Learning Theory Project

Da Wei Zheng*
Department of Computer Science
University of British Columbia

Yihan Zhou*
Department of Computer Science
University of British Columbia

Kevin Dsouza*
Department of Electrical and Computer Engineering
University of British Columbia

Abstract

In this project we look to study the setting under which a learner is trying to query from a database in the presence of an adversary who is capable of eavesdropping on the queries but not the responses. We focus on two main extensions from previous work on private sequential learning. To start with, we investigate query complexity in a general Bayesian setting and in the later part provide bounds for querying in higher dimensions.

1 Introduction

The query complexity ($N^*(\epsilon, \delta, L)$) of the learner in the presence of an adversary is investigated in [1]. In this work, it is assumed that the adversary gets to observe all the learners queries but is oblivious to the responses that the learner gets for each query. [1] also introduces the same problem under the bayesian setting and provides bounds on the query complexity when operating under a uniform prior over the queried values. We provide tighter bounds for this case thus managing to improve both the upper and lower bound on the query complexity. For the upper bound, we propose a new algorithm that is better than replicated bisection search under certain assumptions of the value of ϵ, δ. In addition we improve the lower bound with an additive factor. We also investigate the private query model in higher dimension, which is mentioned as one of the open problems in [1]. We successfully extend the opportunistic bisection strategy to higher dimension and derive an upper bound based on the strategy. We use the same method as in [1] to prove the lower bound with a strengthened accuracy constraint and derive a less tighter lower bound with the original accuracy constraint.

1.1 Definitions of Private Sequential Learning

Here we present the original set cover definitions used in Tsitsiklis et al's orignal paper and a modification of the Bayesian version included in their appendix [1]. Let Φ be the set of learner strategies that make decisions sequentially to find some target $x^* \in [0, 1]$.

Definition 1.1. ***Set cover definition.***

Fix $\epsilon > 0, \delta > 0, L \in \mathbb{N}$. A learner strategy $\phi \in \Phi$ is (ϵ, δ, L)-private if it satisfies both of the following for finding any $x^* \in [0, 1]$:

1. *Accuracy constraint: The strategy produces a guess x that is close to x^* with probability one:*

$$\mathbb{P}\left(|x - x^*| \leq \frac{\epsilon}{2}\right) = 1 \tag{1}$$

2. *Privacy Constraint: for every adversary $\psi \in \Psi$, we have*

$$C_\delta(\mathcal{I}) \geq L \tag{2}$$

Where $\mathcal{I}$ is the set of all possible regions that x could lie in from the perspective of the adversary, and $C_\delta(A)$ is the minimum number of closed intervals of length at most δ to cover the set A.

One drawback of this definition is that $\mathcal{I}$ can contain large regions that may have very low probabilities for x^* to lie in. This leads us to the Bayesian sequential learning model which is given by the following definition.

Definition 1.2. ***Bayesian interval definition.***

Fix $\epsilon > 0, \delta > 0, L \in \mathbb{N}$ and some prior distribution $\mathcal{D}$ over $\mathbb{R}$ known to both the learner and the adversary. A learner strategy $\phi \in \Phi$ is $(\epsilon, \delta, \gamma)$-private if it satisfies both of the following for finding any $x^* \sim \mathcal{D}$:

1. *Accuracy constraint: The strategy produces a guess x that is close to x^* with probability one:*

$$\mathbb{P}\left(|x - x^*| \leq \frac{\epsilon}{2}\right) = 1 \tag{3}$$

2. *Privacy Constraint: for every adversary $\psi \in \Psi$ that outputs some x_A, we have that*

$$\mathbb{P}\left(|x_A - x^*| \leq \frac{\delta}{2}\right) \leq \gamma \tag{4}$$

The definition 1.2.2 is very similar to definition 1.1.2 with $\mathcal{D}$ uniform on $[0, 1]$ and $\gamma = \frac{1}{L}$. Although definition 1.2 fixes the drawback that comes with low probability regions of $\mathcal{I}$, it works poorly on highly concentrated distributions, since there could be a region in the prior with a window of size δ where the probability is already greater than γ, so that the privacy constraint is already breached. It should also be noted that enquiring about query complexity is ill defined on wide distributions over all of $\mathbb{R}$, since it may take arbitrary number of queries for a learner to satisfy the accuracy constraint.

2 Improvements on bounds

We start off with a setting of querying over a uniform distribution on $[0, 1]$. For the upper bound, we proposed a new algorithm that is better than replicated bisection search under certain assumptions of the value of ϵ, δ. In addition we improve the lower bound up-to an additive factor.

2.1 Improvement of upper bound

The upper bound on N^* in [1] was given by replicated bisection search. We acknowledge that there exists a trade of between the efficiency of the bisection search and the and privacy-preserving capability of the ϵ-dense algorithm. Thus, we combine these two algorithms together to satisfy both the accuracy and privacy constraint under our setting. The algorithm involves the following 2 steps:

1. Run bisection search on $[0, 1]$ to find an interval with length δL and that contains x^*.

2. Let's denote the interval containing x^* as I. Then ϵ-dense is run on I.

The accuracy constraint is automatically satisfied by this algorithm because of the nature of the ϵ-dense algorithm. Since ϵ-dense gives no additional information to the adversary, all the information the adversary gets is that $x^* \in I$. Thus, without any concrete knowledge, the adversary can only guess the position of x^* in I. The probability of choosing any random interval of length δ is equal to δ, and because the length of I is δL the adversary can at best get the guess right with probability $\frac{1}{L}$. Thus the privacy constraint is also satisfied.

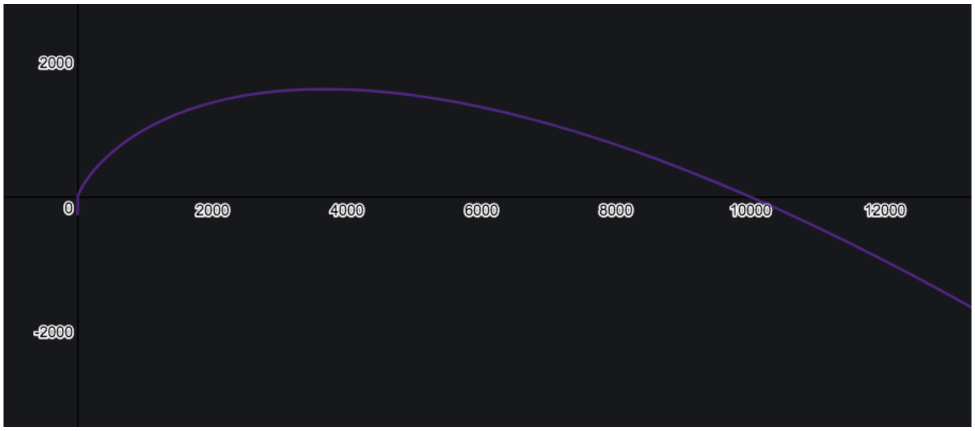

Figure 1: Interval of L for which bound is tighter when k = 2 and $\epsilon = 10^{-5}$. Plot of equation (7).

The query complexity of this algorithm is

$$N^*(\epsilon, \delta, L) = \log(\frac{1}{\delta L}) + \frac{\delta L}{\epsilon} \tag{5}$$

where the first part comes from bisection search and the second part comes from ϵ-dense algorithm run on the interval of length δL.

We now compare the upper bound of the replicated bisection with the one detailed above. We take $\epsilon = 10^{-5}$ and vary δ as a multiple of ϵ. Let $k = \delta/\epsilon$.

In order for our bound to outperform the bisection bound, we need to have,

$$L \log \frac{1}{L\epsilon} + L - 1 \geq \log \frac{1}{Lk\epsilon} + kL \tag{6}$$

where k is a real number and $k \geq 1$. Thus we have,

$$\log L\epsilon(1 - L) + L(1 - k) + \log k - 1 \geq 0 \tag{7}$$

We note that for a given value of k, we have an interval of L satisfying the above inequality. As k grows from 1, the interval keeps shrinking until when $k = 4.5678$, after which, no such interval can be obtained. Therefore, our bound is tighter when δ is close to ϵ. The interval of values of L for which the bound is tighter shrinks as δ grows and is no longer tighter for any value L, when $k > 4.5678$.

This is only a slight improvement that holds when k is small.

The interval of L for which the bound is tighter when $\delta = 2\epsilon$ is shown in Fig. 1.

2.2 Improvement of the lower bound

As mentioned in [1], improvement of the lower bound is non-trivial. In this section we improve it by additive factors and in our proof, we do not use the fact that the adversary knows the learner's algorithm. Instead, we assume that both the learner and the adversary only have the queries and try to estimate the position of x^* (essentially the adversary won't be able to separate out the special interval that resulted out of the bisection search from the other special intervals).

Note that by the accuracy constraint, x^* must lie in a interval with length less than ϵ. We adopt the terminology in [1] and call these intervals special intervals. Following [1], we use $\bar{I}(q)$ to denote the union of all special intervals under queries q. Note that to satisfy the privacy constraint, the length of the information set $I(q)$ has to be greater than δL. Otherwise, the adversary can just randomly pick a point in the information set and break the privacy constraint. Since $I(q) \subseteq \bar{I}(q)$, the length of $\bar{I}(q)$ is also greater than δL. By the optimality of bisection search, at least $\log(\frac{1}{\epsilon})$ queries need to be submitted by the learner to satisfy the accuracy constraint.

After the bisection search, there are only two special intervals and the length of the union of these two intervals is 2ϵ. Thus, there should be at least $(\delta L - 2\epsilon)/\epsilon$ more special intervals to make the length of the union of special unions greater than δL. Given that at least one query is required to make a new interval, there should be at least $\frac{\delta L}{\epsilon} - 2$ more queries (this would mean that one special interval immediately follows the next until the last). Taking this into account the lower bound improves to

$$N^*(\epsilon, \delta, L) = \log(\frac{1}{\epsilon}) + \frac{\delta L}{\epsilon} - 2 \tag{8}$$

3 Query Complexity in the Bayesian setting

In the previous section we discussed Bayesian setting under uniform distribution. In this section we consider Bayesian setting under some concentrated distributions. Note that under this setting, the previous definition for the private learner strategy is not meaningful. A simple example is to let the support of the distribution $\mathcal{D}$ be $[0, \delta]$, i.e.

$$\mathop{\mathbb{P}}_{x^* \sim \mathcal{D}}[x^* \in [0, \delta)] > 0$$
$$\mathop{\mathbb{P}}_{x^* \sim \mathcal{D}}[x^* \in [\delta, 1]] = 0.$$

Then clearly the privacy constraint cannot be satisfied because the adversary knows the true point lies in an interval of size δ. Even if we put some constraints on the support of the distribution, this problem still exists. Consider the Gaussian distribution $\mathcal{N}\big(0, (\frac{\delta}{4})^2\big)$. The support of Gaussian distribution is $\mathbb{R}$ and here we exclude the assumption that the true values lies in $[0, 1]$. If the adversary just returns the estimation $\hat{x}^a = 0$, $\mathbb{P}[|x^* - \hat{x}^a| \leq \delta] \approx 0.95$ and therefore the privacy constraint fails for any smaller γ. It can be observed that the adversary is more powerful if the prior distribution of the true value is concentrated. The privacy constraint cannot be satisfied under the definition of [1] if the distribution is concentrated enough and this motivates us to come with an alternate definition for bayesian learning in a continuous setting.

3.1 Private Bayesian Learning in a Continuous Setting

We now present an alternative definition of learning in a general continuous setting, where we allow x^* to follow a distribution $\mathcal{D}$ on $\mathbb{R}$, that is known to both the learner and the adversary.

Definition 3.1. ***Bayesian Sequential Learning***

Let ϵ, δ, γ be non-negative real numbers. We define a learner strategy ϕ to be $(\epsilon, \delta, \gamma)$-private if the following is satisfied.

1. *Accuracy constraint: The learner picks some interval $[a, b]$, where $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$ where the initial probability is less than ϵ. Formally this is $\mathop{\mathbb{P}}_{\mathcal{D}}(x^* \in (a, b)) \leq \epsilon$. This interval must satisfy:*
$$\mathop{\mathbb{P}}_{\mathcal{D}}(x^* \in [a, b] \mid q_1, r_1, q_2, r_2, \dots) = 1 \tag{9}$$
2. *Privacy constraint: for every $x \in \mathbb{R}$, and every possible sequence of queries q, we consider any adversary's guess of an interval (c, d) with $c, d \in \mathbb{R} \cup \{-\infty, \infty\}$ that has initial probability less than δ. Formally this is $\mathop{\mathbb{P}}_{\mathcal{D}}(x^* \in [c, d)] \leq \delta$. For ϕ to be private we must have the following hold.*
$$\mathop{\mathbb{P}}_{\mathcal{D}}(x^* \in [c, d] \mid q_1, q_2, \dots) \leq \gamma \tag{10}$$

In this definition the learner and adversary are trying to narrow down the region x^* to a small region, where small is in probability density. The goal of the learner now is not to guess a good approximation to x^* but instead to find a region of initially low probability ϵ that x^* is sure to lie in. The goal of the adversary is similar, but we allow the adversary a potentially bigger interval that has probability mass δ. The learner is private if no adversary can do better than picking a window that contains probability mass less than γ.

Note the similarity and differences between this definition and the Bayesian interval definition 1.2. They are the same when applied to the uniform distribution on $[0, 1]$. However this definition is more flexible and can be applied to both concentrated distributions and distributions that span all of $\mathbb{R}$.

3.1.1 Universality of Uniform Bayesian Learning on the Unit Interval

In this section we prove that Bayesian sequential learning in the general continuous Bayesian setting with definition 3.1 is equivalent to the learning on the uniform distribution on $[0, 1]$.

Theorem 1. *Any $\phi^{\mathcal{P}}$ that is $(\epsilon, \delta, \gamma)$-private Bayesian learner with a uniform prior over $[0, 1]$ is solvable with a ϕ that is also an $(\epsilon, \delta, \gamma)$-private Bayesian learner with a continuous prior $\mathcal{P}$ over $\mathbb{R}$.*

Consider any learning algorithm ϕ that achieves $(\epsilon, \delta, \gamma)$-privacy with a prior of the uniform distribution on $[0, 1]$ in accordance with the Bayesian sequential learning in definition 3.1 by giving queries $q_1, q_2, \ldots, q_n$. Consider the cumulative distribution of $\mathcal{P}$, $F : \mathbb{R} \to [0, 1]$ and its inverse distribution function F^{-1} which must exists by the continuity of $\mathcal{P}$ and hence F. Note that F^{-1} need not be unique due to flat regions in the cumulative distribution function, but that doesn't matter much as they correspond to regions with probability zero.

Now let $\phi^{\mathcal{P}}$ be a learner that runs ϕ. Every time ϕ would query q_i, $\phi^{\mathcal{P}}$ would instead query $q_i^{\mathcal{P}} = F^{-1}(q_i)$ and hand the same response $r_i^{\mathcal{P}}$ to ϕ. At the end, $\phi^{\mathcal{P}}$ would find a window between some queries q_i and q_j such that $q_j - q_i \leq \epsilon$ where

$$\mathcal{P}(x^* \in (F^{-1}(q_i), F^{-1}(q_j))|\mathcal{F}) = 1$$
$$\mathcal{P}(x^* \in (F^{-1}(q_i), F^{-1}(q_j))) = q_j - q_i \leq \epsilon$$

Consider if there existed adversary $\psi^{\mathcal{P}}$ with some interval $[c, d] \subset R$ where

$$\mathcal{P}(x^* \in [c, d]) \leq \delta$$
$$\mathcal{P}(x^* \in [c, d] \mid \bar{q}) \geq \gamma.$$

However, this would mean that the following would hold.

$$\mathcal{P}(\tilde{x}^* \in [F(c), F(d)]) \leq \delta$$
$$\mathcal{P}(\tilde{x}^* \in [F(c), F(d)] \mid \bar{q}) \geq \gamma.$$

Since $\phi^{\mathcal{P}}$ simply translates the queries of ϕ, with F^{-1}, an adversary ψ can simulate $\phi^{\mathcal{P}}$ then translate the coordinates back. This would contradict how ϕ is an $(\epsilon, \delta, \gamma)$-private learner.

Thus no such adversary exists which means that $\phi^{\mathcal{P}}$ is an $(\epsilon, \delta, \gamma)$-private learner on $\mathcal{P}$, hence proving Theorem 1.

Theorem 2. *Any algorithm $\phi^{\mathcal{P}}$ that is $(\epsilon, \delta, \gamma)$-private Bayesian learner with a prior $\mathcal{P}$ over $\mathbb{R}$ is also a ϕ that is $(\epsilon, \delta, \gamma)$-private learner on the uniform learning problem on $[0, 1]$.*

Theorem 2 is the converse of theorem 1, and can be proved in the same way using the cumulative distribution function.

This shows how any bound of the query complexity on the uniform distribution on $[0, 1]$ is a bound on the query complexity in the general Bayesian setting.

4 Query model in higher dimensions

In this section, first we give a formal definition of the model higher dimensions. Then we extend the opportunistic bisection strategy to higher dimensions which gives an upper bound on the number of queries. For the lower bound, we derive a tight lower bound with a strengthened accuracy constraint and a less tighter bound with the original definition of accuracy constraint. We prove the following theorem.

Theorem 3. ***Query Complexity of Private Sequential Learning in higher dimensions.***

Fix $\epsilon > 0$, $\delta > 0$, and a positive integer $L \geq 2$, such that $2\epsilon < \delta < 1/L$. Then,

$$d \log(\delta/\epsilon) + 2d\sqrt[d]{L-1} - 2d \leq N * (\epsilon, \delta, L) \leq d \log \frac{\sqrt{d}}{\sqrt[d]{L}\epsilon} + 2d\sqrt[d]{L}$$

with a strengthened accuracy constraint, and

$$d \log(\delta/\epsilon) + d\sqrt[d]{L-1} \leq N * (\epsilon, \delta, L) \leq d \log \frac{\sqrt{d}}{\sqrt[d]{L}\epsilon} + 2d\sqrt[d]{L}$$

with the original accuracy constraint and without considering some exceptional cases.

4.1 Extension of the private query model

In the higher dimensional query model, the true value $x^* \in [0,1)^d$. At each step k, the learner submits a query $q_k \in \mathbb{R}^d, a_k \in \mathbb{R}$, and receives a response from the database $\mathbb{I}(q_k^T \cdot x^* + a_k \geq 0)$, indicating whether the true value point is above (to the right of) or below (to the left of) the hyperplane $q_k^T \cdot x + a_k = 0$. Note that to specify a hyperplane in $\mathbb{R}^d$, only d parameters are required. Thus, one of the parameter of the query is redundant. We use $d+1$ parameters here for simplicity and interpretability of the notation. One can observe that in the single dimension query model, the query parameter is bounded in $[0,1)$, however, in the higher dimension query model, the query parameter is not bounded. This difference could make the learning task harder for the learner. We keep the definition of learning strategy, except the domain of queries becomes $\mathbb{R}^d \times \mathbb{R}$. Denote by $\mathcal{Q}(x)$ the set of query sequences that have a positive probability of appearing under ϕ, when the true x^* is equal to x:

$$\mathcal{Q}(x) = \{\bar{q} \in (\mathbb{R}^d \times \mathbb{R})^N : \mathbb{P}_\phi(Q = \bar{q}) > 0\}, \bar{q} \in (\mathbb{R}^d \times \mathbb{R})^N$$

Consequently, the definition of information set for the adversary, $\mathcal{I}(\bar{q})$, is

$$\mathcal{I}(\bar{q}) = \{x \in [0,1)^d : \bar{q} \in \mathcal{Q}(x)\}, \;\; \bar{q} \in (\mathbb{R}^d \times \mathbb{R})^N$$

We can extend the metric of accuracy to higher dimensions easily by using l_2-norm to measure the distance between the true value point and the estimation. The metric of privacy in higher dimension needs to be considered more carefully. In [1], the authors use set coverability to measure the size of the information set. In the 1d case, we can use intervals of length δ to cover $[0,1)$. And in each such interval, if the true value lies in that interval and the adversary picks its estimation to be the middle of the interval, $|\hat{x}^a - x^*| \leq \frac{\delta}{2}$. Similarly, we can first extend the notion of these intervals to higher dimensions:

Definition 4.1. *Let's define an a-cube to be a hypercube in $\mathbb{R}^d$ such that the length of every side of this hypercube is a.*

Then we can use $\frac{\delta}{\sqrt{d}}$-cubes as the covering "intervals" in $\mathbb{R}^d$. We choose $\frac{\delta}{\sqrt{d}}$ as the side length because if the adversary picks its estimation to be at the center of the hypercube, $\|x^* - \hat{x}^a\|_2 \leq \sqrt{d \cdot (\frac{\delta}{\sqrt{d}2})^2} = \frac{\delta}{2}$. It can also be viewed as using l_∞-norm to measure distance instead of using l_2-norm. However, a natural question would arise in this definition: can the shape of $\frac{\delta}{\sqrt{d}}$-cubes capture the effective size of an information set? Since the query hyperplane may not be axis aligned, it is possible to have a set of small volume but needs many $\frac{\delta}{\sqrt{d}}$-cubes to cover as with a very thin but long hyperrectangle. In addition, we use ℓ_2-norm to measure the distance so it seems to be more natural to use balls instead of hypercubes in the definition. But balls do not tile the space and they leave gaps between them. Let's put aside this debate for now and continue to address the definition of set coverability in higher dimensions.

Definition 4.2. *Fix $\delta > 0$, $L \in \mathbb{N}$, and a set $\mathcal{E} \subseteq \mathbb{R}$. We say that a collection of L $\frac{\epsilon}{\sqrt{d}}$-cubes $a_1, \cdots, a_L$, is a (δ, L) cover for $\mathcal{E}$ if $\mathcal{E} \subseteq \bigcup_{1 \leq j \leq L} a_j$.*
We say that a set $\mathcal{E}$ is (δ, L)-coverable if it admits a (δ, L)-cover. In addition, we define the δ-cover number of a set $\mathcal{E}$, $C_\delta(\mathcal{E})$, as

$$C_\delta(\mathcal{E}) \triangleq \min\{L \in \mathbb{N} : \mathcal{E} \textit{ is } (\epsilon, L)\textit{-coverable}\}$$

The definition of $(\epsilon, \delta, L)-$private learner strategy follows:

Definition 4.3. *Fix $\epsilon > 0$, $\delta > 0$, $L \geq 2$, with $L \in \mathbb{N}$. A learner strategy $\phi \in \Phi_N$ is (ϵ, δ, L)-private if it satisfies the following:*
1. Accuracy constraint: the learner estimate accurately recovers the true value, with probability one:

$$\mathbb{P}[\|\hat{x}(x^*, Y) - x^*\|_2 \leq \epsilon/2] = 1, \;\; \forall x^* \in [0,1)^d,$$

where the probability is measured with respect to the randomness in Y.
2. Privacy constraint: for every $x \in [0,1)$ and every possible sequence of queries $\bar{q} \in \mathcal{Q}(x^*)$, the δ-cover number of the information set for the adversary, $C_\delta\big(\mathcal{I}(\bar{q})\big)$, is at least L, i.e.,

$$C_\delta\big(\mathcal{I}(\bar{q})\big) \geq L, \;\; \forall \bar{q} \in \mathcal{Q}(x^*)$$

4.2 Opportunistic bisection strategy in higher dimension

After settling the formal definition of the private query model in higher dimensions, we extend the opportunistic bisection strategy to higher dimensions.

Coordinate opportunistic bisection strategy: We first try to apply opportunistic bisection strategy to each dimension with the length of the guess equal to $\frac{\epsilon}{\sqrt{d}}$. In phase 1 of the opportunistic bisection strategy of each dimension, the corresponding coordinate value of the true value point could be in L disjoint intervals with size $\frac{\epsilon}{\sqrt{d}}$. We modify Claim 6.3 in [1] to the following:

Claim 4.1. *Let $\mathcal{G}$ denote the union of L disjoint intervals made in phase 1 of opportunistic bisection search in a coordinate. Then at least L intervals of length $\frac{\delta}{\sqrt{d}}$ are required to cover $\mathcal{G}$.*

Proof. Let J be the any interval in $[0, 1)$ with length $\frac{\delta}{\sqrt{d}}$. By construction, the gap between two consecutive disjoint intervals of length $\frac{\epsilon}{\sqrt{d}}$ is $1/L - \frac{\epsilon}{\sqrt{d}}$. Note that we assume that $\delta \leq 1/L$, so $\frac{\delta}{\sqrt{d}} \leq \delta \leq 1/L$. Thus, the Lebesgue measure of $J \cap \mathcal{G}$ is at most $\frac{\epsilon}{\sqrt{d}}$. The Lebesgue measure of $\mathcal{G}$ is $L\frac{\epsilon}{\sqrt{d}}$, so at least L intervals with length $\frac{\delta}{\sqrt{d}}$ are required to cover $\mathcal{G}$. □

After running through all dimensions, there are L^d hypercubes with side length $\frac{\epsilon}{\sqrt{d}}$ and by the above claim we need L^d $\frac{\delta}{\sqrt{d}}$-cubes to cover these. To the adversary, each one of these L^d hypercubes is possible to contain x^*, so these hypercubes are in $\mathcal{I}(\bar{q})$. Thus, $C_\delta\big(\mathcal{I}(\bar{q})\big) \geq L^d$. The learner knows the true point lies in a hypercube with length $\frac{\epsilon}{\sqrt{d}}$. If the learner picks its estimation to be at the center of that hypercube, the accuracy constraint is satisfied. Note that a bisection search on one coordinate takes $\log \frac{\sqrt{d}}{L\epsilon} + 2L$ queries so the algorithm needs $d \log \frac{\sqrt{d}}{L\epsilon} + 2dL$ queries. Although this proves the correctness of the algorithm, it should be noted that it achieves excessive level of privacy. The algorithm that we propose next looks to trade some of this privacy for efficiency and results in a tighter upper bound.

Fast Coordinate opportunistic bisection strategy: This algorithm still apply opportunistic bisection strategy to each dimension with the length of guess equal to $\frac{\epsilon}{\sqrt{d}}$, but reduce the number of guesses on each dimension to $\sqrt[d]{L}$. We modify claim 3.1 to the following:

Claim 4.2. *Let $\mathcal{G}$ denote the union of $\sqrt[d]{L}$ disjoint intervals made in phase 1 of opportunistic bisection search in a coordinate. Then at least $\sqrt[d]{L}$ intervals of length $\frac{\delta}{\sqrt{d}}$ are required to cover $\mathcal{G}$.*

We can use the same method in the proof of Claim 3.1 to prove Claim 3.2 as lesser number of guesses leave behind larger gaps between these guesses. Therefore, after running on all coordinates, there will be $(\sqrt[d]{L})^d = L$ $\frac{\epsilon}{\sqrt{d}}$-cubes and the adversary cannot tell which one contains the true value point. The proof follows as before from then on-wards. This approach is much more efficient and reduces the running time to $d \log \frac{\sqrt{d}}{\sqrt[d]{L}\epsilon} + 2d\sqrt[d]{L}$.

4.3 Proof of lower bound with strengthened accuracy constraint

We now use a similar method as in [1] to prove the lower bound of any (ϵ, δ, L)-private learner strategy. But before that, we need to modify the accuracy constraint of the (ϵ, δ, L)-private learner strategy to the following:

Definition 4.4. *Strengthened accuracy constraint: the learner estimates accurately a hypercube $A \in [0, 1)^d$ such that A can be contained in some $\frac{\epsilon}{\sqrt{d}}$-cube and:*

$$\mathbb{P}[x^* \in A] = 1, \ \ \forall x^* \in [0, 1)^d$$

As we mentioned before, if we use l_2-norm as the measure of accuracy, the estimation of the learner should be inside the $\frac{\epsilon}{2}$-ball centered at the true value point. Since all the query planes are axis aligned, the learner will finally pick its estimation as the center of a hypercube, which is contained in the $\frac{\epsilon}{2}$-ball centered at the true value point. But the shape of this hypercube is indefinite, thus we strengthen the accuracy constraint for simplicity of proof.

We call a hypercube that can be contained in a $\frac{\epsilon}{\sqrt{d}}$-cube as a special cube. We denote the union of all special cubes as $\bar{\mathcal{I}}(\bar{q})$. From the learner's perspective, every hypercube containing the true value point x^* is required to be contained in a special cube. Otherwise, the accuracy constraint cannot be satisfied by the learner. Consequently, $\mathcal{I}(\bar{q}) \subseteq \bar{\mathcal{I}}(\bar{q})$. We then modify Lemma 7.1 in [1] and get:

Lemma 1. *Fix a learner strategy ϕ that satisfies the strengthened accuracy constraint. For every $y \in \{1, 2, \cdots, \mathcal{Y}\}$, there exists $x^* \in [0, \frac{\delta}{\sqrt{d}}]^d$ such that there are at least $d \log(\delta/\epsilon)$ of the queries in $\bar{q}(x^*, y)$ that belong to $[0, \frac{\delta}{\sqrt{d}}]^d$.*

In [2], Waeber et al. proved the bisection strategy is optimal for the unit interval in one coordinate. By the strengthened accuracy constraint, in each coordinate, we need to locate the corresponding coordinate value of x^* in an interval of length $\frac{\epsilon}{\sqrt{d}}$. By optimality of bisection strategy, the best strategy is to run bisection strategy on each coordinate and it requires $d \log(\frac{\delta}{\sqrt{d}} / \frac{\epsilon}{\sqrt{d}}) = d \log(\delta/\epsilon)$ queries. The intuitive explanation is that the $\frac{\delta}{\sqrt{d}}$-cube contains many $\frac{\epsilon}{\sqrt{d}}$-cubes and an accurate learner strategy needs to distinguish the $\frac{\epsilon}{\sqrt{d}}$-cube containing the true value point.

Now consider the number of queries required in the region $D = [0, 1)^d \setminus [0, \frac{\delta}{\sqrt{d}}]^d$. Among them, we restrict our attention to endpoints of these special cubes. We call these queries special queries. We decompose the set $\bar{\mathcal{I}}(\bar{q}) \cap D$ as the union of infinite many connected components. Now consider the projection of the components on dimension i. The projection is a union of connected intervals, we use l_i to denote the Lebesgue measure of these intervals on dimension i. Let K_i denote the number of special queries on dimension i. Since the gap between two special queries in one component cannot be greater than $\frac{\epsilon}{\sqrt{d}}$, we have $l_i \leq (K_i - 1)\frac{\epsilon}{\sqrt{d}}$. So the number of $\frac{\delta}{\sqrt{d}}$-interval required to cover the projection on this dimension is $\leq \frac{l_i}{\frac{\delta}{\sqrt{d}}} \leq (K_i - 1)\frac{\epsilon}{\delta} \leq \frac{K_i}{2}$ by the assumption $\delta \geq 2\epsilon$. If we consider the exceptional case mentioned in [1], it is safer to make the upper bound be $\frac{K_i+1}{2}$. Then the number of $\frac{\delta}{\sqrt{d}}$-cubes needed to cover $\bar{\mathcal{I}}(\bar{q})$ is $\leq \prod_{i=1}^{d} \frac{K_i+1}{2} + 1$. The constant one is an additional factor due to the one $\frac{\delta}{\sqrt{d}}$ cube needed to cover $[0, \frac{\delta}{\sqrt{d}}]^d$. Since $C_\delta(\mathcal{I}(\bar{q})) = L$ and $\mathcal{I}(\bar{q}) \subseteq \bar{\mathcal{I}}(\bar{q})$, $L \leq \prod_{i=1}^{d} \frac{K_i+1}{2} + 1$.

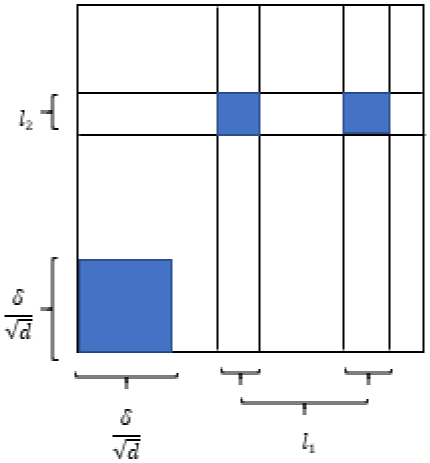

Figure 2: Illustration of the proof idea in two dimensions.

Considering the inequality mentioned above, we observe the following:

$$\prod_{i=1}^{d} \frac{K_i+1}{2} + 1 \geq L$$

$$\prod_{i=1}^{d} \frac{K_i+1}{2} \geq L-1$$

$$\sum_{i=1}^{d} \frac{K_i+1}{2d} \geq \sqrt[d]{\prod_{i=1}^{d} \frac{K_i+1}{2}} \geq \sqrt[d]{L-1} \quad \text{By inequality of arithmetic and geometric means}$$

$$\sum_{i=1}^{d} K_i \geq 2d\sqrt[d]{L-1} - d$$

As mentioned in [1], there is an implicit query at 1 for each dimension, so after removing these queries, we have the total number of queries to be $d\log(\delta/\epsilon) + 2d\sqrt[d]{L-1} - 2d$, which is the lower bound.

4.4 Proof of lower bound

We now investigate whether the strengthened accuracy constraint is necessary for proving the lower bound. First, we try to prove the lower bound with the original definition of accuracy constraint. We observe that Lemma 1 is not trivial if we use the l_2-norm as the measure of accuracy. Instead of considering that the adversary picks an estimation point, we say that the adversary is picking an hypercube whose center is the estimation point. In the strengthened definition of accuracy constraint, the adversary will always pick a $\frac{\epsilon}{\sqrt{d}}$-cube, while in the original definition, the adversary can pick any hypercube that can be contained in the $\frac{\epsilon}{2}$-ball. We want to show that picking the $\frac{\epsilon}{\sqrt{d}}$-cube will still be optimal for the adversary. We formalize this idea with the following lemma, which may be of separate interest to someone:

Lemma 2. *Let $x^* \in [0, \frac{\delta}{\sqrt{d}}]^d$. The best strategy for the adversary to find a hypercube A such that $\|a^* - x^*\| \leq \frac{\epsilon}{2}$ where a^* is the center of the hypercube is to decompose $[0, \frac{\delta}{\sqrt{d}}]^d$ into $\frac{\epsilon}{\sqrt{d}}$-cubes and use bisection search on each dimension to locate the $\frac{\epsilon}{\sqrt{d}}$-cube containing x^*.*

Proof. Let's say the adversary returns a hypercube A. Let $[a_i, b_i]$ denote the projection onto the axis i and let $l_i = b_i - a_i$. Again by [2], the optimal strategy to find the interval $[a_i, b_i]$ is to do a bisection search on coordinate i. A bisection search on a single coordinate will take $\log(\frac{\delta}{\sqrt{d}l_i})$ queries. Then the total number of queries is $\sum_{i=1}^{d} \log(\frac{\delta}{\sqrt{d}l_i})$. Since the hypercube is contained in a $\frac{\epsilon}{2}$-ball, we have $\frac{1}{2}\sqrt{\sum_{i=1}^{d} l_i^2} \leq \frac{\epsilon}{2}$. Then by inequality of arithmetic and geometric means, $\epsilon^2 \geq \sum_{i=1}^{d} l_i^2 \geq d\sqrt[d]{\prod_{i=1}^{d} l_i^2}$. Then the total number of queries $\sum_{i=1}^{d} \log(\frac{\delta}{\sqrt{d}l_i}) = \log(\prod_{i=1}^{d} \frac{\delta}{\sqrt{d}l_i}) = \log(\frac{\delta^d}{\sqrt{d}^d} \prod_{i=1}^{1} \frac{1}{l_i}) \geq d\log\frac{\delta}{\epsilon}$. To get a lower bound we need, $l_1 = \cdots = l_i = \frac{\epsilon}{\sqrt{d}}$. This proves that the optimal strategy is to decompose the higher dimensional interval into $\frac{\epsilon}{\sqrt{d}}$-cubes and therefore lemma 1 still holds in the original definition of accuracy constraint. □

Now we continue our proof with the original accuracy constraint. The special cubes are now hypercubes that can be contained in the $\frac{\epsilon}{2}$-ball. The assumption of gaps between two special queries now does not hold anymore because the hypercube can be of any shape. More precisely, we can say the gap is less than $\frac{\epsilon}{2}$ but this bound is loose when d is big. But since $C_\delta(\bar{\mathcal{I}}(\bar{q})) \geq L$, we need at least $(L-1)$ $\frac{\delta}{\sqrt{d}}$-cubes to cover $\bar{\mathcal{I}}(\bar{q}) \cap D$, which means that $\bar{\mathcal{I}}(\bar{q}) \cap D$ contains at least $L-1$ special cubes. Then $\prod_{i=1}^{d} l_i \geq (L-1) \cdot (\frac{\epsilon}{\sqrt{d}})^d$. K_i queries on a coordinate can divide the coordinate into $K_i - 1$ intervals. The max interval(measured by length) is the projection of the max special hypercube(measured by the distance from a corner to its center) by the nature of geometry, so we only need to consider the max interval. Note that the length of the max interval can be lower bounded by

$\frac{l_i}{K_i-1}$ and the minimum is achieved when we query uniformly. This means that the corner-to-center distance of any max hypercube is greater than $\frac{1}{2}\sqrt{\sum_{i=1}^{d}(\frac{l_i}{K_i-1})^2}$. Since the max hypercube can be contained in an $\frac{\epsilon}{2}$-ball, at least we need to satisfy:

$$
\begin{aligned}
\frac{1}{2}\sqrt{\sum_{i=1}^{d}(\frac{l_i}{K_i-1})^2} &\leq \frac{\epsilon}{2} \\
\epsilon^2 \geq \sum_{i=1}^{d}(\frac{l_i}{K_i-1})^2 &\geq d\sqrt[d]{\prod_{i=1}^{d}\frac{l_i^2}{(K_i-1)^2}} && \text{by AM-GM inequality} \\
\epsilon^2 &\geq d\sqrt[d]{(L-1)^2\cdot(\frac{\epsilon}{\sqrt{d}})^{2d}\prod_{i=1}^{d}\frac{1}{(K_i-1)^2}} \\
\epsilon^2 &\geq d\sqrt[d]{(L-1)^2}\frac{\epsilon^2}{d}\sqrt[d]{\prod_{i=1}^{d}\frac{1}{(K_i-1)^2}} \\
\sqrt[d]{\prod_{i=1}^{d}(K_i-1)^2} &\geq \sqrt[d]{(L-1)^2} \\
\sqrt[d]{\prod_{i=1}^{d}(K_i-1)} &\geq \sqrt[d]{L-1} \\
\sum_{i=1}^{d}(K_i-1) \geq d\sqrt[d]{\prod_{i=1}^{d}(K_i-1)} &\geq d\sqrt[d]{L-1} \\
\sum_{i=1}^{d}K_i &\geq d\sqrt[d]{L-1}+d
\end{aligned}
$$

After removing d implicit queries at 1 and combining the result from lemma 2, we get the lower bound $d\log(\delta/\epsilon)+d\sqrt[d]{L-1}$. Note that here we do not consider the exceptional case due to time constraints, but it should not have a conspicuous effect on the derived bound. It can be seen that, without considering the exceptional case, the lower bound can get rid of the addition of d but is not as tight as the bound with the strengthened definition of accuracy constraint.

5 Future work

Due to time constraints, many problems that we identified during this research project could not be investigated in detail. Here we list these problems for interested readers.

5.1 Accuracy and privacy constraint under other measures

In this paper we used l_2-norm to measure accuracy and the cube notion to measure privacy. As mentioned in the previous section, the cube measure actually corresponds l_∞-norm. In the above proof we change the accuracy measure and obtained a tighter bound, so it would be interesting to investigate different measures of accuracy and privacy, like l_1-norm distance.

5.2 Optimality of axis aligned queries

In this paper, all queries we made are axis aligned and we have derived a lower bound, which seems to be quite close to the upper bound. Thus, it is reasonable to speculate that there exists an optimal strategy that only uses axis aligned queries, i.e., non axis aligned queries are not useful.

5.3 Other query algorithms in higher dimension

There may exist some other query algorithms that achieve similar or even better performance in higher dimensions, especially the ones involving non axis-aligned queries. We believe that investigating the behaviour of these algorithms could be meaningful. An algorithm we think may be promising is to query using concentric circles and hyperplanes passing the origin. We could do a bisection along the radial direction and also rotate the hyperplane by some angle according to the response each time.

5.4 Discrete distributions

We considered the Bayesian setting with a continuous distribution, but it would be interesting to investigate the setting in which the point we are looking for, x^*, can only be found within some discrete set of n points $S = \{a_1, a_2, ..., a_n\}$, possibly with a discrete distribution over points in S. The accuracy constraint needs to be strengthened, since the absolute distance between the points in S could be arbitrarily small or large. Instead, it would be more meaningful to consider consecutive elements $a_i, ..., a_j$ for some $1 \leq i \leq j \leq n$.

References

[1] John N. Tsitsiklis, Kuang Xu, and Zhi Xu. Private sequential learning. In *COLT*, 2018.

[2] Rolf Waeber, Peter I. Frazier, and Shane Henderson. Bisection search with noisy responses. *SIAM Journal on Control and Optimization*, 51, 05 2013.