows:

$$M := \left(\frac{\alpha(J)}{D_1'} + \frac{C'(\eta_0, \eta) + \log(J/\delta)}{D_0'} \right) \tag{3.7}$$

where, $C'(\eta_0, \eta) = \log(1 + \eta^2/\eta_0^2)$, $c > 0$ is a constant, and we define the term D_1' as follows:

$$D_1' := \min_{\boldsymbol{\theta} \neq \boldsymbol{\theta}', \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^{n} u_{\boldsymbol{\theta}\boldsymbol{\theta}'}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$$

and the term D_0' as follows:

$$D_0' := \min_{\boldsymbol{\theta}, \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^{n} u_{\boldsymbol{\theta}^*\boldsymbol{\theta}}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2.$$

It then follows that

$$\begin{aligned}
\mathbb{E}[\tau_\delta] \leq \mathbb{E}[\tau_{\boldsymbol{\theta}^*}] &= \sum_{t=0}^{\infty} \mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) \overset{(a)}{\leq} 1 + (1+c)M + \sum_{t: t > (1+c)M} \mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) \\
&\overset{(b)}{\leq} 1 + (1+c)\left(\frac{\ell \log J}{D_1'} + \frac{C'(\eta_0, \eta) + \log(J/\delta)}{D_0'} \right) + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t: t > \left(\frac{\alpha(J)}{D_1'} + \frac{4\log(J/\delta)}{D_0'} \right)(1+c)} C_1 \exp(-C_2 t) \\
&\leq 1 + (1+c)\left(\frac{\ell \log J}{D_1'} + \frac{C'(\eta_0, \eta) + \log(J/\delta)}{D_0'} \right) + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} C_1 \frac{\exp\left(-C_2 \left(\frac{\alpha(J)}{D_1'} + \frac{4\log(J/\delta)}{D_0'} \right) \right)}{1 - \exp(-C_2)} \\
&\overset{(c)}{\leq} 1 + \underbrace{2\left(\frac{\ell \log J}{D_1'} + \frac{C'(\eta_0, \eta) + \log(J/\delta)}{D_0'} \right)}_{\textbf{Term A}} + \underbrace{\left(44 + \frac{2\eta^2}{\eta_0^2} \right)^2 J^{1 - \frac{\eta_0^2(\ell+4)}{2\eta^3}} \underbrace{\delta \frac{D_0'}{\eta^2}}_{\textbf{Term C}}}_{\textbf{Term B}}
\end{aligned} \tag{3.8}$$

where, (a) follows from definition of M in (3.7), (b) follows from Lemma 7.11 where $C_1 := 23 + 11 \max\left\{ 1, \frac{\eta^2}{2D_1^2} \right\}$, $C_2 := \frac{2D_1^2 \min\{(\frac{c}{2} - 1)^2, c\}}{\eta^2}$, and (c) follows the same steps in Theorem 2 by setting $c = \frac{1}{2}$ in C_1 and C_2.

Again, note that in (3.8) the term C ≤ 1 as $\delta \in (0, 1)$. So for a constant $\ell > 1$ such that if ℓ satisfies the following condition

$$\ell = \eta \log\left(44 + \frac{\eta^2}{\eta_0^2} \right) > \log\left(1 + \frac{\eta^2}{\eta_0^2} \right) \tag{3.9}$$

we have that Term B $\leq J^{1-\frac{\eta_0^2(\ell+4)}{2\eta^3}+\frac{\ell}{\eta \log J}}$. Hence, Plugging this in (3.8) we get that the expected sample complexity is of the order of

$$\mathbb{E}[\tau_\delta] \leq O\left(\frac{\eta \log(C(\eta_0,\eta)) \log J}{D_1'} + \frac{\log(J/\delta)}{D_0'} + JC(\eta_0,\eta)^{1/\eta}\delta^{D_0'/\eta^2}\right).$$

where, $C(\eta_0,\eta) = 44 + \frac{\eta^2}{\eta_0^2}$. The claim of the theorem follows. □

3.2 GC with Batch Updates

We can solve $\max\min$ optimization for $\mathbf{p}_{\widehat{\boldsymbol{\theta}}}(t)$ every B rounds instead of at each round. This reduces the computational load while increasing the sample complexity by only an additive term as shown below. We state the proposition below and provide the corresponding proof. The result show that the sample complexity scales with additive factor of B instead of multiplicative.

Proposition 1. *Let τ_δ, D_0, and D_1 be defined as in Theorem 2. Let B be the batch size. Then the sample complexity bound of δ-PAC Batch-GC is given by*

$$\mathbb{E}[\tau_\delta] \leq O\left(B + \frac{\eta \log(C) \log J}{D_1} + \frac{\log(J/\delta)}{D_0} + BJC^{\frac{1}{\eta}}\delta^{\frac{D_0}{\eta^2}}\right).$$

Proof. We follow the same proof technique as in Theorem 2. We define the last phase after which the algorithm stops as m_δ defined as follows:

$$m_\delta = \min\{m : L_{Y^{mB}}(\boldsymbol{\theta}') - L_{Y^{mB}}(\widehat{\boldsymbol{\theta}}(t)) > \beta(J,\delta), \forall \boldsymbol{\theta}' \neq \widehat{\boldsymbol{\theta}}(t)\}.$$

We further define the phase $m_{\boldsymbol{\theta}^*}$ as follows:

$$m_{\boldsymbol{\theta}^*} = \min\{m : L_{Y^{mB}}(\boldsymbol{\theta}') - L_{Y^{mB}}(\widehat{\boldsymbol{\theta}}(t)) > \beta(J,\delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*\}.$$

Then we can show that the expected stopping phase m_δ is bounded as follows:

$$\mathbb{E}[m_\delta] \leq \mathbb{E}[m_{\boldsymbol{\theta}^*}] = \sum_{m=1}^{\infty} \mathbb{P}(m_{\boldsymbol{\theta}^*} > m) \leq 1 + \underbrace{(1+c)M_1}_{\textbf{Part A}} + \sum_{m':m'>(1+c)M_1} \underbrace{\mathbb{P}(m_{\boldsymbol{\theta}^*} > m')}_{\textbf{Part B}} \quad (3.10)$$

where, in Part A we define the critical number of phases

$$M_1 = \frac{\alpha(J)}{BD_1} + \frac{\beta(J,\delta)}{BD_0}.$$

and D_0, D_1 as defined in Theorem 2. Note that this definition of M_1 is different that the critical number of samples defined in Theorem 2. Now we control the Part B. As like Lemma 7.5 we define the following bad events

$$\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(m) = \{L_{Y^{mB}}(\boldsymbol{\theta}') - L_{Y^{mB}}(\boldsymbol{\theta}^*) < \beta(J,\delta)\}$$
$$\tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(m) = \{L_{Y^{mB}}(\boldsymbol{\theta}') - L_{Y^{mB}}(\boldsymbol{\theta}^*) < \alpha(J)\}$$

We further define the good stopping phase $\tilde{m}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}$ as follows:

$$\tilde{m}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*} = \min\{m : L_{Y^{m'B}}(\boldsymbol{\theta}') - L_{Y^{m'B}}(\boldsymbol{\theta}^*) > \alpha(J), \forall m' > m\}$$
$$\text{and } \tilde{m}_{\boldsymbol{\theta}^*} = \max_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \{\tilde{m}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}\}$$

denote the last phase after which in all subsequent phases we have $\boldsymbol{\theta}^*$ is $\widehat{\boldsymbol{\theta}}(t)$. We further define $\tilde{m}_{\boldsymbol{\theta}^*}$ as

$$\tilde{m}_{\boldsymbol{\theta}^*} = \frac{\alpha(J)}{BD_1} + \frac{mc}{2}.$$

Note that this definition of $\tilde{m}_{\boldsymbol{\theta}^*}$ is different that $\tilde{\tau}_{\boldsymbol{\theta}^*}$ in Lemma 7.5. Using Lemma 7.6 we can further show the event $\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}$ is bounded as follows:

$$\begin{aligned}
\mathbb{P}(\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(m)) &= \mathbb{P}\left(\underbrace{L_{Y^{mB}}(\boldsymbol{\theta}') - L_{Y^{mB}}(\boldsymbol{\theta}^*)}_{:=T_{Y^{mB}}(\boldsymbol{\theta}',\boldsymbol{\theta}^*)} < \beta(J,\delta)\right) \\
&\leq \mathbb{P}\left(T_{Y^{mB}}(\boldsymbol{\theta}',\boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^{mB}}(\boldsymbol{\theta}',\boldsymbol{\theta}^*)] < D_0\left(\frac{\beta(J,\delta)}{D_0} - B(m - \tilde{m}_{\boldsymbol{\theta}^*})\right)\right) \\
&\overset{(a)}{\leq} \exp(4B) \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1^2 Bm\left(\frac{c}{2} - 1\right)^2}{\eta^2}\right) \qquad (3.11)
\end{aligned}$$

where, (a) follows usinf the same steps as in Lemma 7.6. Similarly we can show that,

$$\mathbb{P}\left(\tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(m)\right) = \sum_{\boldsymbol{\theta}\neq\boldsymbol{\theta}'} \sum_{m':m'>\frac{\alpha(J)}{BD_1}+\frac{mc}{2}} \mathbb{P}\left(T_{Y^{m'B}} - \mathbb{E}[T_{Y^{m'B}}] < \alpha(J) - \mathbb{E}[T_{Y^{m'B}}]\right)$$

$$\overset{(a)}{\leq} \sum_{\boldsymbol{\theta}\neq\boldsymbol{\theta}'} \sum_{m':m'>\frac{\alpha(J)}{BD_1}+\frac{mc}{2}} \mathbb{P}\left(T_{Y^{m'B}} - \mathbb{E}[T_{Y^{m'B}}] < \alpha(J) - m'BD_1\right) \overset{(b)}{\leq} \exp(4) \sum_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*} \frac{\exp\left(-\frac{2D_1^2mcB}{2}\right)}{1-\exp\left(-\frac{2D_1^2cB}{2}\right)} \tag{3.12}$$

where, (a) follows as $\mathbb{E}[T_{Y^{m'B}}] \geq m'BD_1$ for all $m' > \frac{\alpha(J)}{BD_1} + \frac{mc}{2}$, and (b) follows using the same steps as in Lemma 7.7.

Finally using eq. (3.11) and eq. (3.12) we can show that the Part B is bounded as follows:

$$\begin{aligned}
\mathbb{P}(m_{\boldsymbol{\theta}^*} > m) &\leq \mathbb{P}\left(\left\{\bigcup_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*} \xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(m)\right\} \bigcap \left\{\tilde{m}_{\boldsymbol{\theta}^*} < \frac{\alpha}{BD_1} + \frac{mc}{2}\right\}\right) + \sum_{\boldsymbol{\theta}\neq\boldsymbol{\theta}'} \sum_{m':m'>\frac{\alpha(J)}{BD_1}+\frac{mc}{2}} \mathbb{P}\left(\tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(m')\right) \\
&\overset{(a)}{\leq} \sum_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*} \underbrace{\left[23 + 11\max\{1, \frac{\eta^2}{2D_1^2B}\}\right]}_{:=C_1} \exp\left(-\underbrace{\frac{2D_1^2B\min\{(c/2-1)^2, c\}}{\eta^2}}_{:=C_2} m\right) \\
&\leq JC_1\exp(-C_2m)
\end{aligned}$$

where, (a) follows from the same steps as in Lemma 7.5, and using eq. (3.11) and eq. (3.12). Plugging this back in eq. (3.10) we get that

$$\begin{aligned}
\mathbb{E}[m_\delta] &\leq 1 + (1+c)\left(\frac{\alpha(J)}{BD_1} + \frac{\beta(J,\delta)}{BD_0}\right) + J \sum_{m:m>\frac{\alpha(J)}{BD_1}+\frac{\beta(J,\delta)}{BD_0}} C_1\exp(-C_2m) \\
&\overset{(b)}{\leq} 1 + 2\left(\frac{\eta\log\left(44+\eta^2/B\eta_0^2\right)\log J}{BD_1} + \frac{\log\left(1+\eta^2/\eta_0^2B\right) + \log\left(\frac{J}{\delta}\right)}{BD_0}\right) + J^{1+\frac{\log\left(44+\eta^2/B\eta_0^2\right)}{\log J}}\delta^{D_0/\eta^2} \\
&\overset{(b)}{=} O\left(1 + \frac{\eta\log C\log J}{BD_1} + \frac{\log(J/\delta)}{BD_0} + JC^{1/\delta}\delta^{D_0/\eta^2}\right)
\end{aligned}$$

where, (a) follows using the same steps as in Step 6 of Theorem 2, and in (b) we substitute $C := \log\left(44 + \eta^2/B\eta_0^2\right)$. Finally, the expected total number of samples is given by

$$\mathbb{E}[\tau_\delta] \leq B\mathbb{E}[m_\delta] = O\left(B + \frac{\eta \log C \log J}{D_1} + \frac{\log(J/\delta)}{D_0} + BJC^{1/\delta}\delta^{D_0/\eta^2}\right).$$

The claim of the proposition follows. □

3.3 GC with Exploration

Recall that $D_1 > 0$ (assumption 2) is required to prove Theorem 2. We now relax this assumption with the policy GCE which at every round t follows the proportion $p_{\widehat{\boldsymbol{\theta}}(t)}$ with probability $1 - \epsilon_t$ and uniform randomly explores any other arm $i \in [n]$ with probability ϵ_t. Note that the exploration parameter ϵ_t reduces with time.

Define $D_e := \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n \frac{1}{n}(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$ as the objective value of uniform sampling optimization, and note that $D_e > 0$ by definition.

Proposition 2. *Let τ_δ, D_0, $C(\eta_0, \eta)$ be defined as in Theorem 2 and D_e be defined as above and $\epsilon_t := 1/\sqrt{t}$. Then the sample complexity bound of δ-PAC GCE with ϵ_t exploration is given by*

$$\mathbb{E}[\tau_\delta] \leq O\left(\frac{\eta \log(C(\eta_0, \eta)) \log J}{D_e} + \frac{\log(J/\delta)}{D_0} + J(C(\eta_0, \eta))^{1/\eta}\delta^{D_0/\eta^2}\right).$$

From the result above we can see that the non-asymptotic term does not depend on D_1 and scales with D_e. Note that $D_0 > D_e > D_1$ by definition. When $\delta \to 0$ then the asymptotic term dominates and so GCE is asymptotically optimal. In Example 1 we can calculate that $\log(J)/D_e \approx 2.1$ and $\log(J/\delta)/D_0 = 3.4$. So GCE performs similar to Unif (see Figure 3.1a) as well as theoretically enjoy asymptotic and moderate confidence guarantee similar to GC . Nest we provide the proof of Proposition 2.

Proof. Recall that

$$D_1 := \min_{\{\mathbf{p}_{\boldsymbol{\theta}}: \boldsymbol{\theta} \in \boldsymbol{\Theta}\}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n p_{\boldsymbol{\theta}}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2.$$

Define $D_e := \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n \frac{1}{n}(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$ as the objective value of uniform sampling optimization. Finally define the quantity at round s as

$$D_1^{\epsilon_s} := (1 - \epsilon_s)D_1 + \epsilon_s D_e \tag{3.13}$$

Let $T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) := L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*)$. Note that following Assumption 1 we can show that

$$T_{Y_s}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - D_1^{\epsilon_s} \overset{(a)}{\leq} 4\eta, \quad T_{Y_s}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - D_0 \leq 4\eta$$

where, (a) follows as $D_1 \leq \eta$ and $D_e \leq \eta$ which implies $(1 - \epsilon)D_1 + \epsilon D_e \leq \eta$ as $\epsilon \in (0, 1)$. Now define the quantity

$$\alpha(J) := \frac{C(\eta_0, \eta) \log J}{D_e},$$
$$\beta(J, \delta) := \frac{C(\eta_0, \eta) \log(J/\delta)}{D_0}$$

where, the constant $C(\eta_0, \eta) := \log\left(44 + {}^{\eta^2}/_{\eta_0^2}\right)$. Now define the failure events

$$\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t) := \{L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*) < \beta(J, \delta)\},$$
$$\tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t) := \{L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*) < 2\alpha(J)\}$$

where, for a constant $C(\eta_0, \eta) := \log\left(44 + {}^{\eta^2}/_{\eta_0^2}\right)$ we define $\alpha(J) := \frac{C(\eta_0, \eta) \log J}{D_1^e}$, and $\beta(J, \delta) = \frac{C(\eta_0, \eta) \log(J/\delta)}{D_0}$. Then we define the stopping times as follows:

$$\tau_{\boldsymbol{\theta}^*} := \min\{t : L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*\}$$
$$\tilde{\tau}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*} := \min\{t : L_{Y^{t'}}(\boldsymbol{\theta}') - L_{Y^{t'}}(\boldsymbol{\theta}^*) > 2\alpha(J), \forall t' > t\}$$
$$\tilde{\tau}_{\boldsymbol{\theta}^*} := \max_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \{\tilde{\tau}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}\}.$$

Then we have that Assumption 2 is no longer required which can be shown as follows

$$\begin{aligned}
\mathbb{E}_{I^t, Y^t}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] = \mathbb{E}_{I^t, Y^t}[L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*)] &= \mathbb{E}_{I^t, Y^t}\left[\sum_{s=1}^t (Y_s - \mu_{I_s}(\boldsymbol{\theta}^*))^2 - \sum_{s=1}^t (Y_s - \mu_{I_s}(\boldsymbol{\theta}))^2\right] \\
&= \sum_{s=1}^t \mathbb{E}_{I_s} \mathbb{E}_{Y_s | I_s}\left[(\mu_{I_s}(\boldsymbol{\theta}^*) - \mu_{I_s}(\boldsymbol{\theta}))^2 | I_s\right] \\
&= \sum_{s=1}^t \sum_{i=1}^n \mathbb{P}(I_s = i)\left(\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta})\right)^2
\end{aligned}$$

In the case when $D_1 \geq 0$, we can show that

$$\begin{aligned}
\mathbb{E}_{I^t,Y^t}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] &\overset{(a)}{\geq} \sum_{s=1}^{t} ((1-\epsilon_s)D_1 + \epsilon_s D_e) \\
&= \sum_{s=1}^{\tilde{\tau}_{\boldsymbol{\theta}^*}} (1-\epsilon_s)D_1 + \sum_{s=1}^{\tilde{\tau}_{\boldsymbol{\theta}^*}} \epsilon_s D_e + \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}+1}^{t} (1-\epsilon_s)D_0 + \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}+1}^{t} \epsilon_s D_e \\
&\overset{(b)}{\geq} \sum_{s=1}^{\tilde{\tau}_{\boldsymbol{\theta}^*}} \epsilon_s D_e + \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}+1}^{t} D_e - \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}+1}^{t} \epsilon_s D_0 + \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}+1}^{t} \epsilon_s D_e
\end{aligned}$$

$$\begin{aligned}
&\overset{(c)}{\geq} 2\sqrt{\tilde{\tau}_{\boldsymbol{\theta}^*}} D_e - 2D_e + (t - \tilde{\tau}_{\boldsymbol{\theta}^*} - 1)D_e - kD_0 \\
&= (t-1)D_e + 2\sqrt{\tilde{\tau}_{\boldsymbol{\theta}^*}} D_e - \tilde{\tau}_{\boldsymbol{\theta}^*} D_e - (kD_0 + 2D_e) \\
&\overset{(d)}{\geq} (t-1)D_e - \left(\frac{\alpha(J)}{D_e} + \frac{tc}{2}\right) D_e \\
&= (t-1)D_e - \alpha(J) - \frac{tc}{2} D_e \\
&= \left(t - 1 - \frac{tc}{2}\right) D_e - \alpha(J) = tD_e \left(1 - \frac{1}{t} - \frac{c}{2}\right) - \alpha(J) \\
&\overset{(e)}{\geq} tD_e \underbrace{\left(\frac{1}{2} - \frac{c}{2}\right)}_{c_1} - \alpha(J) \\
&= c_1 t D_e - \alpha(J)
\end{aligned}$$

where, (a) follows due to the forced exploration definition, (b) follows by dropping D_1, (c) follows $\epsilon_s > \frac{1}{\sqrt{s}}$ and $\int_1^t \frac{1}{\sqrt{s}} ds = 2\sqrt{s} - 2$ and $D_0 \geq D_e$, and (d) follows by definition of $\tilde{\tau}_{\boldsymbol{\theta}^*} = \frac{\alpha(J)}{D_e} + \frac{tc}{2}$ and trivially assuming that $2\sqrt{\tilde{\tau}_{\boldsymbol{\theta}^*}} D_e - (kD_0 + 2D_e) > 0$ for large enough $\tilde{\tau}_{\boldsymbol{\theta}^*}$, and (e) follows as

$t \geq 2$. Next we can show that for $t \geq \tilde{\tau}_{\boldsymbol{\theta}^*}$

$$\begin{aligned}
\mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] &\overset{(a)}{=} \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}}^{t} (1-\epsilon_s) D_0 + \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}}^{t} \epsilon_s D_e \overset{(b)}{\geq} \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}}^{t} (1-\epsilon_s) D_0 + \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}}^{t} \epsilon_s \frac{D_0}{n} \\
&= (t - \tilde{\tau}_{\boldsymbol{\theta}^*}) D_0 - \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}}^{t} \epsilon_s D_0 + \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}}^{t} \epsilon_s \frac{D_0}{n} \\
&= (t - \tilde{\tau}_{\boldsymbol{\theta}^*}) D_0 - D_0 \frac{n-1}{n} \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}}^{t} \epsilon_s \\
&\geq (t - \tilde{\tau}_{\boldsymbol{\theta}^*}) D_0 - k D_0 \\
&= (t - \tilde{\tau}_{\boldsymbol{\theta}^*} - k) D_0
\end{aligned}$$

where, (a) follows from the definition of exploration, and (b) follows from as $D_0 \leq nD_e$, It follows that

$$\begin{aligned}
\mathbb{P}(\xi_{\boldsymbol{\theta}', \boldsymbol{\theta}^*}(t)) &= \mathbb{P}(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) < \beta(J, \delta)) \\
&= \mathbb{P}(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < \beta(J, \delta) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)]) \\
&\overset{(a)}{\leq} \mathbb{P}\left(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < \beta(J, \delta) - (t - \tilde{\tau}_{\boldsymbol{\theta}^*} - k) D_0\right) \\
&\overset{(b)}{=} \mathbb{P}\left(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < D_0 \left(\frac{\beta(J, \delta)}{D_0} - t + \tilde{\tau}_{\boldsymbol{\theta}^*} + k\right)\right)
\end{aligned}$$

Once we define the failure events and stopping times we can follow the same proof technique as Theorem 2 and show that

$$\mathbb{P}(\tilde{\xi}_{\boldsymbol{\theta}' \boldsymbol{\theta}^*}(t)) \overset{(a)}{\leq} \mathbb{P}(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < 2\alpha(J) - c_1 t D_e)$$

where, in (a) the choice of $c_1 t D_1$ follows as $\mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] \geq c_1 t D_e$ as $D_e > 0$ and noting that $|T_{Y_s}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - D_e| \leq 4\eta$ by (7.4) for any round $s \in [t]$. It follows from Lemma 7.5 since $D_0 \geq D_e$

and $D_e > 0$ that the probability of the failure event is bounded as follows:

$$\mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) \leq \mathbb{P}(\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t))$$
$$= \mathbb{P}\left(\left\{\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\right\} \bigcap \{\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_e} + \frac{tc}{2}\}\right) + \mathbb{P}\left(\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \{\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\} \bigcap \{\tilde{\tau}_{\boldsymbol{\theta}^*} \geq \frac{\alpha(J)}{D_e} + \frac{tc}{2}\}\right)$$
$$\leq \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(\{\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\} \bigcap \{\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_e} + \frac{tc}{2}\}\right) + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t': t' \geq \frac{\alpha(J)}{D_e} + \frac{tc}{2}} \mathbb{P}\left(\tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t')\right) \overset{(a)}{\leq} JC_1 \exp\left(-C_2 t\right)$$

where, in (a) we substitute $C_1 := 23 + 11 \max\left\{1, \frac{\eta^2}{2(D_e)^2}\right\}$, $C_2 := \frac{2(D_e)^2 \min\{(c/2 - 1)^2, c\}}{\eta^2}$, $c > 0$ is a constant, and $J := |\boldsymbol{\Theta}|$. Now we define the critical number of samples as follows:

$$M := \left(\frac{C(\eta_0, \eta) \log J}{D_e} + \frac{C(\eta_0, \eta) + \log(J/\delta)}{D_0}\right)$$

where, $C(\eta_0, \eta) = \log\left(44 + \eta^2/\eta_0^2\right)$. Then we can bound the sample complexity for some constant $c > 0$ as follows:

$$\mathbb{E}[\tau_\delta] \leq \mathbb{E}[\tau_{\boldsymbol{\theta}^*}] = \sum_{t=0}^{\infty} \mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) \overset{(a)}{\leq} 1 + (1 + c)M + \sum_{t: t > (1+c)M} \mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t)$$
$$\overset{(a)}{\leq} O\left(\frac{\eta \log(C(\eta_0, \eta)) \log J}{D_e} + \frac{\log(J/\delta)}{D_0} + J(C(\eta_0, \eta))^{1/\eta} \delta^{D_0/\eta^2}\right)$$

where, (a) follows as $C(\eta_0, \eta)/D_0 < C(\eta_0, \eta)/D_e$. □

3.4 Applications of Active Testing

In this section we present several applications of active testing.

In all the active testing experiments we use the threshold function for the Gaussian distribution as proved in Lemma 7.10. Hence the threshold function used is

$$\beta = \log(J/\delta).$$

Note that this threshold function is smaller than the general sub-Gaussian threshold function proved in Lemma 7.9.

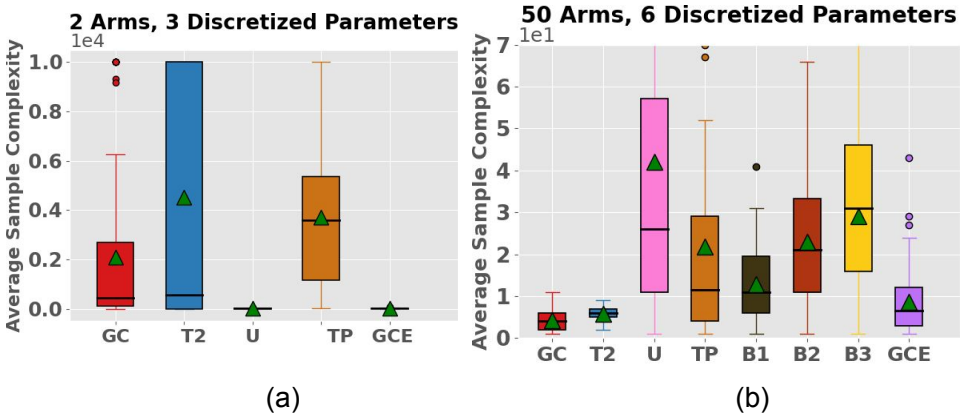

Figure 3.1: **(a)** Sample complexity over Example 1 with $\delta = 0.1$. **(b)** Sample complexity over 3 Group setting with $\delta = 0.1$. U = Unif , B1 = Batch-GC ($B = 5$), B2 = Batch-GC ($B = 10$), B3 = Batch-GC ($B = 15$). The green triangle marker represents the mean stopping time. The horizontal line across each box represents the median.

3.4.1 Experiment on Example 1

Recall that the Example 1 is given by the following table under the three different hypotheses $\{\boldsymbol{\theta}^*, \boldsymbol{\theta}', \boldsymbol{\theta}''\}$

$\boldsymbol{\theta}$	$=$	$\boldsymbol{\theta}^*$	$\boldsymbol{\theta}'$	$\boldsymbol{\theta}''$
$\mu_1(\boldsymbol{\theta})$	$=$	1	0.001	0
$\mu_2(\boldsymbol{\theta})$	$=$	1	1.002	0.998

We can show that under $p_{\boldsymbol{\theta}^*}$ we have the following optimization problem

$$\begin{aligned} &\max \ell \\ \textbf{s.t. } &p(1)0.999^2 + p(2)0.002^2 \geq \ell \\ &p(1)1^2 + p(2)0.002^2 \geq \ell. \end{aligned}$$

The solution to the above optimization is given by $p_{\boldsymbol{\theta}^*} = [p(1), p(2)] = [1, 0]$. Similarly we can show that $p_{\boldsymbol{\theta}'}$ we have the following optimization problem

$$\begin{aligned} &\max \ell \\ \textbf{s.t. } &p(1)0.001^2 + p(2)0.002^2 \geq \ell \\ &p(1)0.001^2 + p(2)0.004^2 \geq \ell. \end{aligned}$$

The solution to the above optimization is given by $p_{\boldsymbol{\theta}'} = [p(1), p(2)] = [0, 1]$. Finally we can show that $p_{\boldsymbol{\theta}''}$ we have the following optimization problem

$$\begin{aligned} &\max \ell \\ \textbf{s.t. } &p(1)0.001^2 + p(2)0.004^2 \geq \ell \\ &p(1)1^2 + p(2)0.002^2 \geq \ell. \end{aligned}$$

The solution to the above optimization is given by $p_{\boldsymbol{\theta}''} = [p(1), p(2)] = [0, 1]$. Hence, $D_1 := \min_{p_{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n p_{\boldsymbol{\theta}}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2 = 0.002^2 = 4 \times 10^{-6}$. Similarly, we can compute that $D_0 := \max_{\text{l}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2 = 0.999^2$. Hence the non-asymptotic term $(\log J)/D_1 = 0.3 \times 10^6$ and the asymptotic term $\log(J/\delta)/D_0 = 3.4$.

3.4.2 Active testing experiment (3 Group setting)

Consider an environment of 50 actions and 6 parameters. The actions can be divided into three groups. The first group contains a single action that most effectively discriminates $\boldsymbol{\theta}^*$ from all other hypotheses. The second group consists of 5 actions, each of which can discriminate one hypothesis from the others. Finally, the third group of 44 actions are barely informative as their means are similar under all hypotheses. The mean values under all hypotheses are given below. We show the empirical performance of different policies in the Figure 3.1b. Both GC and T2 outperforms TP which conducts a uniform exploration over the actions in the second group in its first phase. T2 performs well as it samples the action in the first group when $\boldsymbol{\theta}^*$ is either the most-likely or the second most-likely hypothesis. Unif performs the worst as it uniformly samples all actions, including the uninformative actions in the third group. We further show the performance

of Batch-GC with batch sizes of $B = 5, 10, 15$ and see that there is a increase in sample complexity with larger batches.

3 **Group setting Implementation Details:** In this setting there are three groups of actions. In first group there is a single action that discriminates best between all pair of hypotheses. In second group there are 5 actions which can discriminate one hypotheses from others. Finally in the third group there are 44 actions which cannot discriminate between any pair of hypotheses. The following table describes the $\mu_1(\cdot), \mu_2(\cdot), \ldots, \mu_{50}(\cdot)$ under different hypotheses as follows:

$\boldsymbol{\theta}$	$=$	$\boldsymbol{\theta}^*$	$\boldsymbol{\theta}_2$	$\boldsymbol{\theta}_3$	$\boldsymbol{\theta}_4$	$\boldsymbol{\theta}_5$	$\boldsymbol{\theta}_6$
$\mu_1(\boldsymbol{\theta})$	$=$	3	0	0	0	0	0
$\mu_2(\boldsymbol{\theta})$	$=$	2	3	2	2	2	2
$\mu_3(\boldsymbol{\theta})$	$=$	2	2	3	2	2	2
$\mu_4(\boldsymbol{\theta})$	$=$	2	2	2	3	2	2
$\mu_5(\boldsymbol{\theta})$	$=$	2	2	2	2	3	2
$\mu_6(\boldsymbol{\theta})$	$=$	2	2	2	2	2	3
$\mu_7(\boldsymbol{\theta})$	$=$	$1+\epsilon_{7,1}$	$1+\epsilon_{7,2}$	$1+\epsilon_{7,3}$	$1+\epsilon_{7,4}$	$1+\epsilon_{7,5}$	$1+\epsilon_{7,6}$
$\vdots$			$\vdots$				
$\mu_{50}(\boldsymbol{\theta})$	$=$	$1+\epsilon_{50,1}$	$1+\epsilon_{50,2}$	$1+\epsilon_{50,3}$	$1+\epsilon_{50,4}$	$1+\epsilon_{50,5}$	$1+\epsilon_{50,6}$

In the above setting,we define $\epsilon_{i,j}$ for the i-th action and j-th hypothesis as a small value close to 0 and $\epsilon_{i,j} \neq \epsilon_{i',j'}$ for any pair of hypotheses $j, j' \in [J]$ and actions $i, i' \in [n]$.

In this chapter we introduced several variations of GC sampling and the T2 sampling rule to reduce the computational load of calculating the Chernoff proportion. We established the sample complexity bounds of these policies and show their empirical performance in several active testing environments. In the next chapter we extend GC to smoothly parameterized hypothesis space.

Chapter 4

Active Regression with GC

In this chapter we extend the original GC policy for finite hypotheses to the smoothly parameterized continuous hypotheses space. This enables us to apply the GC policy in the active regression setting. We formulate a new sampling rule for the original Chernoff policy and draw its connection to the e-optimality setting in design of experiments. We further provide the sample complexity bounds for this new GC policy for smoothly parameterized hypotheses space. We first discuss the related works in active regression setting below.

4.1 Related Works in Active Regression

In active regression the hypothesis space is a compact set, and (1.2) is equivalent to locally optimum experiment design with the E-optimality criterion (see e.g. Chap. 5 in Pronzato and Pázman (2013)). We consider noisy oracle responses in the setup of pool-based active learning (Settles, 2009, Section 2.3). A lot of recent work in this setup has been for active classification (Dasgupta, 2005; Hanneke, 2007; Dasgupta et al., 2008; Balcan et al., 2009; Balcan and Long, 2013; Zhang and Chaudhuri, 2014; Katz-Samuels et al., 2021); instead our focus here is on active regression Cai et al. (2016); Wu (2018); Wu et al. (2019); Bu et al. (2019). Theoretical results for active regression using maximum likelihood estimates were given by Chaudhuri and Mykland (1993) and Chaudhuri et al. (2015). We establish a direct connection between the E-optimal criterion to the optimality of Chernoff's method in active testing. Most theoretical work on active learning has focused on the PAC or the agnostic PAC model, where the goal is to learn binary classifiers that belong to a particular hypothesis class (Dasgupta, 2005; Hanneke, 2007; Dasgupta et al., 2008; Balcan et al., 2009; Balcan and Long, 2013; Zhang and Chaudhuri, 2014; Katz-Samuels et al., 2021). In contrast the pool-based active regression setting has been studied relatively less from the theoretical standpoint. Chaudhuri et al. (2015) analyze a maximum likelihood estimate of the true parameter. They propose a two-stage process ActiveS that first samples from the unlabelled sample space uniform randomly to build an estimation of $\boldsymbol{\theta}^*$. It then solves an SDP to get a sampling proportion over the sample space. They show that ActiveS have polynomial time complexity. In contrast, our method is a fully adaptive process as opposed to stage-based ActiveS . As opposed to us Chaudhuri et al. (2015) is not motivated by the experimental design approach and directly go

after the MLE. Sabato and Munos (2014) provides an active learning algorithm for linear regression problems under model mismatch. The same setting under heteroscedastic noise has been studied by Chaudhuri et al. (2017) where they propose a two-stage process adapted to the noise. Fontaine et al. (2019) also studies the linear regression setting under heteroscedastic noise but proposes a per-step adaptive algorithm that is similar to A-optimal design. Bu et al. (2019) studies a different setting where $\boldsymbol{\theta}^*$ is also changing with time. They modify the algorithm of Chaudhuri et al. (2015) to per step process where the optimization needs to be solved at every round. Wu (2018) studies the linear regression setting where the goal is to maximize the diversity of the samples. (Cai et al., 2016) studies both the linear and non-linear regression setting and proposes the heuristic EMCM without any convergence guarantee. Similarly, (Wu et al., 2019) studies the active regression for noiseless setting, with no convergence proof. As opposed to these works we propose a fully adaptive policy GC with convergence guarantee and show that our method works well in several real-world applications. Other forms of optimal experiment design have been explored in the context of active learning by Yu et al. (2006), and in different bandit problems by Soare et al. (2014); Fiez et al. (2019); Degenne et al. (2020a). Different approaches to converge to the optimal $\max\min$ sampling proportions in the context of best-action identification (pure exploration) have been obtained by Ménard (2019); Degenne et al. (2019); Jedra and Proutière (2020). Note that our objective of identifying $\boldsymbol{\theta}^*$ is a strictly difficult objective than these works. The $\max\min$ optimization also appears in Combes et al. (2017); Degenne et al. (2020b); Tirinzoni et al. (2020) under the regret minimization framework which is not our focus.

4.2 Extending GC to Continuous Hypotheses Setting

In this section, we extend the Chernoff sampling policy to smoothly parameterized hypothesis spaces, such as $\boldsymbol{\Theta} \subseteq \mathbb{R}^d$. The original sampling rule in (1.3) asks to solve a $\max\min$ optimization, where the $\min$ is over all possible choices of the parameter that are not equal to the parameter $\boldsymbol{\theta}$ being verified. An extension of the rule for when $\boldsymbol{\theta}^*$ can take infinitely many values was first given by Albert (1961). In it, they want to identify which of two partitions $\boldsymbol{\Theta}_1 \cup \boldsymbol{\Theta}_2 = \boldsymbol{\Theta}$ does the true $\boldsymbol{\theta}^*$ belong to. For any given $\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}_1$, their verification sampling rule (specialized to the case of

Gaussian noise) is

$$\mathbf{p}_{\boldsymbol{\theta}_1} = \arg\max_{\mathbf{p}} \inf_{\boldsymbol{\theta}_2 \in \boldsymbol{\Theta}_2} \sum_{i=1}^{n} p(i)(\mu_i(\boldsymbol{\theta}_1) - \mu_i(\boldsymbol{\theta}_2))^2. \tag{4.1}$$

Recall that $\widehat{\boldsymbol{\theta}}(t) := \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L_{Y^t}(\boldsymbol{\theta})$. Suppose we want to find the optimal verification proportion for testing $\widehat{\boldsymbol{\theta}}(t)$, the current best estimate of $\boldsymbol{\theta}^*$. Let $\mathcal{B}^{\complement}_{\epsilon}(\widehat{\boldsymbol{\theta}}(t)) := \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)\| > \epsilon\}$ denote the complement of a ball of radius $\epsilon > 0$ centered at $\widehat{\boldsymbol{\theta}}(t)$. Instantiate (4.1) with $\boldsymbol{\theta}_1 = \widehat{\boldsymbol{\theta}}(t), \boldsymbol{\Theta}_1 = \boldsymbol{\Theta} \setminus \mathcal{B}^{\complement}_{\epsilon}(\widehat{\boldsymbol{\theta}}(t))$ and $\boldsymbol{\Theta}_2 = \mathcal{B}^{\complement}_{\epsilon}(\widehat{\boldsymbol{\theta}}(t))$. Denote the solution of this optimization as $\mathbf{p}_{\widehat{\boldsymbol{\theta}}(t),\epsilon}$ and let $\mathbf{p}_{\widehat{\boldsymbol{\theta}}(t)} := \lim_{\epsilon \to 0} \mathbf{p}_{\widehat{\boldsymbol{\theta}}(t),\epsilon}$. We show in Theorem 5 that $\mathbf{p}_{\widehat{\boldsymbol{\theta}}(t)}$ is well-defined and can be computed efficiently. For any $i \in [n]$ the gradient of $\mu_i(\cdot)$ evaluated at $\boldsymbol{\theta}$ is a column vector denoted as $\nabla\mu_i(\boldsymbol{\theta})$.

Theorem 5. *Assume that $\mu_i(\boldsymbol{\theta})$ for all $i \in [n]$ is a differentiable function, and the set $\{\nabla\mu_i(\widehat{\boldsymbol{\theta}}(t)) : i \in [n]\}$ of gradients evaluated at $\widehat{\boldsymbol{\theta}}(t)$ span $\mathbb{R}^d$. Consider the p.m.f. $\mathbf{p}_{\widehat{\boldsymbol{\theta}}(t),\epsilon}$ from* (4.1) *for verifying $\widehat{\boldsymbol{\theta}}(t)$ against all alternatives in $\mathcal{B}^{\complement}_{\epsilon}(\widehat{\boldsymbol{\theta}}(t))$. The limiting value of $\mathbf{p}_{\widehat{\boldsymbol{\theta}}(t),\epsilon}$ as $\epsilon \to 0$*

$$\mathbf{p}_{\widehat{\boldsymbol{\theta}}(t)} := \arg\max_{\mathbf{p}} \text{min-eigenvalue} \left(\sum_{i=1}^{n} p(i) \nabla\mu_i(\widehat{\boldsymbol{\theta}}(t)) \nabla\mu_i(\widehat{\boldsymbol{\theta}}(t))^T \right)$$

Discussion 1. We first discuss the main ideas to prove this theorem. Define $f_i(\boldsymbol{\theta}) := (\mu_i(\boldsymbol{\theta}) - \mu_i(\widehat{\boldsymbol{\theta}}))^2$ for any generic $\boldsymbol{\theta}$. Introducing a probability density function $\mathbf{q}$ over $\boldsymbol{\Theta}$, we can rewrite the optimization for $\mathbf{p}_{\widehat{\boldsymbol{\theta}}(t)}$ from (4.1) as

$$\mathbf{p}_{\widehat{\boldsymbol{\theta}}(t)} = \arg\max_{\mathbf{p}} \inf_{\mathbf{q}: q(\hat{\boldsymbol{\theta}})=0} \int_{\boldsymbol{\Theta}} q(\boldsymbol{\theta}) \sum_{i=1}^{n} p(i) f_i(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{4.2}$$

We consider a family of p.d.f.s supported on the boundary of a small ball $\mathcal{Q}_\eta$ around $\widehat{\boldsymbol{\theta}}(t)$, whose radius is determined by a scalar $\eta > 0$. We show that the value of (4.2) is equal to

$$\lim_{\eta \to 0} \max_{\mathbf{p}} \inf_{\mathbf{q} \in \mathcal{Q}_\eta} \int_{\boldsymbol{\Theta}} q(\boldsymbol{\theta}) \sum_{i=1}^{n} p_i f_i(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{4.3}$$

When η is small, we can use the Taylor series for $f_i(\boldsymbol{\theta})$ around $\widehat{\boldsymbol{\theta}}(t)$ in (4.3). By the definition of f_i we have $\nabla f_i(\widehat{\boldsymbol{\theta}}(t)) = 0$ and the second-order term in the Taylor series is $0.5(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))^T \nabla^2 f_i(\widehat{\boldsymbol{\theta}}(t))(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)) = (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))^T \nabla\mu_i(\widehat{\boldsymbol{\theta}}(t)) \nabla\mu_i(\widehat{\boldsymbol{\theta}}(t))^T (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))$. Using this in (4.3) gives us the result.

Proof. Consider the $\max\min$ optimization in (4.1) for verifying $\widehat{\boldsymbol{\theta}}(t)$ against all alternatives in $\boldsymbol{\Theta}\backslash\{\widehat{\boldsymbol{\theta}}(t)\}$. Denoting $f_i(\boldsymbol{\theta}) := (\mu_i(\boldsymbol{\theta})-\mu_i(\widehat{\boldsymbol{\theta}}(t)))^2$ for any generic $\boldsymbol{\theta}$, we can rewrite the optimization using a probability density function (p.d.f.) $\mathbf{q}$ over parameters in $\boldsymbol{\Theta}$, as explained subsequently.

$$\begin{aligned}\mathbf{p}_{\widehat{\boldsymbol{\theta}}(t)} &= \arg\max_{\mathbf{p}} \inf_{\boldsymbol{\theta}\neq\widehat{\boldsymbol{\theta}}(t)} \sum_{i=1}^{n} p(i) f_i(\boldsymbol{\theta}) \\ &= \arg\max_{\mathbf{p}} \inf_{\mathbf{q}:q(\widehat{\boldsymbol{\theta}}(t))=0} \int_{\boldsymbol{\Theta}} q(\boldsymbol{\theta}) \sum_{i=1}^{n} p(i) f_i(\boldsymbol{\theta}) d\boldsymbol{\theta}. \end{aligned} \tag{4.4}$$

The above equality is true because if in the LHS, the inner infimum was attained at a $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, then the same value can be attained in the RHS by a degenerate p.m.f. $\mathbf{q}$ that puts all its mass on that $\boldsymbol{\theta}$. Conversely, suppose the inner infimum in the RHS was attained at a p.m.f. $\mathbf{q}^*$. Since the objective function is a linear in $\mathbf{q}$, the objective value is the same for a degenerate pmf that puts all its mass on one of the support points of $\mathbf{q}^*$, and this value can also be attained by the LHS infimum.

For any given $\eta > 0$ and all $i \in [n]$, define

$$\begin{aligned}&\epsilon_\eta(i) := \max\{\epsilon > 0 : |f_i(\boldsymbol{\theta})| \leq \eta \text{ if } ||\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)|| \leq \epsilon\}, \text{ and} \\ &\epsilon_\eta := \min_i \epsilon_\eta(i). \end{aligned} \tag{4.5}$$

Existence of $\epsilon_\eta(i)$ is ensured by the continuity of $f_i(\boldsymbol{\theta})$ in $\boldsymbol{\theta}$ and the fact that $f_i(\widehat{\boldsymbol{\theta}}(t)) = 0$, and thus $\epsilon_\eta > 0$. Consider a family $\mathcal{Q}_\eta$ of pdfs supported on the boundary of an ϵ_η-ball around $\widehat{\boldsymbol{\theta}}(t)$, i.e.,

$$\mathcal{Q}_\eta := \left\{ \mathbf{q} : \int_{\boldsymbol{\Theta}} q(\boldsymbol{\theta}) d\boldsymbol{\theta} = 1, q(\boldsymbol{\theta}') = 0 \text{ if } ||\boldsymbol{\theta}' - \widehat{\boldsymbol{\theta}}(t)|| \neq \epsilon_\eta \right\}.$$

For any pmf $\mathbf{p}$, if the infimum in (4.4) is attained at $\mathbf{q}^*_{\mathbf{p}}$, then the suboptimality gap by restricting the infimum to be over the set $\mathcal{Q}_\eta$ is non-negative, and is upper bounded by

$$\begin{aligned}&\inf_{\mathbf{q}\in\mathcal{Q}_\eta} \sum_{i=1}^{n} p_i \int_{\boldsymbol{\Theta}} q(\boldsymbol{\theta}) f_i(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &\leq \sum_{i=1}^{n} p_i \int_{\boldsymbol{\theta}:||\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}(t)||=\epsilon_\eta} q(\boldsymbol{\theta}) \eta d\boldsymbol{\theta} \leq \eta. \end{aligned}$$

As $\eta \to 0$, the suboptimality gap also tends to zero. Hence for any pmf $\mathbf{p}$,

$$\lim_{\eta\to 0} \inf_{\mathbf{q}\in\mathcal{Q}_\eta} \int_{\boldsymbol{\Theta}} q(\boldsymbol{\theta}) \sum_{i=1}^{n} p_i f_i(\boldsymbol{\theta}) d\boldsymbol{\theta} = \inf_{\mathbf{q}:q(\widehat{\boldsymbol{\theta}}(t))=0} \int_{\boldsymbol{\Theta}} q(\boldsymbol{\theta}) \sum_{i=1}^{n} p_i f_i(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

The objective function in (4.4) is linear in both $\mathbf{p}$ and $\mathbf{q}$, so using Sion's minimax theorem we can interchange the order of $\inf$ and $\max$ to get the same value, i.e., the optimum value is equal to

$$\begin{aligned}
&\inf_{\mathbf{q}:q(\widehat{\boldsymbol{\theta}}(t))=0} \max_{\mathbf{p}} \sum_{i=1}^{n} p(i) \int_{\boldsymbol{\Theta}} q(\boldsymbol{\theta}) f_i(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \lim_{\eta \to 0} \inf_{\mathbf{q} \in \mathcal{Q}_\eta} \max_{\mathbf{p}} \sum_{i=1}^{n} p_i \int_{\boldsymbol{\Theta}} q(\boldsymbol{\theta}) f_i(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \lim_{\eta \to 0} \max_{\mathbf{p}} \inf_{\mathbf{q} \in \mathcal{Q}_\eta} \int_{\boldsymbol{\Theta}} q(\boldsymbol{\theta}) \sum_{i=1}^{n} p_i f_i(\boldsymbol{\theta}) d\boldsymbol{\theta}.
\end{aligned}$$

For any choice of $\eta > 0$ in the convergent series to 0, define $\mathbf{p}^*_\eta$ to be the maximizer in the above display. Consider the multivariable Taylor series of f_i around $\widehat{\boldsymbol{\theta}}(t)$, for $\boldsymbol{\theta}$ that satisfies $||\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)|| < \epsilon_\eta$.

$$\begin{aligned}
f_i(\boldsymbol{\theta}) = f_i(\widehat{\boldsymbol{\theta}}(t)) + (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))^T \nabla f_i(\widehat{\boldsymbol{\theta}}(t)) \\
+ 0.5(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))^T \nabla^2 f_i(\widehat{\boldsymbol{\theta}}(t))(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)) + o(||\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)||^3).
\end{aligned}$$

For indices $j, k \in [d]$ we can evaluate

$$\begin{aligned}
\nabla f_i(\boldsymbol{\theta}) &= \left[\tfrac{\partial f_i}{\partial \boldsymbol{\theta}_j}(\boldsymbol{\theta})\right]_j = \left[2(\mu_i(\boldsymbol{\theta}) - \mu_i(\widehat{\boldsymbol{\theta}}(t)))\tfrac{\partial \mu_i}{\partial \boldsymbol{\theta}_j}(\boldsymbol{\theta})\right]_j, \text{ and} \\
\nabla^2 f_i(\boldsymbol{\theta}) &= \left[\tfrac{\partial^2 f_i}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k}(\boldsymbol{\theta})\right]_{j,k} \\
&= \left[2(\mu_i(\boldsymbol{\theta}) - \mu_i(\widehat{\boldsymbol{\theta}}(t)))\tfrac{\partial^2 \mu_i}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k}(\boldsymbol{\theta}) + 2\tfrac{\partial \mu_i}{\partial \boldsymbol{\theta}_j}(\boldsymbol{\theta})\tfrac{\partial \mu_i}{\partial \boldsymbol{\theta}_k}(\boldsymbol{\theta})\right]_{j,k}
\end{aligned}$$

giving that $\nabla f_i(\widehat{\boldsymbol{\theta}}(t)) = 0$ and $\nabla^2 f_i(\widehat{\boldsymbol{\theta}}(t)) = 2\nabla \mu_i(\widehat{\boldsymbol{\theta}}(t)) \nabla \mu_i(\widehat{\boldsymbol{\theta}}(t))^T$. Thus for small ϵ_η we have the approximation that $f_i(\boldsymbol{\theta}) \approx ((\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))^T \nabla \mu_i(\widehat{\boldsymbol{\theta}}(t)))^2$. By the assumption that the set of gradients at $\widehat{\boldsymbol{\theta}}(t)$ span $\mathbb{R}^d$, choosing a small η implies that ϵ_η defined in (4.5) is also small. When

$\mathbf{q} \in \mathcal{Q}_\eta$ for small η, the optimization of (4.4) becomes the following.

$$\begin{aligned}
&\max_{\mathbf{p}} \inf_{\mathbf{q}\in\mathcal{Q}_\eta} \int_{\boldsymbol{\Theta}} q(\boldsymbol{\theta}) \sum_{i=1}^{n} p_i f_i(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \max_{\mathbf{p}} \inf_{\mathbf{q}\in\mathcal{Q}_\eta} \int_{\boldsymbol{\theta}:||\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}(t)||=\epsilon_\eta} q(\boldsymbol{\theta}) \sum_{i=1}^{n} p_i (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))^T \nabla\mu_i(\widehat{\boldsymbol{\theta}}(t)) \nabla\mu_i(\widehat{\boldsymbol{\theta}}(t))^T (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)) d\boldsymbol{\theta} \\
&= \max_{\mathbf{p}} \inf_{\mathbf{q}\in\mathcal{Q}_\eta} \int_{\boldsymbol{\theta}:||\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}(t)||=\epsilon_\eta} \frac{(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))^T \sum_{i=1}^{n} p_i \nabla\mu_i(\widehat{\boldsymbol{\theta}}(t)) \nabla\mu_i(\widehat{\boldsymbol{\theta}}(t))^T (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))}{(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))^T (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))} q(\boldsymbol{\theta}) ||\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)||^2 d\boldsymbol{\theta} \\
&= \max_{\mathbf{p}} \text{min eigenvalue} \left(\sum_{i=1}^{n} p_i \nabla\mu_i(\widehat{\boldsymbol{\theta}}(t)) \nabla\mu_i(\widehat{\boldsymbol{\theta}}(t))^T \right) \epsilon_\eta^2.
\end{aligned}$$

Since the above is true for all values of η, as $\eta \to 0$ the maximizer $\mathbf{p}_\eta^*$ for the above optimization converges to the desired p.m.f. $\mathbf{p}_{\widehat{\boldsymbol{\theta}}(t)}$, having the definition as claimed. □

Remark 1. (Maximum Derivative) If $\boldsymbol{\Theta}$ is known to be a low-dimensional subspace, then the appropriate p.m.f. would minimize the weighted correlation of the gradients *projected* onto the $\boldsymbol{\Theta}$ space. In the case of $d = 1$, the gradients are scalars, and the p.m.f. would put all its mass on the action that has the maximum absolute value of the gradient at $\widehat{\boldsymbol{\theta}}(t)$. One can apply Theorem 5 to settings where the (smooth) loss function is different from squared error loss. The sampling strategy would then be chosen to maximize the minimum eigenvalue of the weighted sum of Hessians of the losses for each of the points evaluated at $\widehat{\boldsymbol{\theta}}(t)$.

4.3 How to solve the optimization

The optimization in Theorem 5 can be solved using convex optimization software. This is because the objective function, i.e., the minimum eigenvalue function is a concave function of the matrix argument, and the domain of the optimization $\{\mathbf{p} : \sum_{i=1}^{n} p(i) = 1, p(i) \geq 0 \forall i \in [n]\}$ is a convex set. Hence we can maximize the objective over the domain. Since the set of gradients $\{\nabla\mu_i(\widehat{\boldsymbol{\theta}}(t)) : i \in [n]\}$ span $\mathbb{R}^d$, the optimal objective value is positive. Since, the verification proportions are the solution to a convex optimization problem, so we can terminate it early to get an approximate solution. Computing to specified accuracy takes $O(n^3 + nd^2 + d^3)$ operations (Fan and Nekooie, 1995).

4.4 Sample Complexity of GC for Continuous Hypotheses Setting

To illustrate the use of Theorem 5, consider the problem of active learning in a hypothesis space of parametric functions $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$. The target function is $f_{\boldsymbol{\theta}^*}$ for an unknown $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$. Assume that the learner may query the value of $f_{\boldsymbol{\theta}^*}$ at points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in its domain. If point $\mathbf{x}_i$ is queried, then the learner receives the value $f_{\boldsymbol{\theta}^*}(\mathbf{x}_i)$ plus a realization of a zero-mean sub-Gaussian noise. This coincides with the setting above by setting $\mu_i(\boldsymbol{\theta}) := f_{\boldsymbol{\theta}}(\mathbf{x}_i)$.

From Theorem 5 we have the following simple active learning policy. First, make an initial (possibly random) guess $\widehat{\boldsymbol{\theta}}(1)$ of $\boldsymbol{\theta}^*$, and then compute the Chernoff sampling proportions according to the theorem. Sample accordingly and then repeat the Chernoff sampling process to obtain a sequence of estimates $\widehat{\boldsymbol{\theta}}(t)$, $t = 1, 2, \ldots$. The results of Albert (1961) imply that for any $\epsilon > 0$ the iterates will converge to within ϵ of $\boldsymbol{\theta}^*$ within an optimal number of samples in the high confidence ($\delta \to 0$) regime. In the following theorem we show that the expected risk is converging to the optimal value. Define the average empirical loss as $\widehat{P}_t(\boldsymbol{\theta}) = \frac{1}{t}\sum_{s=1}^{t} L_s(\boldsymbol{\theta})$ and λ_1 is an upper bound (assumptions in Appendix).

Theorem 6. ***(Dense GC Sample Complexity)*** *Suppose $L_{Y^1}(\boldsymbol{\theta}), L_{Y^2}(\boldsymbol{\theta}), \cdots, L_{Y^t}(\boldsymbol{\theta}) : \mathbb{R}^d \to \mathbb{R}$ are squared loss functions from a distribution that satisfies Assumption 3 and Assumption 4. Further define $P_t(\boldsymbol{\theta}) = \frac{1}{t}\sum_{s=1}^{t} \mathbb{E}_{\substack{I_s \sim \mathbf{p}_{\widehat{\boldsymbol{\theta}}_{s-1}} \\ Y_s \sim \mathcal{N}(\mu_{I_s}(\boldsymbol{\theta}^*), \frac{1}{2})}} [L_s(\boldsymbol{\theta}) | \mathcal{F}^{s-1}]$ where, $\widehat{\boldsymbol{\theta}}_t = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{s=1}^{t} L_s(\boldsymbol{\theta})$. If t is large enough such that $t \geq \frac{\left(C_1 C_3 + 2\eta^2 \lambda_1^2\right) p \log(dt)}{c'}$ then we can show that for a constant p*

$$(1 - \epsilon_t)\frac{\sigma^2}{t} - \frac{C_1^2}{t^{p/2}} \leq \mathbb{E}\left[P_t(\widehat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*)\right] \leq (1 + \epsilon_t)\frac{\sigma^2}{t} + \frac{\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left(P_t(\boldsymbol{\theta}) - P_t(\boldsymbol{\theta}^*)\right)}{t^p}$$

where, $\sigma^2 := \mathbb{E}\left[\frac{1}{2}\left\|\nabla \widehat{P}_t(\boldsymbol{\theta}^*)\right\|^2_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}\right]$, and $\epsilon_t := (C_1 C_3 + 2\eta^2\lambda_1^2)\sqrt{\frac{p \log(dt)}{t}}$.

The quantity $\mathbb{E}_{I_s \sim \mathbf{p}_{\boldsymbol{\theta}}}[L_s(\boldsymbol{\theta})]$ characterizes the worst-case loss we could suffer at time s due to estimation error. This is because $\mathbb{E}[L_s(\boldsymbol{\theta})] = \mathbb{E}[(Y_s - \mu_{I_s}(\boldsymbol{\theta})^2] = \mathbb{E}[(\mu_{I_s}(\boldsymbol{\theta}^*) - \mu_{I_s}(\boldsymbol{\theta})^2 + 1/2]$, and the definition of $\mathbf{p}_{\boldsymbol{\theta}}$ ensures that $\sum_{i=1}^{n} \mathbf{p}_{\boldsymbol{\theta}}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}))^2$ is maximized for the most confusing $\boldsymbol{\theta}' \notin \mathcal{B}_\epsilon(\boldsymbol{\theta})$ at small enough ϵ. We contrast this with the average-case loss $\mathbb{E}_{I_s \sim \text{Uniform}([n])}[(\mu_{I_s}(\boldsymbol{\theta}') - \mu_{I_s}(\boldsymbol{\theta})^2]$, which is not larger than the expected value under $I_s \sim \mathbf{p}_{\boldsymbol{\theta}}$ due to the $\arg\max$ in (4.1). It

can thus be seen that generalized Chernoff sampling allows us to bound a more stringent notion of risk than traditionally looked at in the literature. The theorem bounds the running average of the worst-case losses at each time step.

4.4.1 Discussion on Definitions and Assumptions

In this section we discuss several definitions and assumptions required to prove the main theorem for active regression.

Definition 2. *We define the following star-norm quantity at round t as*

$$\|A\|_* = \left\| \left(\nabla^2 P_t\left(\boldsymbol{\theta}^*\right)\right)^{-1/2} \cdot A \cdot \left(\nabla^2 P_t\left(\boldsymbol{\theta}^*\right)\right)^{-1/2} \right\|.$$

Now we state the two following assumptions required by the Theorem 6. Also note that we define the squared loss function $L_{Y_s}(\boldsymbol{\theta}) = (\mu_{I_s}(\boldsymbol{\theta}) - Y_s)^2$, the cumulative loss function $L_{Y^s}(\boldsymbol{\theta}) = \sum_{s'=1}^{s} L_{Y_{s'}}(\boldsymbol{\theta}^*) = \sum_{s'=1}^{s}(\mu_{I_{s'}}(\boldsymbol{\theta}) - Y_{s'})^2$, and $L_{Y^1}(\boldsymbol{\theta}), L_{Y^2}(\boldsymbol{\theta}), \ldots, L_{Y^t}(\boldsymbol{\theta})$ for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ are not independent. In contrast (Chaudhuri et al., 2015) assumes that the loss functions are independent for any time $s \in [t]$. Next we state the assumptions used for the proof of Theorem 6.

Assumption 3. *We assume that $\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\mu_{I_s}(\boldsymbol{\theta})$ is bounded such that $\lambda_{\max}\left(\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\mu_{I_s}(\boldsymbol{\theta})\right) \leq \lambda_1$ for each $s \in [1, t]$ and for all $\boldsymbol{\theta}' \in \boldsymbol{\Theta}$.*

The Assumption 3 is a mild assumption on the bounded nature of the eigenvalues of the Hessian matrix $\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\mu_{I_s}(\boldsymbol{\theta})$ for any $\theta \in \boldsymbol{\Theta}$ evaluated at $\boldsymbol{\theta}'$. Then following assumption states a few regularity assumptions required for Theorem 6. A similar assumption has also been used by Chaudhuri et al. (2015); Bu et al. (2019).

Assumption 4. ***(Assumptions for Theorem 6):*** *We assume the following assumptions hold with probability 1:*

1. ***(Convexity of L_{Y_s}):*** *The loss function L_{Y_s} is convex for all time $s \in [t]$.*
2. ***(Smoothness of L_{Y_s}):*** *The L_{Y_s} is smooth such that the first, second, and third derivatives exist at all interior points in $\boldsymbol{\Theta}$.*

3. ***(Regularity Conditions):***

   (a) *$\boldsymbol{\Theta}$ is compact and $L_{Y_s}(\boldsymbol{\theta})$ is bounded for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and for all $s \in [t]$.*

   (b) *$\boldsymbol{\theta}^*$ is an interior point in $\boldsymbol{\Theta}$.*

   (c) *$\nabla^2 L_{Y_s}(\boldsymbol{\theta}^*)$ is positive definite, for all $s \in [t]$.*

   (d) *There exists a neighborhood B of $\boldsymbol{\theta}^*$ and a constant C_3, such that $\nabla^2 L_{Y_s}(\boldsymbol{\theta})$ is C_3 -Lipschitz. Hence, we have that $\|\nabla^2 L_{Y_s}(\boldsymbol{\theta}) - \nabla^2 L_{Y_s}(\boldsymbol{\theta}')\|_* \leq C_3 \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_{\nabla^2 P_s(\boldsymbol{\theta}^*)}$, for $\boldsymbol{\theta}, \boldsymbol{\theta}'$ in this neighborhood.*

4. ***(Concentration at $\boldsymbol{\theta}^*$):*** *We further assume that $\|\nabla L_{Y_s}(\boldsymbol{\theta}^*)\|_{(\nabla^2 P_s(\boldsymbol{\theta}^*))^{-1}} \leq C_1$ and $\|\nabla^2 L_{Y_s}(\boldsymbol{\theta}^*)\|_* \leq C_2$ hold with probability one.*

Assumption 4 (c) is different from that of (Chaudhuri et al., 2015), where they assumed that $\nabla^2 \mathbb{E}[\psi_s](\boldsymbol{\theta}^*)$ is positive definite, where ψ_s are i.i.d. loss functions from some distribution. In our case the loss functions are not i.i.d., which is why we make the assumption on the loss at every time s. Now we show the main proof of Theorem 6.

4.4.2 Main Proof of GC for Smooth Hypotheses

Proof. **Step 1:** We first bound the $\left\|\nabla^2 \widehat{P}_t(\boldsymbol{\theta}) - \nabla^2 P_t(\boldsymbol{\theta}^*)\right\|_*$ as follows

$$\begin{aligned}\left\|\nabla^2 \widehat{P}_t(\boldsymbol{\theta}) - \nabla^2 P_t(\boldsymbol{\theta}^*)\right\|_* &\overset{(a)}{\leq} \left\|\nabla^2 \widehat{P}_t(\boldsymbol{\theta}) - \nabla^2 \widehat{P}_t(\boldsymbol{\theta}^*)\right\|_* + \left\|\nabla^2 \widehat{P}_t(\boldsymbol{\theta}^*) - \nabla^2 P_t(\boldsymbol{\theta}^*)\right\|_* \\ &\overset{(b)}{\leq} C_3 \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} + \sqrt{\frac{2\eta^2 \lambda_1^2 cp \log(dt)}{t}} \end{aligned} \tag{4.6}$$

where, (a) follows from triangle inequality, and (b) follows Lemma 7.14.

Step 2 (Approximation of $\nabla^2 P_t(\boldsymbol{\theta}^*)$): By choosing a sufficiently smaller ball B_1 of radius of $\min\{1/(10C_3),\ \text{diameter}\ (B)\}$), the first term in (4.6) can be made small for $\boldsymbol{\theta} \in B_1$. Also, for sufficiently large t, the second term in (4.6) can be made arbitrarily small (smaller than $1/10$), which occurs if $\sqrt{\frac{p\log(dt)}{t}} \leq \frac{c'}{\sqrt{2\eta^2\lambda_1^2}}$. Hence for large t and $\boldsymbol{\theta} \in B_1$ we have

$$\frac{1}{2}\nabla^2 \widehat{P}_t(\boldsymbol{\theta}) \preceq \nabla^2 P_t(\boldsymbol{\theta}^*) \preceq 2\nabla^2 \widehat{P}_t(\boldsymbol{\theta}) \tag{4.7}$$

Step 3 (Show $\widehat{\boldsymbol{\theta}}_t$ in B_1): Fix a $\widetilde{\boldsymbol{\theta}}$ between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ in B_1. Apply Taylor's series approximation

$$\widehat{P}_t(\boldsymbol{\theta}) = \widehat{P}_t\left(\boldsymbol{\theta}^*\right) + \nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)^\top \left(\boldsymbol{\theta} - \boldsymbol{\theta}^*\right) + \frac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\theta}^*\right)^\top \nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}}) \left(\boldsymbol{\theta} - \boldsymbol{\theta}^*\right)$$

We can further reduce this as follows:

$$\begin{aligned}
\widehat{P}_t(\boldsymbol{\theta}) - \widehat{P}_t\left(\boldsymbol{\theta}^*\right) &\overset{(a)}{=} \nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)^\top \left(\boldsymbol{\theta} - \boldsymbol{\theta}^*\right) + \frac{1}{2}\left\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\right\|^2_{\nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}})} \\
&\overset{(b)}{\geq} \nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)^\top \left(\boldsymbol{\theta} - \boldsymbol{\theta}^*\right) + \frac{1}{4}\left\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\right\|^2_{\nabla^2 P_t(\boldsymbol{\theta}^*)} \\
&\geq -\left\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} \left\|\nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)\right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} + \frac{1}{4}\left(\left\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)}\right)^\top \left(\left\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)}\right) \\
&= \left\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} \left(-\left\|\nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)\right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} + \frac{1}{4}\left\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)}\right) \qquad (4.8)
\end{aligned}$$

where, (a) follows as $\left\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\right\|^2_{\nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}})} := \left(\boldsymbol{\theta} - \boldsymbol{\theta}^*\right)^\top \nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}}) \left(\boldsymbol{\theta} - \boldsymbol{\theta}^*\right)$, and (b) follows as $\widetilde{\boldsymbol{\theta}}$ is in between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ and then using (4.7). Note that in (4.8) if the right hand side is positive for some $\boldsymbol{\theta} \in B_1$, then $\boldsymbol{\theta}$ is not a local minimum. Also, since $\left\|\nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)\right\| \to 0$, for a sufficiently small value of $\left\|\nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)\right\|$, all points on the boundary of B_1 will have values greater than that of $\boldsymbol{\theta}^*$. Hence, we must have a local minimum of $\widehat{P}_t(\boldsymbol{\theta})$ that is strictly inside B_1 (for t large enough). We can ensure this local minimum condition is achieved by choosing an t large enough so that $\sqrt{\frac{p \log(dt)}{t}} \leq c' \min\left\{\frac{1}{L_1 L_3}, \frac{\text{diameter}(B)}{L_1}\right\}$, using Lemma 7.12 (and our bound on the diameter of B_1). By convexity, we have that this is the global minimum, $\widehat{\boldsymbol{\theta}}_t$, and so $\widehat{\boldsymbol{\theta}}_t \in B_1$ for t large enough. We will assume now that t is this large from here on.

Step 4 (Bound $\left\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)}$): For the ERM, $0 = \nabla \widehat{P}_t(\widehat{\boldsymbol{\theta}}_t)$. Again, using Taylor's theorem if $\widehat{\boldsymbol{\theta}}_t$ is an interior point, we have:

$$0 = \nabla \widehat{P}_t(\widehat{\boldsymbol{\theta}}_t) = \nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right) + \nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t)\left(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right) \qquad (4.9)$$

for some $\widetilde{\boldsymbol{\theta}}_t$ between $\boldsymbol{\theta}^*$ and $\widehat{\boldsymbol{\theta}}_t$. Now observe that $\widetilde{\boldsymbol{\theta}}_t$ is in B_1 (since, for t large enough, $\widehat{\boldsymbol{\theta}}_t \in B_1$). Thus it follows from (4.9) that,

$$\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* = \left(\nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t)\right)^{-1} \nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right) \qquad (4.10)$$

where the invertibility is guaranteed by (4.7) and the positive definiteness of $\nabla^2 P_t(\boldsymbol{\theta}^*)$ (by Assumption 4 (3c)). We finally derive the upper bound to $\left\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)}$ as follows

$$\begin{aligned}\left\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} &\overset{(a)}{\leq} \left\|\left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{1/2}\left(\nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t)\right)^{-1}\left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{1/2}\right\| \left\|\nabla \widehat{P}_t(\boldsymbol{\theta}^*)\right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} \\ &\overset{(b)}{\leq} cL_1\sqrt{\frac{p\log(dt)}{t}} \end{aligned} \tag{4.11}$$

where (a) follows from Lemma 7.18, and (b) from Lemma 7.12.

Step 5 (Introducing $\widetilde{\mathbf{z}}$): Fix a $\widetilde{\mathbf{z}}_t$ between $\boldsymbol{\theta}^*$ and $\widehat{\boldsymbol{\theta}}_t$. Apply Taylor's series

$$P_t(\widehat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*) = \frac{1}{2}\left(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right)^\top \nabla^2 P_t(\widetilde{\mathbf{z}}_t)\left(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right) \tag{4.12}$$

Now note that both $\widetilde{\boldsymbol{\theta}}_t$ and $\widetilde{\mathbf{z}}_t$ are between $\widehat{\boldsymbol{\theta}}_t$ and $\boldsymbol{\theta}^*$, which implies $\widetilde{\boldsymbol{\theta}}_t \to \boldsymbol{\theta}^*$ and $\widetilde{\mathbf{z}}_t \to \boldsymbol{\theta}^*$ since $\widehat{\boldsymbol{\theta}}_t \to \boldsymbol{\theta}^*$. By (4.6) and (4.11) and applying the concentration inequalities give us

$$\left\|\nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t) - \nabla^2 P_t(\boldsymbol{\theta}^*)\right\|_* \leq \epsilon_t \tag{4.13}$$

$$\left\|\nabla^2 P_t(\widetilde{\mathbf{z}}_t) - \nabla^2 P_t(\boldsymbol{\theta}^*)\right\|_* \leq L_3\left\|\widetilde{\mathbf{z}}_t - \boldsymbol{\theta}^*\right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} \leq \epsilon_t \tag{4.14}$$

where $\epsilon_t = c\left(L_1 L_3 + 2\eta^2\lambda_1^2\right)\sqrt{\frac{p\log(dt)}{t}}$.

Step 6 (Define $\mathbf{M}_{1,t}$ and $\mathbf{M}_{2,t}$): It follows from the above inequalities (4.13) and (4.14) that

$$\begin{aligned}(1-\epsilon_t)\nabla^2 P_t(\boldsymbol{\theta}^*) \preceq \nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t) \preceq (1+\epsilon_t)\nabla^2 P_t(\boldsymbol{\theta}^*) \\ (1-\epsilon_t)\nabla^2 P_t(\boldsymbol{\theta}^*) \preceq \nabla^2 P_t(\widetilde{\mathbf{z}}_t) \preceq (1+\epsilon_t)\nabla^2 P_t(\boldsymbol{\theta}^*)\end{aligned}$$

Now we define the two quantities $\mathbf{M}_{1,t}$ and $\mathbf{M}_{2,t}$ as follows:

$$\begin{aligned}\mathbf{M}_{1,t} &= \left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{1/2}\left(\nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t)\right)^{-1}\left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{1/2} \\ \mathbf{M}_{2,t} &= \left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{-1/2}\nabla^2 P_t(\widetilde{\mathbf{z}}_t)\left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{-1/2}\end{aligned}$$

Step 7 (Lower bound $P_t(\widehat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*)$): Now for the lower bound it follows from Equation (4.12) that

$$\begin{aligned}
P_t(\widehat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*) &= \frac{1}{2}\left(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right)^\top \nabla^2 P_t(\widetilde{\mathbf{z}}_t)\left(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right) \\
&= \frac{1}{2}\left(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right)^\top \nabla^2 P_t(\boldsymbol{\theta}^*)\nabla^2 P_t(\boldsymbol{\theta}^*)^{-1}\nabla^2 P_t(\widetilde{\mathbf{z}}_t)\left(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right) \\
&= \frac{1}{2}\nabla^2 P_t(\boldsymbol{\theta}^*)^{-1}\nabla^2 P_t(\widetilde{\mathbf{z}}_t)\left(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right)^\top \nabla^2 P_t(\boldsymbol{\theta}^*)\left(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right) \\
&= \frac{1}{2}\nabla^2 P_t(\boldsymbol{\theta}^*)^{-1/2}\nabla^2 P_t(\widetilde{\mathbf{z}}_t)\nabla^2 P_t(\boldsymbol{\theta}^*)^{-1/2}\left\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right\|^2_{\nabla^2 P_t(\boldsymbol{\theta}^*)} \\
&\overset{(a)}{=} \frac{1}{2}\mathbf{M}_{2,t}\left\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right\|^2_{\nabla^2 P_t(\boldsymbol{\theta}^*)}
\end{aligned}$$

where, (a) follows from definition of $\mathbf{M}_{2,t}$. Then using the min-max theorem we can show that

$$\begin{aligned}
P_t(\widehat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*) &\geq \frac{1}{2}\lambda_{\min}(\mathbf{M}_{2,t})\left\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right\|^2_{\nabla^2 P_t(\boldsymbol{\theta}^*)} \\
&= \frac{1}{2}\lambda_{\min}(\mathbf{M}_{2,t})\left\|\nabla^2\widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t)\left(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right)\right\|^2_{\left(\nabla^2\widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t)\right)^{-1}\nabla^2 P_t(\boldsymbol{\theta}^*)\left(\nabla^2\widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t)\right)^{-1}} \\
&\geq \frac{1}{2}\left(\lambda_{\min}(\mathbf{M}_{1,t})\right)^2\lambda_{\min}(\mathbf{M}_{2,t})\left\|\nabla^2\widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t)\left(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right)\right\|^2_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} \\
&\overset{(a)}{=} \frac{1}{2}\left(\lambda_{\min}(\mathbf{M}_{1,t})\right)^2\lambda_{\min}(\mathbf{M}_{2,t})\left\|\nabla\widehat{P}_t(\boldsymbol{\theta}^*)\right\|^2_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}
\end{aligned}$$

where, in (a) we use the eq. (4.10).

Step 8: Define $I(\mathcal{E})$ as the indicator that the desired previous events hold, which we can ensure with probability greater than $1 - \frac{2}{t^p}$. Then we can show that:

$$\begin{aligned}
\mathbb{E}\left[P_t(\widehat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*)\right] &\geq \mathbb{E}\left[\left(P_t(\widehat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*)\right)I(\mathcal{E})\right] \\
&\geq \frac{1}{2}\mathbb{E}\left[\left(\lambda_{\min}(\mathbf{M}_{1,t})\right)^2\lambda_{\min}(\mathbf{M}_{2,t})\left\|\nabla\widehat{P}_t(\boldsymbol{\theta}^*)\right\|^2_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} I(\mathcal{E})\right] \\
&\geq (1 - c'\epsilon_t)\frac{1}{2}\mathbb{E}\left[\left\|\nabla\widehat{P}_t(\boldsymbol{\theta}^*)\right\|^2_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} I(\mathcal{E})\right] \\
&= (1 - c'\epsilon_t)\frac{1}{2}\mathbb{E}\left[\left\|\nabla\widehat{P}_t(\boldsymbol{\theta}^*)\right\|^2_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} (1 - I(\text{not}\,\mathcal{E}))\right] \\
&\overset{(a)}{=} (1 - c'\epsilon_t)\left(\sigma^2 - \frac{1}{2}\mathbb{E}\left[\left\|\nabla\widehat{P}_t(\boldsymbol{\theta}^*)\right\|^2_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} I(\text{not}\,\mathcal{E})\right]\right) \\
&\geq (1 - c'\epsilon_t)\sigma^2 - \mathbb{E}\left[\left\|\nabla\widehat{P}_t(\boldsymbol{\theta}^*)\right\|^2_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} I(\text{not}\,\mathcal{E})\right]
\end{aligned}$$

where, in (a) we have $\sigma^2 := \left\|\nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)\right\|^2_{\left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{-1}}$, and c' is an universal constant.

Step 9: Define the random variable $Z = \left\|\nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)\right\|_{\left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{-1}}$. With a failure event probability of less than $\dfrac{1}{2t^p}$, for any z_0, we have:

$$\begin{aligned}
\mathbb{E}\left[Z^2 I(\text{not}\,\mathcal{E})\right] &= \mathbb{E}\left[Z^2 I(\text{not}\,\mathcal{E}) I\left(Z^2 < z_0\right)\right] + \mathbb{E}\left[Z^2 I(\text{not}\,\mathcal{E}) I\left(Z^2 \geq z_0\right)\right] \\
&\leq z_0 \mathbb{E}[I(\text{not}\,\mathcal{E})] + \mathbb{E}\left[Z^2 I\left(Z^2 \geq z_0\right)\right] \\
&\leq \frac{z_0}{2t^p} + \mathbb{E}\left[Z^2 \frac{Z^2}{z_0}\right] \\
&\leq \frac{z_0}{2t^p} + \frac{\mathbb{E}\left[Z^4\right]}{z_0} \\
&\leq \frac{\sqrt{\mathbb{E}\left[Z^4\right]}}{t^{p/2}}
\end{aligned}$$

where $z_0 = t^{p/2}\sqrt{\mathbb{E}\left[Z^4\right]}$.

Step 10 (Upper Bound): For an upper bound we have that:

$$\begin{aligned}
\mathbb{E}\left[P_t(\widehat{\boldsymbol{\theta}}_t) - P_t\left(\boldsymbol{\theta}^*\right)\right] &= \mathbb{E}\left[\left(P_t(\widehat{\boldsymbol{\theta}}_t) - P_t\left(\boldsymbol{\theta}^*\right)\right) I(\mathcal{E})\right] + \mathbb{E}\left[\left(P_t(\widehat{\boldsymbol{\theta}}_t) - P_t\left(\boldsymbol{\theta}^*\right)\right) I(\text{not}\,\mathcal{E})\right] \\
&\leq \mathbb{E}\left[\left(P_t(\widehat{\boldsymbol{\theta}}_t) - P_t\left(\boldsymbol{\theta}^*\right)\right) I(\mathcal{E})\right] + \frac{\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left(P_t(\boldsymbol{\theta}) - P_t\left(\boldsymbol{\theta}^*\right)\right)}{t^p}
\end{aligned}$$

since the probability of not $\mathcal{E}$ is less than $\dfrac{1}{t^p}$. Now for an upper bound of the first term, observe that

$$\begin{aligned}
\mathbb{E}\left[\left(P_t(\widehat{\boldsymbol{\theta}}_t) - P_t\left(\boldsymbol{\theta}^*\right)\right) I(\mathcal{E})\right] &\leq \frac{1}{2}\mathbb{E}\left[\left(\lambda_{\max}\left(\mathbf{M}_{1,t}\right)\right)^2 \lambda_{\max}\left(\mathbf{M}_{2,t}\right) \left\|\nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)\right\|^2_{\nabla^2 P_t(\boldsymbol{\theta}^*)} I(\mathcal{E})\right] \\
&\leq \left(1 + c'\epsilon_t\right)\frac{1}{2}\mathbb{E}\left[\left\|\nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)\right\|^2_{\left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{-1}} I(\mathcal{E})\right] \\
&\leq \left(1 + c'\epsilon_t\right)\frac{1}{2}\mathbb{E}\left[\left\|\nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)\right\|^2_{\left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{-1}}\right] \\
&= \left(1 + c'\epsilon_t\right)\frac{\sigma^2}{t}
\end{aligned}$$

where, c' is another universal constant. □

4.5 Theoretical Comparisons for Active Regression

We now comapre our result of Theorem 6 with existing works. From the result of Chaudhuri et al. (2015) we can show that the ActiveS algorithm enjoys a convergence guarantee as follows:

$$\mathbb{E}\left[L_U\left(\widehat{\boldsymbol{\theta}}(t)\right) - L_U\left(\theta^*\right)\right] \leq O\left(\frac{\sigma_U^2\sqrt{\log(dt)}}{t} + \frac{R}{t^2}\right) \tag{4.15}$$

where, L_U is the loss under uniform measure, R is the maximum loss under any measure for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and σ_U^2 is defined as

$$\begin{aligned}
\sigma_U^2 &\overset{(a)}{=} \frac{1}{t^2}\,\text{Trace}\left[\left(\sum_{s=1}^{t}\sum_{s'=1}^{t}\mathbb{E}_{L_{Y_s}\sim D}\left[\nabla L_{Y_s}\left(\boldsymbol{\theta}^*\right)\right]\mathbb{E}_{L_{Y_{s'}}\sim D}\left[\nabla L_{Y_{s'}}\left(\boldsymbol{\theta}^*\right)^\top\right]\right) I_\Gamma\left(\boldsymbol{\theta}^*\right)^{-1} I_U\left(\boldsymbol{\theta}^*\right) I_\Gamma\left(\boldsymbol{\theta}^*\right)^{-1}\right] \\
&= \frac{1}{t^2}\,\text{Trace}\left[\left(\sum_{s=1}^{t}\sum_{s'=1}^{t}\mathbb{E}_{L_{Y_s}\sim D}\left[\nabla L_{Y_s}\left(\boldsymbol{\theta}^*\right)\right]\mathbb{E}_{L_{Y_{s'}}\sim D}\left[\nabla L_{Y_{s'}}\left(\boldsymbol{\theta}^*\right)^\top\right]\right) I_\Gamma\left(\boldsymbol{\theta}^*\right)^{-1} I_U\left(\boldsymbol{\theta}^*\right) I_\Gamma\left(\boldsymbol{\theta}^*\right)^{-1}\right] \\
&\overset{(b)}{=} \frac{1}{t^2}\,\text{Trace}\left[\left(\sum_{s=1}^{t}\sum_{s'=1}^{t} I_\Gamma\left(\boldsymbol{\theta}^*\right)\right) I_\Gamma\left(\boldsymbol{\theta}^*\right)^{-1} I_U\left(\boldsymbol{\theta}^*\right) I_\Gamma\left(\boldsymbol{\theta}^*\right)^{-1}\right] \\
&= \text{Trace}\left[I_U\left(\boldsymbol{\theta}^*\right) I_\Gamma\left(\boldsymbol{\theta}^*\right)^{-1}\right]
\end{aligned} \tag{4.16}$$

where, in (a) the loss L_{Y_s} and $L_{Y_{s'}}$ are i.i.d drawn from the same distribution D, and I_Γ, I_U are the Fisher information matrix under the sampling distribution Γ and U, and (b) follows from Lemma 5 of Chaudhuri et al. (2015). Note that ActiveS is a two-stage process that samples according to the uniform distribution U to build an estimate of $\boldsymbol{\theta}^*$ and then solves an SDP to build the sampling proportion Γ that minimizes the quantity σ_U^2 and follows that sampling proportion Γ for the second stage. It follows that

$$\begin{aligned}
\sigma_U^2 &= \text{Trace}\left[I_U\left(\boldsymbol{\theta}^*\right) I_\Gamma\left(\boldsymbol{\theta}^*\right)^{-1}\right] \\
&\overset{(a)}{\leq} \text{Trace}\left[I_U\left(\boldsymbol{\theta}^*\right)\right]\text{Trace}\left[I_\Gamma\left(\boldsymbol{\theta}^*\right)^{-1}\right] \\
&\leq \lambda_1 d C_4 \eta
\end{aligned}$$

where, (a) follows as for any positive semi-definite matrices $\mathrm{Trace}[AB] \leq \mathrm{Trace}[A]\,\mathrm{Trace}[B]$, and for (b) we can show that

$$\begin{aligned} I_U(\boldsymbol{\theta}^*) &= \mathbb{E}_{I_s \sim U} \nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} L_{Y_s}(\boldsymbol{\theta}) = \mathbb{E}_{I_s \sim U} \nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (Y_s - \mu_{I_s}(\boldsymbol{\theta}))^2 \\ &= \mathbb{E}_{I_s \sim U} 2\left(Y_s - \mu_i(\boldsymbol{\theta}^*)\right) \nabla^2 \mu_{I_s}(\boldsymbol{\theta}^*) - 2\nabla \mu_{I_s}(\boldsymbol{\theta}^*) \nabla \mu_{I_s}(\boldsymbol{\theta}^*)^T \\ &= 2\sum_{i=1}^{n} p_U(i) \left[\left(Y_s - \mu_i(\boldsymbol{\theta}^*)\right) \nabla^2 \mu_{I_s}(\boldsymbol{\theta}^*) - \nabla \mu_{I_s}(\boldsymbol{\theta}^*) \nabla \mu_{I_s}(\boldsymbol{\theta}^*)^T\right] \end{aligned}$$

This leads to the following bound on the trace of the matrix $I_U(\boldsymbol{\theta}^*)$

$$\mathrm{Trace}[I_U(\boldsymbol{\theta}^*)] \leq \mathrm{Trace}[\sum_{i=1}^{n} p_U(i) \left(Y_s - \mu_i(\boldsymbol{\theta}^*)\right) \nabla^2 \mu_{I_s}(\boldsymbol{\theta}^*)] \leq \lambda_1 d C_4 \sqrt{\eta}$$

where $\left(Y_s - \mu_i(\boldsymbol{\theta}^*)\right) \nabla^2 \mu_{I_s}(\boldsymbol{\theta}^*) \leq \lambda_1 d C_4 \sqrt{\eta}$ and similarly,

$$I_\Gamma(\boldsymbol{\theta}^*) = \mathbb{E}_{I_s \sim \Gamma} \nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} L_{Y_s}(\boldsymbol{\theta}) = \mathbb{E}_{I_s \sim \Gamma} \nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (Y_s - \mu_{I_s}(\boldsymbol{\theta}))^2.$$

which leads to an upper bound for $\mathrm{Trace}[I_\Gamma(\boldsymbol{\theta}^*)^{-1}] \leq \lambda_1 d C_4 \sqrt{\eta}$. Plugging this in the statement of the result in eq. (4.15) we get that

$$\mathbb{E}\left[L_U\left(\widehat{\boldsymbol{\theta}}(t)\right) - L_U\left(\theta^*\right)\right] \leq O\left(\frac{d\sqrt{\log(dt)}}{t} + \frac{R}{t^2}\right).$$

Comparing this to our result in Theorem 6 we have the following convergence rate

$$\mathbb{E}\left[P_t(\widehat{\boldsymbol{\theta}}_t) - P_t\left(\boldsymbol{\theta}^*\right)\right] \leq \left(1 + \epsilon_t\right) \frac{\sigma^2}{t} + \frac{R}{t^2} \tag{4.17}$$

where P_t is a worst case measure over the data points. Next, we can show that,

$$
\begin{aligned}
\sigma^2 := \left\|\nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)\right\|^2_{\left(\nabla^2 P_t\left(\boldsymbol{\theta}^*\right)\right)^{-1}} &= \nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)^T \nabla^2 P_t\left(\boldsymbol{\theta}^*\right)^{-1} \nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right) \\
&\overset{(a)}{\leq} \lambda_{\max}\left(\nabla^2 P_t\left(\boldsymbol{\theta}^*\right)^{-1}\right) \|\nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)\|^2 \\
&\overset{(b)}{\leq} \lambda_{\max}\left(\nabla^2 P_t\left(\boldsymbol{\theta}^*\right)^{-1}\right) dC_4\eta \\
&= \lambda_{\min}\left(\nabla^2 P_t\left(\boldsymbol{\theta}^*\right)\right) dC_4\eta \\
&= \lambda_{\min}\left(\frac{2}{t}\sum_{s=1}^{t}\sum_{i=1}^{n} p_{\widehat{\boldsymbol{\theta}}_s}(i)\nabla\mu_{I_s}(\boldsymbol{\theta}^*)\nabla\mu_{I_s}(\boldsymbol{\theta}^*)^T\right) dC_4\eta \\
&\overset{(c)}{\leq} \left(\frac{2}{t}\sum_{s=1}^{t}\lambda_{\min}\left(\sum_{i=1}^{n} p_{\widehat{\boldsymbol{\theta}}_s}(i)\nabla\mu_{I_s}(\boldsymbol{\theta}^*)\nabla\mu_{I_s}(\boldsymbol{\theta}^*)^T\right)\right) dC_4\eta \\
&\leq 2\lambda_1 dC_4\eta
\end{aligned}
$$

where, (a) follows from the min-max theorem (variational characterization of the maximum eigenvalue), in (b) the quantity $\|\nabla \widehat{P}_t\left(\boldsymbol{\theta}^*\right)\|^2 \leq d^2 C_4\eta$ by assumption and C_4 is a constant, and (c) follows from triangle inequality. Plugging this in our result in eq. (4.17) we get

$$
\mathbb{E}\left[P_t(\widehat{\boldsymbol{\theta}}_t) - P_t\left(\boldsymbol{\theta}^*\right)\right] \leq O\left(\frac{d\sqrt{\log(dt)}}{t} + \frac{R}{t^2}\right).
$$

The updated convergence result is summarized below in this table:

Sample Complexity Bound	**Comments**
$\mathbb{E}\left[L_U\left(\widehat{\boldsymbol{\theta}}(t)\right) - L_U\left(\theta^*\right)\right] \leq O\left(d\frac{\sqrt{\log(dt)}}{t} + \frac{R}{t^2}\right)$	(Chaudhuri et al., 2015) L_U is loss under uniform measure over data points in pool.
$\mathbb{E}\left[P_t(\widehat{\boldsymbol{\theta}}_t) - P_t\left(\boldsymbol{\theta}^*\right)\right] \leq O\left(d\frac{\sqrt{\log(dt)}}{t} + \frac{R}{t^2}\right)$	GC (Ours) P_t is loss under a worst-case measure over data points.

Table 4.1: Active Regression comparison.

The two upper bounds have the same scaling, even though P_t is a different loss measure than L_U. The proof has steps similar to that of Chaudhuri et al. (2015), with some additional arguments to handle the fact that our loss measure varies with time.

4.6 Applications of Active Regression

In this section we present several real-world applications of GC in active regression setting.

4.6.1 Active Regression Experiment for Non-linear Reward Model

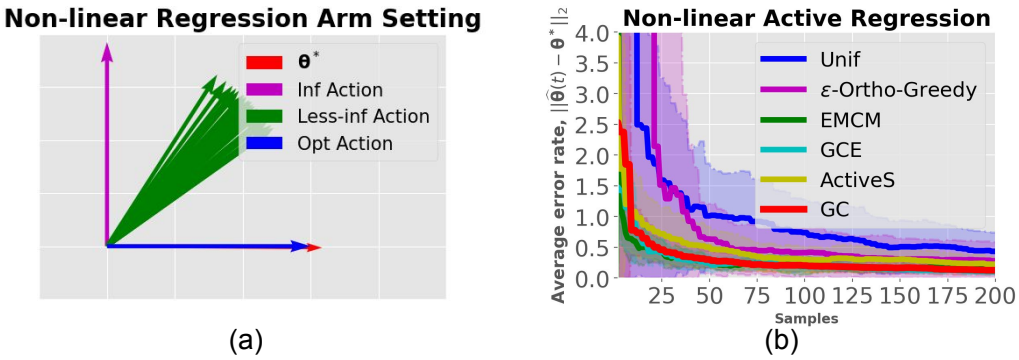

Figure 4.1: **(a)** The action feature vectors in $\mathbb{R}^2$ for active regression experiment. Inf = Informative action, Opt = Optimal action, Less-inf = Less informative action. **(b)** Shows average error rate over 50 trials for Active Regression.

Consider a non-linear class of functions parameterized as $\mu_i(\boldsymbol{\theta}) := 1/(1+\exp(-\mathbf{x}_i^T\boldsymbol{\theta}))$, where $\|\boldsymbol{\theta}\|_2 = 1$, each action has an associated feature vector $\mathbf{x}_i \in \mathbb{R}^2$ and it returns a value whose expectation is $\mu_i(\boldsymbol{\theta}^*)$. The goal is to choose actions such that the estimation error $\|\widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}^*\|_2$ reduces using as few samples as possible. The setting consist of three groups of actions: **a)** the optimal action, **b)** the informative action (orthogonal to optimal action) that maximally reduces the uncertainty of $\widehat{\boldsymbol{\theta}}(t)$ and **c)** the 48 less-informative actions as shown in Figure 4.1a. Appendix 4.6.1 contains more implementation details. We apply GC and GCE to this problem and compare it to baselines Unif , ϵ-Ortho-Greedy , ActiveS . Note that GCE uses the sampling rule in Theorem 5

with $1 - \epsilon_t$ probability or samples a random action with ϵ_t probability. Figure 4.2b shows that GC outperforms ActiveS and is able to find $\boldsymbol{\theta}^*$ quickly. The ϵ-Ortho-Greedy baseline chooses the most orthogonal action to its least square estimate $\widehat{\boldsymbol{\theta}}(t)$ at each round for maximally reducing the uncertainty at $\widehat{\boldsymbol{\theta}}(t)$.

Algorithmic Details: We describe each of the algorithm used in this setting in detail as follows:

1. **EMCM :** The EMCM algorithm of Cai et al. (2016) first quantifies the change as the difference between the current model parameters and the new model parameters learned from enlarged training data, and then chooses the data examples that result in the greatest change.

2. **ϵ-Ortho-Greedy :** The ϵ-Ortho-Greedy algorithm with $(1 - \epsilon)$ probability at each round chooses the most orthogonal action to its least square estimate $\widehat{\boldsymbol{\theta}}(t)$ to maximally reduce the uncertainty of $\widehat{\boldsymbol{\theta}}(t)$. Additionally it explores other actions with ϵ probability. We set $\epsilon = 0.1$.

3. **GC :** The GC policy used is stated as in Algorithm 1. To calculate the least square estimate $\widehat{\boldsymbol{\theta}}(t)$ we use the python scipy.optimize least-square function which solves a nonlinear least-squares problem.

4. **Unif :** The Unif policy samples each action uniform randomly at every round.

5. **ActiveS :** The ActiveS policy in Chaudhuri et al. (2015) is a two-stage algorithm. It first samples all actions uniform randomly to build an initial estimate of $\boldsymbol{\theta}^*$. It then solves an Semi-definite Programming (SDP) to obtain a new sampling distribution that minimizes the quantity σ_U^2 as defined in Equation (4.16). In the second stage ActiveS follows this new sampling distribution to sample actions.

Implementation Details: This setting consist of 50 measurement actions divided into three groups. The first group consist of the optimal action $\mathbf{x}_{i^*} := (1, 0)$ in the direction of $\boldsymbol{\theta}^* := (1, 0)$. The second group consist of the informative action $\mathbf{x}_2 := (0, 1)$ which is orthogonal to $\mathbf{x}_{i^*}$ and selecting it maximally reduces the uncertainty of $\widehat{\boldsymbol{\theta}}(t)$. Finally the third group consist of 48 actions such that $\mathbf{x}_i := (0.71 \pm \epsilon_i, 0.71 \mp \epsilon_i)$ for $i \in [3, 50]$ where ϵ_i is a small value close to 0 and

$\epsilon_i \neq \epsilon_{i'}$. Note that these 48 actions are less informative in comparison to action 2. This is shown in Figure 4.1a.

4.6.2 Active Regression Experiment for Neural Networks

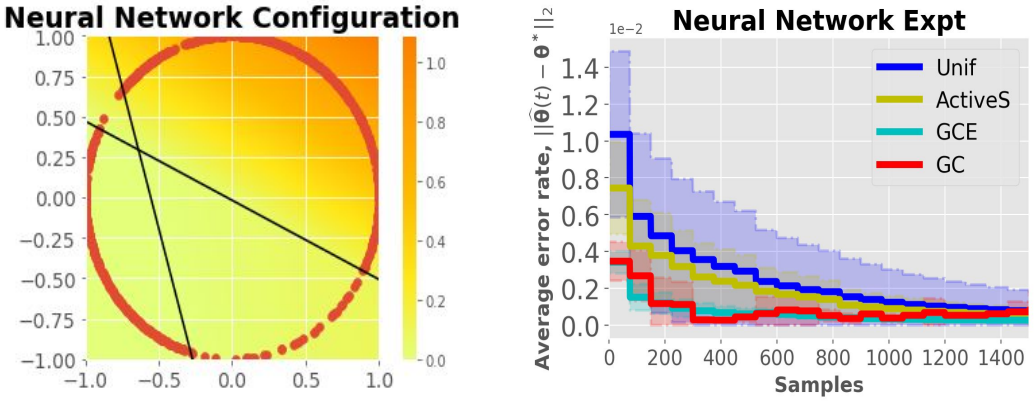

Figure 4.2: **(a)** Training data for neural network example shown as red dots. Yellow to red colors indicate increasing values for the target mean function. Black lines denote "activation" boundaries of the two hidden layer neurons. **(b)** Learning a neural network model. Plots show mean $\pm$ std. dev. of estimation error over 10 trials.

Consider a collection of data points $\{\mathbf{x}_i \in \mathbb{R}^2 : i \in [n]\}$, each of which is assigned a ground truth scalar mean value by a non-linear function. The particular form for the mean function of action i is the following: $\mu_i(\boldsymbol{\theta}^*) = c_1\sigma(\mathbf{w}_1^T\mathbf{x}_i + b_1) + c_2\sigma(\mathbf{w}_2^T\mathbf{x}_i + b_2)$, where $\boldsymbol{\theta}^* = (\mathbf{w}_1, b_1, \mathbf{w}_2, b_2, c_1, c_2)$ is the parameter characterizing the ground truth function, and $\sigma(\cdot) := \max\{0, \cdot\}$ is the non-linear ReLU activation function. This is a single hidden-layer neural network with input layer weights $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^2$, biases $b_1, b_2 \in \mathbb{R}$, and output weights $c_1, c_2 \in \{-1, 1\}$. Our objective in the experiment is to learn the neural network from noisy observations: $\{(\mathbf{x}_{I_s}, \mu_{I_s}(\boldsymbol{\theta}^*) + \epsilon) : s \in [t], \epsilon \text{ is Gaussian noise}\}$ collected by sampling $(I_1, I_2, \ldots, I_t)$ according to the Chernoff proportions defined in Theorem 5. The architecture of the network is known, but the weights and biases must be learned. The data

points are shown in a scatter plot in Figure 4.2a. More implementation details are in Section 4.6.2. The non-uniformity of the data distribution increases the difficulty of the learning task. The performance of a learning algorithm is measured by tracking the estimation error $\|\widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}^*\|_2$ during the course of training. The plot in Figure 4.2b shows the average and standard deviation of the estimation error over 10 trials for GC , GCE , ActiveS and Unif baseline. Unif uses $8\times$ more samples to achieve accuracy comparable to the active learning method. We note that other works (Cohn, 1996; Fukumizu, 2000) have also considered active training of neural networks. Other approaches for active sampling have been described in the survey by Settles (2009). Our neural network learning experiment is intended to show the generality of our approach in active sampling.

Implementation Details: At every time step, we use the the least squares optimizer of scipy to find $\widehat{\boldsymbol{\theta}}_t$. Since $c_1, c_2 \in \{-1, 1\}$, we solve four different least squares problems at each step corresponding to all (c_1, c_2) choices, and use the values returned by the problem having the smallest sum of squares as our current estimate for $(\mathbf{w}_1, \mathbf{w}_2, b_1, b_2)$. The derivative with respect to any parameter is found by the backward pass of automatic differentiation.

4.6.3 Active Regression for the UCI Datasets

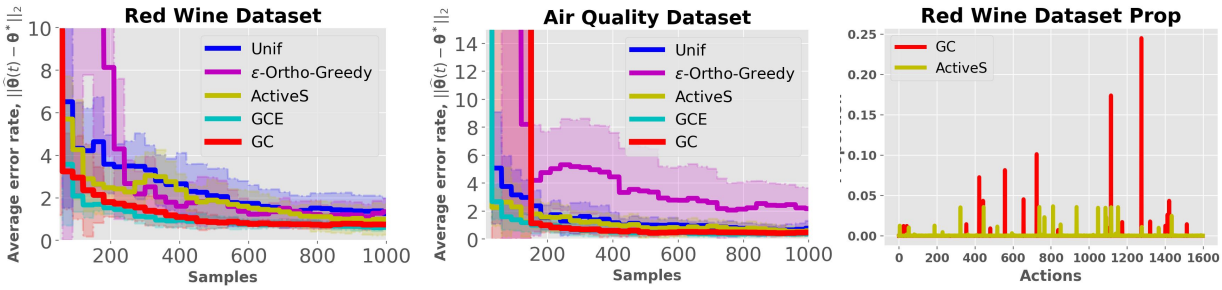

Figure 4.3: **(a)** Experiment with Red Wine Dataset, **(b)** Experiment with Air Quality Dataset. **(c)** GC Proportions over 1600 actions in Red Wine Dataset.

Finally we show the performance of GC , GCE against ActiveS and Unif in two real world datasets from UCI called Red Wine (Cortez et al., 2009) (1600 actions) and Air Quality (De Vito et al., 2008) (1500 actions). The performance is shown in Figure 4.3a and Figure 4.3b respectively where GC and GCE outperforms ActiveS and Unif . We also show in Figure 4.3c that in the real-world dataset GC proportion is actually sparse which validates our theory that Chernoff proportion should be sparse.

Implementation Details: The UCI Red Wine Quality dataset consist of 1600 samples of red wine with each sample i having feature $\mathbf{x}_i \in \mathbb{R}^{11}$. We first fit a least square estimate to the original dataset and get an estimate of $\boldsymbol{\theta}^*$. The reward model is linear and given by $\mathbf{x}_{I_t}^T \boldsymbol{\theta}^* + \epsilon_t$ where x_{I_t} is the observed action at round t, and ϵ_t is a zero-mean additive noise. Note that we consider the 1600 samples as actions. Then we run each of our benchmark algorithms on this dataset and reward model and show the result in Figure 4.3a. We further show the GC proportion on this dataset in fig. 4.3c and show that it is indeed sparse with proportion concentrated on few actions. The Air quality dataset consist of 1500 samples each of which consist of 6 features. We again build an estimate of $\boldsymbol{\theta}^*$ by fitting a least square regression on this dataset. We use a similar additive noise linear reward model as described before and run all the benchmark algorithms on this dataset.

In this chapter we look into how to extend the GC policy into smoothly parameterized hypotheses space. We further modified the GC policy from active testing to work in the active regression setting and derived its sample complexity bounds. In the next chapter we show how to use the GC policy for active regression to identify the best-arm in the multi-armed bandit setting.

Chapter 5

Active Regression in Bandits

A well-studied objective in many bandit problems is to find the best-arm using as few samples as possible. In this section we describe how we can use the asymptotically efficient exploration strategy of GC to find the best-arm in general structured bandits. We first focus on bandit environments where the collection of arm means $\{\mu_i(\boldsymbol{\theta}^*) : i \in [n]\}$ can be one of J different known sets, where $J := |\boldsymbol{\Theta}|$. By thinking of the unknown $\boldsymbol{\theta}^*$ as taking one of J possible values, it naturally fits the framework of the previous section. If $\boldsymbol{\theta}^*$ is correctly identified, then all arm means are known and we can identify the best-arm. However, as described in the Introduction, finding $\boldsymbol{\theta}^*$ is sufficient but not necessary to find the best arm. Our approach is to sample according to the exploration suggested by GC , and modify the stopping condition so as to exit when the best arm has been identified. Let $i^*(\boldsymbol{\theta})$ denote the arm with the highest expected reward under hypothesis $\boldsymbol{\theta}$. We let τ_δ denote the stopping time under the modified stopping rule that will be described later. Let $\widehat{\boldsymbol{\theta}}(\tau_\delta)$ be the estimate of $\boldsymbol{\theta}^*$ at τ_δ. Then, our policy π must satisfy the δ-PAC criteria for best-arm identification such that $\mathbb{P}(i^*(\widehat{\boldsymbol{\theta}}(\tau_\delta)) \neq i^*(\boldsymbol{\theta}^*)) \leq \delta$.

5.1 Stopping time for GC in Bandit setting

The stopping condition in line 6 of Algorithm 1 has the following interpretation. It is satisfied when we can reject all hypotheses other than $\widehat{\boldsymbol{\theta}}(t)$ from being $\boldsymbol{\theta}^*$. For finding the best-arm, we don't need to reject all competing hypotheses, but only those which have a different best-arm. This intuitively suggests the following modified stopping rule. Let $\boldsymbol{Alt}(\boldsymbol{\theta}) := \{\boldsymbol{\theta}' \in \Theta : i^*(\boldsymbol{\theta}') \neq i^*(\boldsymbol{\theta})\}$, and replace line 6 by

$$\begin{aligned} &\textbf{if } L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\widehat{\boldsymbol{\theta}}(t)) > \beta(J, \delta)\ \forall \boldsymbol{\theta}' \in \boldsymbol{Alt}(\widehat{\boldsymbol{\theta}}(t)) \textbf{ then} \\ &\quad \textbf{Return } i^*(\widehat{\boldsymbol{\theta}}(t)) \text{ as the best arm.} \end{aligned} \tag{5.1}$$

We now prove the following proposition on the stopping condition stated in (5.1).

Proposition 2. *Define τ_δ to be the first time when the stopping condition in* (5.1) *is met. Let $\widehat{i}(\tau_\delta) := \arg\max_{i \in [n]} \mu_i(\widehat{\boldsymbol{\theta}}(\tau_\delta))$ be the best-arm the δ-PAC policy GC suggests at τ_δ. Then $\mathbb{P}(\widehat{i}(\tau_\delta) \neq i^*(\boldsymbol{\theta}^*)) \leq \delta$.*

Remark 2. (Sample complexity) Since (5.1) is a weaker condition than that in line 6, the modified stopping time is less than the stopping time of the original GC with probability one. Hence an upper bound to the sample complexity of best-arm identification by GC is given by Theorem 2.

5.1.1 Proof of Proposition 2

Proof. Recall that $\boldsymbol{Alt}(\boldsymbol{\theta}^*) := \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : i^*(\boldsymbol{\theta}) \neq i^*(\boldsymbol{\theta}^*)\}$. Also note that at round τ_δ the policy GC suggests $\widehat{i}(\tau_\delta) := \arg\max_{i\in[n]} \mu_i(\widehat{\boldsymbol{\theta}}(\tau_\delta))$ as the optimal arm. Note that GC for best-arm identification stops at τ_δ when the event $\xi^\delta(\boldsymbol{\theta}) := \{L_{Y^{\tau_\delta}}(\boldsymbol{\theta}') - L_{Y^{\tau_\delta}}(\boldsymbol{\theta}) > \beta(J,\delta), \forall \boldsymbol{\theta}' \in \boldsymbol{Alt}(\boldsymbol{\theta})\}$ is true. We follow a similar proof technique as in Lemma 7.10. Let for some $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$ we define $T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) := L_{Y^t}(\boldsymbol{\theta}^*) - L_{Y^t}(\boldsymbol{\theta})$. Following this we can show that the probability of the event $\{\widehat{i}(t) \neq i^*(\boldsymbol{\theta}^*)\}$ is upper bounded by

$$\begin{aligned}\mathbb{P}(\exists \tau_\delta < \infty, \widehat{i}(\widehat{\boldsymbol{\theta}}(\tau_\delta)) \neq i^*(\boldsymbol{\theta}^*)) &\leq \mathbb{P}(\exists \tau_\delta < \infty, T_{Y^{\tau_\delta}}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > \beta(J,\delta)) \\ &\overset{(a)}{\leq} \mathbb{P}(\exists \tau_\delta < \infty, T_{Y^{\tau_\delta}}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > \beta(J,\delta)) \\ &\overset{(b)}{\leq} \exp\left(-\beta(J,\delta)\right) \overset{(c)}{\leq} \frac{\delta}{J}.\end{aligned}$$

where, (a) follows as the event

$$\begin{aligned}\{T_{Y^{\tau_\delta}}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}(\tau_\delta)) > \beta(J,\delta), \forall \boldsymbol{\theta} \in \boldsymbol{Alt}(\widehat{\boldsymbol{\theta}}(\tau_\delta))\} &= \bigcap_{\phi \in \boldsymbol{Alt}(\widehat{\boldsymbol{\theta}}(\tau_\delta))} \{T_{Y^{\tau_\delta}}(\phi, \widehat{\boldsymbol{\theta}}(\tau_\delta)) > \beta(J,\delta) : \phi \in \boldsymbol{Alt}(\widehat{\boldsymbol{\theta}}(\tau_\delta))\} \\ &\subseteq \{T_{Y^{\tau_\delta}}(\boldsymbol{\theta}^*, \widehat{\boldsymbol{\theta}}(\tau_\delta)) > \beta(J,\delta)\}\end{aligned}$$

(b) follows from Lemma 7.10, and (c) follows from (7.7). Then we define $\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} := \min\{t : T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > \beta(J,\delta)\}$. Then following the same steps as in Step 5 of Theorem 2 we can show that

$$\begin{aligned}\mathbb{P}\left(\tau_\delta < \infty, \widehat{i}(\widehat{\boldsymbol{\theta}}(\tau_\delta) \neq i^*(\boldsymbol{\theta}^*)\right) &\leq \mathbb{P}\left(\exists \widehat{\boldsymbol{\theta}}(\tau_\delta) \in \boldsymbol{Alt}(\boldsymbol{\theta}^*), \exists t \in \mathbb{N} : T_{Y^t}(\boldsymbol{\theta}^*, \widehat{\boldsymbol{\theta}}(\tau_\delta)) > \beta(J,\delta)\right) \\ &\leq \sum_{\widehat{\boldsymbol{\theta}}(\tau_\delta) \in \boldsymbol{Alt}(\boldsymbol{\theta}^*)} \mathbb{P}\left(\xi^\delta(\widehat{\boldsymbol{\theta}}(\tau_\delta)), \tau_{\boldsymbol{\theta}^*\widehat{\boldsymbol{\theta}}(\tau_\delta)} < \infty\right) \overset{(a)}{\leq} \sum_{\widehat{\boldsymbol{\theta}}(\tau_\delta) \in \boldsymbol{Alt}(\boldsymbol{\theta}^*)} \frac{\delta}{J} \leq \delta\end{aligned}$$

where (a) follows follows from Lemma 7.10 and the definition of $\beta(J,\delta)$. The claim of the lemma follows. □

5.2 Best-arm ID when Θ is dense

Many widely-studied bandit environments correspond to the case when $\boldsymbol{\Theta}$ is compact set, such as a subset of Euclidean space. The stopping rule we give in (5.1) considers all alternatives $\boldsymbol{\theta}'$ for which the best arm is different from that under $\widehat{\boldsymbol{\theta}}$. When Θ is compact, we cannot enumerate all $\boldsymbol{\theta}'$. However, if the noise distribution belongs to the exponential family, we can perform a generalized likelihood ratio test, which only requires us to compare $\widehat{\boldsymbol{\theta}}$ against the "closest" competing $\boldsymbol{\theta}'$ that has a different best-arm. This step is carried out in Line 6 of Algorithm 2 and corresponds to a least squares optimization. If the partitions of $\boldsymbol{\Theta}$ corresponding to a fixed best arm are convex sets (e.g. in linear bandits), then we can efficiently check the stopping condition using convex optimization. The threshold to be compared against in the stopping condition is modified to be the following:

$$\beta(t, \delta) = \log(Ct^2/\delta), \tag{5.2}$$

where $C > 0$ is a constant. This choice of $\beta(t, \delta)$ has been previously studied under exponential family model of noise by Garivier and Kaufmann (2016).

Algorithm 2 GC for best-arm ID: dense $\boldsymbol{\Theta}$

1: **Input:** Confidence parameter δ, $\beta(t, \delta)$ from (5.2).
2: Sample $I_1 \in [n]$ randomly, observe Y_1 and find $\widehat{\boldsymbol{\theta}}(1)$.
3: **for** $t = 2, 3, \ldots$ **do**
4: Sample $I_t \sim \mathbf{p}_{\widehat{\boldsymbol{\theta}}(t-1)}$ from Theorem 5 and get Y_t.
5: **for** $\{i \in [n] : i \neq i^*(\widehat{\boldsymbol{\theta}}(t))\}$ **do**
6: Find $\boldsymbol{\theta}'_i := \arg\min_{\{\boldsymbol{\theta} \in \boldsymbol{\Theta} : i^*(\boldsymbol{\theta}) = i\}} L_{Y^t}(\boldsymbol{\theta})$.
7: **if** $L_{Y^t}(\boldsymbol{\theta}'_i) - L_{Y^t}(\widehat{\boldsymbol{\theta}}(t)) > \beta(t, \delta)$ for all $\{\boldsymbol{\theta}'_i : i \neq i^*(\widehat{\boldsymbol{\theta}}(t))\}$ **then**
8: **Return** $i^*(\widehat{\boldsymbol{\theta}}(t))$ as the best-arm.

5.3 Best-arm Identification in linear bandit

For the case when $\mu_i(\boldsymbol{\theta}^*) = \mathbf{x}_i^T \boldsymbol{\theta}^*$, we can observe that $\mathbf{p}_{\widehat{\boldsymbol{\theta}}(t)}$ in Theorem 5 does not depend on $\widehat{\boldsymbol{\theta}}(t)$, thus the sampling proportions are not adaptive. We restrict $\|\mathbf{x}_i\|_2 = 1$ and the distribution

$\nu_i(Y;\boldsymbol{\theta})$ is Gaussian for all $i \in [n]$. Define the smallest distance between any pair of arms as

$$\Delta_{\min} = \min_{\mathbf{x},\mathbf{x}' \in [n]} \|\mathbf{x} - \mathbf{x}'\|_2. \tag{5.3}$$

Using this definition of $\Delta_{\min}$ the stopping condition threshold $\beta(t,\delta)$ for linear bandits is chosen as follows:

$$\beta(t,\delta) = \log(4t^2/\delta) + d\log(12/\Delta_{\min}), \tag{5.4}$$

such that C in (5.2) is equal to $(12/\Delta_{\min})^d$. Note that the algorithm knows $\Delta_{\min}$ as the arm features are known.

In the following theorem, we give a sample complexity bound for Algorithm 2 when it is used to identify the best-arm in a linear bandit environment. We approach this proof from the perspective of model ID in a discretized parameter space. There exists a $\gamma \in [0, 1/2)$ such that $\boldsymbol{\theta}^* = \mathbf{x}^* + \gamma(\mathbf{x}' - \mathbf{x}^*)$, where $\mathbf{x}^*$ is the optimal arm and $\mathbf{x}'$ is the closest arm to $\mathbf{x}^*$. Intuitively, γ indicates how close $\boldsymbol{\theta}^*$ is to the boundary of the best-arm partition between $\mathbf{x}^*$ and $\mathbf{x}'$. Define the distance between $\boldsymbol{\theta}^*$ and the boundary as:

$$\Delta^* := \|\boldsymbol{\theta}^* - (\mathbf{x}^* + 0.5(\mathbf{x}' - \mathbf{x}^*))\|_2. \tag{5.5}$$

We discretize the parameter space $\boldsymbol{\Theta}$ into an ϵ-cover $\widetilde{\boldsymbol{\Theta}}$ and then follow the proof technique of Theorem 2 on $\widetilde{\boldsymbol{\Theta}}$ to derive the bound. We show in Lemma 7.22 in the Appendix 7.5.2 that the choice of $\beta(t,\delta)$ in (5.4) leads to δ-PAC correctness for this choice of ϵ-cover. The proof of Theorem 6 is given below.

Theorem 6. *Let τ_δ be the stopping time of GC in Algorithm 2 for best-arm ID in a linear bandit. There exists a discrete subset $\widetilde{\boldsymbol{\Theta}} \subset \boldsymbol{\Theta}$ such that for all $\widetilde{\boldsymbol{\theta}} \in \widetilde{\boldsymbol{\Theta}}$, the p.m.f. $\mathbf{p}_{\widetilde{\boldsymbol{\theta}}}$ in (1.3) is equal to the p.m.f. $\mathbf{p}_{\widehat{\boldsymbol{\theta}}(t)}$ used by Algorithm 2 for model identification in $\widetilde{\boldsymbol{\Theta}}$ at all t. $\widetilde{\boldsymbol{\Theta}}$ also satisfies the condition that $\widetilde{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}' \geq \epsilon$ for all $\widetilde{\boldsymbol{\theta}} \neq \widetilde{\boldsymbol{\theta}}'$ and some $\epsilon > 0$. For a constant $C = O((\eta/\eta_0)^2)$ the sample complexity of Algorithm 2 is*

$$\mathbb{E}[\tau_\delta] \leq O\left(\frac{d\eta\log(C)\log(\Delta^{-1})}{\tilde{D}_1} + \frac{\log(1/\delta)}{\tilde{D}_0} + \frac{C^{1/\eta}\delta^{\tilde{D}/\eta^2}}{\Delta^d}\right)$$

where $\tilde{D}_1, \tilde{D}_0$ are defined as in Theorem 2 for $\widetilde{\boldsymbol{\Theta}}$, $\Delta_{\min}, \Delta^*$, are defined in (5.3), (5.5), and $\Delta = \min\{\Delta^*, \Delta_{\min}\}$.

5.3.1 Proof of Main Theorem

Proof. **Step 0 (Discretize Θ):** Θ is the dense parameter space and GC in Algorithm 2 runs in that space. We can imagine a hypothetical model identification experiment playing out over a discrete ϵ-cover of Θ, called $\tilde{\Theta}$, as our algorithm runs. We can simulate the imaginary experiment only using the realizations of our real algorithm acting on Θ, in the following manner. Every point in $\tilde{\Theta}$ is also present in Θ and has a squared error value from the real experiment. That determines the most likely $\widehat{\tilde{\boldsymbol{\theta}}}$ in the imaginary experiment. Whatever sample our real algorithm takes, can also be thought of as the sample that would have been taken in the imaginary experiment, since the sampling proportions for both will be shown next to be the same. Finally, our real algorithm stops (at time τ_δ) when the difference of squared errors between $\widehat{\boldsymbol{\theta}}$ and any $\boldsymbol{\theta}'$ which has a different best arm, crosses a threshold. If we choose cover radius ϵ to be small enough, so that all the neighboring points in $\tilde{\Theta}$ around $\widehat{\boldsymbol{\theta}}$ have the same best arm as that at $\widehat{\boldsymbol{\theta}}$, then we can be sure that our imaginary model identification algorithm would not have stopped at time τ_δ because the difference of errors between $\widehat{\boldsymbol{\theta}}$ and a competing $\boldsymbol{\theta}'$ wouldn't have crossed the threshold (this is due to convexity of difference of squared error function).

Let the static proportion used by the algorithm be denoted as $\mathbf{p}_{\text{lin}}$. Let $\mathbf{X}$ be a matrix in $\mathbb{R}^{d\times t}$ whose columns contain $tp_{\text{lin}}(i)$ copies of arm vector $\mathbf{x}_i$ for all $i \in [n]$. Define $\|\mathbf{x}\|_{\mathbf{S}} = \sqrt{\mathbf{x}^T\mathbf{S}\mathbf{x}}$. Then we discretize Θ by obtaining its ϵ-cover $\tilde{\Theta}$ for small enough ϵ, such that for any $\tilde{\boldsymbol{\theta}} \in \tilde{\Theta}$ all parameters $\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\|_{\mathbf{S}} \leq \epsilon$ close to it have the same best arm, i.e., $i^*(\boldsymbol{\theta}) = i^*(\widetilde{\boldsymbol{\theta}})$. In the linear bandit setting when $\mu_i(\boldsymbol{\theta}) = \mathbf{x}_i^T\boldsymbol{\theta}$ we can set $\epsilon = {}^{\Delta^*}/_{2-2\gamma}$ where $\gamma \in [0, {}^1/_2)$ and Δ^* is defined in (5.5). From Theorem 5, we see that the sampling proportion employed by Algorithm 2 for finding the best arm in a linear bandit is static, i.e., it doesn't depend on the current value of $\widehat{\boldsymbol{\theta}}(t)$. We can choose $\tilde{\Theta}$ so that the sampling proportion for model identification over $\tilde{\Theta}$ is the same as the static proportion used in the Linear bandit experiment. This choice can be made by solving a system of equations corresponding to the constraints of the optimization in (1.4).

Also note that in the linear setting, the sum of squared errors $L_{Y^t}(\boldsymbol{\theta})$ for any $\boldsymbol{\theta} \in \Theta$ is convex which we show in (5.9). Finally we define the complexity parameters $\tilde{D}_0$ and $\tilde{D}_1$ as in Theorem 2 but for $\widetilde{\Theta}$.

Step 1 (Definitions): Define $L_{Y^t}(\widetilde{\boldsymbol{\theta}}')$ as the total sum of squared errors of the hypothesis parameterized by $\widetilde{\boldsymbol{\theta}}'$ till round t. Let, the difference of sum of squared errors between the hypotheses parameterized by $\widetilde{\boldsymbol{\theta}}'$ and $\widetilde{\boldsymbol{\theta}}^*$ till round t be denoted by $T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*) = L_{Y^t}(\widetilde{\boldsymbol{\theta}}') - L_{Y^t}(\widetilde{\boldsymbol{\theta}}^*)$. Note that both $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}'$ belong to the discretized parameter space $\tilde{\boldsymbol{\Theta}}$. Also note that the p.m.f. $\mathbf{p}_{\widetilde{\theta}'}$ is the Chernoff verification proportion for verifying hypothesis parameterized by $\widetilde{\boldsymbol{\theta}}'$.

Step 2 (Define stopping time and partition): We define the stopping time τ_δ for the policy π as follows:

$$\tau_\delta = \min\{t : \exists \widetilde{\boldsymbol{\theta}}' \in \tilde{\boldsymbol{\Theta}}, L_{Y^t}(\widetilde{\boldsymbol{\theta}}) - L_{Y^t}(\widetilde{\boldsymbol{\theta}}') > \beta(t,\delta), \forall \widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}\} \tag{5.6}$$

where, $\beta(t,\delta)$ is the threshold function defined in (7.7).

Step 3 (Define bad event): We define the bad event $\xi^\delta(\widetilde{\boldsymbol{\theta}}')$ for the sub-optimal hypothesis parameterized by $\widetilde{\boldsymbol{\theta}}'$ as follows:

$$\xi^\delta(\widetilde{\boldsymbol{\theta}}') = \{\widehat{\widetilde{\boldsymbol{\theta}}}(t) = \widetilde{\boldsymbol{\theta}}', L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}) - L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}') > \beta(t,\delta), \forall \widetilde{\boldsymbol{\theta}} \neq \widetilde{\boldsymbol{\theta}}'\}. \tag{5.7}$$

The event $\xi^\delta(\widetilde{\boldsymbol{\theta}}')$ denotes that a sub-optimal hypothesis parameterized by $\widetilde{\boldsymbol{\theta}}'$ is declared the optimal hypothesis when it has a smaller sum of squared errors than any other hypothesis parameterized by $\widetilde{\boldsymbol{\theta}} \neq \widetilde{\boldsymbol{\theta}}'$ at τ_δ. Next, we define the event

$$\{L_{Y^\tau}(\boldsymbol{\theta}_i') - L_{Y^\tau}(\widehat{\boldsymbol{\theta}}(t)) > \beta(t,\delta) : i \neq i^*(\widehat{\boldsymbol{\theta}}(t)), \boldsymbol{\theta}_i' := \arg\min_{\boldsymbol{\theta}: i^*(\boldsymbol{\theta}) = i} L_{Y^\tau}(\boldsymbol{\theta})\} \tag{5.8}$$

and show that in the dense hypotheses setting the (5.7) implies (5.8). To this effect we first state the Identifiability assumption required to show this implication.

Assumption 5. ***(Identifiability)*** *Let the static proportion used by the algorithm be denoted as* $\mathbf{p}_{lin}$*. Let* $\mathbf{X}$ *be a matrix in* $\mathbb{R}^{d\times t}$ *whose columns contain* $tp_{lin}(i)$ *copies of arm vector* $\mathbf{x}_i$ *for all* $i \in [n]$*. We assume that* ϵ *is sufficient enough such that*

$$\forall \boldsymbol{\theta} : \left\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right\|_{\mathbf{S}} \leq \epsilon, i^*(\boldsymbol{\theta}) = i^*(\widehat{\boldsymbol{\theta}}).$$

where, $\mathbf{S} = \frac{1}{t}\mathbf{X}\mathbf{X}^T$ *and* $\|\mathbf{x}\|_S = \sqrt{\mathbf{x}^T S \mathbf{x}}$*.*

It follows from Assumption 5 that the stopping time condition in (5.7) is

$$\{\widehat{\widetilde{\boldsymbol{\theta}}}(t) = \widetilde{\boldsymbol{\theta}}', L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}) - L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}') > \beta(t,\delta), \forall \widetilde{\boldsymbol{\theta}} \in\neq \widetilde{\boldsymbol{\theta}}'\}$$
$$\implies \{\widehat{\widetilde{\boldsymbol{\theta}}}(t) = \widetilde{\boldsymbol{\theta}}', L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}) - L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}') > \beta(t,\delta), \forall i^*(\widetilde{\boldsymbol{\theta}}) \neq i^*(\widetilde{\boldsymbol{\theta}}')\}.$$

Next we choose the cover radius $\epsilon = \frac{\Delta^*}{2-2\gamma}$ where $\gamma \in [0, 1/2)$ and Δ^* defined in (5.5). Because of this cover radius choice, we have that

$$\exists \tilde{\boldsymbol{\theta}}' : \|\tilde{\boldsymbol{\theta}}' - \widehat{\boldsymbol{\theta}}\|_{\mathbf{S}} \leq \epsilon \overset{(a)}{\leq} \|\boldsymbol{\theta}_i' - \widehat{\boldsymbol{\theta}}\|_{\mathbf{S}}$$

where, (a) follows by Assumption 5. Now note that $L_{Y^t}(\boldsymbol{\theta})$ is convex. This can be show as follows:

$$L_{Y^t}(\boldsymbol{\theta}) \overset{(a)}{=} \sum_{s=1}^{t} (\mathbf{x}_{I_s}^T \boldsymbol{\theta} - Y_s)^2 \textbf{(b)}{=} (\mathbf{X}^T\boldsymbol{\theta} - \mathbf{Y})^T(\mathbf{X}^T\boldsymbol{\theta} - \mathbf{Y}) = \boldsymbol{\theta}^T\mathbf{X}X^T\boldsymbol{\theta} - 2\boldsymbol{\theta}^T\mathbf{X}\mathbf{Y} + \mathbf{Y}^T\mathbf{Y} \tag{5.9}$$

where, (a) follows as $\mu_{I_s}(\boldsymbol{\theta}) = \mathbf{x}_{I_s}^T\boldsymbol{\theta}$, and (b) follows as $\mathbf{X}$ be a matrix in $\mathbb{R}^{d\times t}$ whose columns contain $tp_{\text{lin}}(i)$ copies of arm vector $\mathbf{x}_i$ for all $i \in [n]$ and $\mathbf{Y}$ is a vector of observed rewards in $\mathbb{R}^t$. Differentiating (5.9) twice with respect to $\boldsymbol{\theta}$ we get that $\frac{\partial^2(L_{Y^t}(\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}^2} = \mathbf{X}\mathbf{X}^T > 0$. Hence, $L_{Y^t}(\boldsymbol{\theta})$ is convex in $\boldsymbol{\theta}$. This means $L_{Y^t}(\boldsymbol{\theta}_i') \overset{(a)}{>} L_{Y^t}(\tilde{\boldsymbol{\theta}}') > L_{Y^t}(\widehat{\boldsymbol{\theta}}) + \beta(t,\delta)$. where, (a) follows by convexity of $L_{Y^t}(\boldsymbol{\theta})$. Hence we get that,

$$\{\widehat{\widetilde{\boldsymbol{\theta}}}(t) = \widetilde{\boldsymbol{\theta}}', L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}) - L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}') > \beta(t,\delta), \forall \widetilde{\boldsymbol{\theta}} \in\neq \widetilde{\boldsymbol{\theta}}'\} \implies \{L_{Y^\tau}(\boldsymbol{\theta}_i') - L_{Y^\tau}(\widehat{\boldsymbol{\theta}}(t)) > \beta(t,\delta) : i \neq i^*(\widehat{\boldsymbol{\theta}}(t))\},$$

where, $\boldsymbol{\theta}_i' := \arg\min_{\boldsymbol{\theta}: i^*(\boldsymbol{\theta})=i} L_{Y^\tau}(\boldsymbol{\theta})$.

Step 4 (Decomposition of bad event): In this step we decompose the bad event to show that only comparing $\widetilde{\boldsymbol{\theta}}'$ against $\widetilde{\boldsymbol{\theta}}^*$ is enough to guarantee a δ-PAC policy. Finally we decompose the bad event $\xi^\delta(\widetilde{\boldsymbol{\theta}}')$ as follows:

$$\begin{aligned}
\xi^\delta(\widetilde{\boldsymbol{\theta}}') =& \{\widehat{\widetilde{\boldsymbol{\theta}}}(\tau_\delta) = \widetilde{\boldsymbol{\theta}}', L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}) - L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}') > \beta(t,\delta), \forall \widetilde{\boldsymbol{\theta}} \neq \widetilde{\boldsymbol{\theta}}'\} \\
\overset{(a)}{=} & \underbrace{\{\widehat{\widetilde{\boldsymbol{\theta}}}(\tau_\delta) = \widetilde{\boldsymbol{\theta}}', L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}) - L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}') > \beta(t,\delta), \forall \widetilde{\boldsymbol{\theta}} \neq \widetilde{\boldsymbol{\theta}}'\}}_{\textbf{part A}} \\
& \bigcap \underbrace{\{\widehat{\widetilde{\boldsymbol{\theta}}}(\tau_\delta) = \widetilde{\boldsymbol{\theta}}', L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}^*) - L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}') > \beta(t,\delta)\}}_{\textbf{part B}} \\
\subseteq & \{\widehat{\widetilde{\boldsymbol{\theta}}}(\tau_\delta) = \widetilde{\boldsymbol{\theta}}', L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}^*) - L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}') > \beta(t,\delta)\}
\end{aligned} \tag{5.10}$$

where, (a) follows by decomposing the event in two parts containing $\widetilde{\boldsymbol{\theta}} \in \tilde{\boldsymbol{\Theta}} \setminus \{\widetilde{\boldsymbol{\theta}}^*\}$ and $\{\widetilde{\boldsymbol{\theta}}^*\}$, and (b) follows by noting that the intersection of events holds by taking into account only the event in part B.

Step 5 (Proof of correctness): In this step we want to show that based on the stopping time definition and the bad event $\xi^{\delta}(\widetilde{\boldsymbol{\theta}}')$ the GC stops and outputs the correct hypothesis $\widetilde{\boldsymbol{\theta}}^*$ with $1-\delta$ probability. As shown in Step 4, we can define the error event $\xi^{\delta}(\widetilde{\boldsymbol{\theta}}')$ as follows:

$$\xi^{\delta}(\widetilde{\boldsymbol{\theta}}') \subseteq \{\widehat{\widetilde{\boldsymbol{\theta}}}(\tau_\delta) = \widetilde{\boldsymbol{\theta}}', L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}^*) - L_{Y^{\tau_\delta}}(\widetilde{\boldsymbol{\theta}}') > \beta(t,\delta)\}$$

Define $\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'} = \min\{t : T_{Y^t}(\widetilde{\boldsymbol{\theta}}^*, \widetilde{\boldsymbol{\theta}}') > \beta(t,\delta)\}$. Then we can show for the stopping time τ_δ, round $\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}$ and the threshold $\beta(t,\delta)$ we have

$$\begin{aligned}
\mathbb{P}\left(\tau_\delta < \infty, \widehat{\widetilde{\boldsymbol{\theta}}}(\tau_\delta) \neq \widetilde{\boldsymbol{\theta}}^*\right) &\leq \mathbb{P}\left(\exists \widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*, \exists t \in \mathbb{N} : T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}_*) > \beta(t,\delta)\right) \\
&\leq \sum_{t=1}^{\infty} \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \mathbb{P}\left(L_{Y^{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}}}(\widetilde{\boldsymbol{\theta}}^*) - L_{Y^{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}}}(\widetilde{\boldsymbol{\theta}}') > \beta(t,\delta), \tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'} < \infty\right) \\
&\overset{(a)}{\leq} \sum_{t=1}^{\infty} |\mathcal{C}_\epsilon| (\exp\left(-\beta(t,\delta)\right) \\
&\overset{(b)}{\leq} \sum_{t=1}^{\infty} \left(\frac{3}{\epsilon}\right)^d (\exp\left(-\beta(t,\delta)\right) \overset{(c)}{=} \sum_{t=1}^{\infty} \left(\frac{12(\frac{1}{2}-\gamma)}{\Delta^*}\right)^d \frac{2\delta}{\left(\frac{12}{\Delta_{\min}}\right)^d 4t^2} \\
&\overset{(d)}{\leq} \sum_{t=1}^{\infty} \frac{\delta}{2t^2} \overset{(e)}{\leq} \frac{\pi^2\delta}{12} < \delta.
\end{aligned}$$

where, (a) follows from Lemma 7.22, (b) follows as $|\mathcal{C}_\epsilon| < (3/\epsilon)^d$ (see chapter 20 in Lattimore and Szepesvári (2020)), (c) follows by setting $\epsilon = \frac{\Delta^*}{4(\frac{1}{2}-\gamma)}$ where Δ^* is defined in (5.5) and

$$\beta(t,\delta) := \log(4t^2/\delta) + d\log(12/\Delta_{\min}), \tag{5.11}$$

(d) follows as $\Delta^*/(\frac{1}{2}-\gamma) \geq \Delta_{\min}$ and (e) follows from Basel's equation.

Step 6 (Sample complexity analysis): In this step we bound the total sample complexity satisfying the δ-PAC criteria. We define the stopping time τ_δ as follows:

$$\tau_\delta := \min\left\{t : L_{Y^t}(\widetilde{\boldsymbol{\theta}}') - L_{Y^t}(\widehat{\widetilde{\boldsymbol{\theta}}}(t)) > \beta(t,\delta), \forall \widetilde{\boldsymbol{\theta}}' \neq \widehat{\widetilde{\boldsymbol{\theta}}}(t)\right\}$$

We further define the stopping time $\tau_{\widetilde{\boldsymbol{\theta}}^*}$ for the hypothesis parameterized by $\widetilde{\boldsymbol{\theta}}^*$ as follows:

$$\tau_{\widetilde{\boldsymbol{\theta}}^*} := \min\left\{t : L_{Y^t}(\widetilde{\boldsymbol{\theta}}') - L_{Y^t}(\widetilde{\boldsymbol{\theta}}^*) \geq \beta(t,\delta), \forall \widetilde{\boldsymbol{\theta}}' \neq \widehat{\widetilde{\boldsymbol{\theta}}}(t)\right\} \tag{5.12}$$

We also define the critical number of samples M as follows:

$$M = (1+c)\left(\frac{\alpha(t)}{\tilde{D}_1} + \frac{\log(4(12/\Delta_{\min})^d t^2/\delta)}{\tilde{D}_0}\right) \tag{5.13}$$

where, $c > 0$ is a constant and $\tilde{D}_1, \tilde{D}_0$ are defined as in Theorem 2 for $\widetilde{\boldsymbol{\Theta}}$. Next, as in Garivier and Kaufmann (2016) we further define $t_0(\delta) = \inf\{t : \frac{\alpha(t)}{\tilde{D}_1} + \frac{\log(4(12/\Delta_{\min})^d t^2/\delta)}{\tilde{D}_0} < t\}$. It then follows that for every $t > t_0(\delta)$ the complement of the event $\xi^\delta(\widetilde{\boldsymbol{\theta}}')$ holds for all $\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*$. Then we can show that:

$$\begin{aligned}
\mathbb{E}[\tau_\delta] \leq \mathbb{E}[\tau_{\widetilde{\boldsymbol{\theta}}^*}] &= \sum_{t=0}^{\infty} \mathbb{P}(\tau_{\widetilde{\boldsymbol{\theta}}^*} > t) \overset{(a)}{\leq} 1 + (1+c)M + \sum_{t:t>(1+c)M} \mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) \\
&\overset{(b)}{\leq} 1 + (1+c)\left(\frac{\alpha(t_0(\delta))}{\tilde{D}_1} + \frac{\log(4(12/\Delta_{\min})^d t_0^2(\delta)/\delta)}{\tilde{D}_0}\right) \\
&\quad + |\mathcal{C}_\epsilon| \sum_{t:t>\left(\frac{\alpha(t_0(\delta))}{\tilde{D}_1} + \frac{\log(4(12/\Delta_{\min})^d t_0^2(\delta)/\delta)}{\tilde{D}_0}\right)(1+c)} C_1 \exp(-C_2 t) \\
&\leq 1 + (1+c)\left(\frac{\alpha(t_0(\delta))}{\tilde{D}_1} + \frac{\log(4(12/\Delta_{\min})^d t_0^2(\delta)/\delta)}{\tilde{D}_0}\right) \\
&\quad + |\mathcal{C}_\epsilon| C_1 \frac{\exp\left(-C_2\left(\frac{\alpha(t_0(\delta))}{\tilde{D}_1} + \frac{\log(4(12/\Delta_{\min})^d t_0^2(\delta)/\delta)}{\tilde{D}_0}\right)\right)}{1-\exp(-C_2)} \\
&\overset{(c)}{\leq} 1 + (1+c)\left(\frac{\ell\log(4(12/\Delta_{\min})^d t_0^2(\delta))}{\tilde{D}_1} + \frac{\log(4(12/\Delta_{\min})^d t_0^2(\delta)/\delta)}{\tilde{D}_0}\right) \\
&\quad + (12(1/2-\gamma)/\Delta^*)^d C_1 \frac{\exp\left(-C_2\left(\frac{\ell\log(4(12/\Delta_{\min})^d t_0^2(\delta))}{\tilde{D}_1} + \frac{\log(4(12/\Delta_{\min})^d t_0^2(\delta)/\delta)}{\tilde{D}_0}\right)\right)}{1-\exp(-C_2)}
\end{aligned}$$

$$
\begin{aligned}
&\overset{(d)}{\leq} 1 + (1+c)\left(\frac{\ell \log(4(12/\Delta)^d t_0^2(\delta))}{\tilde{D}_1} + \frac{\log(4(12/\Delta)^d t_0^2(\delta)/\delta)}{\tilde{D}_0}\right) \\
&+ (12(1/2-\gamma)/\Delta)^d C_1 \frac{\exp\left(-C_2\left(\frac{\ell \log(4(12/\Delta)^d t_0^2(\delta))}{\tilde{D}_1} + \frac{\log(4(12/\Delta)^d t_0^2(\delta)/\delta)}{\tilde{D}_0}\right)\right)}{1-\exp(-C_2)} \\
&\overset{(e)}{\leq} 1 + \underbrace{2\left(\frac{\ell \log(4(12/\Delta)^d t_0^2(\delta))}{\tilde{D}_1} + \frac{\log(4(12/\Delta)^d t_0^2(\delta)/\delta)}{\tilde{D}_0}\right)}_{\textbf{Term A}} \\
&\quad + \underbrace{\left(44 + \frac{\eta^2}{\eta_0^2}\right)^2 \left(4(12(1/2-\gamma)/\Delta^*)^d t_0^2(\delta)\right)^{1-\frac{\eta_0^2(\ell+4)}{2\eta^3}}}_{\textbf{Term B}} \underbrace{\delta^{\frac{\tilde{D}_0}{\eta^2}}}_{\textbf{Term C}}
\end{aligned} \tag{5.14}
$$

where, (a) follows from definition of M in (5.13), (b) follows from Lemma 7.19 where c is a constant such that $0 < c < 1$, $C_1 := 23 + 11\max\left\{1, \frac{\eta^2}{2\tilde{D}_1^2}\right\}$, $C_2 := \frac{2\tilde{D}_1^2 \min\{(\frac{c}{2}-1)^2, c\}}{\eta^2}$, (c) follows by setting the cardinality of cover $|\mathcal{C}_\epsilon| := (12(1/2-\gamma)/\Delta^*)^d$, (d) follows as $\Delta := \min\{\Delta^*, \Delta_{\min}\}$ and (e) follows the same steps in Theorem 2 by setting $c = \frac{1}{2}$ in C_1 and C_2.

Again, note that in (5.14) the term C ≤ 1 as $\delta \in (0,1)$. So for a constant $\ell > 1$ such that if ℓ satisfies the following condition

$$
\ell = \eta \log\left(44 + \frac{\eta^2}{\eta_0^2}\right) > \log\left(1 + \frac{\eta^2}{\eta_0^2}\right) \tag{5.15}
$$

we have that Term B $\leq (4(12/\Delta)^d t_0^2(\delta))^{1-\frac{\eta_0^2(\ell+4)}{\eta^2} + \frac{\ell}{\eta \log(4(12/\Delta)^d t_0^2(\delta))}}$. Hence, plugging the value of ℓ in (5.15) in (5.14) we get that the expected sample complexity is of the order of

$$
\mathbb{E}[\tau_\delta] \leq O\left(\frac{\eta \log(C(\eta_0,\eta)) d \log(1/\Delta)}{\tilde{D}_1} + \frac{d\log(1/\Delta) + \log(1/\delta)}{\tilde{D}_0} + \left(\frac{1}{\Delta}\right)^d C(\eta_0,\eta)^{1/\eta} \delta^{\tilde{D}_0/\eta^2}\right)
$$

where, $\Delta := \min\{\Delta^*, \Delta_{\min}\}$, and $C = 44 + \frac{\eta^2}{\eta_0^2}$. The claim of the Theorem follows. $\square$

5.4 Theoretical Comparisons with other Bandit works

We now compare our GC bound against linear bandit algorithms. We first recall the following problem complexity parameters in best-arm ID setting from Theorem 6:

$$\tilde{D}_0 := \max_{\mathbf{p}} \min_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \sum_{i=1}^{n} p(i)(\mu_i(\widetilde{\boldsymbol{\theta}}') - \mu_i(\widetilde{\boldsymbol{\theta}}^*))^2, \qquad \tilde{D}_1 := \min_{\{\mathbf{p}_{\widetilde{\boldsymbol{\theta}}}: \widetilde{\boldsymbol{\theta}} \in \widetilde{\boldsymbol{\Theta}}\}} \min_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \sum_{i=1}^{n} p_{\widetilde{\boldsymbol{\theta}}}(i)(\mu_i(\widetilde{\boldsymbol{\theta}}') - \mu_i(\widetilde{\boldsymbol{\theta}}^*))^2. \tag{5.16}$$

Following Degenne et al. (2020a) we define the problem complexity for best-arm ID in linear bandit as follows:

$$\begin{aligned}
D_{\boldsymbol{\theta}^*} \overset{(a)}{:=} \max_{\mathbf{p}} \min_{\boldsymbol{\theta} \in \boldsymbol{Alt}(\boldsymbol{\theta}^*)} \|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|_{V_p}^2 &= \max_{\mathbf{p}} \min_{\boldsymbol{\theta} \in \boldsymbol{Alt}(\boldsymbol{\theta}^*)} (\boldsymbol{\theta}^* - \boldsymbol{\theta})^T V_p (\boldsymbol{\theta}^* - \boldsymbol{\theta}) \\
&\overset{(b)}{=} \max_{\mathbf{p}} \min_{\boldsymbol{\theta} \in \boldsymbol{Alt}(\boldsymbol{\theta}^*)} (\boldsymbol{\theta}^* - \boldsymbol{\theta})^T \sum_{i=1}^{n} p(i) \mathbf{x}_i \mathbf{x}_i^T (\boldsymbol{\theta}^* - \boldsymbol{\theta}) \\
&= \max_{\mathbf{p}} \min_{\boldsymbol{\theta} \in \boldsymbol{Alt}(\boldsymbol{\theta}^*)} \sum_{i=1}^{n} p(i) (\boldsymbol{\theta}^* - \boldsymbol{\theta})^T \mathbf{x}_i \mathbf{x}_i^T (\boldsymbol{\theta}^* - \boldsymbol{\theta}) \\
&= \max_{\mathbf{p}} \min_{\boldsymbol{\theta} \in \boldsymbol{Alt}(\boldsymbol{\theta}^*)} \sum_{i=1}^{n} p(i) (\boldsymbol{\theta}^* - \boldsymbol{\theta})^T \mathbf{x}_i \mathbf{x}_i^T (\boldsymbol{\theta}^* - \boldsymbol{\theta}) \\
&= \max_{\mathbf{p}} \min_{\boldsymbol{\theta} \in \boldsymbol{Alt}(\boldsymbol{\theta}^*)} \sum_{i=1}^{n} p(i) \left(\mathbf{x}_i^T \boldsymbol{\theta}^* - \mathbf{x}_i^T \boldsymbol{\theta}\right)^2 \\
&= \max_{\mathbf{p}} \min_{\boldsymbol{\theta} \in \boldsymbol{Alt}(\boldsymbol{\theta}^*)} \sum_{i=1}^{n} p(i) \left(\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta})\right)^2
\end{aligned} \tag{5.17}$$

where, in (a) the notation $\|\mathbf{x}\|_M$ denotes $\sqrt{\mathbf{x}^T M \mathbf{x}}$, and $\boldsymbol{Alt}(\boldsymbol{\theta}) := \{\boldsymbol{\theta}' \in \Theta : i^*(\boldsymbol{\theta}') \neq i^*(\boldsymbol{\theta})\}$, (b) follows by substituting $V_p = \sum_{i=1}^{n} p(i) \mathbf{x}_i \mathbf{x}_i^T$. Note that the quantity $\tilde{D}_0, \tilde{D}_1$ (5.16) and $D_{\boldsymbol{\theta}^*}$ are not easily comparable as $\tilde{D}_0, \tilde{D}_1$ depend on the discretization of the $\boldsymbol{\Theta}$ space and this stems from model identification proof techniques that we used to prove the bound. We also define the quantity Δ_μ as follows:

$$\Delta_\mu = \min_{i \in [n]} \mathbf{x}_{i^*}^T \boldsymbol{\theta}^* - \mathbf{x}_i^T \boldsymbol{\theta}^*. \tag{5.18}$$

We compare the bound of GC in Theorem 6 against XY-Static (Soare et al., 2014), RAGE (Fiez et al., 2019), LinGame (Degenne et al., 2020a) in the Table 5.1. Note that in Table 5.1 the bound of

GC does not scale with the number of arms as our bound is derived from model identification proof technique. This is desirable in the standard linear bandit setting as the number of arms n can be very large. Also note that unlike LinGapE we do not require to sample each arm once for initialization which is again a desirable property for the standard linear bandit setting where n can be very large.

Example 2. (Asymptotically optimal linear bandit setting) We compare the bounds of the algorithm in Table 5.1 in the this example. We give a canonical linear bandit example previously studied in Xu et al. (2018); Fiez et al. (2019). Let there be three arms in $\mathbb{R}^2$ such that $\mathbf{x}_1 = (1,0), \mathbf{x}_2 = \mathbf{x}_1 + \epsilon\mathbf{x}_1$, $x_3 = (0,1)$, and $\boldsymbol{\theta}^* = \mathbf{x}_1 = (1,0)$, where $\epsilon > 0$. Note that the max-min optimization in Degenne et al. (2020a) will choose an alternate $\boldsymbol{\theta}$ with a different optimal arm that is closest to the boundary of the partition between $\mathbf{x}^*$ and $\mathbf{x}$. Let it be denoted by $\lambda \approx \mathbf{x}_1 + 0.5(\mathbf{x}_2 - \mathbf{x}_1)$. It follows that

$$
\begin{aligned}
D_{\boldsymbol{\theta}^*} &= \max_{\mathbf{p}} \sum_{i=1}^{3} p(i)(\mu_i(\boldsymbol{\theta}^*) - \mu_i(\lambda))^2 \overset{(a)}{=} \max_{\mathbf{p}} \sum_{i=1}^{3} p(i)(\mathbf{x}_i^T\boldsymbol{\theta}^* - \mathbf{x}_i^T(\mathbf{x}_1 + 0.5(\mathbf{x}_2 - \mathbf{x}_1)))^2 \\
&\overset{(b)}{=} \max_{\mathbf{p}} \sum_{i=1}^{3} p(i)(\mathbf{x}_i^T\mathbf{x}_1 - \mathbf{x}_i^T(\mathbf{x}_1 + 0.5(\mathbf{x}_2 - \mathbf{x}_1)))^2 = \max_{\mathbf{p}} \sum_{i=1}^{3} p(i)(0.5\mathbf{x}_i^T(\mathbf{x}_1 - \mathbf{x}_2))^2 \\
&\overset{(c)}{=} \frac{1}{4}(\mathbf{x}_3^T(\mathbf{x}_1 - \mathbf{x}_2))^2 = \frac{1}{4}\epsilon^2\mathbf{x}_1^2 = \frac{\epsilon^2}{4}
\end{aligned}
$$

where, (a) follows from definition of λ, (b) follows as $\boldsymbol{\theta}^* = \mathbf{x}_1$, (c) follows as $\mathbf{p} = (0,0,1)$. Now we discretize the space $\boldsymbol{\Theta}$ into 36 hypotheses $\widetilde{\boldsymbol{\theta}}$ as the cardinality of the ϵ-cover is $|\mathcal{C}_\epsilon| := (3/\epsilon)^d = 36$ for $\epsilon = 0.5, d = 2$. It follows that the $\lambda = \arg\min \widetilde{\boldsymbol{\theta}} \neq \widetilde{\boldsymbol{\theta}}^*$ is given by $\lambda = \mathbf{x}_2 = \mathbf{x}_1 + \epsilon\mathbf{x}_1$. Then we can show that,

$$
\begin{aligned}
\tilde{D}_0 &= \max_{\mathbf{p}} \sum_{i=1}^{n} p(i)(\mu_i(\widetilde{\boldsymbol{\theta}}^*) - \mu_i(\lambda))^2 \overset{(a)}{=} \max_{\mathbf{p}} \sum_{i=1}^{n} p(i)(\mathbf{x}_i^T\widetilde{\boldsymbol{\theta}}^* - \mathbf{x}_i^T(\mathbf{x}_1 + \epsilon\mathbf{x}_1))^2 \\
&\overset{(b)}{=} \max_{\mathbf{p}} \sum_{i=1}^{n} p(i)(\mathbf{x}_i^T\mathbf{x}_1 - \mathbf{x}_i^T(\mathbf{x}_1 + \epsilon\mathbf{x}_1))^2 \overset{(c)}{=} \epsilon^2\mathbf{x}_1^2 = \epsilon^2
\end{aligned}
$$

where, (a) follows from definition of λ, (b) follows as $\boldsymbol{\theta}^* = \mathbf{x}_1$, and (c) follows as $p = (0,0,1)$. Now note that in the above example when $\delta \to 0$ we can use Table 5.1 to show that GC , RAGE , and LinGame has the same asymptotically optimal bound as follows:

$$
\mathbb{E}[\tau_\delta] \leq \frac{C\log(1/\delta)}{\epsilon^2} + o(\log\frac{1}{\delta}).
$$

Algo	Sample Complexity Bound	Comments
XY-Static	$O\left(\frac{d}{\Delta_\mu^2}\log\left(\frac{n^2}{\delta}\log\frac{1}{\Delta_\mu}\right)+d^2\right)$, where Δ_μ is from (5.18)	Bound valid for any $\delta \in (0,1]$. Bound is not asymptotically optimal and scales with number of arms. Bound specialized for linear bandit setting. XY-Static fixes its allocation before collecting any samples in a way such that the uncertainty in the most difficult direction is minimized.
RAGE	$O\left(d\log_2\frac{1}{\Delta_\mu} + \frac{1}{D_{\boldsymbol{\theta}^*}}\log\left(\frac{n^2}{\delta}\log_2\frac{1}{\Delta_\mu}\right)\right)$, where $D_{\boldsymbol{\theta}^*}$ is from (5.17)	Bound valid for any $\delta \in (0,1]$. Bound is also asymptotically optimal and scales with number of arms. Bound specialized for linear bandit setting.
LinGame	$\frac{C\log(1/\delta)}{D_{\boldsymbol{\theta}^*}} + o\left(\log\frac{1}{\delta}\right)$	Bound is Optimal for $\delta \to 0$. Bound specialized for linear bandit setting and does *not* scale with number of arms.
Lower Bound	$\liminf_{\delta\to 0}\frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \geq \frac{1}{D_{\boldsymbol{\theta}^*}}$	Bound optimal for $\delta \to 0$.
GC (Ours)	$O\left(\frac{d\log(1/\Delta)}{\tilde{D}_1}+\frac{\log(1/\delta)}{\tilde{D}_0}\right)$, where Δ is in Theorem 6, $\tilde{D}_0, \tilde{D}_1$ in (5.16)	Bound valid for any $\delta \in (0,1]$. Bound specialized for model identification and more general. Extends Chernoff's original algorithm to best-arm ID setting. Bound does *not* scale with number of arms. GC for linear bandit is also a static allocation but different than XY-Static because it minimizes uncertainty in every direction.

Table 5.1: Linear Bandit Comparison

Note that this is one such example where because of the discretization of the $\boldsymbol{\Theta}$ space we have that GC is asymptotically optimal and its bound scales as RAGE and LinGame .

Remark 3. (Linear Bandit Theorem) Note that in the bound of Theorem 6 the learner does not need to know the Δ^* in (5.5). The theorem does not require any additional knowledge by the learner and is quite general. The $\beta(t, \delta)$ (5.4) required by GC to be δ-PAC depends on $\Delta_{\min}$ (5.3) which is known to the learner in the standard linear bandit setting. This stopping threshold $\beta(t, \delta)$ can be quite large if $\Delta_{\min}$ is small. Theoretically even the state-of-the-art algorithm LinGame (Degenne et al., 2020a) also requires a large stopping threshold as shown in (5.19).

Degenne et al. (2019) describes a method to solve pure exploration problems by describing an iterative procedure to solve the $\max \min$ optimization problem that determines a lower bound on the sample complexity. Their method could also be used for best-arm identification in our non-linear experiment. Unlike their method, we do not require finding the saddle point in an iterative manner and hence has lower computational cost. Our computational cost is mentioned in Section 4.3.

5.5 Application to Best-arm ID in Structured bandits

5.5.1 Linear Bandits

This experiment consist of 12 arms randomly generated from unit sphere $\mathcal{C}^{d-1} := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 = 1\}$. We choose the two closest arms $\mathbf{x}$ and $\mathbf{x}'$ and set the $\boldsymbol{\theta}^* = \mathbf{x} + 0.01(\mathbf{x}' - \mathbf{x})$ so that $\mathbf{x}$ is the optimal arm. Such an environment has already been studied in Tao et al. (2018) and Degenne et al. (2020a). We experiment with four algorithms namely GC , Unif , XY-Oracle , LinGame , LinGameC and LinGapE . We run each algorithm over 100 independent trials and show the result in Figure 5.1b for dimension $d = 8$. From the result we observe that GC empirically has a similar stopping time as LinGapE , LinGame , LinGameC and all of them outperform Unif . Unif uses the GC stopping rule (but randomly chooses an arm to sample at every round). XY-Oracle is a deterministic allocation strategy that fixes the proportion of arms to be sampled and does not depend on the random rewards observed in every round. So in our plots we show it as a green dashed line. Note that our result matches to that of Xu et al. (2018) where LinGapE outperforms XY-Oracle in several synthetic environments. We chose the problem instances such that the $\Delta_{\min}$ is not too small,

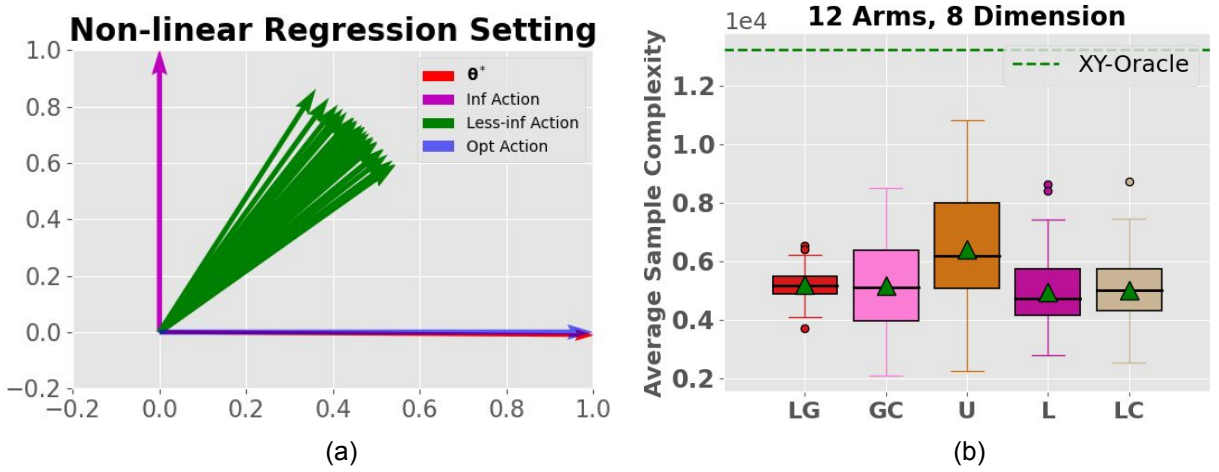

Figure 5.1: **(a)** Sample complexity of linear bandits over unit sphere with $\delta = 0.1$ and dimension $d = 8$. **(b)** Sample complexity over non-linear setting with $\delta = 0.05, d = 1$. LG = LinGapE , GC = GC , U = Unif , L = LinGame , LC = LinGameC , DT = D-Track , KL = KL-LUCB , ST = St-Track .

the choice of $\beta(t, \delta)$ in (5.4) is conservative because we are using a model identification algorithm for best-arm ID.

We show that the GC approach performs comparably to existing linear bandit algorithms such as LinGapE of Xu et al. (2018), LinGame and LinGameC from Degenne et al. (2020a) and the XY-Oracle method of Soare et al. (2014). This experiment consist of 12 arms randomly generated from unit sphere similar to the settings implemented in Tao et al. (2018); Fiez et al. (2019); Degenne et al. (2020a). The sample complexity plots are shown in Figure 5.1b. We see the GC performs comparably to the state-of-the-art algorithms LinGame , LinGameC and LinGapE .

Algorithmic Details: We describe each of the linear bandit algorithms used in the unit ball experiment and the hyper-parameters chosen for these algorithms. The algorithms used are as follows:

1. **XY-Oracle :** The XY-Oracle algorithm (Soare et al., 2014) assumes access to the true parameter $\boldsymbol{\theta}^*$ for selecting the arms to pull. XY-Oracle calculates a static sampling proportion of the arms to pull based on all the possible sampling sequence such that the confidence ellipsoid is contained within one voronoi region (constructed around each arm). Hence it does not depend on the rewards observed at any round t.

2. **LinGapE :** At each round t, LinGapE (Xu et al., 2018) first estimates two arms, the arm with the largest estimated reward $\widehat{i}(t)$ and the most ambiguous arm $\widehat{j}(t)$. It then pulls the most informative arm between $\widehat{i}(t)$ and $\widehat{j}(t)$ to estimate the gap of expected reward between these two arms. We use the greedy version of LinGapE for better performance and set the linear regularizer parameter $\lambda = 1$. Note that LinGapE is designed for finding the ϵ-best arm. To get the best performance we set the $\epsilon = 0.8 \times \Delta_{\min}$ (where $\Delta_{\min}$ is defined in (5.3)) so that LinGapE finds the best arm. In the original code by the authors, they set $\epsilon = \Delta_{\min}$.

3. **Unif :** At every round t the Unif algorithm samples an arm uniform randomly from $[n]$. The Unif uses the same stopping condition as GC (Algorithm 2, Step 8) with the threshold function $\beta(t, \delta)$ as defined in (5.4). We run Unif specifically to show that the sampling rule of Theorem 5 used by GC is better than uniform sampling.

4. **GC :** We run the GC policy as stated in Algorithm 2. The sampling policy is as stated in Theorem 5 and the stopping rule uses the definition of $\beta(t, \delta)$ defined in (5.4).

5. **LinGame :** The LinGame algorithm from Degenne et al. (2020a) is a track-and-stop policy like D-Tracking from Kaufmann et al. (2016). LinGame uses a learner like Adahedge to follow the the sampling proportion that is solution to the optimization problem $\max_{\mathbf{p}} \min_{\lambda \in \neg i^\star(\theta)} \frac{1}{2} \|\boldsymbol{\theta} - \lambda\|^2_{V_w}$ where $\lambda \in \boldsymbol{\Theta}$ is an alternate parameter having a different optimal arm than $i^*(\boldsymbol{\theta})$ and $V_w = \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^T$ is the design matrix. Also note that the optimization of LinGame is more specific to best-arm identification in linear bandits and it has a different sampling procedure than GC . The threshold function is also different than GC

and is as follows:

$$\beta(t,\delta) = \left(\sqrt{\log\left(\frac{1}{\delta}\right) + \frac{d}{2}\log\left(1 + \frac{tL^2}{\eta d}\right)} + \sqrt{\frac{\eta}{2}}M\right)^2 \tag{5.19}$$

where, $\|\boldsymbol{\theta}\| \leq M$, $\|\mathbf{x}\| \leq L$, and η is the regularization parameter. In our unit ball experiment, $M = L = 1$ and we set $\eta = 1.0$.

6. **LinGameC :** The LinGameC algorithm (Degenne et al., 2020a) is similar to LinGame but uses a combined learner Adahedge for all the n arms whereas LinGame uses n Adahedges for each of the arm $i \in [n]$.

5.5.2 Non-linear reward function

Consider 10 points drawn uniformly from $[0, 1]$, these serve as arms whose mean reward is given by $\mu_i := \exp(-(x - \boldsymbol{\theta}^*)^2)\cos(4\pi(x - \boldsymbol{\theta}^*))$ with $\boldsymbol{\theta}^* = 0.7$. The function with the arm means is shown in Figure 5.2a and has two modes which is a more general setting than the unimodal setting studied in Combes and Proutiere (2014) (left). The sample complexity plots are shown in Figure 5.2a (right). GC outperforms the stochastic bandit algorithms D-Track (Garivier and Kaufmann, 2016) and KL-LUCB (Kaufmann and Kalyanakrishnan, 2013) which do not leverage the structure of the reward function. We also compare GC to the general structured tracking algorithm St-Track which constructs a confidence region for $\boldsymbol{\theta}^*$ and runs D-Track for a randomly chosen parameter from the region. The Unif baseline uses the same stopping condition as GC (but uniformly samples arm) and this enables it to use the structure information leading to good performance.

Algorithmic Details: We implement the following algorithms in this setting:

1. **D-Track :** This is the D-tracking algorithm from Kaufmann et al. (2016) for the best-arm ID in stochastic bandit setting. Note that this algorithm does not leverage the non-linear structure of the problem and uses just the plug-in estimate of the empirical mean of the arms to find the optimal proportion $w_i^*(\mu(\boldsymbol{\theta}(\mathbf{t})))$ (see Lemma 3 ofKaufmann et al. (2016)).

2. **KL-LUCB :** This is the KL-LUCB algorithm from Kaufmann and Kalyanakrishnan (2013). This is also an algorithm for best-arm ID stochastic bandit setting that does not leverage

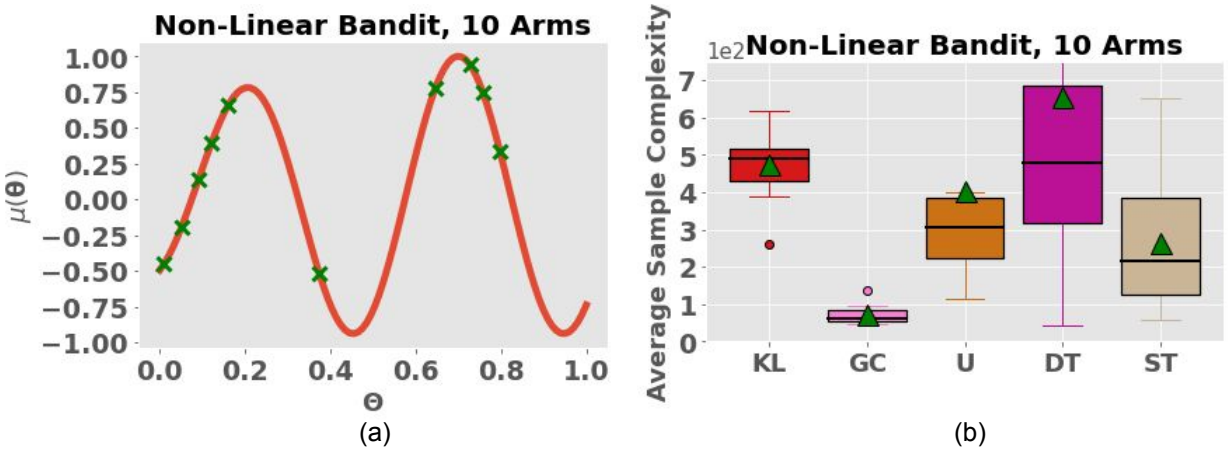

Figure 5.2: **(a)** Sample complexity of linear bandits over unit sphere with $\delta = 0.1$ and dimension $d = 8$. **(b)** Sample complexity over non-linear setting with $\delta = 0.05, d = 1$. LG = LinGapE , GC = GC , U = Unif , L = LinGame , LC = LinGameC , DT = D-Track , KL = KL-LUCB , ST = St-Track .

the structure of the problem. At every round it pulls the empirical best-arm and its closest challenger till the stopping condition is satisfied.

3. **St-Track :** This algorithm is the generalized structured version of D-Track algorithm and a version of the Sticky Track and Stop algorithm from Degenne and Koolen (2019). The algorithm is given below:

   The Algorithm 3 first builds a confidence set $\widetilde{\Theta}_t$ that contains the $\boldsymbol{\theta}^*$ with high probability. This is similar to UCBS algorithm in Lattimore and Munos (2014). Next we sample one $\boldsymbol{\theta}_t$ from the confidence set $\widetilde{\Theta}_t$ and use that as a plug-in estimate to the D-Track algorithm of Kaufmann et al. (2016) to obtain the $w_i^*(\mu(\boldsymbol{\theta}(\mathbf{t})))$ (see Lemma 3 ofKaufmann et al. (2016)) and track that proportion.Finally St-Track stops using the GLRT rule of Kaufmann et al. (2016). Note that this algorithm is different than the general structured algorithm of Gupta et al. (2020); Tirinzoni et al. (2020) which are suitable for regret minimization.

Algorithm 3 Structured Tacking (St-Track)

Input: Confidence parameter δ, functions $\mu_1, \cdots, \mu_n : \Theta \to [0,1]$

Define: $U_t = \left\{i \in [n] : Z_i(t) < \sqrt{t} - n/2\right\}$

for $t = 1, 2, 3, \ldots$ **do**

  Define confidence set $\widetilde{\Theta}_t \leftarrow \left\{\widetilde{\theta} : \forall i, \quad \left|\mu_i(\tilde{\theta}) - \hat{\mu}_i(t)\right| < \sqrt{\frac{2\log(\log t+1)/\delta}{Z_i(t)}}\right\}$

  if $\tilde{\Theta}_t = \emptyset$ **then**

    Choose arm arbitrarily

  else

    Sample $\boldsymbol{\theta}_t$ from $\widetilde{\Theta}_t$

    D-Tracking: $I_{t+1} \in \begin{cases} \arg\min_{a \in U_t} Z_i(t) \text{ if } U_t \neq \emptyset & \text{(forced exploration)} \\ \arg\max_{1 \leq i \leq n} t w_i^*(\mu_i(\boldsymbol{\theta}_t)) - Z_i(t) & \text{(direct tracking)} \end{cases}$

  if $\inf_{\boldsymbol{\lambda} \in \mathbf{Alt}(\hat{\mu}(t))} \sum_{i=1}^n Z_i(t) \mathrm{KL}\left(\widehat{\mu}_i(t), \lambda_i\right) \geq \beta(t, \delta)$ **then**

    Return $i^*(\boldsymbol{\theta}_t)$ as the true hypothesis.

4. **GC :** We implement the GC policy in Algorithm 2. Since this is non-linear setting with $d = 1$ so we use the implementation mentioned in Remark 1. We use python scipy.optmize least-squares function to obtain the least square estimate at every round. The scipy least-squares function solves a non-linear least-squares problem which may give local minima.

5. **Unif :** The Unif policy just uniform randomly samples an arm at every round. The stopping condition of Unif is same as that of GC . So Unif calculates yje $\widehat{\boldsymbol{\theta}}(t)$ using a least squares solver (and so it leverages the structure). We implement Unif to show that the sampling rule of GC is better than Unif .

In this chapter we showed how to use GC for active regression in identifying the best-arm in the multi-armed bandit setting. We also showed that our proposed approach shows competitive performance against other bandit algorithms specifically geared for the linear or non-linear multi-armed bandit setting. In the next chapter we summarize and conclude.

Chapter 6

Conclusions and Future Directions

This thesis provides novel moderate confidence sample complexity bounds for Chernoff's sequential design and extends it to handle smoothly parameterized hypothesis spaces. As a result, we obtain new algorithm for active regression setting that combines experimental design approach as well as have sound theoretical and empirical guarantees. In the structured bandit setting we proved that both of GC can be extended to identify the best arm in smoothly parameterized hypothesis spaces and showed that it empirically outperform specialized linear bandit algorithms for best arm identification.

Future directions include obtaining a lower bound to the active testing setting for the moderate confidence regime and incorporating the geometry of the actions in the sampling strategy for the regression setting. Our work can be potentially extended to new directions which are of interest to both the active learning and the bandit community. Note that our current best-arm ID algorithm samples in a way to identify $\boldsymbol{\theta}^*$ rather than specifically the best-arm. Suppose that at some point we know that one of two arms are the best. If these actions are $\mathbf{x}$ and $\mathbf{x}'$, then we only need to look in "directions" that distinguish $\mathbf{x}$ from $\mathbf{x}'$. In the linear bandit case, this means that we can project the gradients of each arm onto the direction $(\mathbf{x} - \mathbf{x}')$. We can incorporate this projected gradient step before doing the max-min eigenvalue optimization procedure.

While we have looked at the best-arm partition in this work, one can certainly consider other partitions that may be more relevant in different contexts. For example, we can partition $\boldsymbol{\Theta}$ into $n!$ parts, each corresponding to an ordering of the arm means from best to worst. If one believes that the set of arms is representative of all possible actions (including new ones not yet observed), then this could be a reasonable stopping criterion for active learning.

Bibliography

Albert, A. E. (1961). The sequential design of experiments for infinitely many states of nature. *The Annals of Mathematical Statistics*, pages 774–799.

Balcan, M.-F., Beygelzimer, A., and Langford, J. (2009). Agnostic active learning. *Journal of Computer and System Sciences*, 75(1):78–89.

Balcan, M.-F. and Long, P. (2013). Active and passive learning of linear separators under log-concave distributions. In *Conference on Learning Theory*, pages 288–316. PMLR.

Bu, Y., Lu, J., and Veeravalli, V. V. (2019). Active and adaptive sequential learning with per time-step excess risk guarantees. In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pages 1606–1610. IEEE.

Cai, W., Zhang, M., and Zhang, Y. (2016). Batch mode active learning for regression with expected model change. *IEEE transactions on neural networks and learning systems*, 28(7):1668–1681.

Chaudhuri, K., Jain, P., and Natarajan, N. (2017). Active heteroscedastic regression. In *International Conference on Machine Learning*, pages 694–702. PMLR.

Chaudhuri, K., Kakade, S. M., Netrapalli, P., and Sanghavi, S. (2015). Convergence rates of active learning for maximum likelihood estimation. In *Advances in Neural Information Processing Systems*, pages 1090–1098.

Chaudhuri, P. and Mykland, P. A. (1993). Nonlinear experiments: Optimal design and inference based on likelihood. *Journal of the American Statistical Association*, 88(422):538–546.

Chernoff, H. (1959). Sequential design of experiments. *The Annals of Mathematical Statistics*, 30(3):755–770.

Cohn, D. A. (1996). Neural network exploration using optimal experiment design. *Neural Networks*, 9(6):1071 – 1083.

Combes, R., Magureanu, S., and Proutiere, A. (2017). Minimal exploration in structured stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 1763–1771.

Combes, R. and Proutiere, A. (2014). Unimodal bandits: Regret lower bounds and optimal algorithms. In *International Conference on Machine Learning*, pages 521–529.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision support systems*, 47(4):547–553.

Dasgupta, S. (2005). Coarse sample complexity bounds for active learning. In *NIPS*, volume 18, pages 235–242.

Dasgupta, S., Hsu, D. J., and Monteleoni, C. (2008). A general agnostic active learning algorithm. In *International Symposium on Artificial Intelligence and Mathematics, ISAIM 2008, Fort Lauderdale, Florida, USA, January 2-4, 2008*.

De Vito, S., Massera, E., Piga, M., Martinotto, L., and Di Francia, G. (2008). On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B: Chemical*, 129(2):750–757.

Degenne, R. and Koolen, W. M. (2019). Pure exploration with multiple correct answers. In Wallach, H. M., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. B., and Garnett, R., editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 14564–14573.

Degenne, R., Koolen, W. M., and Ménard, P. (2019). Non-asymptotic pure exploration by solving games. In *Advances in Neural Information Processing Systems*, pages 14465–14474.

Degenne, R., Ménard, P., Shang, X., and Valko, M. (2020a). Gamification of pure exploration for linear bandits. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 2432–2442. PMLR.

Degenne, R., Shao, H., and Koolen, W. (2020b). Structure adaptive algorithms for stochastic bandits. In *International Conference on Machine Learning*, pages 2443–2452. PMLR.

Dette, H. and Studden, W. J. (1993). Geometry of e-optimality. *The Annals of Statistics*, 21(1):416–433.

Fan, M. K. and Nekooie, B. (1995). On minimizing the largest eigenvalue of a symmetric matrix. *Linear Algebra and its Applications*, 214:225–246.

Fiez, T., Jain, L., Jamieson, K. G., and Ratliff, L. (2019). Sequential experimental design for transductive linear bandits. In Wallach, H., Larochelle, H., Beygelzimer, A., dÁlché Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32, pages 10667–10677. Curran Associates, Inc.

Fontaine, X., Perrault, P., Valko, M., and Perchet, V. (2019). Online a-optimal design and active linear regression. *arXiv preprint arXiv:1906.08509*.

Fukumizu, K. (2000). Statistical active learning in multilayer perceptrons. *IEEE Transactions on Neural Networks*, 11(1):17–26.

Garivier, A. and Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pages 998–1027.

Gupta, S., Chaudhari, S., Mukherjee, S., Joshi, G., and Yağan, O. (2020). A unified approach to translate classical bandit algorithms to the structured bandit setting. *IEEE Journal on Selected Areas in Information Theory*.

Hanneke, S. (2007). A bound on the label complexity of agnostic active learning. In *Proceedings of the 24th international conference on Machine learning*, pages 353–360.

Hsu, D., Kakade, S., Zhang, T., et al. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17.

Huang, R., Ajallooeian, M. M., Szepesvári, C., and Müller, M. (2017). Structured best arm identification with fixed confidence. In Hanneke, S. and Reyzin, L., editors, *International Conference on Algorithmic Learning Theory, ALT 2017, 15-17 October 2017, Kyoto University, Kyoto, Japan*, volume 76 of *Proceedings of Machine Learning Research*, pages 593–616. PMLR.

Jedra, Y. and Proutière, A. (2020). Optimal best-arm identification in linear bandits. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Katz-Samuels, J., Zhang, J., Jain, L., and Jamieson, K. (2021). Improved algorithms for agnostic pool-based active classification. *arXiv preprint arXiv:2105.06499*.

Kaufmann, E., Cappé, O., and Garivier, A. (2016). On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42.

Kaufmann, E. and Kalyanakrishnan, S. (2013). Information complexity in bandit subset selection. In *Conference on Learning Theory*, pages 228–251. PMLR.

Lattimore, T. and Munos, R. (2014). Bounded regret for finite-armed structured bandits. In *Advances in Neural Information Processing Systems*, pages 550–558.

Lattimore, T. and Szepesvári, C. (2020). *Bandit algorithms*. Cambridge University Press.

Ménard, P. (2019). Gradient ascent for active exploration in bandit problems. *CoRR*, abs/1905.08165.

Naghshvar, M. and Javidi, T. (2013). Active sequential hypothesis testing. *The Annals of Statistics*, 41(6):2703–2738.

Nitinawarat, S., Atia, G. K., and Veeravalli, V. V. (2013). Controlled sensing for multihypothesis testing. *IEEE Transactions on Automatic Control*, 58(10):2451–2464.

Pronzato, L. and Pázman, A. (2013). Design of experiments in nonlinear models. *Lecture notes in statistics*, 212.

Sabato, S. and Munos, R. (2014). Active regression by stratification. *arXiv preprint arXiv:1410.5920*.

Settles, B. (2009). Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison.

Soare, M., Lazaric, A., and Munos, R. (2014). Best-arm identification in linear bandits. In *Advances in Neural Information Processing Systems*, pages 828–836.

Tao, C., Blanco, S., and Zhou, Y. (2018). Best arm identification in linear bandits with linear dimension dependency. In *International Conference on Machine Learning*, pages 4877–4886.

Tirinzoni, A., Pirotta, M., Restelli, M., and Lazaric, A. (2020). An asymptotically optimal primal-dual incremental algorithm for contextual linear bandits. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Tsybakov, A. B. (2008). *Introduction to nonparametric estimation*. Springer Science & Business Media.

Tukey, J. W. (1977). *Exploratory data analysis*, volume 2. Reading, MA.

Vaidhiyan, N. K. and Sundaresan, R. (2017). Learning to detect an oddball target. *IEEE Transactions on Information Theory*, 64(2):831–852.

Wu, D. (2018). Pool-based sequential active learning for regression. *IEEE transactions on neural networks and learning systems*, 30(5):1348–1359.

Wu, D., Lin, C.-T., and Huang, J. (2019). Active learning for regression using greedy sampling. *Information Sciences*, 474:90–105.

Xu, L., Honda, J., and Sugiyama, M. (2018). A fully adaptive algorithm for pure exploration in linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 843–851. PMLR.

Yu, K., Bi, J., and Tresp, V. (2006). Active learning via transductive experimental design. In *Proceedings of the 23rd international conference on Machine learning*, pages 1081–1088.

Zhang, C. and Chaudhuri, K. (2014). Beyond disagreement-based agnostic active learning. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 442–450.

Chapter 7

Appendix

7.1 Probability Tools

We also state a few standard lemmas in Probability Tools that we use to prove our results.

Lemma 7.1. ***(Restatement of Lemma 15.1 in Lattimore and Szepesvári (2020), Divergence Decomposition)*** *Let B and B' be two bandit models having different optimal hypothesis $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}'^*$ respectively. Fix some policy π and round n. Let $\mathbb{P}_{B,\pi}$ and $\mathbb{P}_{B',\pi}$ be two probability measures induced by some n-round interaction of π with B and π with B' respectively. Then*

$$\mathrm{KL}\left(\mathbb{P}_{B,\pi}||\mathbb{P}_{B',\pi}\right) = \sum_{i=1}^{n} \mathbb{E}_{B,\pi}[Z_i(n)] \cdot \mathrm{KL}(\mu_i(\boldsymbol{\theta})||\mu_i(\boldsymbol{\theta}^*))$$

where, $\mathrm{KL}\left(.||.\right)$ *denotes the Kullback-Leibler divergence between two probability measures and* $Z_i(n)$ *denotes the number of times action* i *has been sampled till round* n*.*

Lemma 7.2. ***(Restatement of Lemma 2.6 in Tsybakov (2008))*** *Let $\mathbb{P}, \mathbb{Q}$ be two probability measures on the same measurable space $(\Omega, \mathcal{F})$ and let $\xi \subset \mathcal{F}$ be any arbitrary event then*

$$\mathbb{P}(\xi) + \mathbb{Q}\left(\xi^c\right) \geqslant \frac{1}{2}\exp\left(-\mathrm{KL}(\mathbb{P}||\mathbb{Q})\right)$$

where ξ^c denotes the complement of event ξ and $\mathrm{KL}(\mathbb{P}||\mathbb{Q})$ denotes the Kullback-Leibler divergence between $\mathbb{P}$ and $\mathbb{Q}$.

Lemma 7.3. ***(Hoeffding's Lemma)*** *Let Y be a real-valued random variable with expected value $\mathbb{E}[Y] = \mu$, such that $a \leq Y \leq b$ with probability one. Then, for all $\lambda \in \mathbb{R}$*

$$\mathbb{E}\left[e^{\lambda Y}\right] \leq \exp\left(\lambda\mu + \frac{\lambda^2(b-a)^2}{8}\right)$$

Lemma 7.4. ***(Proposition 2 of (Hsu et al., 2012))*** *Let $u_1, \ldots, u_n$ be a martingale difference vector sequence (i.e., $\mathbb{E}\left[u_i \mid u_1, \ldots, u_{i-1}\right] = 0$ for all $i = 1, \ldots, n$) such that*

$$\sum_{i=1}^{n} \mathbb{E}\left[\|u_i\|^2 \mid u_1, \ldots, u_{i-1}\right] \leq v \quad \textit{and} \quad \|u_i\| \leq b$$

for all $i = 1, \ldots, n$, almost surely. For all $t > 0$

$$\Pr\left[\left\|\sum_{i=1}^{n} u_i\right\| > \sqrt{v} + \sqrt{8vt} + (4/3)bt\right] \leq e^{-t}$$

7.2 Chernoff Sample Complexity Proof

7.2.1 Concentration Lemma

Lemma 7.5. *Define $L_{Y^t}(\boldsymbol{\theta}^*)$ as the sum squared errors for the hypothesis parameterized by $\boldsymbol{\theta}^*$. Let $\tau_{\boldsymbol{\theta}^*} = \min\{t : L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*\}$. Then we can bound the probability that $\tau_{\boldsymbol{\theta}^*}$ is larger than t as*

$$\mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) \leq JC_1 \exp\left(-C_2 t\right)$$

where, $J := |\boldsymbol{\Theta}|$, $C_1 := 23 + 11 \max\left\{1, \dfrac{\eta^2}{2D_1^2 c}\right\}$, $C_2 := \dfrac{2D_1^2 \min\{(c-1)^2, c\}}{\eta^2}$, $\eta > 0$ defined in Definition 1 and $D_1 := \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^{n} p_{\boldsymbol{\theta}}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$.

Proof. We consider the following events when the difference of squared errors crosses certain values:

$$\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t) := \{L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*) < \beta(J, \delta)\},$$
$$\tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t) := \{L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*) < \alpha(J)\}.$$

Then we define the stopping time $\tau_{\boldsymbol{\theta}^*}$ as follows:

$$\tau_{\boldsymbol{\theta}^*} := \min\{t : L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*\}$$

which is the first round when $L_{Y^t}(\boldsymbol{\theta}')$ crosses $\beta(J, \delta)$ threshold against $L_{Y^t}(\boldsymbol{\theta}^*)$ for all $\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*$. We also define the stopping time $\tilde{\tau}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}$ as follows:

$$\tilde{\tau}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*} := \min\{t : L_{Y^{t'}}(\boldsymbol{\theta}') - L_{Y^{t'}}(\boldsymbol{\theta}^*) > \alpha(J), \forall t' > t\} \tag{7.1}$$

which is the first round when $L_{Y^t}(\boldsymbol{\theta}')$ crosses $\alpha(J)$ threshold against $L_{Y^t}(\boldsymbol{\theta}^*)$. We will be particularly interested in the time

$$\tilde{\tau}_{\boldsymbol{\theta}^*} := \max_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \{\tilde{\tau}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}\}. \tag{7.2}$$

Let $T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)$ denote the difference of squared errors between hypotheses $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}^*$ (for Gaussian noise model, it is equal to the log-likelihood ratio between hypotheses parameterized by $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}^*$) shown below.

$$T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) = L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*). \tag{7.3}$$

It follows that $T_{Y^t}(\boldsymbol{\theta}) - D_1$ is upper bounded by 4η which can be shown as follows:

$$\begin{aligned} T_{Y_s}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - D_1 &= L_{Y_s}(\boldsymbol{\theta}') - L_{Y_s}(\boldsymbol{\theta}^*) - D_1 = (Y_s - \mu_{I_s}(\boldsymbol{\theta}'))^2 - (Y_s - \mu_{I_s}(\boldsymbol{\theta}^*))^2 - D_1 \\ &\overset{(a)}{=} 2Y_s(\mu_{I_s}(\boldsymbol{\theta}) - \mu_{I_s}(\boldsymbol{\theta}^*)) + (\mu_{I_s}(\boldsymbol{\theta}^*)^2 - \mu_{I_s}(\boldsymbol{\theta})^2) - \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^{n} p_{\boldsymbol{\theta}}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2 \\ &\overset{(b)}{\leq} 2Y_s\sqrt{\eta} + \eta + \eta \leq 4\eta \end{aligned} \tag{7.4}$$

where, (a) follows from the definition of D_1, and (b) follows from Assumption 1. Similarly, we can show that,

$$\begin{aligned} T_{Y_s}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - D_0 &= L_{Y_s}(\boldsymbol{\theta}') - L_{Y_s}(\boldsymbol{\theta}^*) - D_0 = (Y_s - \mu_{I_s}(\boldsymbol{\theta}'))^2 - (Y_s - \mu_{I_s}(\boldsymbol{\theta}^*))^2 - D_0 \\ &\overset{(a)}{=} 2Y_s(\mu_{I_s}(\boldsymbol{\theta}) - \mu_{I_s}(\boldsymbol{\theta}^*)) + (\mu_{I_s}(\boldsymbol{\theta}^*)^2 - \mu_{I_s}(\boldsymbol{\theta})^2) - \max_{\mathbf{p}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^{n} p(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2 \\ &\overset{(b)}{\leq} 2Y_s\sqrt{\eta} + \eta + \eta \leq 4\eta \end{aligned} \tag{7.5}$$

where, (a) follows from the definition of D_0, and (b) follows from Assumption 1. A key thing to note that for $D_1 > 0$ (Assumption 2) the $\mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] > tD_1$ which is shown as follows:

$$\begin{aligned} \mathbb{E}_{I^t, Y^t}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] = \mathbb{E}_{I^t, Y^t}[L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*)] &= \mathbb{E}_{I^t, Y^t}\left[\sum_{s=1}^{t}(Y_s - \mu_{I_s}(\boldsymbol{\theta}^*))^2 - \sum_{s=1}^{t}(Y_s - \mu_{I_s}(\boldsymbol{\theta}))^2\right] \\ &= \sum_{s=1}^{t} \mathbb{E}_{I_s}\mathbb{E}_{Y_s|I_s}\left[(\mu_{I_s}(\boldsymbol{\theta}^*) - \mu_{I_s}(\boldsymbol{\theta}))^2 \,|I_s\right] \\ &= \sum_{s=1}^{t}\sum_{i=1}^{n} \mathbb{P}(I_s = i)\,(\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta}))^2 \overset{(a)}{\geq} tD_1 \end{aligned}$$

where, (a) follows from the definition of D_1 in Theorem 2. Then it follows that,

$$\mathbb{P}(\tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)) = \mathbb{P}(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) < \alpha(J)) = \mathbb{P}(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < \alpha(J) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)])$$
$$\overset{(a)}{\leq} \mathbb{P}(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < \alpha(J) - tD_1)$$

where, in (a) the choice of tD_1 follows as $\mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] \geq tD_1$ for $D_1 > 0$ and noting that $|T_{Y_s}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - D_1| \leq 4\eta$ by (7.4) for any round $s \in [t]$. Similarly, we can show that $\mathbb{E}[T_{Y_s}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] \geq D_0$ for all rounds $s \geq t - \tilde{\tau}_{\theta^*}$ where $\tilde{\tau}_{\boldsymbol{\theta}^*}$ is defined in (7.2). It follows then that $\mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] \geq (t - \tilde{\tau}_{\theta^*})D_0$ where D_0 is defined in Theorem 2. Then we can show that,

$$\begin{aligned}
\mathbb{P}(\xi_{\boldsymbol{\theta}', \boldsymbol{\theta}^*}(t)) &= \mathbb{P}(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) < \beta(J, \delta)) \\
&= \mathbb{P}(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < \beta(J, \delta) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)]) \\
&\overset{(a)}{\leq} \mathbb{P}\left(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < \beta(J, \delta) - (t - \tilde{\tau}_{\boldsymbol{\theta}^*})D_0\right) \\
&\overset{(b)}{=} \mathbb{P}\left(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < D_0\left(\frac{\beta(J, \delta)}{D_0} - t + \tilde{\tau}_{\boldsymbol{\theta}^*}\right)\right)
\end{aligned}$$

where, in (a) the choice of $(t - \tilde{\tau}_{\boldsymbol{\theta}^*})D_0$ follows as $\mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] \geq (t - \tilde{\tau}_{\theta^*})D_0$ for any round $s > \tilde{\tau}_{\boldsymbol{\theta}^*}$, and finally in (b) the quantity $D_0\left(\frac{\beta(J, \delta)}{D_0} - t + \tilde{\tau}_{\boldsymbol{\theta}^*})\right)$ is negative for $t \geq (1 + c)\left(\frac{\alpha(J)}{D_1} + \frac{\beta(J, \delta)}{D_0}\right)$ and $\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_1} + \frac{tc}{2}$ which allows us to apply the Mcdiarmid's inequality for conditionally independent random variables stated in Lemma 7.8. Then we can show that,

$$\begin{aligned}
&\mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) \leq \mathbb{P}(\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)) \\
&= \mathbb{P}\left(\left\{\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\right\} \bigcap \{\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_1} + \frac{tc}{2}\}\right) + \mathbb{P}\left(\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \{\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\} \bigcap \{\tilde{\tau}_{\boldsymbol{\theta}^*} \geq \frac{\alpha(J)}{D_1} + \frac{tc}{2}\}\right) \\
&\leq \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(\{\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\} \bigcap \{\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_1} + \frac{tc}{2}\}\right) + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t' : t' \geq \frac{\alpha(J)}{D_1} + \frac{tc}{2}} \mathbb{P}\left(\tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t')\right) \\
&\leq \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(\{\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}\} \bigcap \{\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_1} + \frac{tc}{2}\}\right) \\
&\quad + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t' : t' \geq \frac{\alpha(J)}{D_1} + \frac{tc}{2}} \mathbb{P}\left(T_{Y^{t'}}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^{t'}}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < \alpha(J) - t'D_1\right)
\end{aligned}$$

$$\leq \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < D_0\left(\frac{\beta(J,\delta)}{D_0} + \frac{\alpha(J)}{D_1} - t + \frac{tc}{2}\right)\right)$$
$$+ \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t': t' \geq \frac{\alpha(J)}{D_1} + \frac{tc}{2}} \mathbb{P}\left(T_{Y^{t'}}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^{t'}}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < \alpha(J) - t' D_1\right)$$

$$\overset{(a)}{\leq} \exp(4) \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1^2 t(c/2 - 1)^2}{\eta^2}\right) + \exp(4) \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \frac{\exp\left(-\frac{2D_1^2 tc}{\eta^2}\right)}{1 - \exp\left(-\frac{2D_1^2}{\eta^2}\right)}$$

$$\overset{(b)}{\leq} \exp(4) \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1^2 t(\frac{c}{2} - 1)^2}{\eta^2}\right) + \exp(4) \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1^2 tc}{\eta^2}\right)\left(1 + \max\left\{1, \frac{\eta^2}{2D_1^2}\right\}\right)$$

$$\overset{(c)}{\leq} 11 \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1^2 t(\frac{c}{2} - 1)^2}{\eta^2}\right) + 11 \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1^2 tc}{\eta^2}\right) + 11 \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1^2 tc}{\eta^2}\right) \max\left\{1, \frac{\eta^2}{2D_1^2}\right\}$$

$$\leq \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1^2 t \min\{(\frac{c}{2} - 1)^2, c\}}{\eta^2}\right)\left[23 + 11 \max\left\{1, \frac{\eta^2}{2D_1^2}\right\}\right] \overset{(d)}{\leq} JC_1 \exp(-C_2 t)$$

where, (a) follows from Lemma 7.6 and Lemma 7.7, (b) follows from the identity that $1/(1 - \exp(-x)) \leq 1 + \max\{1, 1/x\}$ for $x > 0$, (c) follows for $0 < c < 1$, and in (d) we substitute $C_1 := 23 + 11 \max\left\{1, \frac{\eta^2}{2D_1^2}\right\}$, $C_2 := \frac{2D_1^2 \min\{(c/2 - 1)^2, c\}}{\eta^2}$ and $J := |\boldsymbol{\Theta}|$. □

Lemma 7.6. *Let* $T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) := L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*)$ *from* (7.3)*,* $D_1 := \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n p_{\boldsymbol{\theta}}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$*, and* $\alpha(J)$ *and* $\beta(J, \delta)$ *be the two thresholds. Then we can show that*

$$\sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < D_0\left(\frac{\beta(J,\delta)}{D_0} + \frac{\alpha(J)}{D_1} - t + \frac{tc}{2}\right)\right)$$
$$\leq \exp(4) \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1^2 t(c/2 - 1)^2}{\eta^2}\right).$$

for some constant c *such that* $0 < c < 1$.

Proof. Let us recall that the critical number of samples is given by $(1 + c)M$ where

$$M := \frac{\beta(J,\delta)}{D_0} + \frac{\alpha(J)}{D_1}. \tag{7.6}$$

and c is a constant. Then we can show that for some $0 < c < 1$,

$$\sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < D_0\left(\frac{\beta(J,\delta)}{D_0} + \frac{\alpha(J)}{D_1} - t + \frac{tc}{2}\right)\right)$$

$$\overset{(a)}{\leq} \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_0^2\left(\frac{\beta(J,\delta)}{D_0} + \frac{\alpha(J)}{D_1} + t(c/2 - 1)\right)^2}{t\eta^2}\right) \overset{(b)}{=} \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1^2\left(M + t(c/2 - 1)\right)^2}{t\eta^2}\right)$$

$$\overset{(c)}{\leq} \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1^2t^2(c/2-1)^2 + 4D_1^2tM(c/2-1)}{t\eta^2}\right) \overset{(d)}{\leq} \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1^2t^2(c/2-1)^2 + 4\eta_0^2Mt(c/2-1)}{t\eta^2}\right)$$

$$\overset{(e)}{\leq} \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1^2t^2(c/2-1)^2 + 4\eta_0^2t(c/2-1)}{t\eta^2}\right) \overset{(f)}{=} \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(\frac{4\eta_0^2(1-c/2)}{\eta^2}\right)\exp\left(-\frac{2D_1^2t(c/2-1)^2}{\eta^2}\right)$$

$$\overset{(g)}{\leq} \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(4(1-c/2)\right)\exp\left(-\frac{2D_1^2t(c/2-1)^2}{\eta^2}\right) \leq \exp(4)\sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1^2t(c/2-1)^2}{\eta^2}\right)$$

where (a) follows from Lemma 7.8 and noting that $D_0\left(\frac{\beta(J,\delta)}{D_0} + \frac{\alpha(J)}{D_1} - t + \frac{tc}{2}\right) < 0$ for $t > (1+c)M$, the inequality (b) follows from definition of M and noting that $D_0 \geq D_1$, (c) follows as for $M > 1$ we can show that $M^2 + 2tM(c/2-1) + t^2(c/2-1)^2 \geq 2tM(c/2-1) + t^2(c/2-1)^2$, (d) follows as $D_1 \geq \eta_0$, (e) follows as $M > 1$, (f) follows as $0 < c < 1$, and (g) follows as $\eta_0 \leq \eta$. □

Lemma 7.7. *Let* $T_{Y^{t'}}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) := L_{Y^{t'}}(\boldsymbol{\theta}') - L_{Y^{t'}}(\boldsymbol{\theta}^*)$ *from* (7.3)*,* $D_1 = \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n p_{\boldsymbol{\theta}}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$*, and* $\alpha(J)$ *be the threshold depending only on* J*. Then we can show that*

$$\sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t': t' > \frac{\alpha(J)}{D_1} + \frac{tc}{2}} \mathbb{P}\left(T_{Y^{t'}}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^{t'}}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < \alpha(J) - t'D_1\right) \leq \exp(4)\sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \frac{\exp\left(-\frac{2D_1^2tc}{\eta^2}\right)}{1 - \exp\left(-\frac{2D_1^2}{\eta^2}\right)}$$

for some constant $0 < c < 1$ *and* η *defined in Definition 1.*

Proof. Let us recall that

$$\sum_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*}\sum_{t':t'>\frac{\alpha(J)}{D_1}+\frac{tc}{2}} \mathbb{P}\left(T_{Y^{t'}}(\boldsymbol{\theta}',\boldsymbol{\theta}^*)-\mathbb{E}[T_{Y^{t'}}(\boldsymbol{\theta}',\boldsymbol{\theta}^*)]<\alpha(J)-t'D_1\right)$$

$$\overset{(a)}{\leq}\sum_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*}\sum_{t':t'>\frac{\alpha(J)}{D_1}+\frac{tc}{2}}\exp\left(-\frac{2(\alpha(J)-t'D_1)^2}{t'\eta^2}\right)$$

$$\overset{(b)}{\leq}\sum_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*}\sum_{t':t'>\frac{\alpha(J)}{D_1}+\frac{tc}{2}}\exp\left(-\frac{2D_1^2(t')^2-4D_1\alpha(J)t'}{t'\eta^2}\right)\overset{(c)}{\leq}\sum_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*}\sum_{t':t'>\frac{\alpha(J)}{D_1}+\frac{tc}{2}}\exp\left(-\frac{2D_1^2(t')^2-4\eta^2t'}{t'\eta^2}\right)$$

$$=\sum_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*}\sum_{t':t'>\frac{\alpha(J)}{D_1}+\frac{tc}{2}}\exp(4)\exp\left(-\frac{2D_1^2t'}{\eta^2}\right)\overset{(d)}{\leq}\exp(4)\sum_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*}\frac{\exp\left(-2D_1^2\left(\frac{\alpha(J)}{D_1}+\frac{tc}{2}\right)/\eta^2\right)}{1-\exp\left(-\frac{2D_1^2}{\eta^2}\right)}$$

$$=\exp(4)\sum_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*}\frac{\exp\left(-\frac{2D_1\alpha(J)}{\eta^2}-\frac{D_1^2tc}{\eta^2}\right)}{1-\exp\left(-\frac{2D_1^2}{\eta^2}\right)}\overset{(e)}{\leq}\exp(4)\sum_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*}\frac{\exp\left(-\frac{2\eta_0\alpha(J)}{\eta^2}\right)\exp\left(-\frac{2D_1^2tc}{\eta^2}\right)}{1-\exp\left(-\frac{2D_1^2}{\eta^2}\right)}$$

$$\overset{(f)}{\leq}\exp(4)\sum_{\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*}\frac{\exp\left(-\frac{2D_1^2tc}{\eta^2}\right)}{1-\exp\left(-\frac{2D_1^2}{\eta^2}\right)}$$

where (a) follows from Lemma 7.8 and noting that $\alpha(J)-t'D_1<0$ for $t'>\frac{\alpha(J)}{D_1}+\frac{tc}{2}$, the inequality (b) follows as for $\alpha(J)>0$ we can show that $(\alpha(J)-t'D_1)^2\geq D_1^2(t')^2-2D_1\alpha(J)t'$, (c) follows by noting that $D_1\leq\eta$, (d) follows by applying the infinite geometric progression formula, (e) follows as $D_1\geq\eta_0$, and (f) follows as $\exp\left(-\frac{\eta_0\alpha(J)}{8\eta^2}\right)\leq 1$. □

7.2.2 Concentration of T_{Y^t}

Lemma 7.8. *Define $T_{Y_s}(\boldsymbol{\theta},\boldsymbol{\theta}^*)=L_{Y_s}(\boldsymbol{\theta})-L_{Y_s}(\boldsymbol{\theta}^*)$. Let $\epsilon>0$ be a constant and $\eta>0$ is the constant defined in Assumption 1. Then we can show that,*

$$\mathbb{P}(T_{Y^t}(\boldsymbol{\theta},\boldsymbol{\theta}^*)-\mathbb{E}[T_{Y^t}(\boldsymbol{\theta},\boldsymbol{\theta}^*)]\leq-\epsilon)\leq\exp\left(-\frac{2\epsilon^2}{t\eta^2}\right).$$

Proof. Recall that $T_{Y_s}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = L_{Y_s}(\boldsymbol{\theta}) - L_{Y_s}(\boldsymbol{\theta}^*) = (2Y_s - \mu_{I_s}(\boldsymbol{\theta}) - \mu_{I_s}(\boldsymbol{\theta}^*))(\mu_{I_s}(\boldsymbol{\theta}^*) - \mu_{I_s}(\boldsymbol{\theta}))$. Define $V_s = T_{Y_s}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y_s}(\boldsymbol{\theta}, \boldsymbol{\theta}^*)]$. Note that $\mathbb{E}[V_s] = 0$ which can be shown as follows:

$$\begin{aligned}\mathbb{E}[V_s] &= \mathbb{E}[T_{Y_s}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y_s}(\boldsymbol{\theta}, \boldsymbol{\theta}^*)]] \\ &= \sum_{i=1}^{n} \mathbb{P}(I_s = i)(\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta}))^2 - \sum_{i=1}^{n} \mathbb{P}(I_s = i)(\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta}))^2 = 0.\end{aligned}$$

Also note that $\sum_{s=1}^{t} V_s = T_{Y^t}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}, \boldsymbol{\theta}^*)]$. Next, we show that the moment generating function of the random variable V_s is bounded. First note that the reward Y_s is bounded between $^{-\sqrt{\eta}}/_2$ and $\sqrt{\eta}/_2$. It then follows that:

$$\begin{aligned}V_s &= T_{Y_s}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y_s}(\boldsymbol{\theta}, \boldsymbol{\theta}^*)] \\ &= (2Y_s - \mu_{I_s}(\boldsymbol{\theta}) - \mu_{I_s}(\boldsymbol{\theta}^*))(\mu_{I_s}(\boldsymbol{\theta}^*) - \mu_{I_s}(\boldsymbol{\theta})) - \sum_{i=1}^{n} \mathbb{P}(I_s = i)(\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta}))^2 \leq 2\eta.\end{aligned}$$

Similarly, it can be shown that $V_s \geq -2\eta$. Hence, for the bounded random variable $V_s \in [-2\eta, 2\eta]$ we can show from Hoeffding's lemma in Lemma 7.3 that

$$\mathbb{E}[\exp(\lambda V_s)] \leq \exp\left(\frac{\lambda^2}{8}(2\eta - (-2\eta))^2\right) = \exp\left(2\lambda^2\eta^2\right)$$

for some $\lambda \in \mathbb{R}$. Now for any $\epsilon > 0$ we can show that

$$
\begin{aligned}
\mathbb{P}(T_{Y^t}(\boldsymbol{\theta},\boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta},\boldsymbol{\theta}^*)] \leq -\epsilon) &= \mathbb{P}\left(\sum_{s=1}^{t} V_s \leq -\epsilon\right) = \mathbb{P}\left(-\sum_{s=1}^{t} V_s \geq \epsilon\right) \\
&= \mathbb{P}\left(e^{-\lambda\sum_{s=1}^{t} V_s} \geq e^{\lambda\epsilon}\right) \overset{(a)}{\leq} e^{-\lambda\epsilon}\mathbb{E}\left[e^{-\lambda\sum_{s=1}^{t} V_s}\right] \\
&= e^{-\lambda\epsilon}\mathbb{E}\left[\mathbb{E}\left[e^{-\lambda\sum_{s=1}^{t} V_s}|\hat{\theta}(t-1)\right]\right] \\
&\overset{(b)}{=} e^{-\lambda\epsilon}\mathbb{E}\left[\mathbb{E}\left[e^{-\lambda V_t}|\hat{\theta}(t-1)\right]\mathbb{E}\left[e^{-\lambda\sum_{s=1}^{t-1} V_s}|\hat{\theta}(t-1)\right]\right] \\
&\leq e^{-\lambda\epsilon}\mathbb{E}\left[\exp\left(2\lambda^2\eta^2\right)\mathbb{E}\left[e^{-\lambda\sum_{s=1}^{t-1} V_s}|\hat{\theta}(t-1)\right]\right] \\
&\leq e^{-\lambda\epsilon}e^{2\lambda^2\eta^2}\mathbb{E}\left[e^{-\lambda\sum_{s=1}^{t-1} V_s}\right] \\
&\vdots \\
&\overset{(c)}{\leq} e^{-\lambda\epsilon}e^{2\lambda^2 t\eta^2} \\
&\overset{(d)}{\leq} \exp\left(-\frac{2\epsilon^2}{t\eta^2}\right)
\end{aligned}
$$

where (a) follows by Markov's inequality, (b) follows as V_s is conditionally independent given $\hat{\theta}(s-1)$, (c) follows by unpacking the term for t times and (d) follows by taking $\lambda = \epsilon/4t\eta^2$. $\square$

7.2.3 Proof of correctness for General Sub-Gaussian Case

Lemma 7.9. *Let $L_{Y^t}(\boldsymbol{\theta})$ be the sum of squared errors of the hypothesis parameterized by $\boldsymbol{\theta}$ based on observation vector $\mathbf{Y}^t$ from an underlying sub-Gaussian distribution. Let $\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} := \min\{t : L_{Y^t}(\boldsymbol{\theta}^*) - L_{Y^t}(\boldsymbol{\theta}) > \beta(J,\delta)\}$. Then we can show that*

$$\mathbb{P}\left(L_{Y^{\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}}}}(\boldsymbol{\theta}^*) - L_{Y^{\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}}}}(\boldsymbol{\theta}) > \beta(J,\delta)\right) \leq \frac{\delta}{J}$$

where, $\beta(J,\delta) := \log\left(\dfrac{\left(1+{}^{\eta^2}/_{\eta_0^2}\right)J}{\delta}\right)$.

Proof. Let $T_{Y^t}(\boldsymbol{\theta}^*,\boldsymbol{\theta}) := L_{Y^t}(\boldsymbol{\theta}^*) - L_{Y^t}(\boldsymbol{\theta})$ be the difference of sum of squared errors between hypotheses parameterized by $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}$. As in Lemma 7.10 we again define $\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}}(Y^t) := \min\{t : T_{Y^t}(\boldsymbol{\theta}^*,\boldsymbol{\theta}) > \beta(J,\delta)\}$. Again, for brevity in the following proof we drop the Y^t in $\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}}(Y^t)$. Then

we can show that

$$\begin{aligned}
&\mathbb{P}(\exists t < \infty, T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > \beta(J,\delta)) = \sum_{t=1}^{\infty} \mathbb{P}\left(\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} = t, T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > \beta(J,\delta)\right) \\
&= \sum_{t=1}^{\infty} \mathbb{P}\left(\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} = t, T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta})] > \beta(J,\delta) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta})]\right) \\
&\leq \sum_{t=1}^{\infty} \mathbb{P}\left(\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} = t, T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta})] > \beta(J,\delta) + tD_1\right) \\
&\overset{(a)}{\leq} \sum_{t=1}^{\infty} \exp\left(-\frac{2\left(\beta(J,\delta) + tD_1\right)^2}{t\eta^2}\right) \overset{(b)}{\leq} \sum_{t=1}^{\infty} \exp\left(-\left(\beta(J,\delta) + \frac{t^2 D_1^2}{t\eta^2}\right)\right) \\
&= \sum_{t=1}^{\infty} \exp\left(-\left(\beta(J,\delta) + \frac{tD_1^2}{\eta^2}\right)\right) \\
&= \exp\left(-\beta(J,\delta)\right)\left[1 + \exp(-D_1^2/\eta^2) + \exp(-2D_1^2/\eta^2) + \exp(-3D_1^2/\eta^2) + \ldots\right] \\
&\overset{(c)}{=} \exp\left(-\beta(J,\delta)\right) \frac{1}{1 - \exp\left(-D_1^2/\eta^2\right)} \overset{(d)}{\leq} \exp\left(-\beta(J,\delta)\right)\left(1 + \frac{\eta^2}{D_1^2}\right) \overset{(e)}{\leq} \frac{\delta}{J}.
\end{aligned}$$

where, (a) follows from Lemma 7.8 and noting that $-\mathbb{E}[T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta})] \leq -tD_0$, (b) follows as $(\beta(J,\delta) + tD_0)^2 \geq 2\beta(J,\delta) + t^2 D_0^2$ for $a, b > 0$, (c) follows from the infinite geometric series sum formula, (d) follows as $1/1-\exp(-x) \leq 1 + 1/x$ for $x > 0$, and (e) follows as $\beta(J,\delta) := \log\left(\frac{\left(1 + \eta^2/\eta_0^2\right) J}{\delta}\right)$ and noting that $D_1 \geq \eta_0$. □

7.2.4 Stopping time Correctness Lemma for the Gaussian Case

Lemma 7.10. *Let $L_{Y^t}(\boldsymbol{\theta})$ be the sum of squared errors of the hypothesis parameterized by $\boldsymbol{\theta}$ based on observation vector $\mathbf{Y}^t$ from an underlying Gaussian distribution. Let $\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} := \min\{t : L_{Y^t}(\boldsymbol{\theta}^*) - L_{Y^t}(\boldsymbol{\theta}) > \beta(J,\delta)\}$. Then we can show that*

$$\mathbb{P}\left(L_{Y^{\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}}}}(\boldsymbol{\theta}^*) - L_{Y^{\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}}}}(\boldsymbol{\theta}) > \beta(J,\delta)\right) \leq \frac{\delta}{J}$$

where we define the threshold function as,

$$\beta(J,\delta) := \log(J/\delta) \tag{7.7}$$

Proof. Let $T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) := L_{Y^t}(\boldsymbol{\theta}^*) - L_{Y^t}(\boldsymbol{\theta})$ be the log-likelihood ratio between hypotheses parameterized by $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}$. Define $\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}}(Y^t) := \min\{t : T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > \beta(J,\delta)\}$. For brevity in the

following proof we drop the Y^t in $\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}}(Y^t)$. Then we can show that at time $t \geq \tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}}$ we have

$$\begin{aligned} T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > \beta(J,\delta) &\implies \exp\left(T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta})\right) > \exp\left(\beta(J,\delta)\right) \\ &\implies \left(\frac{\prod_{s=1}^{t} \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta})}{\prod_{s=1}^{t} \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta}^*)} \right) > \exp\left(\beta(J,\delta)\right) \\ &\implies \exp\left(-\beta(J,\delta)\right) \left(\frac{\prod_{s=1}^{t} \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta})}{\prod_{s=1}^{t} \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta}^*)} \right) > 1. \end{aligned} \tag{7.8}$$

Following this we can show that the probability of the event $\{T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > \beta(J,\delta)\}$ is upper bounded by

$$\begin{aligned} &\mathbb{P}(\exists t < \infty, T_{Y^t}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > \beta(J,\delta)) = \sum_{t=1}^{\infty} \mathbb{P}\left(\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} = t\right) = \sum_{t=1}^{\infty} \mathbb{E}[\mathbb{I}\{\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} = t\}] \\ &\overset{(a)}{\leq} \sum_{t=1}^{\infty} \mathbb{E}\left[\mathbb{I}\{\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} = t\} \exp\left(-\beta(J,\delta)\right) \left(\frac{\prod_{s=1}^{t} \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta})}{\prod_{s=1}^{t} \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta}^*)} \right) \right] \\ &= \exp\left(-\beta(J,\delta)\right) \sum_{t=1}^{\infty} \int_{\mathbb{R}^t} \mathbb{I}\{\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} = t\} \frac{\prod_{s=1}^{t} \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta})}{\prod_{s=1}^{t} \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta}^*)} \prod_{s=1}^{t} \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta}^*) dy_1 dy_2 \ldots dy_t \\ &= \exp\left(-\beta(J,\delta)\right) \sum_{t=1}^{\infty} \int_{\mathbb{R}^t} \mathbb{I}\{\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} = t\} \prod_{s=1}^{t} \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta}) dy_1 dy_2 \ldots dy_t \\ &= \exp\left(-\beta(J,\delta)\right) \sum_{t=1}^{\infty} \mathbb{P}(\tau_{\boldsymbol{\theta}^*\boldsymbol{\theta}} = t | I^t, \boldsymbol{\theta}) \leq \exp\left(-\beta(J,\delta)\right) \overset{(b)}{\leq} \frac{\delta}{J}. \end{aligned}$$

where, (a) follows from (7.8), and (b) follows from (7.7). The claim of the lemma follows. □

7.3 Proof of T2 Sample Complexity

7.3.1 Concentration Lemma

This section contains concentration lemma equivalent to the Lemma 7.5 of Appendix 7.2.

Lemma 7.11. *Define $L_{Y^t}(\boldsymbol{\theta}^*)$ as the sum of squared error of the hypothesis parameterized by $\boldsymbol{\theta}^*$. Let $\tau_{\boldsymbol{\theta}^*} = \min\{t : L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*) > \beta(J,\delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*\}$. Then we can bound the probability of the event*

$$\mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) \leq JC_1 \exp(-C_2 t)$$

where, $J := |\boldsymbol{\Theta}|$, $C_1 := 23 + 11\max\left\{1, \dfrac{\eta^2}{2D_1'^2}\right\}$, $C_2 := \dfrac{2D_1'^2 \min\{(c-1)^2, c\}}{\eta^2}$, $\eta > 0$ defined in Definition 1 and $D_1' := \min_{\boldsymbol{\theta}\neq\boldsymbol{\theta}', \boldsymbol{\theta}'\neq\boldsymbol{\theta}^*} \sum_{i=1}^n u_{\boldsymbol{\theta}\boldsymbol{\theta}'}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$.

Proof. We define the event

$$\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t) = \{L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*) < \beta(J,\delta)\}$$
$$\tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t) = \{L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*) < \alpha(J)\}$$

Then we define the stopping time $\tau_{\boldsymbol{\theta}^*}$ as follows:

$$\tau_{\boldsymbol{\theta}^*} = \min\{t : L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*) > \beta(J,\delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*\}$$

which is the first round $L_{Y^t}(\boldsymbol{\theta}')$ crosses $\beta(J,\delta)$ threshold against $L_{Y^t}(\boldsymbol{\theta}^*)$ for all $\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*$. We also define the stopping time $\tilde{\tau}_{\boldsymbol{\theta}^*}$ as follows:

$$\tilde{\tau}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*} = \min\{t : L_{Y^{t'}}(\boldsymbol{\theta}') - L_{Y^{t'}}(\boldsymbol{\theta}^*) > \alpha(J), \forall t' > t\}$$

which is the first round when $L_{Y^t}(\boldsymbol{\theta}^*)$ crosses $\alpha(J)$ threshold against $L_{Y^t}(\boldsymbol{\theta}')$. Then we define the time

$$\tilde{\tau}_{\boldsymbol{\theta}^*} = \max_k \{\tilde{\tau}_{\boldsymbol{\theta}^*\boldsymbol{\theta}'}\}$$

as the last time $\tilde{\tau}_{\boldsymbol{\theta}^*\boldsymbol{\theta}'}$ happens. Define the term D_0' and D_1' as

$$D_0' := \min_{\boldsymbol{\theta},\boldsymbol{\theta}'\neq\boldsymbol{\theta}^*} \sum_{i=1}^n u_{\boldsymbol{\theta}^*\boldsymbol{\theta}}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2, \qquad D_1' := \min_{\boldsymbol{\theta}\neq\boldsymbol{\theta}', \boldsymbol{\theta}'\neq\boldsymbol{\theta}^*} \sum_{i=1}^n u_{\boldsymbol{\theta}\boldsymbol{\theta}'}(i)(\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$$

Let $T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) := L_{Y^t}(\boldsymbol{\theta}') - L_{Y^t}(\boldsymbol{\theta}^*)$ be the sum of squared errors between hypotheses parameterized by $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}^*$. Then it follows that,

$$\begin{aligned}\mathbb{P}(\tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)) &= \mathbb{P}(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - tD_1' < \alpha(J) - tD_1') \\ &\leq \mathbb{P}(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < \alpha(J) - tD_1')\end{aligned}$$

Similarly, we can show that,

$$\begin{aligned}
\mathbb{P}(\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)) &= \mathbb{P}(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < \beta(J, \delta) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)]) \\
&\leq \mathbb{P}(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < \beta(J, \delta) - (t - \tilde{\tau}_{\boldsymbol{\theta}^*})D_0') \\
&= \mathbb{P}\left(T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^t}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < D_0'\left(\frac{\beta(J, \delta)}{D_0'} - t + \tilde{\tau}_{\boldsymbol{\theta}^*}\right)\right)
\end{aligned}$$

Then following the same approach as in Lemma 7.5 we can show that,

$$\begin{aligned}
\mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) &\leq \mathbb{P}\Big(\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\Big) \\
&\leq \mathbb{P}\left(\left\{\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\right\} \bigcap \{\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_1'} + tc\}\right) + \mathbb{P}\left(\{\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\} \bigcap \{\tilde{\tau}_{\boldsymbol{\theta}^*} \geq \frac{\alpha(J)}{D_1'} + tc\}\right) \\
&\leq \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(\{\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\} \bigcap \{\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_1'} + tc\}\right) + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t' : t' \geq \frac{\alpha(J)}{D_1'} + tc} \mathbb{P}\left(\tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t')\right) \\
&\leq \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(\{\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}\} \bigcap \{\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_1'} + tc\}\right) \\
&\quad + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t' : t' \geq \frac{\alpha(J)}{D_1'} + tc} \mathbb{P}\left(T_{Y^{t'}}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - t' D_1' < \alpha(J) - t' D_1'\right) \\
&\leq \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(T_{Y^{t'}}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - \mathbb{E}[T_{Y^{t'}}(\boldsymbol{\theta}', \boldsymbol{\theta}^*)] < D_0'\left(\frac{\beta(J, \delta)}{D_0'} + \frac{\alpha(J)}{D_1'} - t + tc\right)\right) \\
&\quad + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t' : t' \geq \frac{\alpha(J)}{D_1'} + tc} \mathbb{P}\left(T_{Y^{t'}}(\boldsymbol{\theta}', \boldsymbol{\theta}^*) - t' D_1' < \alpha(J) - t' D_1'\right) \\
&\overset{(a)}{\leq} \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp(4) \exp\left(-\frac{2D_1'^2 t(\frac{c}{2} - 1)^2}{\eta^2}\right) + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp(4) \frac{\exp\left(-\frac{2D_1'^2 tc}{\eta^2}\right)}{1 - \exp\left(-\frac{2D_1'^2}{\eta^2}\right)} \\
&\overset{(b)}{\leq} 11 \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1'^2 t(\frac{c}{2} - 1)^2}{\eta^2}\right) + 11 \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1'^2 tc}{\eta^2}\right) \\
&\quad + 11 \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1'^2 tc}{\eta^2}\right) \max\left\{1, \frac{\eta^2}{2D_1'^2}\right\} \\
&\overset{(c)}{\leq} JC_1 \exp(-C_2 t)
\end{aligned}$$

where, (a) follows from Lemma 7.6 and Lemma 7.7 as their result holds for any $D_1', D_0' > 0$, (b) follows from the same steps as in Lemma 7.5 as $D_1' \leq D_0'$, $D_1' > 0$, $D_0' > 0$, and in (c) we substitute $C_1 := 23 + 11 \max\left\{1, \frac{\eta^2}{2D_1^2}\right\}$, $C_2 := \frac{2D_1^2 \min\{(\frac{c}{2}-1)^2, c\}}{\eta^2}$, and $J := |\mathbf{\Theta}|$. □

7.4 GC Convergence Proof for Smooth Hypotheses Space

7.4.1 Concentration Lemmas

Lemma 7.12. *Define* $u_t := \nabla \widehat{P}_t(\boldsymbol{\theta}^*)$. *Then the probabiliity that* $\|u_t\|_{(\nabla^2 P(\boldsymbol{\theta}^*))^{-1}}$ *crosses the threshold* $\sqrt{\frac{cp\log(dt)}{t}} > 0$ *is bounded by*

$$\mathbb{P}\left(\|u_t\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} \geq C_1\sqrt{\frac{cp\log(dt)}{t}}\right) \leq \frac{1}{t^{cp}}.$$

Proof. Define $u_s := \nabla \widehat{P}_s(\boldsymbol{\theta}^*)$. Then we have $u_1, u_2, \ldots, u_t$ as random vectors such that

$$\mathbb{E}\left[\left\|\sum_{s=1}^{t} u_s\right\|^2_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} \middle| u_1, \ldots, u_{s-1}\right] = \mathbb{E}\left[\sum_{s=1}^{t} u_s^\top \left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{-1} u_s \mid u_1, \ldots, u_{s-1}\right] \leq tC_1^2$$

Also we have that $\|u_s\| \leq C_1$. Then following Lemma 7.4 and by setting $\epsilon = cp\log(dt)$ we can show that

$$\begin{aligned}
&\mathbb{P}\left(\|\frac{1}{t}\sum_{s=1}^{t} u_s\|^2_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} - \mathbb{E}\left[\|\frac{1}{t}\sum_{s=1}^{t} u_s\|^2_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}\right] > \frac{1}{t}\sqrt{8tC_1^2\epsilon} + \frac{4C_1}{3\epsilon}\right) \\
&= \mathbb{P}\left(\|\frac{1}{t}\sum_{s=1}^{t} u_s\|^2_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} > C_1^2 + C_1\sqrt{\frac{8\epsilon}{t}} + \frac{4C_1}{3\epsilon}\right) \\
&\leq \mathbb{P}\left(\|\sum_{s=1}^{t} u_s\|^2_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} > C_1\sqrt{\frac{8\epsilon}{t}}\right) = \mathbb{P}\left(\|\sum_{s=1}^{t} u_s\|^2_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} > 4C_1\sqrt{\frac{cp\log(dt)}{t}}\right) \\
&\leq \exp(-cp\log(dt)) = \left(\frac{1}{dt}\right)^{cp} \leq \frac{1}{t^{cp}}
\end{aligned}$$

The claim of the lemma follows. □

Lemma 7.13. *Under Assumption 3 for any* $x > 0$ *we can show that the conditional probability*

$$\mathbb{P}\left(\sum_{s=1}^{t} -2\left(Y_s - \mu_{I_s}(\boldsymbol{\theta}^*)\right)\lambda_{\max}\left(\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\mu_{I_s}(\boldsymbol{\theta})\right) \geq x \mid I^{t-1}\right) \leq 2\exp\left(-\frac{x^2}{2t\eta^2\lambda_1^2}\right).$$

Proof. Recall that the Assumption 3 states that $\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\mu_{I_s}(\boldsymbol{\theta})$ is bounded such that $\lambda_{\max}\left(\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\mu_{I_s}(\boldsymbol{\theta})\right) \leq \lambda_1$ for each $s \in [1,t]$ and for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Following the Assumption 3 we have that

$$-2\left(Y_s - \mu_{I_s}(\boldsymbol{\theta}^*)\right)\lambda_{\max}\left(\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\mu_{I_s}(\boldsymbol{\theta})\right) \in [-2\eta\lambda_1, 2\eta\lambda_1],$$

for all $s \in [t]$ with

$$\mathbb{E}\left[-2\left(Y_s - \mu_{I_s}(\boldsymbol{\theta}^*)\right)\lambda_{\max}\left(\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\mu_{I_s}(\boldsymbol{\theta})\right) \mid I^{s-1}\right] = 0$$

and conditioned on $(\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_t)$, the random variables $-2\left(Y_s - \mu_{I_s}(\boldsymbol{\theta}^*)\right)$ are independent. Hence, we can use Hoeffding's inequality to get

$$\begin{aligned}\mathbb{P}\left(\sum_{s=1}^{t} -2\left(Y_s - \mu_{I_s}(\boldsymbol{\theta}^*)\right)\lambda_{\max}\left(\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\mu_{I_s}(\boldsymbol{\theta})\right) \geq x \mid I_1, \ldots, I_{t-1}\right) &\leq 2\exp\left(-\frac{2x^2}{t4\eta^2\lambda_1^2}\right) \\ &= 2\exp\left(-\frac{x^2}{2t\eta^2\lambda_1^2}\right).\end{aligned}$$

The claim of the lemma follows. □

Lemma 7.14. *Let* $\nabla^2\widehat{P}_t(\boldsymbol{\theta}^*) = \frac{1}{t}\sum_{s=1}^{t} L_{Y_s}(\boldsymbol{\theta}^*)$ *and* $\nabla^2 P_t(\boldsymbol{\theta}^*) = \frac{1}{t}\sum_{s=1}^{t}\nabla^2\mathbb{E}[L_{Y_s}(\boldsymbol{\theta}^*)|\mathcal{F}^{s-1}]$. *Then we can bound the*

$$\mathbb{P}\left(\lambda_{\max}(\nabla^2\widehat{P}_t(\boldsymbol{\theta}^*) - \nabla^2 P_t(\theta^*)) > \sqrt{\frac{2\eta^2\lambda_1^2 cp\log(dt)}{t}}|I^{t-1}\right) \leq \frac{2}{(dt)^{cp}}.$$

Proof. Recall that $\nabla^2\widehat{P}_t(\boldsymbol{\theta}^*) = \frac{1}{t}\sum_{s=1}^{t} L_{Y_s}(\boldsymbol{\theta}^*)$ and $\nabla^2 P_s(\theta^*) = \nabla^2\mathbb{E}[L_{Y_s}(\boldsymbol{\theta}^*)|\mathcal{F}^{s-1}]$. We define $\nabla^2 P_t(\boldsymbol{\theta}^*) = \frac{1}{t}\sum_{s=1}^{t}\nabla^2\mathbb{E}[L_{Y_s}(\boldsymbol{\theta}^*)|\mathcal{F}^{s-1}]$. Then we can show that,

$$
\begin{aligned}
&\mathbb{P}\left(\lambda_{\max}(\nabla^2\widehat{P}_t(\boldsymbol{\theta}^*)-\nabla^2 P_t(\theta^*))>\sqrt{\frac{2\eta^2\lambda_1^2 cp\log(dt)}{t}}|I^{t-1}\right)\\
&=\mathbb{P}\left(\lambda_{\max}\left(\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\frac{1}{t}\sum_{s=1}^{t}L_{Y_s}(\boldsymbol{\theta})-\frac{1}{t}\sum_{s=1}^{t}\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\mathbb{E}[L_{Y_s}(\boldsymbol{\theta})|\mathcal{F}^{s-1}]\right)>\sqrt{\frac{2\eta^2\lambda_1^2 cp\log(dt)}{t}}|I^{t-1}\right)\\
&=\mathbb{P}\left(\lambda_{\max}\left(\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\frac{1}{t}\sum_{s=1}^{t}\left(L_{Y_s}(\boldsymbol{\theta})-\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\mathbb{E}[L_{Y_s}(\boldsymbol{\theta})|\mathcal{F}^{s-1}]\right)\right)>\sqrt{\frac{2\eta^2\lambda_1^2 cp\log(dt)}{t}}|I^{t-1}\right)\\
&\overset{(a)}{\leq}\mathbb{P}\left(\frac{1}{t}\sum_{s=1}^{t}\lambda_{\max}\left(\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}L_{Y_s}(\boldsymbol{\theta})-\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\mathbb{E}[L_{Y_s}(\boldsymbol{\theta})|\mathcal{F}^{s-1}]\right)>\sqrt{\frac{2\eta^2\lambda_1^2 cp\log(dt)}{t}}|I^{t-1}\right)\\
&\overset{(b)}{=}\mathbb{P}\left(\sum_{s=1}^{t}-2\left(Y_s-\mu_{I_s}\left(\boldsymbol{\theta}^*\right)\right)\lambda_{\max}\left(\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\mu_{I_s}\left(\boldsymbol{\theta}'\right)\right)>t\sqrt{\frac{2\eta^2\lambda_1^2 cp\log(dt)}{t}}|I^{t-1}\right)\\
&\overset{(c)}{\leq}2\exp\left(-\frac{t^2 2\eta^2\lambda_1^2 cp\log(dt)}{t}\cdot\frac{1}{2t\eta^2\lambda_1^2}\right)\leq 2\left(\frac{1}{dt}\right)^{cp}.
\end{aligned}
$$

where, (a) follows by triangle inequality, (b) follows by substituting the value of $\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}L_{Y_s}(\boldsymbol{\theta})-\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\mathbb{E}[L_{Y_s}(\boldsymbol{\theta})|\mathcal{F}^{s-1}]$ from Lemma 7.17, and (c) follows from Lemma 7.13. □

7.4.2 Support Lemma

Lemma 7.15. *Let the j-th row and k-th column entry in the Hessian matrix $\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}(L_{Y_s}(\boldsymbol{\theta}))$ be denoted as $[\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}(L_{Y_s}(\boldsymbol{\theta}))]_{jk}$. Then we have that*

$$
[\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}(L_{Y_s}(\boldsymbol{\theta}))]_{jk}=2\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k}+2\left(\mu_{I_s}(\boldsymbol{\theta})-Y_s\right)\frac{\partial^2\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_k}.
$$

Proof. We want to evaluate the Hessian $\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}(L_{Y_s}(\boldsymbol{\theta}))$ at any $\boldsymbol{\theta}'\in\boldsymbol{\Theta}$. We denote the j-th row and k-th column entry in the Hessian matrix as $[\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}(L_{Y_s}(\boldsymbol{\theta}))]_{jk}$. Then we can show that

$$
\begin{aligned}
[\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}(L_{Y_s}(\boldsymbol{\theta}))]_{jk}&\coloneqq\frac{\partial}{\partial\boldsymbol{\theta}_j}\left[\frac{\partial(\mu_{I_s}(\boldsymbol{\theta})-Y_s)^2}{\partial\boldsymbol{\theta}_k}\right]=\frac{\partial}{\partial\boldsymbol{\theta}_j}\left[2(\mu_{I_s}(\boldsymbol{\theta})-Y_s)\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k}\right]\\
&=\frac{\partial}{\partial\boldsymbol{\theta}_j}\left[2\mu_{I_s}(\boldsymbol{\theta})\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k}-2Y_s\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k}\right]\\
&=2\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k}+2\mu_{I_s}(\boldsymbol{\theta})\frac{\partial^2\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_k}-2Y_s\frac{\partial^2\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_k}-2\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}\frac{\partial Y_s}{\partial\boldsymbol{\theta}_k}\\
&=2\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k}+2\left(\mu_{I_s}(\boldsymbol{\theta})-Y_s\right)\frac{\partial^2\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_k}
\end{aligned}
$$

The claim of the lemma follows. □

Lemma 7.16. *Let the j-th row and k-th column entry in the Hessian matrix $\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}(\mathbb{E}[L_{Y_s}(\boldsymbol{\theta})|\mathcal{F}^{s-1}])$ be denoted as $[\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}(\mathbb{E}[L_{Y_s}(\boldsymbol{\theta})|\mathcal{F}^{s-1}])]_{jk}$. Then we have that*

$$\left[\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\sum_{s=1}^{t}\mathbb{E}[L_{Y_s}(\boldsymbol{\theta})|\mathcal{F}^{s-1}]\right]_{jk} = 2\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k} + 2\left(\mu_{I_s}(\boldsymbol{\theta}) - \mu_{I_s}(\boldsymbol{\theta}^*)\right)\frac{\partial^2\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_k}.$$

Proof. Now we want to evaluate the Hessian $\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}(\mathbb{E}[L_{Y_s}(\boldsymbol{\theta})|\mathcal{F}^{s-1}])$ at any $\boldsymbol{\theta}' \in \boldsymbol{\Theta}$. We denote the j-th row and k-th column entry in the Hessian matrix as $[\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}(\mathbb{E}[L_{Y_s}(\boldsymbol{\theta})|\mathcal{F}^{s-1}])]_{jk}$. Then we can show that

$$\begin{aligned}
\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\sum_{s=1}^{t}\mathbb{E}[L_{Y_s}(\boldsymbol{\theta})|\mathcal{F}^{s-1}] &= \nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\sum_{s=1}^{t}\left(\mu^2_{I_s}(\boldsymbol{\theta}) + \mathbb{E}[Y_s^2|\mathcal{F}^{s-1}] - 2\mathbb{E}[Y_s|\mathcal{F}^{s-1}]\mu_{I_s}(\boldsymbol{\theta})\right)\\
&= \nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\sum_{s=1}^{t}\left(\mu^2_{I_s}(\boldsymbol{\theta}) + \mu^2_{I_s}(\boldsymbol{\theta}') + \frac{1}{2} - 2\mu_{I_s}(\boldsymbol{\theta}^*)\mu_{I_s}(\boldsymbol{\theta})\right)\\
&= \sum_{s=1}^{t}\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\left(\left(\mu_{I_s}(\boldsymbol{\theta}^*) - \mu_{I_s}(\boldsymbol{\theta})\right)^2 + \frac{1}{2}\right)\\
&= \sum_{s=1}^{t}\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\left(\left(\mu_{I_s}(\boldsymbol{\theta}^*) - \mu_{I_s}(\boldsymbol{\theta})\right)^2\right) \qquad (7.9)
\end{aligned}$$

We now denote the j-th row and k-th column entry of the Hessian Matrix $\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}((\mu_{I_s}(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}^*))^2)$ as $\left[\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}((\mu_{I_s}(\boldsymbol{\theta}) - \mu_{I_s}(\boldsymbol{\theta}^*))^2)\right]_{jk}$. Then we can show that

$$\begin{aligned}
\left[\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}((\mu_{I_s}(\boldsymbol{\theta}) - \mu_{I_s}(\boldsymbol{\theta}^*))^2)\right]_{jk} &:= \frac{\partial}{\partial\boldsymbol{\theta}_j}\left[\frac{\partial(\mu_{I_s}(\boldsymbol{\theta}) - \mu_{I_s}(\boldsymbol{\theta}^*))^2}{\partial\boldsymbol{\theta}_k}\right] = \frac{\partial}{\partial\boldsymbol{\theta}_j}\left[2(\mu_{I_s}(\boldsymbol{\theta}) - \mu_{I_s}(\boldsymbol{\theta}^*))\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k}\right]\\
&= \frac{\partial}{\partial\boldsymbol{\theta}_j}\left[2\mu_{I_s}(\boldsymbol{\theta})\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k} - 2\mu_i(\boldsymbol{\theta}^*)\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k}\right]\\
&= 2\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k} + 2\mu_{I_s}(\boldsymbol{\theta})\frac{\partial^2\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j\boldsymbol{\theta}_k}\\
&\quad - 2\mu_{I_s}(\boldsymbol{\theta}^*)\frac{\partial^2\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j\boldsymbol{\theta}_k} - 2\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}\frac{\partial\mu_{I_s}(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}_k}\\
&= 2\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k} + 2\left(\mu_{I_s}(\boldsymbol{\theta}) - \mu_{I_s}(\boldsymbol{\theta}^*)\right)\frac{\partial^2\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_k}
\end{aligned}$$

Plugging this back in eq. (7.9) we get that

$$\left[\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'}\sum_{s=1}^{t}\mathbb{E}[L_{Y_s}(\boldsymbol{\theta})|\mathcal{F}^{s-1}]\right]_{jk} = 2\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}\frac{\partial\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k} + 2\left(\mu_{I_s}(\boldsymbol{\theta}) - \mu_{I_s}(\boldsymbol{\theta}^*)\right)\frac{\partial^2\mu_{I_s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_k}.$$

□

Lemma 7.17. *The sum of the difference of the Hessians* $\sum_{s=1}^{t} \nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'} L_{Y_s}(\boldsymbol{\theta}) - \mathbb{E}\left[\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'} L_{Y_s}(\boldsymbol{\theta}) \mid \mathcal{F}^{s-1}\right]$ *is given by*

$$\sum_{s=1}^{t} \nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'} L_{Y_s}(\boldsymbol{\theta}) - \mathbb{E}\left[\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'} L_{Y_s}(\boldsymbol{\theta}) \mid \mathcal{F}^{s-1}\right] = \sum_{s=1}^{t} -2(Y_s - \mu_{I_s}(\boldsymbol{\theta})) \frac{\partial^2 \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k}$$

Proof. First note that the difference $\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'} L_{Y_s}(\boldsymbol{\theta}) - \mathbb{E}\left[\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'} L_{Y_s}(\boldsymbol{\theta}) \mid \mathcal{F}^{s-1}\right]$ is given by

$$\begin{aligned}
\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'} L_{Y_s}(\boldsymbol{\theta}) - \mathbb{E}\left[\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'} L_{Y_s}(\boldsymbol{\theta}) \mid \mathcal{F}^{s-1}\right] &\overset{(a)}{=} 2\frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j}\frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} + 2\left(\mu_{I_s}(\boldsymbol{\theta}) - Y_s\right)\frac{\partial^2 \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} \\
&\quad - 2\frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j}\frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} - 2\left(\mu_{I_s}(\boldsymbol{\theta}) - \mu_{I_s}(\boldsymbol{\theta}^*)\right)\frac{\partial^2 \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} \\
&= -2(Y_s - \mu_{I_s}(\boldsymbol{\theta}))\frac{\partial^2 \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} \qquad (7.10)
\end{aligned}$$

where, (a) follows from Lemma 7.15 and Lemma 7.16. Plugging this equality in Equation (7.10) below we get

$$\sum_{s=1}^{t} \nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'} L_{Y_s}(\boldsymbol{\theta}) - \mathbb{E}\left[\nabla^2_{\boldsymbol{\theta}=\boldsymbol{\theta}'} L_{Y_s}(\boldsymbol{\theta}) \mid \mathcal{F}^{s-1}\right] = \sum_{s=1}^{t} -2(Y_s - \mu_{I_s}(\boldsymbol{\theta})) \frac{\partial^2 \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k}.$$

The claim of the lemma follows. □

Lemma 7.18. *Let* $\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* = \left(\nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t)\right)^{-1} \nabla \widehat{P}_t(\boldsymbol{\theta}^*)$ *where* $\widetilde{\boldsymbol{\theta}}_t$ *is between* $\widehat{\boldsymbol{\theta}}_t$ *and* $\boldsymbol{\theta}^*$*. Then we can show that*

$$\left\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} \leq \left\|\left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{1/2}\left(\nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t)\right)^{-1}\left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{1/2}\right\| \left\|\nabla \widehat{P}_t(\boldsymbol{\theta}^*)\right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}.$$

Proof. We begin with the definition of $\left\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)}$ as follows:

$$\begin{aligned}
\left\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} &\overset{(a)}{=} \sqrt{(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)^T \nabla^2 P_t(\boldsymbol{\theta}^*)(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)} \\
&\overset{(b)}{=} \sqrt{\left(\left(\nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t)\right)^{-1} \nabla \widehat{P}_t(\boldsymbol{\theta}^*)\right)^T \nabla^2 P_t(\boldsymbol{\theta}^*) \left(\left(\nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t)\right)^{-1} \nabla \widehat{P}_t(\boldsymbol{\theta}^*)\right)} \\
&\overset{(c)}{\leq} \left\|\nabla^2 P_t(\boldsymbol{\theta}^*)^{1/2}\left(\nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t)\right)^{-1} \nabla^2 P_t(\boldsymbol{\theta}^*)^{1/2}\right\| \sqrt{\left(\nabla \widehat{P}_t(\boldsymbol{\theta}^*)^T \left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{-1} \nabla \widehat{P}_t(\boldsymbol{\theta}^*)\right)} \\
&= \left\|\left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{1/2}\left(\nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}}_t)\right)^{-1}\left(\nabla^2 P_t(\boldsymbol{\theta}^*)\right)^{1/2}\right\| \left\|\nabla \widehat{P}_t(\boldsymbol{\theta}^*)\right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}.
\end{aligned}$$

where, (a) follows as $\|x\|_M = \sqrt{x^T M x}$, (b) follows as $\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} = \left(\nabla^2 \widehat{P}_t(\widetilde{\boldsymbol{\theta}})\right)^{-1} \nabla \widehat{P}_t(\boldsymbol{\theta}^*)$, and (c) follows from Cauchy Schwarz inequality. The claim of the lemma follows. □

7.5 Proof of GC Sample Complexity in Bandit Setting

7.5.1 Concentration Lemma for Continuous Hypotheses

Lemma 7.19. *Define $L_{Y^t}(\widetilde{\boldsymbol{\theta}}^*)$ as the sum of squared errors of hypothesis parameterized by $\widetilde{\boldsymbol{\theta}}^*$. Let $\tau_{\widetilde{\boldsymbol{\theta}}^*} := \min\{t : L_{Y^t}(\widetilde{\boldsymbol{\theta}}') - L_{Y^t}(\widetilde{\boldsymbol{\theta}}^*) > \beta(t,\delta), \forall \widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*\}$. Then we can bound the probability of the event*

$$\mathbb{P}(\tau_{\widetilde{\boldsymbol{\theta}}^*} > t) \leq |\mathcal{C}_\epsilon| C_1 \exp\left(-C_2 t\right)$$

where, $C_1 := 23 + 11 \max\left\{1, \dfrac{\eta^2}{2\tilde{D}_1^2}\right\}$, $C_2 := \dfrac{2\tilde{D}_1^2 \min\{(\frac{c}{2}-1)^2, c\}}{\eta^2}$, $\eta > 0$ defined in Definition 1, $\tilde{D}_1, \tilde{D}_0$ are defined as in Theorem 2 for $\widetilde{\boldsymbol{\Theta}}$, and $\mathcal{C}_\epsilon$ is the ϵ-cover.

Proof. Recall for a $\widetilde{\boldsymbol{\theta}}$ in the discretized parameter space $\tilde{\boldsymbol{\Theta}}$ we define the alternate hypothesis as $\widetilde{\boldsymbol{\theta}}' \in \tilde{\boldsymbol{\Theta}}$ such that $\widetilde{\boldsymbol{\theta}} \neq \widetilde{\boldsymbol{\theta}}'$. We define the two error event as follows:

$$\xi_{\widetilde{\boldsymbol{\theta}}'\widetilde{\boldsymbol{\theta}}^*}(t) := \{L_{Y^t}(\widetilde{\boldsymbol{\theta}}') - L_{Y^t}(\widetilde{\boldsymbol{\theta}}) < \beta(t,\delta)\}$$
$$\tilde{\xi}_{\widetilde{\boldsymbol{\theta}}'\widetilde{\boldsymbol{\theta}}^*}(t) := \{L_{Y^t}(\widetilde{\boldsymbol{\theta}}') - L_{Y^t}(\widetilde{\boldsymbol{\theta}}^*) < \alpha(t)\}$$

where $L_{Y^t}(\widetilde{\boldsymbol{\theta}}')$ is the sum of squared errors of hypothesis $\widetilde{\boldsymbol{\theta}}'$ and $\beta(t,\delta)$ as defined in (5.4). Then we define the stopping time $\tau_{\widetilde{\boldsymbol{\theta}}^*}$ as follows:

$$\tau_{\widetilde{\boldsymbol{\theta}}^*} := \min\{t : L_{Y^t}(\widetilde{\boldsymbol{\theta}}') - L_{Y^t}(\widetilde{\boldsymbol{\theta}}^*) > \beta(t,\delta), \forall \widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*\}$$

which is the first round $L_{Y^t}(\widetilde{\boldsymbol{\theta}}')$ crosses $\beta(t,\delta)$ threshold against $L_{Y^t}(\widetilde{\boldsymbol{\theta}}^*)$ for all $\widetilde{\boldsymbol{\theta}}' \in \boldsymbol{Alt}(\widetilde{\boldsymbol{\theta}}^*)$. We also define the stopping time $\tilde{\tau}_{\widetilde{\boldsymbol{\theta}}^*, \widetilde{\boldsymbol{\theta}}'}$ as follows:

$$\tilde{\tau}_{\widetilde{\boldsymbol{\theta}}'\widetilde{\boldsymbol{\theta}}^*} := \min\{t : L_{Y^{t'}}(\widetilde{\boldsymbol{\theta}}') - L_{Y^{t'}}(\widetilde{\boldsymbol{\theta}}^*) > \alpha(t), \forall t' > t\}$$

which is the first round when $L_{Y^t}(\widetilde{\boldsymbol{\theta}}')$ crosses $\alpha(t)$ threshold against $L_{Y^t}(\widetilde{\boldsymbol{\theta}}^*)$. Then we define the time

$$\tilde{\tau}_{\widetilde{\boldsymbol{\theta}}^*} := \sup_{\widetilde{\boldsymbol{\theta}}' \in \tilde{\boldsymbol{\Theta}}} \{\tilde{\tau}_{\widetilde{\boldsymbol{\theta}}'\widetilde{\boldsymbol{\theta}}^*}\}$$

as the last time $\tilde{\tau}_{\widetilde{\boldsymbol{\theta}}'\widetilde{\boldsymbol{\theta}}^*}$ happens. Then it follows that,

$$\begin{aligned}\mathbb{P}(\tilde{\xi}_{\widetilde{\boldsymbol{\theta}}'\widetilde{\boldsymbol{\theta}}^*}(t)) &= \mathbb{P}(T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*) - \mathbb{E}[T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*)] < \alpha(t) - \mathbb{E}[T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*)] \\ &\overset{(a)}{\leq} \mathbb{P}(T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*) - \mathbb{E}[T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*)] < \alpha(t) - t\tilde{D}_1)\end{aligned}$$

where, (a) follows by a similar argument as in Lemma 7.5 and noting $\tilde{D}_1 > 0$. Similarly, we can show that,

$$\begin{aligned}\mathbb{P}(\xi_{\widetilde{\boldsymbol{\theta}}'\widetilde{\boldsymbol{\theta}}^*}(t)) &= \mathbb{P}(T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*) - \mathbb{E}[T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*)] < \beta(t,\delta) - \mathbb{E}[T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*)]) \\ &\leq \mathbb{P}(T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*) - \mathbb{E}[T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*)] < \beta(t,\delta) - (t - \tilde{\tau}_{\widetilde{\boldsymbol{\theta}}^*})\tilde{D}_0) \\ &= \mathbb{P}\left(T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*) - \mathbb{E}[T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*)] < \tilde{D}_0\left(\frac{\beta(t,\delta)}{\tilde{D}_0} - t + \tilde{\tau}_{\widetilde{\boldsymbol{\theta}}^*}\right)\right)\end{aligned}$$

Then we can follow the same steps as in Lemma 7.5 and show that,

$$\begin{aligned}\mathbb{P}(\tau_{\widetilde{\boldsymbol{\theta}}^*} > t) &\leq \mathbb{P}\Big(\bigcup_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \xi_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}(t)\Big) \\ &\leq \mathbb{P}\left(\left\{\bigcup_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \xi_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}(t)\right\} \bigcap \{\tilde{\tau}_{\widetilde{\boldsymbol{\theta}}^*} < \frac{\alpha(t)}{\tilde{D}_1} + \frac{tc}{2}\}\right) + \mathbb{P}\left(\bigcup_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \{\xi_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}(t)\} \bigcap \{\tilde{\tau}_{\boldsymbol{\theta}^*} \geq \frac{\alpha(t)}{\tilde{D}_1} + \frac{tc}{2}\}\right) \\ &\leq \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \mathbb{P}\left(\{\xi_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}(t)\} \bigcap \{\tilde{\tau}_{\widetilde{\boldsymbol{\theta}}^*} < \frac{\alpha(t)}{\tilde{D}_1} + \frac{tc}{2}\}\right) + \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \sum_{t': t' \geq \frac{\alpha(t)}{\tilde{D}_1} + \frac{tc}{2}} \mathbb{P}\left(\tilde{\xi}_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}(t')\right) \\ &\leq \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \mathbb{P}\left(\{\xi_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}\} \bigcap \{\tilde{\tau}_{\widetilde{\boldsymbol{\theta}}^*} < \frac{\alpha(t)}{\tilde{D}_1} + \frac{tc}{2}\}\right) + \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \sum_{t': t' \geq \frac{\alpha(t)}{\tilde{D}_1} + \frac{tc}{2}} \mathbb{P}\left(T_{Y^{t'}}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*) - t'\tilde{D}_1 < \alpha(J) - t'\tilde{D}_1\right) \\ &\leq \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \mathbb{P}\left(T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*) - \mathbb{E}[T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*)] < \tilde{D}_0\left(\frac{\beta(t,\delta)}{\tilde{D}_0} + \frac{\alpha(t)}{\tilde{D}_1} - t + \frac{tc}{2}\right)\right) \\ &\quad + \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \sum_{t': t' \geq \frac{\alpha(t)}{\tilde{D}_1} + \frac{tc}{2}} \mathbb{P}\left(T_{Y^{t'}}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*) - t'\tilde{D}_1 < \alpha(t) - t'\tilde{D}_1\right)\end{aligned}$$

$$\overset{(a)}{\leq} \exp(4) \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \exp\left(-\frac{2\tilde{D}_0^2 t(\frac{c}{2}-1)^2}{\eta^2}\right) + \exp(4) \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \frac{\exp\left(-\frac{2\tilde{D}_1^2 tc}{\eta^2}\right)}{1-\exp\left(-\frac{2\tilde{D}_1^2}{\eta^2}\right)}$$

$$\overset{(b)}{\leq} \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \exp\left(-\frac{2\tilde{D}_1^2 t \min\{(\frac{c}{2}-1)^2, c\}}{\eta^2}\right) \left[23 + 11\max\left\{1, \frac{\eta^2}{2\tilde{D}_1^2}\right\}\right] \overset{(c)}{\leq} |\mathcal{C}_\epsilon| C_1 \exp(-C_2 t)$$

where, (a) follows from Lemma 7.20 and Lemma 7.21 as their result holds for any finite $\tilde{D}_1, \tilde{D}_0 > 0$, (b) follows as $\tilde{D}_1 \leq \tilde{D}_0$, and in (c) we substitute $C_1 := 23 + 11\max\left\{1, \frac{\eta^2}{2\tilde{D}_1^2}\right\}$, $C_2 := \frac{2\tilde{D}_1^2 \min\{(\frac{c}{2}-1)^2, c\}}{\eta^2}$ and $\mathcal{C}_\epsilon$ is the ϵ-cover. □

Lemma 7.20. *Let $T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*) = L_{Y^t}(\widetilde{\boldsymbol{\theta}}^*) - L_{Y^t}(\widetilde{\boldsymbol{\theta}}')$. Let $\tilde{D}_1, \tilde{D}_0$ are defined as in Theorem 2 for $\widetilde{\boldsymbol{\Theta}}$, and $\alpha(t)$ and $\beta(t,\delta)$ be the two thresholds. Then we can show that*

$$\sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \mathbb{P}\left(T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*) - \mathbb{E}[T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*)] < \tilde{D}_0\left(\frac{\beta(t,\delta)}{\tilde{D}_0} + \frac{\alpha(t)}{\tilde{D}_1} - t + \frac{tc}{2}\right)\right) \leq \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \exp\left(-\frac{2\tilde{D}_1^2 t(c/2-1)^2}{\eta^2}\right).$$

for some constant $0 < c < 1$.

Proof. Let us recall that the critical number of samples is given by $(1+c)M$ for some constant $0 < c < 1$ and $M := \frac{\beta(t,\delta)}{\tilde{D}_0} + \frac{\alpha(t)}{\tilde{D}_1}$. Then we can show that for the constant $c > 0$ the following

$$\sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \mathbb{P}\left(T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*) - \mathbb{E}[T_{Y^t}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*)] < \tilde{D}_0\left(\frac{\beta(t,\delta)}{\tilde{D}_0} + \frac{\alpha(t)}{\tilde{D}_1} - t + \frac{tc}{2}\right)\right) \tag{7.11}$$

$$\overset{(a)}{\leq} \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \exp\left(-\frac{2\tilde{D}_0^2\left(\frac{\beta(t,\delta)}{\tilde{D}_0} + \frac{\alpha(t)}{\tilde{D}_1} + t(\frac{c}{2}-1)\right)^2}{t\eta^2}\right) \overset{(b)}{\leq} \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \exp\left(-\frac{2\tilde{D}_1^2\left(M + t(\frac{c}{2}-1)\right)^2}{t\eta^2}\right)$$

$$\overset{(c)}{\leq} \exp(4(1-c/2)) \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \exp\left(-\frac{2\tilde{D}_1^2 M t^2 (c/2-1)^2}{t\eta^2}\right) \leq \exp(4) \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \exp\left(-\frac{2\tilde{D}_1^2 M t (c/2-1)^2}{\eta^2}\right) \tag{7.12}$$

$$\overset{(d)}{\leq} \exp(4) \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \exp\left(-\frac{2\tilde{D}_1^2 t(c/2-1)^2}{\eta^2}\right)$$

where (a) follows by applying Lemma 7.8, (b) follows from definition of M, (c) follows by the same argument as in Lemma 7.6, $\square$

Lemma 7.21. *Let $T_{Y^{t'}}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*) := L_{Y^{t'}}(\widetilde{\boldsymbol{\theta}}') - L_{Y^{t'}}(\widetilde{\boldsymbol{\theta}}^*)$ be the difference of sum of squared errors between $\widetilde{\boldsymbol{\theta}}'$ and $\widetilde{\boldsymbol{\theta}}'*$. Let $\tilde{D}_1, \tilde{D}_0$ are defined as in Theorem 2 for $\widetilde{\boldsymbol{\Theta}}$, and $\alpha(t)$ be the threshold depending in t. Then we can show that,*

$$\sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \sum_{t': t' > \frac{\alpha(t)}{\tilde{D}_1} + tc} \mathbb{P}\left(T_{Y^{t'}}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*) - \mathbb{E}[T_{Y^{t'}}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*)] < \alpha(t) - t'\tilde{D}_1\right) \leq \exp(4) \sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \frac{\exp\left(-\frac{2\tilde{D}_1^2 ct}{\eta^2}\right)}{1 - \exp\left(-\frac{2\tilde{D}_1^2}{\eta^2}\right)}$$

for some constant c such that $0 < c < 1$, and η defined in Definition 1.

Proof. Let us recall that for a $\widetilde{\boldsymbol{\theta}}' \in \tilde{\boldsymbol{\Theta}}$ and $\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*$ we have

$$\sum_{\widetilde{\boldsymbol{\theta}}' \neq \widetilde{\boldsymbol{\theta}}^*} \sum_{t': t' > \frac{\alpha(t)}{\tilde{D}_1} + tc} \mathbb{P}\left(T_{Y^{t'}}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*) - \mathbb{E}[T_{Y^{t'}}(\widetilde{\boldsymbol{\theta}}', \widetilde{\boldsymbol{\theta}}^*)] < \alpha(t) - t'\tilde{D}_1\right)$$

$$\overset{(a)}{\leq} \sum_{\widetilde{\boldsymbol{\theta}}' \neq \tilde{\boldsymbol{\theta}}^*} \sum_{t': t' > \frac{\alpha(t)}{\tilde{D}_1} + \frac{tc}{2}} \exp(4) \exp\left(-\frac{2\tilde{D}_1^2 (t')^2}{\eta^2}\right) \overset{(b)}{\leq} \sum_{\widetilde{\boldsymbol{\theta}}' \neq \tilde{\boldsymbol{\theta}}^*} \frac{\exp\left(-\frac{2\tilde{D}_1^2 ct}{\eta^2}\right)}{1 - \exp\left(-\frac{2\tilde{D}_1^2}{\eta^2}\right)}$$

where (a) follows by the same steps as in relation (a) and (b) in Lemma 7.7 which holds as $\tilde{D}_1, \alpha(t) > 0$, finally (b) follows by applying the formula for geometric sums and then following the same steps as in relation (d). (e) and (f) in Lemma 7.7. $\square$

7.5.2 Stopping time Correctness Lemma for Gaussian Case in Bandit setting

Lemma 7.22. *Let $L_{Y^t}(\widetilde{\boldsymbol{\theta}}')$ be the sum of squared errors of the hypothesis parameterized by $\widetilde{\boldsymbol{\theta}}'$ based on observation vector Y^t. Let $\tau_{\widetilde{\boldsymbol{\theta}}^* \widetilde{\boldsymbol{\theta}}'} := \min\{t : L_{Y^t}(L_{Y^t}(\widetilde{\boldsymbol{\theta}}^* - \widetilde{\boldsymbol{\theta}}')) > \beta(t, \delta)\}$. Then we can show that*

$$\mathbb{P}\left(L_{Y^{\tau_{\widetilde{\boldsymbol{\theta}}^* \widetilde{\boldsymbol{\theta}}'}}}(\widetilde{\boldsymbol{\theta}}^*) - L_{Y^{\tau_{\widetilde{\boldsymbol{\theta}}^* \widetilde{\boldsymbol{\theta}}'}}}(\widetilde{\boldsymbol{\theta}}') > \beta(t, \delta)\right) \leq \exp(-\beta(t, \delta))$$

where, $\beta(t, \delta)$ is defined in (7.7)*.*

Proof. Let $T_{Y^t}(\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}') := L_{Y^t}(\widetilde{\boldsymbol{\theta}}^*) - L_{Y^t}(\widetilde{\boldsymbol{\theta}}')$. Define $\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}(Y^t) = \min\{t : T_{Y^t}(\widetilde{\boldsymbol{\theta}}^*, \widetilde{\boldsymbol{\theta}}') > \beta(t,\delta)\}$. Again for brevity we drop the Y^t in $\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}(Y^t)$ in the following proof. Then we can show that at time $t \geq \tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}$ we have

$$
\begin{aligned}
T_{Y^t}(\widetilde{\boldsymbol{\theta}}^*, \widetilde{\boldsymbol{\theta}}') > \beta(t,\delta) \implies & \exp\left(T_{Y^t}(\widetilde{\boldsymbol{\theta}}^*, \widetilde{\boldsymbol{\theta}}')\right) > \exp\left(\beta(t,\delta)\right) \\
\implies & \left(\frac{\prod_{s=1}^{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}} \mathbb{P}(Y_{I_s} = y_s | I_s, \widetilde{\boldsymbol{\theta}}')}{\prod_{s=1}^{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}} \mathbb{P}(Y_{I_s} = y_s | I_s, \widetilde{\boldsymbol{\theta}}^*)}\right) > \exp\left(\beta(t,\delta)\right) \\
\implies & \exp\left(-\beta(t,\delta)\right) \left(\frac{\prod_{s=1}^{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}} \mathbb{P}(Y_{I_s} = y_s | I_s, \widetilde{\boldsymbol{\theta}}')}{\prod_{s=1}^{\tau_{\widetilde{\boldsymbol{\theta}}'\widetilde{\boldsymbol{\theta}}^*}} \mathbb{P}(Y_{I_s} = y_s | I_s, \widetilde{\boldsymbol{\theta}}^*)}\right) > 1. \qquad (7.13)
\end{aligned}
$$

Following this we can show that the probability of the event $\{T_{Y^t}(\widetilde{\boldsymbol{\theta}}^*, \widetilde{\boldsymbol{\theta}}') > \beta(t,\delta)\}$ is upper bounded by

$$
\begin{aligned}
\mathbb{P}(\exists t < \infty, T_{Y^t}(\widetilde{\boldsymbol{\theta}}^*, \widetilde{\boldsymbol{\theta}}') > \beta(t,\delta)) &= \sum_{s=1}^{\infty} \mathbb{P}\left(\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'} = s\right) = \sum_{s=1}^{\infty} \mathbb{E}[\mathbb{I}\{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'} = s\}] \\
&\overset{(a)}{\leq} \sum_{s=1}^{\infty} \mathbb{E}\left[\mathbb{I}\{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'} = s\} \exp\left(-\beta(t,\delta)\right) \left(\frac{\prod_{s=1}^{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}} \mathbb{P}(Y_{I_s} = y_s | I_s, \widetilde{\boldsymbol{\theta}}')}{\prod_{s=1}^{\tau_{\widetilde{\boldsymbol{\theta}}'\widetilde{\boldsymbol{\theta}}^*}} \mathbb{P}(Y_{I_s} = y_s | I_s, \widetilde{\boldsymbol{\theta}}^*)}\right)\right] \\
&= \exp\left(-\beta(t,\delta)\right) \sum_{s=1}^{\infty} \int_{\mathbb{R}^{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}}} \mathbb{I}\{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'} = s\} \frac{\prod_{s=1}^{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}} \mathbb{P}(Y_{I_s} = y_s | I_s, \widetilde{\boldsymbol{\theta}}')}{\prod_{s=1}^{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}} \mathbb{P}(Y_{I_s} = y_s | I_s, \widetilde{\boldsymbol{\theta}}^*)} \prod_{s=1}^{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}} \mathbb{P}(Y_{I_s} = y_s | I_s, \widetilde{\boldsymbol{\theta}}^*) dx_1 dx_2 \ldots dx_{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}} \\
&= \exp\left(-\beta(t,\delta)\right) \sum_{s=1}^{\infty} \int_{\mathbb{R}^{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}}} \mathbb{I}\{\tau_{\widetilde{\boldsymbol{\theta}}'\widetilde{\boldsymbol{\theta}}^*} = s\} \prod_{s=1}^{\tau_{\widetilde{\boldsymbol{\theta}}'\widetilde{\boldsymbol{\theta}}^*}} \mathbb{P}(Y_{I_s} = y_s | I_s, \widetilde{\boldsymbol{\theta}}') dx_1 dx_2 \ldots dx_{\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'}} \\
&= \exp\left(-\beta(t,\delta)\right) \sum_{s=1}^{\infty} \mathbb{P}(\tau_{\widetilde{\boldsymbol{\theta}}^*\widetilde{\boldsymbol{\theta}}'} = s | I_s, \widetilde{\boldsymbol{\theta}}') \\
&\leq \exp\left(-\beta(t,\delta)\right)
\end{aligned}
$$

where, (a) follows from (7.13). The claim of the lemma follows. □