Supplementary Material to *Learner-Private Convex Optimization*

In this supplementary material, we complete the proofs of the main results Theorems 1,2 in the paper. An extension of the results to multi-dimensional separable functions is also included.

1 Proof of Main Results

1.1 Proof of Theorem 2 (Bayesian Setting)

1.1.1 Proof of the lower bound

To complete the lower bound proof of Theorem 2, it remains to prove Lemma 1, 2. We repeat here the statements of the two lemmas.

Lemma 1. *For all $z \geq 1/2$, J, y, i, $\rho^{(i)}$, $\rho_- < 1/2$, $\rho_+ > 1/2$, for the event $\mathcal{B} = \mathcal{B}(z, J, y, i, \rho^{(i)}, \rho_-, \rho_+)$ defined as*

$$\mathcal{B} = \left\{\mathcal{A}_z,\, X^* \in J,\, Y = y,\, r^{(i)} = \rho^{(i)},\, F(q_-) = \rho_-,\, F(q_+) = \rho_+\right\},$$

we have

$$\mathcal{L}\left(X^* \mid \mathcal{B}\right) = \mathrm{Unif}[q_-, q_+],$$

where $\mathcal{L}(\cdot)$ denotes the (conditional) distribution.

Lemma 2. *For all i, we have that*

$$\mathbb{E}\left(\log \frac{|I_{i+1} \cap J^*|}{|I_i \cap J^*|}\Big|\mathcal{A}\right) \geq -\mathbb{P}\left\{q_{i+1} \in J^* \mid \mathcal{A}\right\}. \tag{1}$$

Proof of Lemma 1. Since the gradient of the convex function f^* is defined with $(f^*)' = \gamma_- + (\gamma - \gamma_-)F$, the minimizer of f^* is at the median of F, i.e.,

$$X^* = \inf\left\{x : F(x) \geq \frac{-\gamma_-}{\gamma_+ - \gamma_-} = \frac{1}{2}\right\}.$$

Under our prior construction, the distribution of F follows a Dirichlet process with the uniform base distribution on $[0, 1]$ and scale parameter α. Therefore with probability 1, F is a distribution function with countably many points of discontinuity, which we will refer to as *jumps*. If we characterize F with the stick breaking process, then the locations of the jumps are at $X_1, X_2, ...$ where the X_k's are independently and uniformly distributed on $[0, 1]$. The sizes of the jumps $\beta_1, \beta_2, ...$ correspond to the lengths of the sticks from the stick-breaking process. We have $\sum \beta_k = 1$, and the two sequences $\{X_k\}_{k\geq 1}$ and $\{\beta_k\}_{k\geq 1}$ are independent.

To proceed, we first show that if the size of the largest jumps is larger than $1/2$, then X^* must occur at the largest jump. That is,

$$\mathcal{A} \subset \cup_{i\geq 1}\left\{X^* = X_k, \beta_{(1)} = \beta_k\right\}. \tag{2}$$

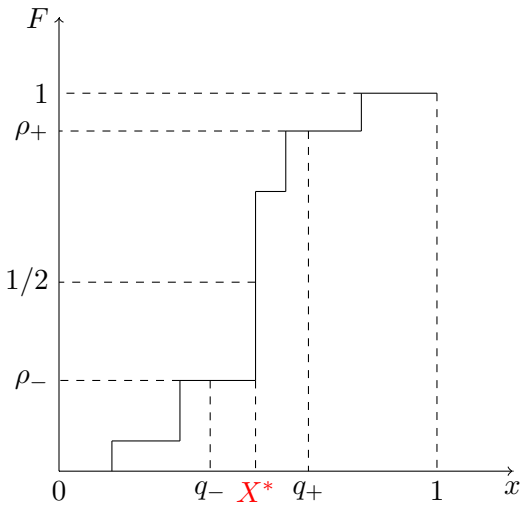

Figure 1: Conditional on $X^* \in J$ and the responses to the first i queries, the range of X^* is narrowed down to $I_i \cap J = [q_-, q_+]$. Further conditioning on $F(q_-) = \rho_-$ and $F(q_+) = \rho_+$, we show that F restricted to $[q_-, q_+]$ also follows a Dirichlet process after appropriate scaling.

To see why, recall that X^* is the median of F. Thus $F(X^*) \geq 1/2$ and $\sup_{x<X^*} F(x) \leq 1/2$. Suppose $\beta_{(1)} = \beta_k$. We consider two cases:

1. if $X^* < X_k$, then $F(X_k) \geq F(X^*) + \beta_{(1)} > 1$;
2. if, on the other hand, $X^* > X_k$, then $F(X_k) \leq \sup_{x<X^*} F(x) - \beta_{(1)} \leq 1/2 - \beta_{(1)} < 0$.

In neither case can F be a distribution function. Therefore we must have $X^* = X_k$ is the location of the largest jump.

For $z \geq 1/2$, conditional on $\mathcal{A}_z$ and $X^* \in [q_-, q_+]$, we know that X^* is at the largest jump in $[q_-, q_+]$. Moreover, since the learner would not have submitted any queries between q_- and q_+ at time i, the events conditioned on do not contain any information on the location of the largest jump. Therefore the conditional distribution of X^* is uniform. To prove the claim rigorously, we need to invoke the self-similarity property of the Dirichlet process.

Recall that F follows a Dirichlet Process is supported on $[0, 1]$ with base distribution $\lambda_{[0,1]}$. The self-similarity property asserts that for any finite partition $0 = x_0 \leq x_1 \leq ... \leq x_{n-1} \leq x_n = 1$ of $[0, 1]$, conditional on the realization of F on $x_1, ..., x_n$, the restriction of F onto each subinterval is also a Dirichlet process scaled. In particular, for each $j \leq n$, we have

$$\mathcal{L}\left(\frac{[F]_{[x_j,x_{j+1}]} - t_j}{t_{j+1} - t_j} \;\middle|\; F(x_1) = t_1, ..., F(x_{n-1}) = t_{n-1}\right) = \mathrm{DP}\left(\lambda_{[x_j,x_{j+1}]}, \alpha\lambda\left([x_j, x_{j+1}]\right)\right),$$

where $[F]_I$ denotes the function F restricted to interval I, λ_I denotes the uniform probability measure on I, and $\lambda(I)$ denotes the Lebesgue measure of I. This property is well-known, and follows from the definition of the Dirichlet process. See Section 1.3 for a proof.

Importantly, the following is a direct consequence of the self-similarity property. For each interval $[a, b] \subset [0, 1]$, conditional on the value of $F(a)$ and $F(b)$, the distribution of F restricted to $[a, b]$ is independent of the realization of F outside of $[a, b]$. As a result, for each interval $I \subset [0, 1]$, given $X^* \in I$, the learner cannot gain any additional information on X^* without querying in I.

This property ensures that the posterior distribution of X^* conditional on $\mathcal{A}$ and the responses is uniform between the two closest queries that sandwich X^*. Therefore, the learner cannot beat the bisection search on the event $\mathcal{A}$.

By definition of the learner's interval I_i, none of the first i queries $q_1, ..., q_i$ can be in $I_i \cap J = [q_-, q_+]$. Since X^* is determined by the values of F inside $[q_-, q_+]$, by the self-similarity property of the Dirichlet process, X^* is independent of the responses to the first i queries conditioning on the values of $F(q_-)$ and $F(q_+)$. Therefore the event $\{r^{(i)} = \rho^{(i)}\}$ can be dropped from $\mathcal{B}$ without changing the conditional distribution of X^*. The indicator $\mathbb{1}\{X^* \in J\}$ is completely determined by whether ρ_- and ρ_+ are above or below $1/2$; and the outside randomness Y is independent of F. Therefore we can drop both events $\{X^* \in J\}$ and $\{Y = y\}$, and obtain

$$\mathcal{L}\left(X^* \mid \mathcal{B}\right) = \mathcal{L}\left(X^* \mid \mathcal{A}_z, F(q_-) = \rho_-, F(q_+) = \rho_+\right).$$

By the self-similarity property of the Dirichlet process, given $F(q_-) = \rho_-$ and $F(q_+) = \rho_+$,the conditional distribution of $(F - \rho_-)/(\rho_+ - \rho_-)$ restricted to $[q_-, q_+]$ is also a Dirichlet process with the uniform base distribution on $[q_-, q_+]$ and scaling parameter $\alpha' = \alpha(q_+ - q_-)$. See Figure 1 for an illustration of this step. In other words, there exist ancillary random vectors $\{X'_k\}_{k\geq 1}$, $\{\beta'_k\}_{k\geq 1}$ generated from a stick-breaking process that characterize the distribution function

$$\widetilde{F} = (F - \rho_-)/(\rho_+ - \rho_-)$$

on $[q_-, q_+]$. In addition, $X'_k \overset{i.i.d.}{\sim} \text{Unif}[q_-, q_+]$, and $(\{X'_k\}_{k\geq 1}, \{\beta'_k\}_{k\geq 1})$ is independent of $(F(q_-), F(q_+))$.

We claim that for all $z \geq 1/2$, the event $\mathcal{A}_z = \{\beta_{(1)} = z\}$ is equivalent to $\{\beta'_{(1)} = z/(\rho_+ - \rho_-)\}$. Suppose $\mathcal{A}_z$ holds, and say $\beta_{(1)} = \beta_j$. Then by (2), $X^* = X_j$. Thus $[q_-, q_+]$ contains the largest jump in F. Since $\widetilde{F}$ is a scaled version of F restricted to $[q_-, q_+]$, the largest jump of $\widetilde{F}$ must be of size $z/(\rho_+ - \rho_-)$. Conversely, if $\beta'_{(1)} = z/(\rho_+ - \rho_-)$, then F contains a jump of size z. When $z \geq 1/2$, this must be the largest jump in F, i.e. $\beta_{(1)} = z$.

Note that conditional on $\mathcal{A}_z$ for $z \geq 1/2$, X^* can be written as the location of the largest jump in $\widetilde{F}$. We have shown that X^* and $\mathcal{A}_z$ can both be expressed as functions that only depend on $\{X'_k, \beta'_k\}$. As a result,

$$\begin{aligned}
&\mathcal{L}(X^* \mid \mathcal{A}_z, F(q_-) = \rho_-, F(q_+) = \rho_+) \\
=&\mathcal{L}\left(\text{location of the largest jump in } F' \mid \beta'_{(1)} = \frac{z}{\rho_+ - \rho_-}, F(q_-) = \rho_-, F(q_+) = \rho_+\right) \\
\overset{(a)}{=}&\mathcal{L}\left(\text{location of the largest jump in } F' \mid \beta'_{(1)} = \frac{z}{\rho_+ - \rho_-}\right) \\
\overset{(b)}{=}&\mathcal{L}\left(\text{location of the largest jump in } F'\right) \\
\overset{(c)}{=}&\mathcal{L}(X'_1) = \text{Unif}[q_-, q_+],
\end{aligned}$$

where (a) is from the independence between $(\{X'_k\}_{k\geq 1}, \{\beta'_k\}_{k\geq 1})$ and $(F(q_-), F(q_+))$; (b) holds because by the stick-breaking characterization of the Dirichlet process, the locations of the jumps $\{\beta_k\}_{k\geq 1}$ and the sizes of the jumps $\{X_k\}_{k\geq 1}$ are independent. More specifically, let j be the index of the largest jump, i.e., $\beta'_{(1)} = \beta'_j$. Then j is only a function of $\{\beta_k\}_{k\geq 1}$ and is therefore independent of $\{X'_k\}_{k\geq 1}$. We have X'_j is independent of $\{\beta'_k\}_{k\geq 1}$, thus we can drop the conditional event which only depends on $\{\beta'_k\}_{k\geq 1}$; (c) is again from the independence of j and $\{\beta'_k\}_{k\leq 1}$. Since $\{X'_k\}_{k\geq 1}$ are distributed *i.i.d.* Unif q_-, q_+, we have $\mathcal{L}(X'_j) = \mathcal{L}(X'_1) = \text{Unif}[q_-, q_+]$.

□

Proof of Lemma 2. From Lemma 1, we have $\mathcal{L}(X^* \mid \mathcal{B}) = \text{Unif}[I_i \cap J]$. We claim that as a consequence,

$$\mathbb{E}\left(\log \frac{|I_{i+1} \cap J|}{|I_i \cap J|} \Big| \mathcal{B}\right) \geq -\mathbb{1}\left\{q_{i+1} = \phi_i(\rho^{(i)}, y) \in I_i \cap J\right\}. \tag{3}$$

(3) can be interpreted as follows. Firstly, the interval $I_i \cap J^*$ is only shortened when querying within $I_i \cap J^*$. Secondly, conditional on all instances of the behavior of F outside of $I_i \cap J^*$, on average, no query can reduce the length of $I_i \cap J^*$ by more than a half. By taking the union of the events $\mathcal{B}$ over all the variables $z > 1/2$, $y \in [-0, 1]$, $\rho_- < 1/2$, $\rho_+ > 1/2$, $\rho^{(i)}$, and J ranging over $J_1, ..., J_{2/\delta}$, we arrive at the event $\mathcal{A}$. Therefore, Lemma 2 follows from (3) by integrating over these variables.

We now prove (3). If $q_{i+1} \notin I_i \cap J$, then $I_{i+1} \cap J = I_i \cap J$ and the claim (3) trivially holds. If $q_{i+1} \in I_i \cap J$, we have

$$\log \frac{|I_{i+1} \cap J|}{|I_i \cap J|} = \mathbb{1}\{X^* \leq q_{i+1}\} \log \frac{q_{i+1} - q_-}{q_+ - q_-} + \mathbb{1}\{X^* > q_{i+1}\} \log \frac{q_+ - q_{i+1}}{q_+ - q_-}.$$

Since the conditional distribution of X^* is uniform, we have

$$\mathbb{E}\left(\log \frac{|I_{i+1} \cap J|}{|I_i \cap J|} \Big| \mathcal{B}\right) \geq \inf_{t \in [0,1]} [t \log t + (1-t) \log(1-t)] = -1.$$

We have finished the proof of (3) and, by consequence, Lemma 2.

□

1.1.2 Proof of the upper bound

To prove the upper bound in Theorem 2, we repeat here our proposed querying strategy under the Bayesian setting. Recall that ν denotes the distribution of X^*. For an interval $I \subset [0, 1]$, ν_I denotes the probability distribution of ν conditioned on I, *i.e.*, $\frac{d\nu_I}{d\nu}(x) = \mathbb{1}\{x \in I\}/\nu(I)$.

Algorithm 1 Querying Strategy under the Bayesian Setting

1: Recursively query the median of the posterior distribution of X^*, until it is supported on an interval I with $\nu(I) \in [2\delta L H_\alpha, 4\delta L H_\alpha]$.
2: Let κ_ℓ be the ℓ/L quantile of ν_I for $\ell = 0, 1, ..., L$ and let $I_\ell = [\kappa_{\ell-1}, \kappa_\ell]$ for $\ell \in [L]$. Query $\kappa_1, ..., \kappa_{L-1}$ and identify j^* for $f'(\kappa_{j-1}) \leq 0$ and $f'(\kappa_j) > 0$ so that I_{j^*} contains X^*.
3: Query the median m_j of ν_{I_j} for $j \in [L]$. If $f'(m_{j^*}) > 0$, let $J_j = [\kappa_{j-1}, m_j]$ for all j; otherwise let $J_j = [m_j, \kappa_j]$.
4: For all $j \neq j^*$, sample $X_j \sim \nu_{J_j}$ independently. Denote $X_{j^*} = X^*$. For $j = 1, ..., L$, run the regular bisection search on J_j to locate X_j up to ϵ-accuracy.

See Figure 2 for an illustration of Algorithm 1. Algorithm 1 is clearly ϵ-accurate by design. In the paper, we proved that it is also (δ, L)-private, assuming that the following claims hold.

(i) $X^* \mid X_1, ..., X_L \sim \text{Unif}\{X_1, ..., X_L\}$.

(ii) With probability 1, $|X_i - X_j| > \delta$ for all $i \neq j$.

It remains to prove the two claims.

Proof of (i): Recall that the index of the subinterval containing X^* is j^*. Since $\nu(I_j)$ are equal for all j, j^* is distributed uniformly in $\{1, ..., L\}$. Therefore the desired claim $X^* \mid X_1, ..., X_L \sim \text{Unif}\{X_1, ..., X_L\}$ is equivalent to j^* and $(X_1, ..., X_L)$ being independent.

Figure 2: Example of phases 2 to 4 of the querying strategy under the Bayesian setting with $L = 3$. In phase 2, the learner queries the 1/3 and 2/3 quantile of ν_I (represented by the dashed lines), and learns that $X^* \in I_2$. In phase 3, she queries the medians m_1,...m_L, and learners that X^* is to the left of m_2. Therefore J_1,...,J_L are defined to be the shaded intervals. In phase 4, X_1 and X_3 are sampled from ν_{J_1} and ν_{J_3} respectively and X_2 is defined to be X^*. Note that the separation of X_1,...,X_L are guaranteed by the separation of J_1,...,J_L.

To show $j^* \perp\!\!\!\perp (X_1, ..., X_L)$, first note that $j^* \perp\!\!\!\perp (J_1, ..., J_L)$, because conditional on j^*, either $J_j = [\kappa_{j-1}, m_j]$ for all j or $J_j = [m_j, \kappa_j]$ for all j, with equal probability. Second, conditional on $(J_1, \ldots, J_L)$, X_j's are independently distributed according to ν_{J_j} across all j. Therefore, we arrive at the conclusion $j^* \perp\!\!\!\perp (X_1, ..., X_L)$.

Proof of (ii): It suffices to show that the intervals $J_1, ..., J_L$ are δ-separated, or equivalently, $|I_j \backslash J_j| \geq \delta$ for all $j \leq L$. Since phase 2 of the querying strategies queries all the medians of $I_1, ..., I_L$, we have $\nu(I_j \backslash J_j) = \nu(I_j)/2 = \nu(I)/(2L) \geq \delta H_\alpha$. Let $\mathbf{m} = d\nu/d\lambda$ be the density of ν. Then

$$|I_j \backslash J_j| \geq \frac{\nu\left(I_j \backslash J_j\right)}{\sup_t \mathbf{m}(t)} = \frac{\delta H_\alpha}{\sup_t \mathbf{m}(t)}. \tag{4}$$

To finish proof of this claim, we only need to bound the density of ν from above. Recall that ν is the distribution of X^*, which is the median of F. Thus the distribution function of ν has the form

$$\nu([0, t]) = \mathbb{P}\left\{X^* \leq t\right\} = \mathbb{P}\left\{F(t) \geq 1/2\right\}.$$

Since $F \sim \mathrm{DP}(\alpha, \lambda_{[0,1]})$, we have $(F(t), 1-F(t)) \sim \mathrm{Dir}(\alpha t, \alpha(1-t))$. Therefore $F(t) \sim \mathrm{Beta}(\alpha t, \alpha(1-t))$. We will use the following Lemma 3 to bound the density of ν. The proof of Lemma 3 is deferred to Section 1.4.

Lemma 3. *Suppose* $X \sim \mathrm{Beta}(\alpha t, \alpha(1-t))$ *for some* $\alpha > 0$, *then for all* $t \in (0, 1)$,

$$h_\alpha \leq \frac{d}{dt}\mathbb{P}\left\{X \geq 1/2\right\} \leq H_\alpha,$$

where $h_\alpha = \frac{1}{3}2^{-\alpha-2}$ *and* $H_\alpha = (3 + 2e^{-1})\alpha + 14$.

By Lemma 3,

$$\mathbf{m}(t) = \frac{d}{dt}\mathbb{P}\left\{F(t) \geq 1/2\right\} \leq H_\alpha, \tag{5}$$

for all $t \in [0, 1]$. Combining (4) and (5) yields that

$$|I_j \backslash J_j| \geq \frac{\delta H_\alpha}{H_\alpha} \geq \delta.$$

We have shown that ν_{j_1}, ... ν_{j_L} are continuous distributions supported on L intervals that are δ-separated from each other. Therefore $|X_i - X_j| > \delta$ for all $i \neq j$ with probability 1.

Finally, we show that Algorithm 1 attains the query complexity upper bound stated in Theorem 2. The number of queries submitted in phase 1 is at most $\log(1/(2\delta L H_\alpha))$. Phase 2 and phase 3 involve $L-1$ and L queries respectively. The number of queries submitted in phase 4 equals

$$\sum_{j\le L}\left\lceil \log\frac{|J_j|}{\epsilon}\right\rceil \le L+\sum_{j\le L}\log\frac{|J_j|}{\epsilon} = L+\log\left(\prod_{j\le L}|J_j|\right)+L\log\frac{1}{\epsilon},$$

To bound the above, note that from Lemma 3 we have

$$\sum_{j\le L}|J_j| \le \frac{\nu(\cup_{j\le L}J_j)}{h_\alpha} \le \frac{2\delta L H_\alpha}{h_\alpha}.$$

Therefore $\prod_{j\le L}|J_j| \le (2\delta H_\alpha/h_\alpha)^L$. Thus the total number of queries submitted by the learner is at most

$$\begin{aligned}
&\log\frac{1}{2\delta L H_\alpha}+(L-1)+L+L\left(\log\frac{\delta}{\epsilon}+\log\frac{4H_\alpha}{h_\alpha}\right)\\
=&L\left(\log\frac{\delta}{\epsilon}+\log\frac{16H_\alpha}{h_\alpha}\right)+\log\frac{1}{\delta L}+\log\frac{1}{4H\alpha}\\
\le&L\left(\log\frac{\delta}{\epsilon}+c_2\right)+\log\frac{1}{\delta L}
\end{aligned}$$

for $c_2=\log(16H_\alpha/h_\alpha)$. The inequality is from $H_\alpha>14$ for all $\alpha>0$.

1.2 Proof under the Minimax Setting

1.2.1 Proof of the lower bound

As a first step, we prove that if $\mathcal{F}$ satisfies Assumption 1, then the learner cannot search faster than the bisection method on any interval $I\subset[0,1]$. The lemma below contains a formal statement of this claim. Note that by taking $I=[0,1]$, Lemma 4 immediately implies a lower bound of $\log(1/\epsilon)$ on the optimal query complexity.

Lemma 4. *Suppose $\mathcal{F}$ satisfies Assumption 1. Let ϕ be an ϵ-accurate querying strategy. Then for each $f\in\mathcal{F}$, each interval $I\subset[0,1]$ that contains the minimizer of f, and each realization of the random seed y, there exists $\widetilde{f}\in\mathcal{F}$, such that*

(1) under ϕ, the query sequence $q(\widetilde{f},y)$ contains at least $\log(|I|/\epsilon)$ queries in I;

(2) the gradient of $\widetilde{f}$ and f coincide outside of I.

Next, we prove the lower bound in Theorem 1 assuming correctness of Lemma 4. The proof of Lemma 4 is deferred to after the lower bound proof.

A key step in this proof is to connect the definition of (δ,L)-privacy with the covering numbers of the information sets. We claim that for a strategy to be (δ,L)-private in the minimax sense, there must be one information set with a large covering number.

Let ϕ be a querying strategy that is both ϵ-accurate and (δ,L)-private. Define the information set of a query sequence q as

$$\mathcal{I}(q)=\{x\in[0,1]:\exists f\in\mathcal{F}\text{ and }y,\ \ s.t.\ \ x=\arg\min f,\ \text{and }q(f,y)=q\}.$$

Denote the $\delta/2$-covering number of $\mathcal{I}(q)$ as $N_c(\mathcal{I}(q), \delta/2)$. Fix the adversary's strategy to be one that samples uniformly from a δ-covering set of $\mathcal{I}(q)$. Since ϕ is (δ, L)-private, there must exist some f minimized at x, for which

$$1/L > \mathbb{P}_f\left\{\left|\widetilde{X} - x\right| \le \delta/2\right\} = \mathbb{E}\left[\mathbb{P}_f\left\{\left|\widetilde{X} - x\right| \le \delta/2 \;\middle|\; q\right\}\right],$$

where the first integration is over q and the second is over the randomness from the adversary's estimation scheme conditional on q. Since x is in $\mathcal{I}(q)$, it must be $\delta/2$-close to at least one of the points in the covering set. Therefore for all q,

$$\mathbb{P}_f\left\{\left|\widetilde{X} - x\right| \le \delta/2 \;\middle|\; q\right\} \ge \frac{1}{N_c(\mathcal{I}(q), \delta/2)}.$$

Taking expected value over q on both sides, we have $\mathbb{E}(1/N_c(\mathcal{I}(q), \delta/2)) < 1/L$. Hence there must exist some query sequence $\bar{q}$ for which $N_c(\mathcal{I}(\bar{q}), \delta/2) > L$. As a result, $\mathcal{I}(\bar{q})$ contains L points $x_1, ..., x_L$ that are at least $\delta/2$-apart.

By definition of the information set, there exist $f_1, ..., f_L \in \mathcal{F}$ and $y_1, ..., y_L \in [0, 1]$, such that f_i is minimized at x_i, and $q(f_i, y_i) = \bar{q}$ for all i. Notice that for each i, $\bar{q}$ must contain a pair of queries at most ϵ-apart that sandwiches x_i. Otherwise suppose the closest pair of queries in $\bar{q}$ that contains x_i forms an interval I of size larger than ϵ. Under Assumption 1, for each $x \in I$, there exists $f \in \mathcal{F}$ for which f is minimized at x and $q(f, y_i)$ is also $\bar{q}$. By taking x to be arbitrarily close to the endpoints of I, the ϵ-accuracy requirement is violated since no estimator $\widehat{X}$ can ensure $|\widehat{X} - x| \le \epsilon/2$ for all $x \in I$. Therefore, the length of I is at most ϵ. Combined with the fact that $x_1, ..., x_L$ are δ-separated, and the assumption $\delta \ge 2\epsilon$, we have shown that $\bar{q}$ contains L pairs of distinct queries. Thus the optimal query complexity is lower bounded by $2L$.

To improve the lower bound to the desired $2L + \log(\delta/\epsilon)$, we would like to argue that aside from the L pairs of queries in $\bar{q}$, the learner must submit enough queries elsewhere to search for X^* in order to fulfill the accuracy requirement. Indeed, the worst-case query complexity is lower bounded by $\log(1/\epsilon)$ for any strategy that is ϵ-accurate. However, the worst-case instance may not be one of $f_1, ..., f_L$. To combine the $2L$ queries used to ensure privacy with the queries used to ensure accuracy therefore becomes the main challenge of the lower bound proof. To address this difficulty, we will again utilize Assumption 1 on the richness of $\mathcal{F}$. On a high level, Assumption 1 allows us to find a large class of functions in $\mathcal{F}$ which can also lead to the query sequence $\bar{q}$. Out of these functions, we show that for at least one of them it takes $\log(\delta/\epsilon)$ extra queries to search for its minimizer. Next we give the rigorous proof of the existence of such a function.

Firstly, note that $\bar{q}$ contains L pairs of ϵ-close queries that sandwich $x_1, ..., x_L$. Since $\delta \ge \epsilon$, we have that for all i, $\bar{q}$ contains at least one query in $[x_i - \delta/2]$, and one query in $[x_i + \delta/2]$. Once at least one query has appeared in each of $[x_i - \delta/2, x_i]$ and $[x_i, x_i + \delta/2]$, we say x_i is "$\delta/2$-localized". Let x_j be the last one to be $\delta/2$-localized out of $x_1, ..., x_L$, and suppose it is $\delta/2$-localized at time T. Without loss of generality, assume a query in $[x_j - \delta/2, x_j]$ appears first, so that $\bar{q}_T \in [x_j, x_j + \delta/2]$. Let $I = [a, b]$ with a defined as the query in $\bar{q}_1, ..., \bar{q}_T$ to the left of x_j that is the closest to x_j, and $b = x_j + \delta/2$. See Figure 3 for an illustration.

Apply Lemma 4 with $I = [a, b]$, $f = f_j$ and $y = y_j$. We can find some $\widetilde{f} \in \mathcal{F}$ that satisfies the two criteria in the statement of Lemma 4. Criterion (2) ensures that the gradient of $\widetilde{f}$ and f_j coincide outside of I. Since x_j is $\delta/2$-localized at time T, $\bar{q}_1, ..., \bar{q}_{T-1}$ do not contain any queries between a and b. Thus $q(\widetilde{f}, y_j)$ and $q(f_j, y_j) = \bar{q}$ agree completely up to time $T - 1$, and contain at least the $2L - 1$ queries outside of I used to sandwich $x_1, ..., x_L$. The reason we need to subtract 1 is because the T'th queries in $\bar{q}$ is in I.

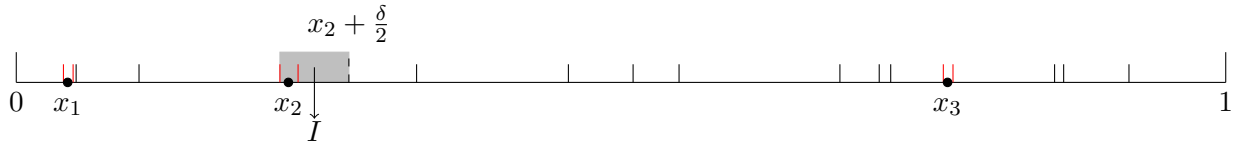

Figure 3: An illustration of the lower bound argument with $L = 3$. The ticks represent all queries in $\bar{q}$. The L pairs of ϵ-close queries that sandwich $x_1, ..., x_L$ are colored red. Suppose x_2 is the last one out of $x_1, ..., x_L$ to be $\delta/2$-localized, and the query in $[x_2 - \delta/2, x_2]$ appears before the one in $[x_2, x_2 + \delta/2]$, then I is defined as the shaded interval. Note that until all of $x_1, ..., x_L$ are $\delta/2$-localized, no query is submitted in I.

By criterion (1) in the statement of Lemma 4, $q(\widetilde{f}, y_j)$ contains at least $\log(|I|/\epsilon) \geq \log(\delta/(2\epsilon))$ queries in I. Combined with the $2L - 1$ queries outside of I, we arrive at the desired lower bound $2L + \log(\delta/\epsilon) - 2$.

Proof of Lemma 4. The lemma is proved by constructing an $\widetilde{f}$ that satisfies both criteria. Our construction scheme in inspired by that of Nemirovski's (See Section 2.1.2 in lecture notes by Iouditski [5])With the querying strategy ϕ fixed, we construct a sequence of functions $\{g_i\}_{i\geq 0} \subset \mathcal{F}$ adapted to the queries and the responses. The construction ensures that for each $i \geq 0$, there is an interval $\Delta_i \subset I$ with $|\Delta_i| \geq |I|/2^i$, such that

1. g_i is minimized at the midpoint of Δ_i;
2. in the query sequence $q(g_i, y)$, the first i queries in I are outside of Δ_i.

By Assumption 1, there exists a function in $\mathcal{F}$ whose gradient of f agrees with that of f outside of I, and is minimized at the midpoint of I. Let this function be g_0 and let $\Delta_0 = I$.

Inductively construct the rest of $\{g_i\}$. Given $g_0, ..., g_i$, by the induction hypothesis in $q(g_i, y)$, the first i queries in I are all outside of $\Delta_i = [a_i, b_i]$. Let q be the $(i + 1)$'th query of $q(g_i, y)$ in I. If q is not in Δ_i, then we can simply let $g_{i+1} = g_i$ and $\Delta_{i+1} = \Delta_i$ to complete the $(i + 1)$'th step of the induction. If $q \in \Delta_i$, depending on whether q lands to the left or right of the midpoint of Δ_i, let Δ_{i+1} be either $[q, b_i]$ or $[a_i, q]$, so that $|\Delta_{i+1}| \geq |\Delta_i|/2$. Let $g_{i+1} \in \mathcal{F}$ be a function whose gradient agrees with g_i outside of Δ_i, and is minimized at the midpoint of Δ_{i+1}. By Assumption 1 such a g_{i+1} always exists.

The construction can be carried out until for some integer K, we cannot find the $(K + 1)$'th query of $q(g_K, y)$ in I. That is, $q(g_K, y)$ contains only K queries in I. By construction, $q(g_K, y)$ does not contain any queries in Δ_K. Therefore under Assumption 1, the learner cannot rule out any member of Δ_K being X^*. For the strategy to be ϵ-accurate, we must have $|\Delta_K| < \epsilon$; hence $K > \log(|I|/\epsilon)$. Taking $\widetilde{f} = g_K$ finishes the proof of the lemma. □

1.2.2 Proof of the upper bound

Define a guess at q as a pair of queries placed at q and $q + \epsilon$. The guess allows the learner to test whether X^* is contained in the ϵ-length interval $[q, q + \epsilon]$. To ensure privacy, we create L potential locations for X^* that are at least δ-separated but induce the same querying sequence. That is achieved by submitting L *guesses* that are δ-separated. Once guessed correctly, the learner's accuracy requirement is automatically fulfilled and the remaining queries can be used to conceal

X^* from the adversary. We consider the cases $\delta \leq 2^{-L}$ and $\delta > 2^{-L}$ separately. The querying strategy is contained in Algorithm 2.

Algorithm 2 Querying Strategy under the Minimax Setting

Let $I = [0, 1]$.
if $\delta \leq 2^{-L}$ **then**
  Submit the first guess at $1/2$.
  Recursively submit the remaining $L - 1$ guesses via bisection: if none of the submitted guesses is correct, update $I = [a, b]$ according the gradient $(f^*)'(q)$ at the previous guess q. If $(f^*)'(q) \leq 0$, then $X^* \geq q$, so we let the updated I be $[q, b]$; otherwise update I to be $[a, q]$. Submit the next guess at the midpoint of the updated I.
  Once a guess is found to be correct, always (do this also for all the remaining guesses) update I to be the right half of I, and submit the next guess at the midpoint of the updated I.
else
  Submit the first guess at 0.
  Let K be an integer solution in $\{0, 1, ..., L-1\}$ such that $\ell_K := 2^{-K}/(L-K) \in [\delta, 2\delta]$. When $\delta > 2^{-L}$, a solution always exists.
  Submit the next K guesses via bisection. Update I accordingly. As in the $\delta \leq 2^{-L}$ case, once any guess is found to be correct, always update I to its right half.
  Divide I into $L - K$ equal length subintervals. Submit the next $L - K - 1$ queries at the endpoints of the subintervals (excluding the 2 endpoints of I).
end if
if none of the guesses is correct **then**
  Run bisection search on I until reaching ϵ-accuracy.
else
  Fill the remaining query sequence with trivial queries at 1.
end if

We first prove the upper bound in the case $\delta \leq 2^{-L}$. In total, $L+\log(1/\epsilon)$ queries are submitted under Algorithm 2. The strategy is clearly ϵ-accurate. To see that it is also (δ, L)-private, note that all f^* whose minimizer lies in one of the L intervals $[1/2, 1/2+\epsilon]$, $[3/4, 3/4+\epsilon]$, ..., $[1-2^{-L}, 1-2^{-L}+\epsilon]$ share exactly the same query sequence. Under Assumption 1, for each i there exists at least one function f_i minimized at some $x_i \in [1 - 2^{-i}, 1 - 2^{-i} + \epsilon]$. When $\delta \leq 2^{-L}$, the x_i's are at least δ apart from each other. Therefore no adversary can achieve $\inf_{f \in \{f_1, ..., f_L\}} \mathbb{P}_f\{|\widetilde{X} - x| \leq \delta/2\} > 1/L$.

When $\delta > 2^{-L}$, the total number of queries is at most $\log(\delta/\epsilon)+2L+1$. Note that the first guess at 0 always contains a trivial query at 0. Removing the trivial query yields a query complexity of $\log(\delta/\epsilon) + 2L$. To prove (δ, L)-privacy, note that for if f^* is minimized in one of the L intervals $[0, \epsilon]$, $[1 - 2^{-i}, 1 - 2^{-i} + \epsilon]$ for $i \leq K$, or $[1 - 2^{-K} + \frac{i\ell_K}{L-K}, 1 - 2^{-K} + \frac{i\ell_K}{L-K} + \epsilon]$ for $i \leq L - K$, then they induce the same query sequence. This completes the proof of the upper bound.

1.3 Self-similarity property of the Dirichlet Process

Proposition 1. *Let μ be a random probability measure on $\mathcal{X}$ that follows a Dirichlet Process with base distribution function μ_0 and concentration parameter α. Let $\mathcal{X} = \cup_{i \leq n} B_i$ be an arbitrary finite partition of $\mathcal{X}$. Then for all $i \leq n$, we have*

$$\mu_{B_i} \mid \mu(B_1), ..., \mu(B_n) \sim DP\left(\mu_{0,B_i}, \alpha\mu_0\left(B_i\right)\right),$$

where μ_{B_i} and μ_{0,B_i} denote the conditional probability measures of μ and μ_0 respectively, conditioned on B_i.

Proof. For simplicity we present the proof only for $i = 1$. The proof for general i is identical. Let $B_1 = \cup_{j\leq m} A_j$ be an arbitrary finite partition of B_1. Then $(A_1, ..., A_m, B_2, ..., B_n)$ is a partition of $\mathcal{X}$. Therefore from the definition of the Dirichlet Process, we have

$$(\mu(A_1), ..., \mu(A_m), \mu(B_2), ..., \mu(B_n)) \sim \mathrm{Dir}(\alpha\mu_0(A_1), ..., \alpha\mu_0(A_m), \alpha\mu_0(B_2), ..., \alpha\mu_0(B_n)).$$

From the density function of the Dirichlet distribution, we can derive that

$$\frac{(\mu(A_1), ..., \mu(A_m))}{1 - \sum_{i\geq 2}\mu(B_i)} \,\Big|\, \mu(B_2), ..., \mu(B_n) \sim \mathrm{Dir}(\alpha\mu_0(A_1), ..., \alpha\mu_0(A_m)).$$

Again by definition of the Dirichlet Process, we have

$$\mu_{B_1} \mid \mu(B_2), ..., \mu(B_n) \sim \mathrm{DP}(\mu_{0,B_1}, \alpha\mu_0(B_1)).$$

□

Consider the special case where $\mathcal{X} = [0, 1]$. As a corollary of Proposition 1, we have for any finite partition $0 = x_0 \leq x_1 \leq ... \leq x_{n-1} \leq x_n = 1$ of $[0, 1]$,

$$\mathcal{L}\left(\frac{[F]_{[x_i,x_{i+1}]} - t_i}{t_{i+1} - t_i} \,\Big|\, F(x_1) = t_1, ..., F(x_{n-1}) = t_{n-1}\right) = \mathrm{DP}\left(\mu_{0,[x_i,x_{i+1}]}, \alpha\mu_0[x_i, x_{i+1}]\right).$$

1.4 Proof of Lemma 3

In this section we prove the technical result Lemma 3 on the Beta distribution.

Proof. We can assume WOLG that $t \in (0, 1/2]$. That is because for $t > 1/2$, $1 - X \sim \mathrm{Beta}(\alpha(1 - t), \alpha t)$ and

$$\frac{d}{dt}\mathbb{P}\{X \geq 1/2\} = \frac{d}{d(1-t)}\mathbb{P}\{1 - X \geq 1/2\}.$$

Let $\phi_t(x) = x^{\alpha t-1}(1-x)^{\alpha(1-t)-1}$ be the unnormalized density of the $\mathrm{Beta}(\alpha t, \alpha(1-t))$ distribution. Since $\frac{d}{dt}\phi_t(x) = \alpha \ln\frac{x}{1-x}\phi_t(x)$, we have

$$\begin{aligned}
\frac{d}{dt}\mathbb{P}\{X \geq 1/2\} =& \frac{d}{dt}\frac{\int_{1/2}^1 \phi_t(x)dx}{\int_0^1 \phi_t(x)dx} \\
=& \alpha\frac{\int_{1/2}^1 \ln\frac{x}{1-x}\phi_t(x)dx \int_0^1 \phi_t(x)dx - \int_{1/2}^1 \phi_t(x)dx \int_0^1 \ln\frac{x}{1-x}\phi_t(x)dx}{\left(\int_0^1 \phi_t(x)dx\right)^2} \\
=& \alpha\left[\mathbb{E}\left(\mathbb{1}\{X \geq 1/2\}\ln\frac{X}{1-X}\right) - \mathbb{P}\{X \geq 1/2\}\mathbb{E}\left(\ln\frac{X}{1-X}\right)\right].
\end{aligned}$$

To prove the lemma, we claim that for $t \leq 1/2$,

$$2^{-\alpha-2}t \leq \alpha\mathbb{E}\left(\mathbb{1}\{X \geq 1/2\}\ln\frac{X}{1-X}\right) \leq \max\{3\alpha, 12\}; \tag{6}$$

$$\left[2^{-\alpha-2}\left(\tfrac{1}{2}-\tfrac{t}{1-t}\right)\right]_+ \le -\alpha\mathbb{P}\{X\ge 1/2\}\mathbb{E}\left(\ln\frac{X}{1-X}\right) \le 2e^{-1}\alpha+2, \tag{7}$$

where $[\cdot]_+ = \max\{\cdot, 0\}$ stands for the positive part.

The upper bound $\frac{d}{dt}\mathbb{P}\{X\ge 1/2\} \le H_\alpha$ follows easily from adding up the two upper bounds. For the lower bound on the derivative, the two lower bounds in (6) and (7) yield

$$\frac{d}{dt}\mathbb{P}\left\{X\ge 1/2\right\} \ge 2^{-\alpha-2}\left(t+\left(\frac{1}{2}-\frac{t}{1-t}\right)_+\right) \ge \frac{1}{3}2^{-\alpha-2} = h_\alpha,$$

where the last equality is achieved at $t = 1/3$.

It remains to prove (6) and (7). Let us start from the cross-product term (6). Since $\mathbb{1}\{X \ge 1/2\}\ln\frac{X}{1-X} \ge 0$, by Tonelli's theorem,

$$\mathbb{E}\left(\mathbb{1}\{X\ge 1/2\}\ln\frac{X}{1-X}\right) = \int_0^\infty \mathbb{P}\left\{\mathbb{1}\{X\ge 1/2\}\ln\frac{X}{1-X} > s\right\}ds = \int_0^\infty \mathbb{P}\left\{X\ge\frac{e^s}{1+e^s}\right\}ds.$$

The density function of X allows us to write

$$\mathbb{E}\left(\mathbb{1}\{X\ge 1/2\}\ln\frac{X}{1-X}\right) = \frac{\int_0^\infty\int_{\frac{e^s}{1+e^s}}^1 x^{\alpha t-1}(1-x)^{\alpha(1-t)-1}dxds}{B(\alpha t,\alpha(1-t))}, \tag{8}$$

where $B(\alpha,\beta) = \int_0^1 s^{\alpha-1}(1-s)^{\beta-1}ds$ is the Beta function. First we prove the upper bound in (6). For the numerator, since $\alpha t - 1 > -1$ and $x \ge \frac{e^s}{1+e^s} \ge 1/2$, we have $x^{\alpha t-1} \le 2$, and

$$\int_{\frac{e^s}{1+e^s}}^1 x^{\alpha t-1}(1-x)^{\alpha(1-t)-1}dx \le 2\int_{\frac{e^s}{1+e^s}}^1 (1-x)^{\alpha(1-t)-1}dx = \frac{2(1+e^s)^{-\alpha(1-t)}}{\alpha(1-t)}.$$

Therefore the numerator of (8) is upper bounded by

$$2\int_0^\infty \frac{e^{-\alpha(1-t)s}}{\alpha(1-t)}ds = \frac{2}{\alpha^2(1-t)^2} \le \frac{8}{\alpha^2}$$

for all $t \le 1/2$. Moreover,

$$B(\alpha t,\alpha(1-t)) = \frac{\Gamma(\alpha t)\Gamma(\alpha(1-t))}{\Gamma(\alpha)}$$

is minimized at $t = 1/2$ by the log-convexity of the Gamma function $\Gamma(z)$ [3], where $\Gamma(z) = \int_0^\infty s^{z-1}e^{-s}ds$ satisifying $\Gamma(z+1) = z\Gamma(z)$ for $z > 0$. Hence it follows from (8) that for all $t \le 1/2$,

$$\alpha\mathbb{E}\left(\mathbb{1}\{X\ge 1/2\}\ln\frac{X}{1-X}\right) \le \frac{8\Gamma(\alpha)}{\alpha\Gamma(\alpha/2)^2}. \tag{9}$$

We claim that the right-hand side of (9) is a non-decreasing function in α on $(0,\infty)$. To see that, let $g(\alpha) = 8\Gamma(\alpha)/(\alpha\Gamma(\alpha/2)^2)$. We have

$$\frac{d}{d\alpha}(\ln g(\alpha)) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - \frac{1}{\alpha} - \frac{\Gamma'(\alpha/2)}{\Gamma(\alpha/2)} = \psi(\alpha) - \psi(\alpha/2) - \frac{1}{\alpha}. \tag{10}$$

Here $\psi(\cdot) = \Gamma'(\cdot)/\Gamma(\cdot)$ is the digamma function with expansion [1, 6.3.16]

$$\psi(1+z) = -\gamma + \sum_{n=1}^\infty \frac{z}{n+z},$$

where γ is the Euler-Mascheroni constant. Applying the expansion on (10) yields

$$\frac{d}{d\alpha}\left(\ln g(\alpha)\right) = \sum_{n=1}^{\infty}\left(\frac{\alpha-1}{n+\alpha-1} - \frac{\alpha/2-1}{n+\alpha/2-1}\right) - \frac{1}{\alpha} \geq \frac{\alpha-1}{1+\alpha-1} - \frac{\alpha/2-1}{1+\alpha/2-1} - \frac{1}{\alpha} = 0.$$

We have shown that g is a non-decreasing function on $\mathbb{R}^+$. It follows from (9) that for all $\alpha \leq 4$, $\alpha\mathbb{E}(\mathbb{1}\{X \geq 1/2\}\ln\frac{X}{1-X}) \leq g(4) = 12$.

Next we show that for all $\alpha > 4$, the cross-product term in (6) is upper bounded by 3α. By Markov's inequality,

$$\mathbb{P}\left\{X \geq \frac{e^s}{1+e^s}\right\} = \mathbb{P}\left\{1-X \leq \frac{1}{1+e^s}\right\} = \mathbb{P}\left\{\frac{1}{1-X} \geq 1+e^s\right\} \leq \frac{1}{1+e^s}\mathbb{E}\left[\frac{1}{1-X}\right].$$

Since $1-X \sim \text{Beta}(\alpha(1-t), \alpha t)$, we have

$$\mathbb{E}\left[\frac{1}{1-X}\right] = \frac{\int_0^1 x^{\alpha(1-t)-2}(1-x)^{\alpha t-1}dx}{\int_0^1 x^{\alpha(1-t)-1}(1-x)^{\alpha t-1}dx}.$$

For all $\alpha \geq 4$ and $t \leq 1/2$, $\alpha(1-t)-1 \geq 0$, hence both integrals converge, and

$$\mathbb{E}\left[\frac{1}{1-X}\right] = \frac{B(\alpha(1-t)-1, \alpha t)}{B(\alpha(1-t), \alpha t)} = \frac{\Gamma(\alpha(1-t)-1)\Gamma(\alpha t)/\Gamma(\alpha-1)}{\Gamma(\alpha(1-t))\Gamma(\alpha t)/\Gamma(\alpha)} = \frac{\alpha-1}{\alpha(1-t)-1} \leq 3$$

when $\alpha \geq 4$. Therefore

$$\alpha\mathbb{E}\left(\mathbb{1}\{X \geq 1/2\}\ln\frac{X}{1-X}\right) \leq 3\alpha\int_0^\infty \frac{1}{1+e^s}ds \leq 3\alpha.$$

That finishes the proof of the upper bound in (6). Next we prove the lower bound in (6). Since $x^{\alpha t-1} \geq \min\{(1/2)^{\alpha t-1}, 1\}$ for all $x \geq e^s/(1+e^s) \geq 1/2$, we have that the numerator in (8) is lower bounded by

$$\begin{aligned}
&\min\left\{\left(\tfrac{1}{2}\right)^{\alpha t-1}, 1\right\}\int_0^\infty \int_{\frac{e^s}{1+e^s}}^1 (1-x)^{\alpha(1-t)-1}dxds \\
=&\frac{\min\left\{\left(\tfrac{1}{2}\right)^{\alpha t-1}, 1\right\}}{\alpha(1-t)}\int_0^\infty \left(\frac{1}{1+e^s}\right)^{\alpha(1-t)}ds \\
\geq&\frac{\min\left\{\left(\tfrac{1}{2}\right)^{\alpha t-1}, 1\right\}\left(\tfrac{1}{2}\right)^{\alpha(1-t)}}{\alpha(1-t)}\int_0^\infty e^{-s\alpha(1-t)}ds \\
=&\frac{\left(\tfrac{1}{2}\right)^{\max\{\alpha-1, \alpha(1-t)\}}}{\alpha^2(1-t)^2} \geq \frac{2^{-\alpha}}{\alpha^2}.
\end{aligned} \tag{11}$$

To handle the denominator in (6), note that $(1-x)^{\alpha(1-t)-1} \leq 2$ for all $x \leq 1/2$ and $x^{\alpha t-1} \leq 2$ for all $x \geq 1/2$. Therefore the denominator in (6)

$$B(\alpha t, \alpha(1-t)) \leq 2\int_0^{1/2} x^{\alpha t-1}dx + 2\int_{1/2}^1 (1-x)^{\alpha(1-t)-1}dx = 2\left[\frac{2^{-\alpha t}}{\alpha t} + \frac{2^{-\alpha(1-t)}}{\alpha(1-t)}\right] \leq \frac{2}{\alpha t(1-t)}. \tag{12}$$

Combining (8), (11) and (12) yields

$$\alpha\mathbb{E}\left(\mathbb{1}\{X \geq 1/2\} \ln \frac{X}{1-X}\right) \geq \alpha \frac{2^{-\alpha}\alpha t(1-t)}{2\alpha^2} \geq 2^{-\alpha-2}t.$$

Next let us prove (7). Firstly, write

$$\mathbb{E}\left(\ln \frac{X}{1-X}\right) = \psi(\alpha t) - \psi(\alpha) - (\psi(\alpha(1-t)) - \psi(\alpha)) = \psi(\alpha t) - \psi(\alpha(1-t))$$

where we recall that $\psi(z) = \frac{d}{dz}\ln\Gamma(z)$ is the digamma function. Since Γ is log-convex on $\mathbb{R}^+$, ψ is non-decreasing. Therefore for all $t \leq 1/2$, we have

$$-\alpha\mathbb{P}\{X \geq 1/2\}\mathbb{E}\left(\ln \frac{X}{1-X}\right) \geq 0.$$

Furthermore, it has been shown in [2, Eq (2.2)] that for all $z > 0$, the digamma function satisfies

$$\frac{1}{2z} < \ln z - \psi(z) < \frac{1}{z}. \tag{13}$$

Therefore

$$-\mathbb{E}\left(\ln \frac{X}{1-X}\right) = \psi(\alpha(1-t)) - \psi(\alpha t) \geq \ln(\alpha(1-t)) - \frac{1}{\alpha(1-t)} - \ln(\alpha t) + \frac{1}{2\alpha t} \geq \frac{1}{\alpha}\left(\frac{1}{2t} - \frac{1}{1-t}\right) \tag{14}$$

when $t \leq 1/2$.

We still need to bound $\mathbb{P}\{X \geq 1/2\}$ from below. As in the proof of (6), we can write

$$\mathbb{P}\left\{X \geq \tfrac{1}{2}\right\} = \frac{\int_{1/2}^1 x^{\alpha t-1}(1-x)^{\alpha(1-t)-1}dx}{B(\alpha t, \alpha(1-t))}. \tag{15}$$

Again from $x^{\alpha t-1} \geq \min\{(1/2)^{\alpha t-1}, 1\}$ for all $x \geq 1/2$, we have that the numerator of (15) is bounded from below by

$$\max\left\{\left(\tfrac{1}{2}\right)^{\alpha t-1}, 1\right\} \int_{1/2}^1 (1-x)^{\alpha(1-t)-1}dx \geq \frac{2^{-\alpha}}{\alpha}.$$

Combining the last displayed equation with (12) and (15) yields that

$$\mathbb{P}\left\{X \geq \tfrac{1}{2}\right\} \geq \frac{2^{-\alpha}}{\alpha} \times \frac{\alpha t(1-t)}{2} \geq 2^{-\alpha-2}t$$

for all $t \leq 1/2$. In view of (14), it follows that

$$-\alpha\mathbb{P}\left\{X \geq \tfrac{1}{2}\right\}\mathbb{E}\left(\ln \frac{X}{1-X}\right) \geq 2^{-\alpha-2}\left(\frac{1}{2} - \frac{t}{1-t}\right).$$

That concludes the proof of the lower bound in (7). Next we move to the upper bound in (7). By Markov's inequality,

$$\mathbb{P}\{X \geq 1/2\} \leq 2\mathbb{E}X = 2t. \tag{16}$$

Again from (13) we have that for all $t \le 1/2$,

$$\begin{aligned}-\mathbb{E}\left(\ln\frac{X}{1-X}\right) =& \psi(\alpha(1-t)) - \psi(\alpha t)\\ \le& \ln(\alpha(1-t)) - \frac{1}{2\alpha(1-t)} - \left(\ln(\alpha t) - \frac{1}{\alpha t}\right)\\ =& \ln\frac{1-t}{t} + \frac{2-3t}{2\alpha t(1-t)}.\end{aligned}$$

Combining the last displayed equation with (16) yields that

$$\begin{aligned}-\alpha\mathbb{P}\{X \ge 1/2\}\mathbb{E}\left(\ln\frac{X}{1-X}\right) \le& 2\alpha t\left(\ln\frac{1-t}{t} + \frac{2-3t}{2\alpha t(1-t)}\right)\\ \le& (2t\ln(1/t))\alpha + \frac{2-3t}{1-t} \le 2e^{-1}\alpha + 2.\end{aligned}$$

We have thus established the inequalities (6) and (7). □

2 Extension to Multidimensions

In this section we extend our results under the minimax setting to optimization of convex separable functions in $\mathbb{R}^d$. Separable convex optimization arises in a variety applications such as inventory control in operation research, resource allocation in networking, and distributed optimization in multi-agent networks [6, 7, 4], when the global objection function is a sum of the local objective functions and each local objective function depends only on one component of the decision variable. Here, separability ensures that there is no cross-coordinate information leakage. Further generalizing our result to allow for general (non-separable) functions in $\mathbb{R}^d$ is left as future work.

Suppose the true function $f^* : [0,1]^d \to \mathbb{R}$ belongs to a family of convex separable functions

$$\mathcal{F} = \left\{ f : f(x) = \sum_{i=1}^{d} f_i(x_i),\ f_i \in \mathcal{F}_i \right\},$$

where each $\mathcal{F}_i$ is a family of one-dimensional convex functions. For each query $q \in [0,1]^d$ submitted, the learner receives the gradient vector $\nabla f(q) = (f_1'(q_1), ..., f_d'(q_d))$ as the response. We say a querying strategy is ϵ-accurate if

$$\inf_{f\in\mathcal{F}} \mathbb{P}_f\left\{\left\|\widehat{X} - x\right\|_\infty \le \epsilon/2\right\} = 1,$$

where x is the minimizer of f. We say ϕ is (δ, L)-private if

$$\sup_{\widetilde{X}} \inf_{f\in\mathcal{F}} \mathbb{P}_f\left\{\left\|\widetilde{X} - x\right\|_\infty \le \delta/2\right\} \le 1/L.$$

In other words, we declare privacy breach if the adversary's estimator is within a $\delta/2$-neighborhood around the true minimizer with probability higher than $1/L$. As in the one-dimensional case, we need to impose some assumption on the complexity of the function class $\mathcal{F}$. Since $\mathcal{F}$ contains only separable functions, we can simply impose the one-dimensional assumption onto each of the d one-dimensional function classes $\mathcal{F}_1, ..., \mathcal{F}_d$. Below is the extension of our one-dimensional result to d dimensions.

Theorem 3. *Let $N_d(\epsilon, \delta, L)$ denote the optimal query complexity in dimension d under the minimax setting. Suppose $\mathcal{F}_i$ all satisfy Assumption 1 for all $i = 1, ..., d$. If $2\epsilon \le \delta \le L^{-1/d}$, then*

$$2L^{1/d} + \log \frac{\delta}{\epsilon} - 2 \le N_d(\epsilon, \delta, L) \le \begin{cases} 2L^{1/d} + \log \frac{\delta}{\epsilon} & \text{if } L^{1/d} \ge \log \frac{1}{\delta} \\ L^{1/d} + \log \frac{1}{\epsilon} & o.w. \end{cases}.$$

Remark 1. *We choose to quantify the error of the learner and the adversary with respect to the $\|\cdot\|_\infty$ norm because $\|x - y\|_\infty \le \epsilon/2$ is equivalent to $|x_i - y_i| \le \epsilon/2$ for all $i \le d$, so the analysis can be elegantly reduced to the one-dimensional case. However our result does not crucially depend on the choice of the norm. From the basic inequality $\|x\|_\infty \le \|x\|_2 \le \sqrt{d}\|x\|_\infty$, we have that the optimal query complexity can differ by at most a d-dependent additive constant if the Euclidian norm were used instead.*

Proof of the upper bound. Under the minimax privacy framework, to make a strategy private, we only need to find L functions $f^{(1)}, .., f^{(L)} \in \mathcal{F}$ whose minimizers are δ-apart, such that the query sequence for $f^{(1)}, ..., f^{(L)}$ are identical. That would ensure that the adversary who only observes the query sequence cannot succeed with probability higher than $1/L$.

To construct such L functions, we design a querying strategy that submits $L^{1/d}$ *guesses* δ-apart along each dimension. To recap, in Section 1.2 we defined a guess at x to be a pair of ϵ-apart queries $(x, x + \epsilon)$. The guesses across the d dimensions intersect with each other in $[0, 1]^d$ to create $(L^{1/d})^d = L$ cubes of diameter ϵ that potentially contain the minimizer of the true function f^*. The guesses are submitted following the same algorithm as in the one-dimensional case (see the upper bound proof of Theorem 1), except with L replaced by $L^{1/d}$.

Note that since each query is a d-dimensional vector and the function f^* is separable, we can run the search algorithms along the d directions in parallel. More concretely, write $f^*(x) = \sum_{i \le d} f_i^*(x_i)$, and let $q = (q_1, q_2, ..., q_n)$ be the query sequence where $q_j = (q_{j,1}, ..., q_{j,d}) \in [0, 1]^d$. Each time the learner submits a query q_j, she receives the gradient vector

$$\nabla f^*(q_j) = \left((f_1^*)'(q_{j,1}), ..., (f_d^*)'(q_{j,d})\right).$$

For each dimension i, the learner leverages the gradient information $(f_i^*)'(q_{j,i})$ and constructs the next query $q_{j+1,i}$ in dimension i, as if she were learning the minimizer of f_i^* in one-dimension.

In particular, fix any dimension $1 \le i \le d$. The first $2L^{1/d}$ queries $q_{1,i}, ..., q_{2L^{1/d},i}$ consist of $L^{1/d}$ pairs of queries (guesses) that are δ-apart. When $\delta \le 2^{-L^{1/d}}$, these guesses are submitted along the bisection search path:

1. The first guess is at $1/2$, i.e., $q_{1,i} = 1/2$ and $q_{2,i} = 1/2 + \epsilon$. The learner's interval I is initialized to be $[0, 1]$.

2. For each $1 \le j \le L^{1/d} - 1$, submit the $(j + 1)$'th guess at follows: if none of the previous guesses is correct, then inspect the gradient $(f_i^*)'(q_{2j-1,i})$ from the j'th guess to deduce which half of I contains the minimizer X_i^* of f_i^*. Update the learner's interval I accordingly so that it contains X_i^*. Submit the $(j + 1)$'th guess at the midpoint of the updated I. If one of the first j guesses is correct, then update I to its right half, and submit the $(j + 1)$'th guess at its midpoint.

When $\delta > 2^{-L^{1/d}}$, only the first K guesses are submitted along the bisection path, and the remaining $L^{1/d} - K$ guesses are submitted via a grid search on the interval I generated from the first K guesses. Here K is the largest integer for which all the guesses are δ-apart. Under the assumption $\delta \le L^{-1/d}$ such a K always exists.

After all the guesses are submitted, if none of the guesses is correct, the learner runs a simple bisection search on a $\max\{2^{-L^{1/d}}, \delta\}$-length interval until reaching ϵ-accuracy; otherwise the learner simply fills the remaining queries along this dimension with trivial queries $q_{i,j} = 1$ for all $j \geq 2L^{1/d}$. The total number of queries is exactly the desired upper bound $2L^{1/d} + \log(\max\{2^{-L^{1/d}}, \delta\}/\epsilon)$.

Next we show this querying strategy is (δ, L)-private. Here we give the proof in the $\delta \leq 2^{-L^{1/d}}$ case. The proof for the $\delta > 2^{-L^{1/d}}$ case follows analogously. For each i, it is easy to see that if

$$X_i^* \in \cup_{j \leq L^{1/d}} [1 - 2^{-j}, 1 - 2^{-j} + \epsilon]$$

then the queries along the i'th dimension would always be L guesses at $1/2, 3/4, ..., 1 - 2^{L^{1/d}}$, followed by trivial queries at 1. As a result, for all $f^* \in \mathcal{F}$ such that

$$X^* \in \prod_{i \leq d} \left(\cup_{j \leq L^{1/d}} [1 - 2^{-j}, 1 - 2^{-j} + \epsilon] \right) \triangleq J,$$

share the same query sequence. Clearly J contains $(L^{1/d})^d$ members that are separated by at least δ in $\|\cdot\|_\infty$ distance. Hence the strategy is (δ, L)-private.

□

Proof of the lower bound. Let ϕ be a querying strategy that is ϵ-accurate and (δ, L)-private. Via the same argument in one-dimension, we can show that there is at least one query sequence q whose information set $\mathcal{I}(q)$ has a $\delta/2$-covering number at least L. For each $i = 1, ..., d$, let

$$\mathcal{I}_i(q) = \left\{ x_i : x = (x_1, ..., x_i, ..., x_d) \in \mathcal{I}(q) \text{ for some } x \in [0,1]^d \right\}$$

be the projection of $\mathcal{I}(q)$ to dimension i. Then we have $\mathcal{I}(q) \subset \prod_{i \leq d} \mathcal{I}_i(q)$, thus

$$L \leq N_c \left(\mathcal{I}(q), \delta/2, \|\cdot\|_\infty \right) \leq N_c \left(\prod_{i \leq d} \mathcal{I}_i(q), \delta/2, \|\cdot\|_\infty \right) = \prod_{i \leq d} N_c \left(\mathcal{I}_i(q), \delta/2, |\cdot| \right).$$

Therefore for at least one $i \leq d$, we must have that the $\delta/2$-covering number of the projection $\mathcal{I}_i(q)$ is no less than $L^{1/d}$. It follows that $\mathcal{I}_i(q)$ contains $x_i^{(1)}, ..., x_i^{(L^{1/d})}$ that are at least $\delta/2$-apart. For the strategy to be ϵ-accurate, the queries in q along this dimension i must contain at least $L^{1/d}$ pairs of ϵ-apart queries sandwiching $x_i^{(1)}, ..., x_i^{(L^{1/d})}$. The rest of the proof exactly follows the one-dimensional case. □

References

[1] Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. US Government printing office, 1948.

[2] Horst Alzer. On some inequalities for the gamma and psi functions. *Mathematics of computation*, 66(217):373–389, 1997.

[3] Emil Artin. *The gamma function*. Courier Dover Publications, 2015.

[4] Stephen Boyd, Neal Parikh, and Eric Chu. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.

[5] Anatoli Iouditski. Efficient methods in optimization, 2007.

[6] Angelia Nedic, Asuman Ozdaglar, et al. Convex optimization in signal processing and communications, chapter cooperative distributed multi-agent optimization. eds., eldar, y. and palomar, d. *Eds. Eldar Y. and Palomar D*, 2008.

[7] Arun Padakandla and Rajesh Sundaresan. Separable convex optimization problems with linear ascending constraints. *SIAM Journal on Optimization*, 20(3):1185–1204, 2010.