A.1 Notations

We first define some notations in the context of the model (1). For $p \geq 1$ and $d \geq 1$, define

$$\mathcal{A}_{p,d} := \left\{ \prod_{j=1}^{p} [\ell_j, u_j] \in \mathcal{A} \mid \# \{ j \in [p] \mid [\ell_j, u_j] \neq [0,1] \} \leq d \right\}$$

(21)

That is, each rectangle in $\mathcal{A}_{p,d}$ has at most $d$ dimensions that are not the full interval $[0,1]$. Note that for a decision tree with depth $d$, each leaf node represents a rectangle in $\mathcal{A}_{p,d}$. Furthermore, for $\delta \in (0,1)$, define values

$$\bar{\ell}_1(\delta) = \bar{\ell}_1(\delta, n, d) := \frac{4}{n} \log(2p^d(n+1)^{2d}/\delta)$$

$$\bar{\ell}_2(\delta) = \bar{\ell}_2(\delta, n, d) := \frac{2\bar{\theta}e^2d}{n} \vee \log(p^d(n+1)^{2d}/\delta)$$

$$\bar{\ell}(\delta) = \bar{\ell}(\delta, n, d) := \bar{\ell}_1(\delta, n, d) \vee \bar{\ell}_2(\delta, n, d)$$

where $\bar{\theta}$ is the constant in Assumption 2.1. Note that we have $\bar{\ell}(\delta) \leq O(d \log(np/\delta)/n)$.

For two values $a, b > 0$, we write $a \lesssim b$ if there is a universal constant $C > 0$ such that $a \leq Cb$. We write $a \lesssim_{r} b$ if there is a constant $C_r$ that only depends on $r$ such that $a \leq C_r b$.

A.2 Technical lemmas

Now we can introduce the major technical results to establish the error bound.

Lemma A.1 Suppose Assumption [2,7] holds true. Suppose $\bar{\ell}_2(\delta/12) < 3/4$. Then with probability at least $1 - \delta$, it holds

$$\sup_{A \in \mathcal{A}_{p,d}} \sqrt{\mathbb{P}(X \in A)} \left| \mathbb{E}(f^*(X)|X \in A) - \bar{y}_{X \in A} \right| \leq 20U \sqrt{\bar{\ell}(\delta/12/12)}$$

(23)

The proof of Lemma A.1 is presented in Section [A.1.1]. Note that Lemma A.1 provides a uniform bound on the gap between the populational mean $\mathbb{E}(f^*(X)|X \in A)$ and the sample mean $\bar{y}_{X \in A}$. This is used to derive the geometric decrease of the bias, using the SID assumption.

Lemma A.2 Suppose Assumption [2,7] holds true. Given any $\delta \in (0,1)$, suppose $\bar{\ell}_2(\delta/4) < 3/4$. Then with probability at least $1 - \delta$ it holds

$$\sup_{A \in \mathcal{A}_{p,d}} \left| \sqrt{\mathbb{P}(X \in A)} - \sqrt{|A|/n} \right| \leq 5\sqrt{\bar{\ell}(\delta/4)}$$

(24)

The proof of Lemma A.2 is presented in Section [A.3]. Lemma A.2 provides a uniform deviation gap between the square root of probability and sample frequency over all sets in $\mathcal{A}_{p,d}$. Note that this uniform bound is stronger than a result without a square root (which can be obtained easily via Hoeffding’s inequality and a union bound), and is useful to prove the final error bound in Theorem 2.3.

For any rectangle $A \in \mathcal{A}$, $j \in [p]$ and $b \in \mathbb{R}$, define

$$\Delta_L(A, j, b) := \mathbb{P}(X \in A_L) \left( \mathbb{E}(f^*(X)|X \in A) - \mathbb{E}(f^*(X)|X \in A_L) \right)^2$$

$$\Delta_R(A, j, b) := \mathbb{P}(X \in A_R) \left( \mathbb{E}(f^*(X)|X \in A) - \mathbb{E}(f^*(X)|X \in A_R) \right)^2$$

$$\tilde{\Delta}_L(A, j, b) := \frac{|A_L|}{n} \left( \bar{y}_{X \in A_L} - \bar{y}_{X \in A} \right)^2$$

$$\tilde{\Delta}_R(A, j, b) := \frac{|A_R|}{n} \left( \bar{y}_{X \in A_R} - \bar{y}_{X \in A} \right)^2$$

We have the following identity regarding the impurity decrease of each split.

Lemma A.3 For any rectangle $A \in \mathcal{A}$, $j \in [p]$ and $b \in \mathbb{R}$, it holds

$$\Delta(A, j, b) = \Delta_L(A, j, b) + \Delta_R(A, j, b)$$

$$\tilde{\Delta}(A, j, b) = \tilde{\Delta}_L(A, j, b) + \tilde{\Delta}_R(A, j, b)$$

(25)
Proof. We just present the proof of the second equality. The proof of the first equality can be proved similarly.

Note that

\[
\hat{\Delta}(A, j, b) = \frac{1}{n} \sum_{i \in I_A} (y_i - \bar{y}_A)^2 - \frac{1}{n} \sum_{i \in I_{AL}} (y_i - \bar{y}_{AL})^2 - \frac{1}{n} \sum_{i \in I_{AR}} (y_i - \bar{y}_{AR})^2
\]

\[
= \frac{1}{n} \sum_{i \in I_{AL}} [(y_i - \bar{y}_A)^2 - (y_i - \bar{y}_{AL})^2] + \frac{1}{n} \sum_{i \in I_{AR}} [(y_i - \bar{y}_A)^2 - (y_i - \bar{y}_{AR})^2]
\]

(26)

For the first term, we have

\[
\frac{1}{n} \sum_{i \in I_{AL}} [(y_i - \bar{y}_A)^2 - (y_i - \bar{y}_{AL})^2] = \Delta_L(A, j, b)
\]

(27)

Similarly, we have

\[
\hat{\Delta}(A, j, b) = \frac{1}{n} \sum_{i \in I_{AR}} [(y_i - \bar{y}_A)^2 - (y_i - \bar{y}_{AR})^2] = \Delta_R(A, j, b)
\]

(28)

The proof is complete by combining (26), (27) and (28).

Lemma A.4 Suppose Assumption 2.1 holds true. Given a constant \(\alpha > 0\). Given any \(\delta \in (0, 1)\), suppose \(\bar{I} \delta /36 < 3/4\). Then with probability at least \(1 - \delta\), it holds

\[
\Delta(A, j, b) \leq (1 + \alpha)\Delta(A, j, b) + (1 + 1/\alpha) \cdot 5000U^2 \bar{t}(\delta/36) \quad \forall A \in A_{p, d-1}, j \in [p], b \in R
\]

(29)

and

\[
\hat{\Delta}(A, j, b) \leq (1 + \alpha)\Delta(A, j, b) + (1 + 1/\alpha) \cdot 5000U^2 \bar{t}(\delta/36) \quad \forall A \in A_{p, d-1}, j \in [p], b \in R
\]

(30)

Proof. For \(A \in A_{p, d-1}, j \in [p]\) and \(a \in R\), by Lemma A.3 we have

\[
\Delta(A, j, b) = \Delta_L(A, j, b) + \Delta_R(A, j, b)
\]

(31)

Define the events \(E_1\) and \(E_2\):

\[
E_1 := \left\{ \sup_{A \in A_{p, d}} \sqrt{P(X \in A)} |E^*(X)| X \in A - \bar{y}_A | \leq 20U \sqrt{\bar{t}(\delta/36)} \right\}
\]

\[
E_2 := \left\{ \sup_{A \in A_{p, d}} \sqrt{P(X \in A)} - \sqrt{|A|/n} \leq 5\sqrt{\bar{t}(\delta/12)} \right\}
\]

Then by Lemmas A.1 and A.2 we have \(P(E_i) \geq 1 - \delta/3\) for \(i = 1, 2\), so we have \(P(\cap_{i=1}^2 E_i) \geq 1 - \delta\). Below we prove (29) and (30) conditioned on the events \(E_1\) and \(E_2\).

Note that

\[
\sqrt{\Delta_L(A, j, a)} = \sqrt{P(X \in A)} |E(f^*(X))| X \in A - \bar{y}_A | \leq 20U \sqrt{\bar{t}(\delta/36)}
\]

(32)

To bound \(J_1\), we have

\[
J_1 \leq \sqrt{P(X \in A)} |E(f^*(X))| X \in A - \bar{y}_A | \leq 20U \sqrt{\bar{t}(\delta/36)}
\]

(33)

where the second inequality is by event \(E_1\). Similarly, to bound \(J_3\), we have

\[
J_3 = \sqrt{P(X \in A)} |\bar{y}_{AL} - E(f^*(X))| X \in A_L | \leq 20U \sqrt{\bar{t}(\delta/36)}
\]

(34)
To bound $J_2$, note that
\[
J_2 \leq \left| \frac{\mathbb{P}(X \in A_L)}{\sqrt{n} \Delta L} \right| \cdot |\bar{y}_L - \bar{y}_{L_A}| + \sqrt{\frac{\Delta L}{n}} \cdot |\bar{y}_L - \bar{y}_{L_A}| \tag{35}
\]
where the second inequality made use of the event $E_2$. Combining (32) – (35), we have
\[
\sqrt{\Delta L(A, j, b)} \leq 40U \sqrt{\tilde{t}(\delta/36)} + 10U \sqrt{\tilde{t}(\delta/12)} + \sqrt{\frac{\Delta L}{n}} \cdot |\bar{y}_L - \bar{y}_{L_A}| \leq 50U \sqrt{\tilde{t}(\delta/36)} + \sqrt{\frac{\Delta L}{n}} \cdot |\bar{y}_L - \bar{y}_{L_A}|
\]
which implies (by Young’s inequality)
\[
\Delta L(A, j, a) \leq (1 + 1/\alpha) \cdot 2500U^2 \tilde{t}(\delta/36) + (1 + \alpha) \frac{\Delta L}{n} \cdot |\bar{y}_L - \bar{y}_{L_A}|^2 \tag{36}
\]
By a similar argument, we have
\[
\Delta L(A, j, a) \leq (1 + 1/\alpha) \cdot 5000U^2 \tilde{t}(\delta/36) + (1 + \alpha) \Delta L(A, j, a)
\]
Summing up (36) and (37), and by (31), we have
\[
\Delta L(A, j, a) \leq (1 + 1/\alpha) \cdot 5000U^2 \tilde{t}(\delta/36) + (1 + \alpha) \Delta L(A, j, a)
\]
This completes the proof of (29). The proof of (30) is by a similar argument.

\[\square\]

Lemma \textbf{A.4} provides upper bounds between $\Delta L(A, j, a)$ and $\Delta L(A, j, b)$, which serves as a link to translate the population impurity decrease to sample impurity decrease. With all these technical lemmas at hand, we are ready to present the proof Theorem \textbf{2.3} as shown in the next subsection.

\subsection*{A.3 Completing the proof of Theorem \textbf{2.3}}

Define events
\[
E_1 := \left\{ \sup_{A \in A_{p,d}} \sqrt{\frac{\mathbb{P}(X \in A)}{\mathbb{E}[f^*(X)|X \in A]}} \right\} \leq 20U \sqrt{\frac{\tilde{t}(\delta/24)}}
\]
\[
E_2 := \left\{ \Delta(A, j, a) \leq (1 + 1/\alpha) \tilde{\Delta}(A, j, a) + (1 + 1/\alpha) \cdot 5000U^2 \tilde{t}(\delta/72) \quad \forall A \in A_{p,d-1}, j \in [p], a \in \mathbb{R} \right\}
\]
\[
E_3 := \left\{ \tilde{\Delta}(A, j, a) \leq (1 + 1/\alpha) \Delta(A, j, a) + (1 + 1/\alpha) \cdot 5000U^2 \tilde{t}(\delta/72) \quad \forall A \in A_{p,d-1}, j \in [p], a \in \mathbb{R} \right\}
\]
Then by Lemmas \textbf{A.1} and \textbf{A.4} and note that from the statement of Theorem \textbf{2.3}, $l_2(\delta/72) < 3/4$, so we have
\[
\mathbb{P}(E_1) \geq 1 - \delta/2 \quad \text{and} \quad \mathbb{P}(E_2 \cup E_3) \geq 1 - \delta/2,
\]
which implies $\mathbb{P}(U_{t=1}^T \mathcal{E}_t) \geq 1 - \delta$. In the following, we prove (10) using a deterministic argument conditioned on $U_{t=1}^T \mathcal{E}_t$.

For any $k \in [d]$ and any leave node $t$ of $\hat{f}(k)$ (recall that $\hat{f}(k)$ is the decision tree by CART with depth $k$), let $A_t^{(k)}$ be the corresponding cube, that is, for any $x \in \mathbb{R}^p, x \in A_t^{(k)}$ if and only if $x$ is routed to $t$ in $\hat{f}(k)$. Let $\mathcal{L}^{(k)}$ be the set of all leave nodes of $\hat{f}(k)$. Then we have
\[
\hat{f}(k)(x) := \sum_{t \in \mathcal{L}^{(k)}} \bar{y}_t A_t^{(k)} 1_{x \in A_t^{(k)}} \tag{38}
\]
Define a function
\[
\bar{f}(k)(x) := \sum_{t \in \mathcal{L}^{(k)}} \mathbb{E}\left[f^*(X)|X \in A_t^{(k)}, X_t^n\right] 1_{x \in A_t^{(k)}} \tag{39}
\]
where $X_t^n$ is the set of iid random variables $\{x_1, ..., x_n\}$, and $X$ is a random variable having the same distribution as $x_1$ but independent of $X_t^n$. In other words, $\hat{f}(k)$ is a tree with the same splitting structure as $\bar{f}(k)$ and replaces the prediction value of each leave node as the populational conditional mean of $f^*$.

First, using Cauchy-Schwarz inequality, we have
\[
||\hat{f}(k) - f^*||^2_{L_2(X)} \leq 2||f^* - \bar{f}(k)||^2_{L_2(X)} + 2||\bar{f}(k) - \bar{f}(k)||^2_{L_2(X)} := 2J_1(k) + 2J_2(k) \tag{40}
\]
To bound $J_1(d)$, we derive recursive inequalities between $J_1(k)$ and $J_1(k+1)$ for all $0 \leq k \leq d-1$. Note that
\[
J_1(k) = \mathbb{E}\left[ (f^*(X) - \bar{f}(k)(X))^2 | X_t^n\right] = \sum_{t \in \mathcal{L}^{(k)}} \mathbb{P}(X \in A_t^{(k)}) \cdot \text{Var}(f^*(X)|X \in A_t^{(k)}, X_t^n) \tag{41}
\]
For each \( t \in \mathcal{L}^{(k)} \), let \( t_L \) and \( t_R \) be the two children of \( t \), then we have
\[
\mathbb{P}(X \in A_t | \mathcal{X}_1^n) \cdot \text{Var}(f^*(X) | X \in A_t, \mathcal{X}_1^n) \\
= \mathbb{P}(X \in A_{t_L} | \mathcal{X}_1^n) \cdot \text{Var}(f^*(X) | X \in A_{t_L}, \mathcal{X}_1^n) + \mathbb{P}(X \in A_{t_R} | \mathcal{X}_1^n) \cdot \text{Var}(f^*(X) | X \in A_{t_R}, \mathcal{X}_1^n) + \Delta(A_t, \hat{J}_t, \hat{b}_t)
\]
where
\[
(\hat{J}_t, \hat{b}_t) \in \arg\max_{j \in [p], b \in R} \Delta(A_t, j, b)
\]
Let us define
\[
(\hat{J}_t, \hat{b}_t) \in \arg\max_{j \in [p], b \in R} \Delta(A_t, j, b)
\]
Then we have
\[
\Delta(A_t, \hat{J}_t, \hat{b}_t) \geq \frac{1}{1 + \alpha} \Delta(A_t, \hat{J}_t, \hat{b}_t) - (5000/\alpha)U^2i(\delta/72)
\]
\[
\geq \frac{1}{1 + \alpha} \Delta(A_t, \hat{J}_t, \hat{b}_t) - (5000/\alpha)U^2i(\delta/72)
\]
\[
\geq \frac{1}{(1 + \alpha)^2} \Delta(A_t, \hat{J}_t, \hat{b}_t) - \frac{2 + \alpha}{\alpha(1 + \alpha)} 5000U^2i(\delta/72)
\]
where the first inequality is by event \( \mathcal{E}_3 \), the second inequality is by the definition of \( (\hat{J}_t, \hat{b}_t) \), and the third inequality is because of event \( \mathcal{E}_2 \). By Assumption 2.2, we have
\[
\Delta(A_t, j, b) = \sup_{j \in [p], b \in R} \Delta(A_t, j, b) \geq \lambda \cdot \mathbb{P}(X \in A_t | \mathcal{X}_1^n) \text{Var}(f^*(X) | X \in A_t, \mathcal{X}_1^n)
\]
Combining (42), (43) and (44), we have
\[
\mathbb{P}(X \in A_{t_L} | \mathcal{X}_1^n) \cdot \text{Var}(f^*(X) | X \in A_{t_L}, \mathcal{X}_1^n) + \mathbb{P}(X \in A_{t_R} | \mathcal{X}_1^n) \cdot \text{Var}(f^*(X) | X \in A_{t_R}, \mathcal{X}_1^n)
\]
\[
\leq \left(1 - \frac{\lambda}{(1 + \alpha)^2}\right) \mathbb{P}(X \in A_t | \mathcal{X}_1^n) \cdot \text{Var}(f^*(X) | X \in A_t, \mathcal{X}_1^n) + \frac{2 + \alpha}{\alpha(1 + \alpha)} 5000U^2i(\delta/72)
\]
Summing up the inequality above for all \( t \in \mathcal{L}^{(k)} \), we have
\[
J_1(k + 1) \leq \left(1 - \frac{\lambda}{(1 + \alpha)^2}\right) J_1(k) + 2^k \cdot \frac{2 + \alpha}{\alpha(1 + \alpha)} 5000U^2i(\delta/72)
\]
Using the inequality above recursively for \( k = 0, 1, \ldots, d - 1 \), we have
\[
J_1(d) \leq \left(1 - \frac{\lambda}{(1 + \alpha)^2}\right)^d J_1(0) + \frac{2 + \alpha}{\alpha(1 + \alpha)} 5000U^2i(\delta/72) \sum_{k=1}^{d} 2^{k-1} \leq \left(1 - \frac{\lambda}{(1 + \alpha)^2}\right)^d \text{Var}(f^*(X)) + 2^d \cdot \frac{2 + \alpha}{\alpha(1 + \alpha)} 5000U^2i(\delta/72)
\]
To bound \( J_2(d) \), we have
\[
J_2(d) = \sum_{t \in \mathcal{L}^{(d)}} \mathbb{P}(X \in A_t) \left( \mathbb{E}(f^*(X) | X \in A_t, \mathcal{X}_1^n) - \hat{g}_{\mathcal{A}_t} \right)^2 \leq 2^{2d} \cdot 400U^2i(\delta/24)
\]
where the inequality made use of event \( \mathcal{E}_1 \). Using (45) and (46), and recalling (40), we have
\[
\|\hat{f}^{(k)} - f^*\|_{L^2(X)}^2 \leq 2 \left(1 - \frac{\lambda}{(1 + \alpha)^2}\right)^d \text{Var}(f^*(X)) + 2^{d+1} \cdot \frac{2 + \alpha}{\alpha(1 + \alpha)} 5000U^2i(\delta/72) + 2^{d+1} \cdot 400U^2i(\delta/24)
\]
\[
\leq \text{Var}(f^*(X)) \cdot (1 - \lambda/(1 + \alpha)^2)^d + \frac{2 + \alpha}{\alpha(1 + \alpha)} \frac{2^d(d \log(np) + \log(1/\delta))}{n} U^2
\]
\[
\leq \text{Var}(f^*(X)) \cdot (1 - \lambda/(1 + \alpha)^2)^d + \frac{2^d(d \log(np) + \log(1/\delta))}{\alpha n} U^2
\]
This completes the proof of (10). To prove (11), by taking \( \alpha = 1/d \) and \( d = [\log_2(n)/(1 - \log_2(1 - \lambda))] \), we have
\[
\left(1 - \frac{\lambda}{(1 + \alpha)^2}\right)^d = (1 - \lambda)^d \left(1 + \frac{\lambda}{1 - \lambda} (1 - (1 + \alpha)^{-2})\right)^d
\]
\[
= (1 - \lambda)^d \left(1 + \frac{2/d + 1/d^2}{1 - \lambda (1 + d)}\right)^d \leq (1 - \lambda)^d
\]
Note that for \( s = \log_2(n)/(1 - \log_2(1 - \lambda)) \) we have \((1 - \lambda)^s = 2^s/n\), hence by taking \( d = \lceil \log_2(n)/(1 - \log_2(1 - \lambda)) \rceil \), we have

\[
(1 - \lambda)^d \leq \frac{\lambda^d}{n} \leq 2n^{-1 + \frac{1}{\log_2(1 - \lambda)}} = 2n^{-\phi(\lambda)}.
\]

Combining (47), (48) and (49) and note that \( \text{Var}(f^*(X)) \leq M < U \), we have

\[
\|f(k) - f^*\|_{\ell^2(X)} \leq \lambda, U \leq \frac{\lambda}{2n} d \log(np) + d \log(1/\delta)
\]

\[
\leq \lambda, U \leq \frac{\lambda}{2n} d \log(np) + \log(n) \log(1/\delta)
\]

this completes the proof of (11).

### A.4 Proof of Lemma A.1

The main idea of proving Lemma A.1 is to find a proper finite net of the set \( A \), control the gap on this net, and finally prove the result for all \( A \in A_{p,d} \) based on the approximation gap of the net. We need a few auxiliary results. Let \( S := \{0, 1/n, 2/n, \ldots, (n - 1)/n, 1\} \), and define

\[
\tilde{A}_{p,d} := \left\{ \prod_{j=1}^p [\ell_j, u_j] \in A_{p,d} \mid \ell_j, u_j \in S \text{ for all } j \in [p] \right\}
\]

For any \( A = \prod_{j=1}^p [\ell_j, u_j] \in A_{p,d} \), define

\[
A' = \prod_{j=1}^p [\ell'_j, u'_j]
\]

where \( \ell'_j := \max \{ s \in S \mid s \leq \ell_j \} \), and \( u'_j := \min \{ s \in S \mid s \geq u_j \} \). Roughly speaking, \( A' \) is the smallest box with all edges in \( S \) that contains \( A \). For any \( A = \prod_{j=1}^p [\ell_j, u_j] \in \tilde{A}_{p,d} \) with \( \tilde{u}_j - \tilde{\ell}_j \geq 2/n \) for all \( j \in [p] \), define

\[
B(\tilde{A}) := \tilde{A} \setminus \prod_{j=1}^p \left[ \tilde{\ell}_j + (1/n) \cdot 1_{\{\tilde{\ell}_j \neq 0\}}, \tilde{u}_j - (1/n) \cdot 1_{\{\tilde{u}_j \neq 1\}} \right].
\]

and define \( B_{p,d} \) to be the set of all such sets, that is

\[
B_{p,d} := \left\{ B(\tilde{A}) \mid \tilde{A} = \prod_{j=1}^p [\ell_j, u_j] \in \tilde{A}_{p,d} \text{ with } \tilde{u}_j - \tilde{\ell}_j \geq 2/n \right\}
\]

The following lemma can be easily verified from the definitions of \( \tilde{A}_{p,d} \) and \( B_{p,d} \).

**Lemma A.5** (1) For any \( A \in A_{p,d} \), there exists \( B \in B_{p,d} \) such that \( A' \setminus A \subseteq B \).

(2) \( \mathbb{P}(X \in B) \leq 2\delta d^d/n \) for all \( B \in B_{p,d} \).

(3) The cardinality

\[
|B_{p,d}| \leq |\tilde{A}_{p,d}| \leq \binom{p}{d}(n + 1)^{2d} \leq p^d(n + 1)^{2d}
\]

Finally, for any \( t \geq 0 \), we define

\[
A_{p,d}(t) := \left\{ A \in A_{p,d} \mid \mathbb{P}(X \in A) \leq t \right\}, \quad \text{and} \quad \tilde{A}_{p,d}(t) := \left\{ A \in \tilde{A}_{p,d} \mid \mathbb{P}(X \in A) \leq t \right\}
\]

**Lemma A.6** Suppose Assumption 2.1 holds true. Let \( z_1, \ldots, z_n \) be i.i.d. bounded random variables with \( |z_i| \leq V < \infty \) almost surely. Assume that for each \( i \in [n] \), \( z_i \) is independent of \( \{x_j\}_{j \neq i} \), but may be dependent on \( x_i \). Given any \( \delta \in (0, 1) \), with probability at least \( 1 - \delta \), it holds

\[
\max_{A \in \tilde{A}_{p,d} \setminus \tilde{A}_{p,d}(t_1)} \frac{1}{\mathbb{P}(X \in A)} \left| \frac{1}{n} \sum_{i=1}^{n} z_i 1_{\{x_i \in A\}} - \mathbb{E}(1_{\{x_1 \in A\}}) \right| \leq 2V \sqrt{t_1(\delta)}
\]

where \( U = M + n \).

**Proof.** For each fixed \( A \in \tilde{A}_{p,d} \setminus \tilde{A}_{p,d}(t_1(\delta)) \), note that

\[
\mathbb{E}\left[ \left( z_1 1_{\{x_1 \in A\}} - \mathbb{E}(z_1 1_{\{x_1 \in A\}}) \right)^k \right] \leq (2V)^k \mathbb{P}(X \in A) \ \forall \ k \geq 2
\]
so by Lemma [D.1] with $t = 2V \sqrt{\mathbb{P}(X \in A) \sqrt{\ell_1(\delta)}}$, $\gamma^2 = (2V)^2 \mathbb{P}(X \in A)$ and $b = 2V$, we have

$$
\mathbb{P} \left( \frac{1}{n} \sum_{i=1}^{n} z_{i1_{\{x_i \in A\}}} - \mathbb{E}(z_{i1_{\{x_i \in A\}}}) > 2V \sqrt{\ell_1(\delta)} \right) 
\leq \exp \left( -\frac{n}{4} \left( 4V^2 \mathbb{P}(X \in A) \ell_1(\delta) \right) \right) 
\leq \exp \left( -\frac{n}{4} \ell_1(\delta) \right) = \delta / (p^d(n+1)^{2d})
$$

where $(i)$ is because $\mathbb{P}(X \in A) \geq \ell_1(\delta)$ (since $A \in A_{p,d} \setminus \bar{A}_{p,d}(\ell_1(\delta))$). As a result, we have

$$
\mathbb{P} \left( \max_{A \in A_{p,d} \setminus \bar{A}_{p,d}(\ell_1(\delta))} \frac{1}{n} \sum_{i=1}^{n} z_{i1_{\{x_i \in A\}}} - \mathbb{E}(z_{i1_{\{x_i \in A\}}}) > 2V \sqrt{\ell_1(\delta)} \right) 
\leq \sum_{A \in A_{p,d} \setminus \bar{A}_{p,d}(\ell_1(\delta))} \mathbb{P} \left( \frac{1}{n} \sum_{i=1}^{n} z_{i1_{\{x_i \in A\}}} - \mathbb{E}(z_{i1_{\{x_i \in A\}}}) > 2V \sqrt{\ell_1(\delta)} \right) 
\leq |\bar{A}_{p,d} \setminus \bar{A}_{p,d}(\ell_1(\delta))| \cdot \frac{\delta}{(p^d(n+1)^{2d})} \leq \delta
$$

where the last inequality makes use of Lemma [A.5](3).

\[\square\]

**Lemma A.7** Let $D$ be a finite collection of measurable subsets of $[0, 1]^p$ satisfying $\mathbb{P}(X \in D) \leq \bar{\alpha}$ for all $D \in D$ (for some constant $\alpha \in (0, 1)$). Given any $\delta \in (0, 1)$, if

$$
w(\bar{\alpha}, \delta) := (e^2 \bar{\alpha}) \sqrt{\log \left( \frac{|D|}{\delta} \right)} \leq 3/4
$$

then with probability at least $1 - \delta$ it holds

$$
\max_{D \in D} \left\{ \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in D\}} \right\} \leq w(\bar{\alpha}, \delta)
$$

**Proof.** For any fixed $D \in D$, denote $\alpha = \mathbb{P}(X \in D)$, then by Lemma [D.2] for any $t \in (0, 3/4]$, we have

$$
\mathbb{P} \left( \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in D\}} > t \right) \leq \exp \left( -n \left( t \log(t/\alpha) + (1-t) \log \left( \frac{1-t}{1-\alpha} \right) \right) \right)
\leq \exp \left( -n \left( t \log(t/\alpha) + (1-t) \log(1-t) \right) \right)
\leq \exp \left( -n \left( t \log(t/\alpha) + (1-t)(1-t) \right) \right)
= \exp \left( -n \left( t \log(t/\alpha) - 1 \right) \right)
\leq \exp \left( -nt \log(t/\alpha) \right)
$$

where the third inequality makes use of Lemma [D.3] and the assumption $t \leq 3/4$. Take $t = w(\bar{\alpha}, \delta)$, and note that

$$
\log(w(\bar{\alpha}, \delta)/\alpha) - 1 \geq \log(w(\bar{\alpha}, \delta)/\bar{\alpha}) - 1 \geq \log(e^2) - 1 \geq 1
$$

we have

$$
\mathbb{P} \left( \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in D\}} > w(\bar{\alpha}, \delta) \right) \leq \exp \left( -nw(\bar{\alpha}, \delta) \right) \leq \delta / |D|
$$

where the last inequality is because of the definition of $w(\bar{\alpha}, \delta)$. Taking the union bound we have

$$
\mathbb{P} \left( \max_{D \in D} \left\{ \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in D\}} \right\} > w(\bar{\alpha}, \delta) \right) \leq |D| \cdot \delta / |D| = \delta
$$

\[\square\]

**Corollary A.8** Under Assumption [2.7] and given $\delta \in (0, 1)$, suppose $\tilde{\ell}_2(\delta) < 3/4$, then with probability at least $1 - \delta$, it holds

$$
\max_{B \in B_{p,d}} \left\{ \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in B\}} \right\} \leq \tilde{\ell}_2(\delta)
$$
Proof. Apply Lemma A.7 with $D = B_{p,d}$ and $\bar{\alpha} = 2\bar{d}/n$, and note that $|B_{p,d}| \leq (n + 1)^{2d}/n$ (by Lemma A.5 (3)) and the definition $\bar{t}_1(\delta) = \frac{2\bar{d}^2d}{n} + \log(p(n + 1)^{2d}/\delta)$.

**Lemma A.9** Suppose Assumption [A.7] holds true. Let $z_1, \ldots, z_n$ be i.i.d. bounded random variables with $|Z| \leq V < \infty$ almost surely. Assume that for each $i \in [n]$, $z_i$ is independent of $(x_j)_{j \neq i}$, but may be dependent on $x_i$. Given any $\delta \in (0, 1)$, suppose $\bar{t}_2(\delta/2) < 3/4$, then with probability at least $1 - \delta$, it holds

$$\sup_{A \in A_{p,d} \setminus A_{p,d}(\bar{t}_1(\delta/2))} \frac{1}{\sqrt{\mathbb{P}(X \in A)}} \left| \frac{1}{n} \sum_{i=1}^{n} z_i 1_{\{x_i \in A\}} - \mathbb{E}(z_1 1_{\{x_1 \in A\}}) \right| \leq 5\sqrt{V \bar{t}_2(\delta/2)}. \quad (50)$$

**Proof.** Define events $\mathcal{E}_1$ and $\mathcal{E}_2$:

$$\mathcal{E}_1 := \left\{ \max_{B \in B_{p,d}} \left\{ \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in B\}} \right\} \leq \bar{t}_2(\delta/2) \right\}$$

$$\mathcal{E}_2 := \left\{ \max_{A \in \bar{A}_{p,d} \setminus A_{p,d}(\bar{t}_1(\delta/2))} \frac{1}{\sqrt{\mathbb{P}(X \in A)}} \left| \frac{1}{n} \sum_{i=1}^{n} z_i 1_{\{x_i \in A\}} - \mathbb{E}(z_1 1_{\{x_1 \in A\}}) \right| \leq V \sqrt{\bar{t}_1(\delta/2)} \right\}$$

Then by Lemma A.6 and Corollary A.8 we have $\mathbb{P}(\mathcal{E}_1) \geq 1 - \delta/2$ and $\mathbb{P}(\mathcal{E}_2) \geq 1 - \delta/2$, hence $\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) \geq 1 - \delta$. Below we prove that when $\mathcal{E}_1$ and $\mathcal{E}_2$ hold true, inequality (50) holds true.

Note that for any $A \in A_{p,d} \setminus A_{p,d}(\bar{t}_1(\delta/2))$,

$$\frac{1}{\sqrt{\mathbb{P}(X \in A)}} \left| \frac{1}{n} \sum_{i=1}^{n} z_i 1_{\{x_i \in A\}} - \mathbb{E}(z_1 1_{\{x_1 \in A\}}) \right| \leq \frac{1}{\sqrt{\mathbb{P}(X \in A)}} \left| \frac{1}{n} \sum_{i=1}^{n} z_i 1_{\{x_i \in A\}} - 1 \frac{1}{n} \sum_{i=1}^{n} z_i 1_{\{x_i \in A'\}} \right|$$

$$+ \frac{1}{\sqrt{\mathbb{P}(X \in A)}} \left| \frac{1}{n} \sum_{i=1}^{n} z_i 1_{\{x_i \in A'\}} - \mathbb{E}(z_1 1_{\{x_1 \in A'\}}) \right|$$

$$+ \frac{1}{\sqrt{\mathbb{P}(X \in A)}} \left| \mathbb{E}(z_1 1_{\{x_1 \in A'\}}) - \mathbb{E}(z_1 1_{\{x_1 \in A\}}) \right|$$

$$:= T_1 + T_2 + T_3$$

To bound $T_1$, we have

$$T_1 = \frac{1}{\sqrt{\mathbb{P}(X \in A)}} \left| \frac{1}{n} \sum_{i=1}^{n} z_i 1_{\{x_i \in A \setminus A'\}} \right| \leq \frac{V}{\sqrt{\mathbb{P}(X \in A)}} \left( \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in A \setminus A'\}} \right)$$

$$\leq \frac{V}{\sqrt{\mathbb{P}(X \in A)}} \max_{B \in B_{p,d}} \left\{ \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in B\}} \right\} \leq \frac{V}{\sqrt{\mathbb{P}(X \in A)}} \bar{t}_2(\delta/2) \leq V \sqrt{\bar{t}_2(\delta/2)} \quad (52)$$

where the second inequality is because $A' \in \bar{A}_{p,d}$ and $\mathbb{P}(X \in A') \geq \mathbb{P}(X \in A) \geq \bar{t}_1(\delta/2)$. The third inequality is by inequality (50). The third inequality is by $\mathcal{E}_1$.

To bound $T_2$, note that

$$T_2 = \sqrt{\frac{\mathbb{P}(X \in A')}{\mathbb{P}(X \in A)}} \sqrt{\frac{1}{\mathbb{P}(X \in A')}} \left| \frac{1}{n} \sum_{i=1}^{n} z_i 1_{\{x_i \in A'\}} - \mathbb{E}(z_1 1_{\{x_1 \in A'\}}) \right|$$

$$\leq \sqrt{\frac{\mathbb{P}(X \in A')}{\mathbb{P}(X \in A)}} \frac{2V}{\sqrt{\bar{t}_1(\delta/2)}} \quad (53)$$

where the inequality is by event $\mathcal{E}_2$ and because $A' \in \bar{A}_{p,d}$ and $\mathbb{P}(X \in A') \geq \mathbb{P}(X \in A) \geq \bar{t}_1(\delta/2)$. Note that

$$\mathbb{P}(X \in A') \leq \frac{2\bar{d}d}{n} \leq \bar{t}_2(\delta/2) \leq \mathbb{P}(X \in A) \quad (54)$$

where the first inequality is by Lemma A.5 (2); the second inequality is by the definition of $\bar{t}_2(\delta/2)$ in (22); the third inequality is because $A \in A_{p,d} \setminus A_{p,d}(\bar{t}_1(\delta/2))$. As a result of (53) and (54), we have

$$T_2 \leq \sqrt{\frac{\mathbb{P}(X \in A')}{\mathbb{P}(X \in A)}} \mathbb{P}(X \in A') \frac{2V}{\sqrt{\bar{t}_1(\delta/2)}} \leq 2\sqrt{2V \sqrt{\bar{t}_1(\delta/2)}} \quad (55)$$
To bound $T_3$, note that
\[
T_3 = \frac{1}{\sqrt{\mathbb{P}(X \in A)}} \left| \mathbb{E}(z_1 \mathbf{1}_{\{x_1 \in A \setminus \Lambda\}}) \right| \leq \frac{V}{\sqrt{\mathbb{P}(X \in A)}} \mathbb{P}(X \in A') \setminus A \right)
\leq V \sqrt{\mathbb{P}(X \in A') \setminus A \setminus A} \leq V \sqrt{\frac{2\delta d}{n}} \leq V \sqrt{t_2(\delta/2)}
\]

The proof is complete by combining inequalities (51), (52), (55) and (56), and note that
\[
2V \sqrt{t_2(\delta/2)} + 2\sqrt{2V \sqrt{t_1(\delta/2)}} \leq 5V \sqrt{t(\delta/2)}.
\]

Now we are ready to wrap up the proof of Lemma A.1.

**Completing the proof of Lemma A.1**

Define events $\mathcal{E}_1$ and $\mathcal{E}_2$:
\[
\mathcal{E}_1 := \left\{ \sup_{A \in \mathcal{A}_{p,d} \setminus \mathcal{A}_{p,d}(\alpha)} \frac{1}{\mathbb{P}(X \in A)} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{x_i \in A\}} - \mathbb{P}(X \in A) \right| \leq 5\sqrt{t(\delta/8)} \right\}
\]
\[
\mathcal{E}_2 := \left\{ \sup_{A \in \mathcal{A}_{p,d} \setminus \mathcal{A}_{p,d}(\alpha)} \frac{1}{\mathbb{P}(X \in A)} \left| \frac{1}{n} \sum_{i=1}^{n} y_i \mathbf{1}_{\{x_i \in A\}} - \mathbb{E}(y_i \mathbf{1}_{\{x_i \in A\}}) \right| \leq 5U \sqrt{t(\delta/8)} \right\}
\]

Then by Lemma A.9 with $z_i = y_i$ and $z_i = 1$ respectively, we know that $\mathbb{P}(\mathcal{E}_i) \geq 1 - \delta/4$ for all $i = 1, 2$. So we know $\mathbb{P}(\cap_{i=1}^{2} \mathcal{E}_i) \geq 1 - \delta$. Below we prove that inequality (23) is true when $\cap_{i=1}^{2} \mathcal{E}_i$ hold.

Define $\alpha := 100\bar{t}(\delta/8)$. Then it holds
\[
\sup_{A \in \mathcal{A}_{p,d}(\alpha)} \mathbb{P}(X \in A) \left| \mathbb{E}(f^*(X)|X \in A) - \bar{y}X_A \right| \leq 2U \sqrt{\alpha} = 20U \sqrt{t(\delta/8)} \tag{57}
\]

On the other hand, for any $A \in \mathcal{A}_{p,d} \setminus \mathcal{A}_{p,d}(\alpha)$, by event $\mathcal{E}_1$, we have
\[
\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{x_i \in A\}} \geq \mathbb{P}(X \in A) - 5\sqrt{t(\delta/8)} \mathbb{P}(X \in A) \geq \frac{1}{2} \mathbb{P}(X \in A) \tag{58}
\]
where the second inequality is because $\mathbb{P}(X \in A) \geq \alpha = 100\bar{t}(\delta/8)$. Therefore we know $\sum_{i=1}^{n} \mathbf{1}_{\{x_i \in A\}} > 0$, and we can write
\[
\mathbb{E}(f^*(X)|X \in A) - \bar{y}X_A = \frac{\mathbb{E}(y_i \mathbf{1}_{\{x_i \in A\}})}{\mathbb{P}(X \in A)} - \frac{1}{n} \sum_{i=1}^{n} y_i \mathbf{1}_{\{x_i \in A\}} = \frac{1}{\mathbb{P}(X \in A)} \left( \mathbb{E}(y_i \mathbf{1}_{\{x_i \in A\}}) - \frac{1}{n} \sum_{i=1}^{n} y_i \mathbf{1}_{\{x_i \in A\}} \right)
\]
\[+ \sum_{i=1}^{n} y_i \mathbf{1}_{\{x_i \in A\}} \frac{1}{\mathbb{P}(X \in A)} - \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{x_i \in A\}} - \mathbb{P}(X \in A) \right) := H_1(A) + H_2(A) \tag{59}
\]

By event $\mathcal{E}_2$, and note that $\alpha \geq \bar{t}(\delta/8)$, we have
\[
\sup_{A \in \mathcal{A}_{p,d} \setminus \mathcal{A}_{p,d}(\alpha)} \mathbb{P}(X \in A) |H_1(A)| \leq 5U \sqrt{t(\delta/8)} \tag{60}
\]

By event $\mathcal{E}_1$, and note that $\alpha \geq \bar{t}(\delta/8)$, we have
\[
\sup_{A \in \mathcal{A}_{p,d} \setminus \mathcal{A}_{p,d}(\alpha)} \mathbb{P}(X \in A) |H_2(A)| \leq 5U \sqrt{t(\delta/8)} \tag{61}
\]

Combining (59), (60) and (61), we have
\[
\sup_{A \in \mathcal{A}_{p,d} \setminus \mathcal{A}_{p,d}(\alpha)} \mathbb{P}(X \in A) \left| \mathbb{E}(f^*(X)|X \in A) - \bar{y}X_A \right| \leq 10U \sqrt{t(\delta/8)} \tag{54}
\]

Combining the inequality above with (57) we have
\[
\sup_{A \in \mathcal{A}_{p,d}} \mathbb{P}(X \in A) \left| \mathbb{E}(f^*(X)|X \in A) - \bar{y}X_A \right| \leq 20U \sqrt{t(\delta/8)} \leq 20U \sqrt{t(\delta/12)}
\]
543 A.5 Proof of Lemma [A.2]

544 Define \( a := \bar{t}(\delta/4) \) and \( b := a + \frac{2\delta d}{n} \). Define events \( E_1 \) and \( E_2 \):

\[
E_1 := \left\{ \max_{A \in \mathcal{A}_{p,d}(b)} \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in A\}} \leq (\varepsilon^2 b) + \frac{\log (2(n+1)^2d\delta/n)}{n} \right\}
\]

\[
E_2 := \left\{ \sup_{A \in \mathcal{A}_{p,d}(a) \setminus \mathcal{A}_{p,d}(b)} \frac{1}{\sqrt{\mathbb{P}(X \in A)}} \left| \mathbb{P}(X \in A) - \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in A\}} \right| \leq 5\sqrt{\bar{t}(\delta/4)} \right\}
\]

Then by Lemmas [A.7] and [A.9] we know that \( \mathbb{P}(E_1) \geq 1 - \delta/2 \) and \( \mathbb{P}(E_2) \geq 1 - \delta/2 \), so \( \mathbb{P}(E_1 \cap E_2) \geq 1 - \delta \).

Below we prove \((24)\) when \( E_1 \cap E_2 \) holds.

547 For \( A \in \mathcal{A}_{p,d} \), if \( \mathbb{P}(X \in A) \leq a \), then \( \mathbb{P}(A') \leq a + \frac{2\delta d}{n} = b \). So we have

\[
\sup_{A \in \mathcal{A}_{p,d}(a)} \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in A\}} \leq \sup_{A \in \mathcal{A}_{p,d}(a)} \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in A'\}} \leq \sup_{A \in \mathcal{A}_{p,d}(a)} \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in A\}}
\]

\[
\leq \left( e^{2\bar{t}(\delta/4)} + 2e^2 \bar{b}d/n \right) + \frac{\log (2(n+1)^2d\delta/n)}{n}
\]

\[
\leq (e^{2} + 1)\bar{t}(\delta/4) \leq 25\bar{t}(\delta/4)
\]

where the third inequality is by event \( E_1 \) and the definition of \( b \); the fourth inequality is because

\[
\bar{t}(\delta/4) \geq \bar{t}_1(\delta/4) \geq \frac{1}{n} \log (2p^d(n+1)^2d\delta/n) \quad \text{and} \quad \bar{t}(\delta/4) \geq \bar{t}_2(\delta/4) \geq 2e^2 \bar{b}d/n.
\]

548 As a result, we have

\[
\sup_{A \in \mathcal{A}_{p,d}(a)} \left| \sqrt{\mathbb{P}(X \in A)} - \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in A\}} \right| \leq \sup_{A \in \mathcal{A}_{p,d}(a)} \max \left\{ \sqrt{\mathbb{P}(X \in A)}, \left| \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in A\}} \right| \right\} \leq 5\sqrt{\bar{t}(\delta/4)}
\]

where the second inequality made use of \((62)\).

549 On the other hand,

\[
\sup_{A \in \mathcal{A}_{p,d} \setminus \mathcal{A}_{p,d}(a)} \left| \sqrt{\mathbb{P}(X \in A)} - \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in A\}} \right| = \sup_{A \in \mathcal{A}_{p,d} \setminus \mathcal{A}_{p,d}(a)} \left| \frac{\mathbb{P}(X \in A) - \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in A\}}}{\sqrt{\mathbb{P}(X \in A)} + \sqrt{\frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in A\}}}} \right|
\]

\[
\leq \sup_{A \in \mathcal{A}_{p,d} \setminus \mathcal{A}_{p,d}(a)} \frac{1}{\sqrt{\mathbb{P}(X \in A)}} \left| \mathbb{P}(X \in A) - \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \in A\}} \right| \leq 5\sqrt{\bar{t}(\delta/4)}
\]

where the last inequality is by event \( E_2 \).

550 Combining \((63)\) and \((64)\) the proof is complete.
For any interval $E \in \mathcal{E}$ and any univariate function $g$ on $[0, 1]$, let $V_g(E)$ be the total variation of $g$ on $E$. For the additive model \cite{14} and a rectangle $A = \prod_{j=1}^{p} E_j \in \mathcal{A}$, define $V_{f^*}(A) = \sum_{j=1}^{p} V_{f^*_j}(E_j)$. Recall that $X$ is a random variable with the same distribution as $x_i$, and $X^{(j)}$ is the $j$-th coordinate of $X$.

### B Proofs in Section 3

#### B.1 Technical lemmas

**Lemma B.1** For any rectangle $A \subseteq [0, 1]^p$, any $j \in [p]$ and any $b \in \mathbb{R}$, it holds

\[
\Delta(A, j, b) = \left( \mathbb{E}(f^*(X)1_{X \in A}) - \mathbb{E}(f^*(X)|X \in A)\mathbb{P}(X \in A) \right)^2 \frac{\mathbb{P}(X \in A)}{\mathbb{P}(X \in A)\mathbb{P}(X \in A)}
\]

where $A_L = A_{L}(j, b)$ and $A_R = A_{R}(j, b)$.

**Proof.** We use the notations $\nu := \mathbb{E}(f^*(X)|X \in A), \nu_L := \mathbb{E}(f^*(X)|X \in A_L)$ and $\nu_R := \mathbb{E}(f^*(X)|X \in A_R)$. First, note that

\[
\mathbb{E} \left( (f^*(X) - \nu)^2 1_{X \in A_L} \right) = \mathbb{E} \left( (f^*(X) - \nu_L + \nu_L - \nu)^2 1_{X \in A_L} \right)
\]

Similarly, we have

\[
\mathbb{E} \left( (f^*(X) - \nu_R)^2 1_{X \in A_R} \right) + (\nu_R - \nu)^2 \mathbb{P}(X \in A_R)
\]

Summing up (65) and (66) we have

\[
(\nu_L - \nu)^2 \mathbb{P}(X \in A) + (\nu_R - \nu)^2 \mathbb{P}(X \in A_R)
\]

Noting that

\[
(\nu_L - \nu)^2 \mathbb{P}(X \in A_L) = \left( \mathbb{E}(f^*(X)1_{X \in A_L}) - \mathbb{E}(f^*(X)|X \in A_L)\mathbb{P}(X \in A_L) \right)^2 \mathbb{P}(X \in A_L)^{-1}
\]

Combining (67) and (68) we have

\[
\Delta(A, j, b) = (\nu_L - \nu)^2 \frac{\mathbb{P}(X \in A_L)^2}{\mathbb{P}(X \in A)} + (\nu_R - \nu)^2 \mathbb{P}(X \in A_R) = (\nu_R - \nu)^2 \mathbb{P}(X \in A \cap A_R) \frac{\mathbb{P}(X \in A)}{\mathbb{P}(X \in A_L)}
\]

**Lemma B.2** Suppose Assumption 2.7 holds true, and $f^*$ has the additive structure in (14). Then for any $A = \prod_{j=1}^{p} [\ell_j, u_j] \subseteq [0, 1]^p$, it holds

\[
\max_{j \in [p], k \in \mathbb{R}} \sqrt{\Delta(A, j, b)} \geq \frac{\sqrt{\mathbb{P}(X \in A)\text{Var}(f^*(X)|X \in A)}}{\sum_{k=1}^{p} \int_{\ell_k}^{u_k} \sqrt{q^{(b)}_A(t)(1 - q^{(b)}_A(t))} dV_{f^*}(\{\ell_j, t\})}
\]

where $q^{(b)}_A(t) := \mathbb{P}(X^{(k)} \leq t|x_i \in A)$.

**Proof.** For a fixed $A = \prod_{j=1}^{p} [\ell_j, u_j] \subseteq [0, 1]^p$, without loss of generality, assume $\mathbb{E}(f^*(X)|X \in A) = 0$.

Note that for any $j \in [p]$,

\[
\max_{b \in \mathbb{R}} \sqrt{\Delta(A, j, b)} \geq \frac{\int_{\ell_j}^{u_j} \sqrt{q^{(b)}_A(s)(1 - q^{(b)}_A(s))} dV_{f^*}(\{\ell_j, s\})}{\int_{\ell_j}^{u_j} \sqrt{q^{(b)}_A(s)(1 - q^{(b)}_A(s))} dV_{f^*}(\{\ell_j, s\})}
\]

(69)
where \( s \) is the integration variable. Because \( q_A^{(j)}(s) = \mathbb{P}(X \in A_L(j,s)) / \mathbb{P}(X \in A) \), using Lemma B.1 and recall that we have assumed \( \mathbb{E}(f^*(X)|X \in A) = 0 \), we have

\[
\int_{\ell_j}^{u_j} \sqrt{q_A^{(j)}(s)(1 - q_A^{(j)}(s))} \sqrt{\Delta(A, j, s)} \, dV_{f^*}([\ell_j, s])
\]

\[
= \frac{1}{\mathbb{P}(X \in A)} \int_{\ell_j}^{u_j} \left| \mathbb{E}(f^*(X)1_{X \in AR(j,s)}) \right| \, dV_{f^*}([\ell_j, s])
\]

\[
= \frac{1}{\mathbb{P}(X \in A)} \int_{\ell_j}^{u_j} \left| \mathbb{E}(f^*(X)1_{X(\cdot) > s}) \right| \, dV_{f^*}([\ell_j, s])
\]

\[
\geq \frac{1}{\mathbb{P}(X \in A)} \int_{\ell_j}^{u_j} \mathbb{E}(f^*(X)(f^*_j(X)^{(j)} - f^*_j(\ell_j))1_{X(\cdot) > s}) \, dV_{f^*}([\ell_j, s])
\]

\[
= \frac{1}{\mathbb{P}(X \in A)} \mathbb{E}(f^*(X)f^*_j(X)^{(j)}1_{X(\cdot) > s})
\]

where the last equality makes use of the assumption that \( \mathbb{E}(f^*(X)|X \in A) = 0 \). Combining the inequality above with (69), we have

\[
\max_{b \in \mathbb{R}} \sqrt{\Delta(A, j, b)} \geq \frac{1}{\mathbb{P}(X \in A)} \int_{\ell_j}^{u_j} \sqrt{q_A^{(j)}(s)(1 - q_A^{(j)}(s))} \, dV_{f^*}([\ell_j, s])
\]

As a result, we have

\[
\max_{j \in [p], b \in \mathbb{R}} \sqrt{\Delta(A, j, b)} \geq \frac{1}{\mathbb{P}(X \in A)} \sum_{j=1}^{p} \mathbb{E}(f^*(X)f^*_j(X)^{(j)}1_{X(\cdot) > s}) \geq \mathbb{P}(X \in A) \mathbb{E}(f^*(X)|X \in A)
\]

Combining (70) and (71), the proof is complete. \( \square \)

Lemma B.3 Suppose Assumption 2.1 holds true, and \( f^* \) has the additive structure in (13). If for any \( A = \prod_{j=1}^{p} [\ell_j, u_j] \subseteq [0, 1]^p \) and any \( k \in [p] \) it holds

\[
\left( \int_{\ell_k}^{u_k} \sqrt{q_k^{(b)}(t)(1 - q_k^{(b)}(t))} \, dV_k^*([\ell_k, t]) \right)^2 \leq \frac{\tau^2}{\ell_k - \ell_k} \inf_{w \in \mathbb{R}} \int_{\ell_k}^{u_k} |f_k^*(t) - w|^2 \, dt
\]

Then Assumption 2.2 is satisfied with \( \lambda = \theta/(pr^2 \theta) \).

Proof. Given \( A = \prod_{j=1}^{p} [\ell_j, u_j] \subseteq [0, 1]^p \), without loss of generality, assume \( \mathbb{E}(f^*(X)|X \in A) = 0 \). Let \( p_X(\cdot) \) be the density of \( X \) on \([0,1]^p\). Then we have

\[
\mathbb{E}(f^*(X)|X \in A) = \frac{1}{\mathbb{P}(X \in A)} \int_A (f^*(z))^2 p_X(z) \, dz \geq \frac{\theta}{\mathbb{P}(X \in A)} \int_A (f^*(z))^2 \, dz
\]

where the second inequality made use of Assumption 2.1 (i). Denote \( c_j := \frac{1}{u_j - \ell_j} \int_{\ell_j}^{u_j} f_j^*(t) \, dt \) and \( c := \sum_{j=1}^{p} c_j \), then we have

\[
\int_A (f^*(z))^2 \, dz = \int_A \left( c + \sum_{j=1}^{p} f_j^*(z) - c_j \right)^2 \, dz_1 \cdot \ldots \cdot dz_p
\]

\[
= c^2 + \sum_{j=1}^{p} (f_j^*(z) - c_j)^2 \, dz_1 \cdot \ldots \cdot dz_p
\]

\[
\geq \sum_{j=1}^{p} \prod_{k=1}^{p} \frac{u_k - \ell_k}{u_j - \ell_j} \int_{\ell_j}^{u_j} (f_j^*(t) - c_j)^2 \, dt
\]

\[
\geq \frac{\mathbb{P}(X \in A)}{\theta} \sum_{j=1}^{p} \frac{1}{u_j - \ell_j} \int_{\ell_j}^{u_j} (f_j^*(t) - c_j)^2 \, dt
\]
Combining the inequality above with (73) we have
\[ \text{Var}(f^*(X)|X \in A) \geq \frac{\theta}{\theta^2} \sum_{j=1}^{p} \frac{1}{u_j - \ell_j} \int_{\ell_j}^{u_j} (f_j(t) - c_j)^2 \, dt \]
(74)

We use $H_k^2$ to denote the LHS of (72), then (72) implies
\[ \frac{1}{u_j - \ell_j} \int_{\ell_j}^{u_j} (f_j(t) - c_j)^2 \, dt \geq \frac{1}{\tau^2} H_j^2 \]
(75)

As a result of (74) and (75), we have
\[ \text{Var}(f^*(X)|X \in A) \geq \frac{\theta}{\theta^2} \sum_{j=1}^{p} H_k^2 \]
(76)

By Lemma B.2 we have
\[ \Delta(A, b, \ell) \geq \frac{\mathbb{P}(X \in A) \text{Var}(f^*(X)|X \in A)^2}{\left(\sum_{k=1}^{p} H_k\right)^2} \]
\[ \geq \frac{\theta}{\theta^2} \sum_{j=1}^{p} H_k^2 \mathbb{P}(X \in A) \text{Var}(f^*(X)|X \in A) \]
\[ \geq \frac{\theta}{\theta^2} \mathbb{P}(X \in A) \text{Var}(f^*(X)|X \in A) \]
where the second inequality is by (76), and the last inequality made use of the Cauchy-Schwarz inequality.

\[ \square \]

B.2 Proof of Proposition 3.1

For any $A = \prod_{j=1}^{p} [\ell_j, u_j] \subseteq [0, 1]^p$ and any $k \in [p]$ it holds
\[ \left( \int_{\ell_k}^{u_k} \sqrt{q_k^*(t)}(1 - q_k^*(t)) \, dV^*_{\ell_k}([\ell_k, t]) \right)^2 \leq \frac{1}{4} \left( \int_{\ell_k}^{u_k} |f_k^*(t)| \, dt \right)^2 \]
\[ \leq \frac{\tau^2}{4} \inf_{w \in \mathbb{R}} \int_{\ell_k}^{u_k} |f_k^*(t) - w|^2 \, dt \]
(77)

where the first inequality is by Cauchy-Schwarz inequality, and the second is because $f_k^* \in LRP([0, 1], \tau)$.

Using Lemma B.3 the proof of complete.

B.3 Proof of Proposition 3.2

For any $A = \prod_{j=1}^{p} [\ell_j, u_j] \subseteq [0, 1]^p$ and any $k \in [p]$, we prove that
\[ \left( \int_{\ell_k}^{u_k} \sqrt{q_k^*(t)}(1 - q_k^*(t)) \, dV^*_{\ell_k}([\ell_k, t]) \right)^2 \leq 2r \max \left\{ \frac{\theta}{\theta} \frac{r^2}{\alpha} \right\} \inf_{w \in \mathbb{R}} \int_{a}^{b} (g(t) - w)^2 \, dt \]
(78)

Then the conclusion follows Lemma B.3.

For fixed $A$ and $k \in [p]$, to simplify the notation, we denote $g := f_k^*, a := \ell_k, b := u_k, q(t) := q_k^*(t)$ for all $t \in [\ell_k, u_k]$, and $t_j := t^{(k)}_j$ for $j = 0, 1, \ldots, r$. Then (77) can be written as
\[ \left( \int_{a}^{b} \sqrt{q(t)(1 - q(t))} \, dV_{g}([a, t]) \right)^2 \leq 2r \max \left\{ \frac{\theta}{\theta} \frac{r^2}{\alpha} \right\} \inf_{w \in \mathbb{R}} \int_{a}^{b} (g(t) - w)^2 \, dt \]
(78)

For any $s \in (0, 1)$, define $\Delta g(s) := \lim_{t \to s^+} g(t) - \lim_{t \to s^-} g(t)$. Let $j', j'' \in [r]$ such that $t_{j'-1} \leq a < t_{j'}$ and $t_{j''-1} < b \leq t_{j''}$, and define $r' = j'' - j' + 1$, and
\[ z_0 = a, z_1 = t_{j'}, z_2 = t_{j'+1}, \ldots, z_{r'-1} = t_{j''-1}, z_{r'} = b. \]

Then we have
\[ \left( \int_{a}^{b} \sqrt{q(t)(1 - q(t))} \, dV_{g}([a, t]) \right)^2 \]
\[ = \left( \sum_{j=1}^{r'} \int_{j-1}^{j} \sqrt{q(t)(1 - q(t))} \, dt \right)^2 \]
\[ \leq 2r' \sum_{j=1}^{r'} \left( \int_{j-1}^{j} \sqrt{q(t)(1 - q(t))} \, dt \right)^2 \]
(79)

\[ + 2(r' - 1) \sum_{j=1}^{r' - 1} q(z_j)(1 - q(z_j)) |\Delta g(z_j)|^2 \]

We have the following 4 claims bounding the terms in the last line of the display above.

**Claim B.4** For $j \in \{1, r'\}$, it holds
\[
\left( \int_{z_{j-1}}^{z_j} \sqrt{q(t)(1-q(t))} |g'(t)| \, dt \right)^2 \leq \frac{\bar{\theta} \beta^2}{\theta (b-a)} \inf_{w \in \mathbb{R}} \int_{z_{j-1}}^{z_j} (g(t) - w)^2 \, dt
\]

**Proof of Claim B.4** We just prove the claim for $j = 1$. The proof for $j = r'$ follows a similar argument. To prove the claim for $j = 1$, we discuss two cases.

(Case 1) $q(z_1) \leq 1/2$. Then we have $\sqrt{q(t)(1-q(t))} \leq \sqrt{q(z_1)(1-q(z_1))}$, hence
\[
\left( \int_{a}^{z_1} \sqrt{q(t)(1-q(t))} |g'(t)| \, dt \right)^2 \leq q(z_1)(1-q(z_1)) \left( \int_{a}^{z_1} |g'(t)| \, dt \right)^2
\]
\[
\leq q(z_1)(1-q(z_1)) \frac{\beta^2}{z_1-a} \inf_{w \in \mathbb{R}} \int_{a}^{z_1} (g(t) - w)^2 \, dt
\]
\[
\leq \frac{\bar{\theta} \beta^2}{\theta (b-a)} \inf_{w \in \mathbb{R}} \int_{a}^{z_1} (g(t) - w)^2 \, dt
\]
where the second inequality is because $g \in LRP((a, z_1), \beta)$; and the last inequality makes use of the fact $q(z_1) \leq \bar{\theta} (z_1 - a)/(\theta (b-a))$.

(Case 2) $q(z_1) > 1/2$. Then we have
\[
z_1 - a \geq \frac{\theta (b-a)}{\bar{\theta}} q(z_1) \geq \frac{\theta (b-a)}{2\theta}
\]
As a result,
\[
\left( \int_{a}^{z_1} \sqrt{q(t)(1-q(t))} |g'(t)| \, dt \right)^2 \leq \frac{1}{4} \left( \int_{a}^{z_1} |g'(t)| \, dt \right)^2 \leq \frac{\beta^2}{4(z_1-a)} \inf_{w \in \mathbb{R}} \int_{a}^{z_1} (g(t) - w)^2 \, dt
\]
\[
\leq \frac{\bar{\theta} \beta^2}{2\theta (b-a)} \inf_{w \in \mathbb{R}} \int_{a}^{z_1} (g(t) - w)^2 \, dt
\]
where the first inequality is by Cauchy-Schwarz inequality; the second inequality is because $g \in LRP((a, z_1), \beta)$; the third inequality is by (81).

Combining (Case 1) and (Case 2), the proof of Claim B.4 is complete.

**Claim B.5** For $j \in \{1, r' - 1\}$, it holds
\[
q(z_j)(1-q(z_j)) |\Delta g(z_j)|^2 \leq \max \left\{ \frac{4 \bar{\theta}}{\theta}, \frac{r}{\alpha} \right\} \max \{\beta^2, 4\} \inf_{w \in \mathbb{R}} \int_{x_0}^{x_j} (g(t) - w)^2 \, dt
\]

**Proof of Claim B.5** We just prove the claim for $j = 1$. The proof for $j = r' - 1$ follows a similar argument. To prove the claim for $j = 1$, we discuss two cases.

(Case 1) $|\Delta g(z_1)| > 4 \max \{\int_{x_0}^{x_1} |g'(t)| \, dt, \int_{z_1}^{x_1} |g'(t)| \, dt\}$. Then by Lemma D.6 we have
\[
\inf_{w \in \mathbb{R}} \int_{x_0}^{x_1} (g(t) - w)^2 \, dt \geq \min \{z_1 - z_0, z_2 - z_1\} \cdot \frac{(\Delta g(z_1))^2}{16}
\]
\[
(82)
\]
Note that
\[
\min \{z_1 - z_0, z_2 - z_1\} \geq \min \left\{ \frac{\theta q(z_1)}{\bar{\theta}}, \frac{\alpha}{r} (b-a) \right\} \geq \min \left\{ \frac{\theta q(z_1)}{\bar{\theta}}, \frac{\alpha}{r} \right\} (b-a)
\]
\[
(83)
\]
So by (82) and (83) we have
\[
|\Delta g(z_1)|^2 \leq \max \left\{ \frac{4 \bar{\theta}}{\theta}, \frac{r}{\alpha} \right\} \frac{16}{b-a} \inf_{w \in \mathbb{R}} \int_{x_0}^{x_1} (g(t) - w)^2 \, dt
\]
As a result,
\[
q(z_1)(1-q(z_1)) |\Delta g(z_1)|^2
\]
\[
\leq \max \left\{ \frac{\bar{\theta}}{\theta}, \frac{r}{\alpha} q(z_1)(1-q(z_1)) \right\} \frac{16}{b-a} \inf_{w \in \mathbb{R}} \int_{x_0}^{x_1} (g(t) - w)^2 \, dt
\]
\[
\leq \max \left\{ \frac{\bar{\theta}}{\theta}, \frac{r}{4\alpha} \right\} \frac{16}{b-a} \inf_{w \in \mathbb{R}} \int_{x_0}^{x_1} (g(t) - w)^2 \, dt
\]
where the second inequality is by Cauchy-Schwarz inequality.

\[ (\text{Case 2}) \quad |\Delta g(z_t)| \leq 4 \max\{ \int_{z_0}^{z_1} |g'(t)| \, dt : \int_{z_1}^{z_2} |g'(t)| \, dt \} \]

Then we have

\[ q(z_t)(1 - q(z_t))|\Delta g(z_t)|^2 \leq 4q(z_t)(1 - q(z_t)) \max\left\{ \int_{z_0}^{z_1} |g'(t)| \, dt , \int_{z_1}^{z_2} |g'(t)| \, dt \right\}^2 \tag{84} \]

By the same argument in [80], we have

\[ 4q(z_t)(1 - q(z_t))\left( \int_{z_0}^{z_1} |g'(t)| \, dt \right)^2 \leq \frac{4\delta^2}{\theta(b - a)} \inf_{w \in \bar{R}} \int_{z_0}^{z_1} (g(t) - w)^2 \, dt \tag{85} \]

On the other hand,

\[ 4q(z_t)(1 - q(z_t))\left( \int_{z_0}^{z_1} |g'(t)| \, dt \right)^2 \leq \left( \int_{z_1}^{z_2} |g'(t)| \, dt \right)^2 \]

\[ \leq \frac{\beta^2}{\alpha^2} \inf_{w \in \bar{R}} \int_{z_1}^{z_2} (g(t) - w)^2 \, dt \leq \frac{r\beta^2}{\alpha(b - a)} \inf_{w \in \bar{R}} \int_{z_1}^{z_2} (g(t) - w)^2 \, dt \tag{86} \]

where the second inequality is because \( g \in LRP((z_1, z_2), \beta) \); the last inequality is because \( z_2 - z_1 \geq \alpha/r \geq (\alpha/r)(b - a). \) By (84), (85) and (86), we have

\[ q(z_t)(1 - q(z_t))|\Delta g(z_t)|^2 \leq \max\left\{ \frac{4\delta^2}{\theta} \frac{r}{\alpha} \frac{\beta^2}{b - a} \inf_{w \in \bar{R}} \int_{z_0}^{z_1} (g(t) - w)^2 \, dt \right\} \]

Combining (Case 1) and (Case 2), the proof of Claim B.5 is complete.

\[ \square \]

**Claim B.6** For \( 2 \leq j \leq r' - 1 \), it holds

\[ \left( \int_{z_{j-1}}^{z_j} \sqrt{q(t)(1 - q(t))|g'(t)|} \, dt \right)^2 \leq \frac{r\beta^2}{4\alpha(b - a)} \inf_{w \in \bar{R}} \int_{z_{j-1}}^{z_j} (g(t) - w)^2 \, dt \]

**Proof of Claim B.6** Note that

\[ \left( \int_{z_{j-1}}^{z_j} \sqrt{q(t)(1 - q(t))|g'(t)|} \, dt \right)^2 \leq \frac{1}{4} \left( \int_{z_{j-1}}^{z_j} |g'(t)| \, dt \right)^2 \leq \frac{\beta^2}{4(z_j - z_{j-1})} \inf_{w \in \bar{R}} \int_{z_{j-1}}^{z_j} (g(t) - w)^2 \, dt \]

\[ \leq \frac{r\beta^2}{4\alpha(b - a)} \inf_{w \in \bar{R}} \int_{z_{j-1}}^{z_j} (g(t) - w)^2 \, dt \]

where the first inequality is by Cauchy-Schwarz inequality; the second inequality is because \( g \in LRP((z_{j-1}, z_j), \beta) \); the last inequality is by the assumption that \( t_j - t_{j-1} \geq \alpha/r \).

\[ \square \]

**Claim B.7** For \( 2 \leq j \leq r' - 2 \), it holds

\[ q(z_j)(1 - q(z_j))|\Delta g(z_j)|^2 \leq \frac{r \max\{\beta^2, 4\}}{\alpha} \inf_{w \in \bar{R}} \int_{z_{j-1}}^{z_{j+1}} (g(t) - w)^2 \, dt \]

**Proof of Claim B.7** We discuss two cases.

**Case 1** \( |\Delta g(z_j)| > 4 \max\{ \int_{z_{j-1}}^{z_j} |g'(t)| \, dt : \int_{z_j}^{z_{j+1}} |g'(t)| \, dt \} \). Then by Lemma D.6, we have

\[ \inf_{w \in \bar{R}} \int_{z_{j-1}}^{z_{j+1}} (g(t) - w)^2 \, dt \geq \min\{ z_j - z_{j-1}, z_{j+1} - z_j \} \cdot \frac{(\Delta g(z_j))^2}{16} \geq \frac{\alpha}{r} \frac{(\Delta g(z_j))^2}{16} \]

As a result,

\[ q(z_j)(1 - q(z_j))|\Delta g(z_j)|^2 \leq \frac{16r}{\alpha} q(z_j)(1 - q(z_j)) \inf_{w \in \bar{R}} \int_{z_{j-1}}^{z_{j+1}} (g(t) - w)^2 \, dt \]

\[ \leq \frac{4r}{\alpha} \inf_{w \in \bar{R}} \int_{z_{j-1}}^{z_{j+1}} (g(t) - w)^2 \, dt \]
By Claims B.5 and B.7, we have
\[ g(z_j)(1 - q(z_j)) |\Delta g(z_j)|^2 \]
\[ \leq 4q(z_j)(1 - q(z_j)) \max \left\{ \int_{z_{j-1}}^{z_j} |g'(t)| \, dt, \int_{z_j}^{z_{j+1}} |g'(t)| \, dt \right\}^2 \]
\[ \leq \max \left\{ \int_{z_{j-1}}^{z_j} |g'(t)| \, dt, \int_{z_j}^{z_{j+1}} |g'(t)| \, dt \right\}^2 \]
\[ \leq \max \left\{ \int_{z_{j-1}}^{z_j} |g'(t)| \, dt, \int_{z_j}^{z_{j+1}} |g'(t)| \, dt \right\} \leq \frac{\beta^2}{\alpha} \max \left\{ \inf_{w \in \mathbb{R}} \int_{z_{j-1}}^{z_j} (g(t) - w)^2 \, dt, \inf_{w \in \mathbb{R}} \int_{z_j}^{z_{j+1}} (g(t) - w)^2 \, dt \right\} \]
\[ \leq \frac{\beta^2 r}{\alpha} \max \left\{ \inf_{w \in \mathbb{R}} \int_{z_{j-1}}^{z_j} (g(t) - w)^2 \, dt, \inf_{w \in \mathbb{R}} \int_{z_j}^{z_{j+1}} (g(t) - w)^2 \, dt \right\} \]
\[ \leq \frac{\beta^2 r}{\alpha} \int_{z_{j-1}}^{z_{j+1}} (g(t) - w)^2 \, dt \]
where the second inequality is by Cauchy-Schwarz inequality; the third inequality is because \( g \in LRP((z_{j-1}, z_j), \beta) \) and \( g \in LRP((z_j, z_{j+1}), \beta) \).

Combining (Case 1) and (Case 2), and note that \( b - \alpha \leq 1 \), the proof of Claim B.7 is complete.

\[ \square \]

### Completion of the Proof of Proposition 3.2

By (79) and note that \( r' \leq r \), we have
\[ \left( \int_a^b \sqrt{q(t)(1 - q(t))} \, dV_g([a, t]) \right)^2 \]
\[ \leq 2r \sum_{j=1}^{r'} \left( \int_{z_{j-1}}^{z_j} \sqrt{q(t)(1 - q(t))} |g'(t)| \, dt \right)^2 + 2r \sum_{j=1}^{r'-1} q(z_j)(1 - q(z_j)) |\Delta g(z_j)|^2 \]

By Claims B.4 and B.6 we have
\[ \sum_{j=1}^{r'} \left( \int_{z_{j-1}}^{z_j} \sqrt{q(t)(1 - q(t))} |g'(t)| \, dt \right)^2 \]
\[ \leq \max \left\{ \frac{\beta^2}{\alpha} \sum_{j=1}^{r'} \left( \int_{z_{j-1}}^{z_j} (g(t) - w)^2 \, dt \right) \right\} \]
\[ \leq \max \left\{ \frac{\beta^2 r}{\alpha} \int_{z_{j-1}}^{z_j} (g(t) - w)^2 \, dt \right\} \]

By Claims B.5 and B.7 we have
\[ \sum_{j=1}^{r'-1} q(z_j)(1 - q(z_j)) |\Delta g(z_j)|^2 \]
\[ \leq \max \left\{ \frac{\beta^2}{\alpha} \int_{z_{j-1}}^{z_j} (g(t) - w)^2 \, dt \right\} \]
\[ \leq \max \left\{ \frac{\beta^2 r}{\alpha} \int_{z_{j-1}}^{z_j} (g(t) - w)^2 \, dt \right\} \]
\[ = \max \left\{ \frac{\beta^2 r}{\alpha} \int_{z_{j-1}}^{z_j} (g(t) - w)^2 \, dt \right\} \]

By (87), (88) and (89) we have
\[ \left( \int_a^b \sqrt{q(t)(1 - q(t))} \, dV_g([a, t]) \right)^2 \]
\[ \leq 2r \max \left\{ \frac{\beta^2 r}{\alpha} \int_{z_{j-1}}^{z_j} (g(t) - w)^2 \, dt \right\} \]

Hence (75) is true, and the proof of Proposition 3.2 is complete.
We first prove the conclusion when \( a \), where the last step is by Cauchy-Schwarz inequality.

**B.5 Proof of Example 3.3**

It suffices to prove that for any \( a, b \),

\[
\int_{a}^{b} (g(t))^{2} \, dt \geq \int_{a}^{b} (c_{1}(t - t_{0}))^{2} \, dt = \frac{c_{1}^{2}}{3} (b - t_{0})^{3}
\]

Similarly,

\[
\int_{a}^{t_{0}} (g(t))^{2} \, dt \geq \int_{a}^{t_{0}} (c_{1}(t_{0} - t))^{2} \, dt = \frac{c_{1}^{2}}{3} (t_{0} - a)^{3}
\]

As a result, we have

\[
\int_{a}^{b} (g(t))^{2} \, dt \geq \frac{c_{1}^{2}}{3} ((b - t_{0})^{3} + (t_{0} - a)^{3}) \geq \frac{2c_{1}^{2}}{3} \frac{b-a}{2}^{3} = \frac{c_{1}^{2}}{12} (b-a)^{3}
\]

(90)

On the other hand, since \( |g'(t)| \leq c_{2} \), we have

\[
\left( \int_{a}^{b} |g'(t)| \, dt \right)^{2} \leq c_{2}^{2} (b-a)^{2}
\]

(91)

Combining (90) and (91), we have

\[
\left( \int_{a}^{b} |g'(t)| \, dt \right)^{2} \leq \frac{12c_{2}^{2}}{c_{1}^{2}(b-a)^{3}} \int_{a}^{b} (g(t))^{2} \, dt
\]

**B.6 Proof of Example 3.2**

It suffices to prove that there exists a constant \( C_{r} \) such that for any univariate polynomial with a degree at most \( r \) and for any \( a < b \),

\[
\left( \int_{a}^{b} |g'(t)| \, dt \right)^{2} \leq \frac{C_{r}}{b-a} \int_{a}^{b} |g(t)|^{2} \, dt
\]

(92)

We first prove the conclusion when \( a = 0 \) and \( b = 1 \). Let \( \mathcal{P}(r) \) be the set of all univariate polynomials with degree at most \( r \). Note that \( \mathcal{P}(r) \) is a finite-dimensional linear space, and the differential operator \( \Phi : g \mapsto g' \) is a linear mapping on \( \mathcal{P}(r) \). As a result, there exists \( C_{r} \) such that

\[
\int_{0}^{1} |g'(t)| \, dt \leq \sqrt{C_{r}} \int_{0}^{1} |g(t)| \, dt
\]

for all \( g \in \mathcal{P}(r) \).

For general \( a < b \), given \( g \in \mathcal{P}(r) \), define \( h(s) := g(a + (b-a)s) \), then \( h \in \mathcal{P}(r) \). So we have

\[
\int_{0}^{1} |h'(s)| \, ds \leq \sqrt{C_{r}} \int_{0}^{1} |h(s)| \, ds
\]

(93)

Note that

\[
\int_{0}^{1} |h'(s)| \, ds = (b-a) \int_{0}^{1} |g'(a+(b-a)s)| \, ds = \int_{a}^{b} |g'(t)| \, dt
\]

(94)

and

\[
\int_{0}^{1} |h(s)| \, ds = \int_{0}^{1} |g(a+(b-a)s)| \, ds = \frac{1}{b-a} \int_{a}^{b} |g(t)| \, dt
\]

(95)

Combining (93), (94) and (95), we know that

\[
\left( \int_{a}^{b} |g'(t)| \, dt \right)^{2} \leq \left( \sqrt{\frac{C_{r}}{b-a}} \int_{a}^{b} |g(t)| \, dt \right)^{2} \leq \frac{C_{r}}{b-a} \int_{a}^{b} |g(t)|^{2} \, dt
\]

where the last step is by Cauchy-Schwarz inequality.

**B.7 Proof of Example 3.1**

It suffices to prove that for any \( a, b \in [0, 1] \) with \( a < b \), it holds

\[
\int_{a}^{b} |g'(t)| \, dt \leq \frac{10(L/\sigma)}{b-a} \int_{a}^{b} |g(t)| \, dt
\]

(96)

Once (96) is proved, the conclusion is true via Jensen’s inequality.
Below we prove (96). Denote $C = L/10$. For given $a, b \in [0, 1]$, without loss of generality, we assume the median of $g$ on $[a, b]$ is 0, i.e., \( \int_a^b 1_{(g(t) \geq 0)} \, dt = \int_a^b 1_{(g(t) < 0)} \, dt = (b - a)/2 \) (otherwise translate by a constant).

We discuss two different cases. We denote $m := \min_{t \in [a, b]} \{|g'(t)|\}$ and $M := \max_{t \in [a, b]} \{|g'(t)|\}$.

(Case 1) $m \geq C(b - a)$. Since $g$ is $L$-smooth on $[0, 1]$, we have

\[ M \leq m + L(b - a) \]  

Hence we have

\[ \frac{M}{m} \leq 1 + \frac{L(b - a)}{m} \leq 1 + \frac{L}{C} \]  

where the second inequality is by the assumption of (Case 1). Since $\min_{t \in [a, b]} \{|g'(t)|\} = m > 0$, without loss of generality, we assume that $g'(t) > 0$ for all $t \in [a, b]$. Denote $t_0 = (a + b)/2$. By our assumption that the median of $g$ on $[a, b]$ is 0, we know that $g(t_0) = 0$. Since $g$ is convex, for any $t \in [a, t_0]$, we have

\[ g(t_0) - g(t) \geq \frac{t_0 - t}{t_0 - a} (g(t_0) - g(a)) \]

which implies $g(t) \leq \frac{t_0 - t}{t_0 - a} (g(a) - g(t_0)) \leq 0$. As a result, we have

\[ \int_a^{t_0} |g(t)| \, dt \geq |g(t_0) - g(a)| \frac{t_0 - t}{2} \geq \frac{(t_0 - a)^2}{2} m = \frac{(b - a)^2}{8} m \]  

(98)

On the other hand,

\[ \int_a^b |g'(t)| \, dt \leq M(b - a) \]  

(99)

Combining (98) and (99) we have

\[ (b - a) \int_a^b |g'(t)| \, dt \leq \frac{8M}{m} \int_a^b |g(t)| \, dt \leq 8(1 + L/C) \int_a^b |g(t)| \, dt = 88 \int_a^b |g(t)| \, dt \]

where the second inequality made use of (97).

(Case 2) $m < C(b - a)$. Then by the $L$-smoothness of $g$ we have

\[ M \leq (C + L)(b - a) \]

so we have

\[ \int_a^b |g'(t)| \, dt \leq M(b - a) \leq (C + L)(b - a)^2 \]  

(100)

Define interval $[t_1, t_2] := \{ t \in [a, b] \mid g(t) \leq 0 \}$. By our assumption that the median of $g$ on $[0, 1]$ is 0, we have $t_2 - t_1 = (b - a)/2$. Denote $t_0 = \arg\min_{t \in [a, b]} g(t)$. Define function $f$ on $[t_1, t_2]$:

\[ f(t) := \begin{cases} g(t_0) \cdot (t - t_1)/(t_0 - t_1) & t \in [t_1, t_0] \\ g(t_0) \cdot (t_2 - t)/(t_2 - t_0) & t \in [t_0, t_2] \end{cases} \]

Then $0 \geq f(t) \geq g(t)$ for all $t \in [t_1, t_2]$ (because $g$ is convex). Note that

\[ \int_{t_1}^{t_2} f(t) \, dt = \frac{1}{2} g(t_0) (t_2 - t_1) + \frac{1}{2} g(t_0) (t_2 - t_1 - t_0) = \frac{1}{2} g(t_0) (t_2 - t_1) = \frac{1}{2} \int_{t_1}^{t_2} g(t_0) \, dt \]  

(101)

As a result,

\[ \int_{t_1}^{t_2} |g(t)| \, dt \geq \int_{t_1}^{t_2} |f(t)| \, dt = - \int_{t_1}^{t_2} f(t) \, dt = \int_{t_1}^{t_2} f(t) - g(t_0) \, dt \geq \int_{t_1}^{t_2} g(t) - g(t_0) \, dt \]  

(102)

where the first and last inequalities are because $0 \geq f(t) \geq g(t)$ for all $t \in [t_1, t_2]$; the second equality is by (101). Note that for any $t \in [t_1, t_2]$,

\[ g(t) - g(t_0) \geq g'(t_0)(t - t_0) + \frac{\sigma}{2}(t - t_0)^2 \geq \frac{\sigma}{2}(t - t_0)^2 \]  

(103)

where the first inequality is because $g$ is $\sigma$-strongly-convex, and the second is because $t_0$ is the minimizer of $g$ on $[t_1, t_2]$. By (102) and (103), we have

\[ \int_{t_1}^{t_2} |g(t)| \, dt \geq \frac{\sigma}{2} \int_{t_1}^{t_2} (t - t_0)^2 \, dt \geq 2 \cdot \frac{\sigma}{2} \int_0^{(t_2 - t_1)/2} s^2 \, ds = \frac{\sigma}{24} (t_2 - t_1)^3 = \frac{\sigma}{192} (b - a)^3 \]

Since $g$ is convex on $[a, b]$, the median of $g$ on $[a, b]$ is 0, and $[t_1, t_2] = \{ t \in [a, b] \mid g(t) \leq 0 \}$, it is not hard to check that

\[ \int_a^b |g(t)| \, dt \geq 2 \int_{t_1}^{t_2} |g(t)| \, dt \geq \frac{\sigma}{96} (b - a)^3 \]  

(104)

Combining (100) and (104) we have

\[ \int_a^b |g'(t)| \, dt \leq \frac{1}{b - a} \int_a^b |g(t)| \, dt \leq \frac{96(C + L)}{\sigma} \int_a^b |g(t)| \, dt \leq \frac{110(L/\sigma)}{b - a} \int_a^b |g(t)| \, dt \]

where the last inequality made use of $C = L/10$.

The proof is complete by combining the discussions in (Case 1) and (Case 2).
C Comparison of Theorem 2.3 and Theorem 1 of [10]

We first restate Theorem 1 of [10] in the setting of fitting a single tree (note that [10] discussed random forest).

**Proposition C.1** (Theorem 1 of [10]) Suppose Assumptions 2.2, 2.1 and 2.1 hold true. Let \( \hat{f}^{(d)}(\cdot) \) be the tree estimated by CART with depth \( d \). Fixed constants \( \alpha_2 > 1, 0 < \eta < 1/8, 0 < c < 1/4 \) and \( \delta > 0 \) with \( 2\eta < \delta < 1/4 \). Then there exists constant \( C > 0 \) such that for all \( n, d \) satisfying \( 1 < d < c \log_2(n) \), it holds

\[
\mathbb{E}(|\hat{f}^{(d)} - f^*|^2 L^2(\mu)) \leq C \left( n^{-\eta} + (1 - \alpha_2^{-1}) \delta + n^{-\delta + c} \right)
\]

(105)

In particular, the RHS of (105) is lower bounded by

\[
\Omega(n^{-\eta} + n^{-\delta + c} + n^{c \log_2(1 - \delta)})
\]

(106)

Note that (106) follows (105) by the fact that noises are bounded. In addition, the dependence on \( p \) was not explicitly stated in the bound (105), which seems to be hidden in the constant \( C \).

To compare our error bound with the error bound in (106), since the \( \| \hat{f}^{(d)} - f^* \|_2 L^2(\mu) \) is bounded almost surely, it is not hard to transform the high-probability bound in (11) to an bound in expectation, and we have

\[
\mathbb{E}(|\hat{f}^{(d)} - f^*|^2 L^2(\mu)) \leq O(n^{-\phi(\lambda)} \log(np) \log^2(n))
\]

(107)

Below we discuss two different cases.

- (Case 1) \( \lambda \geq 1/2 \). Then it holds

\[
\phi(\lambda) = \frac{-\log_2(1 - \lambda)}{1 - \log_2(1 - \lambda)} \geq \frac{-\log_2(1/2)}{1 - \log_2(1/2)} = 1/2
\]

(108)

So our convergence rate in (107) is \( O(n^{-1/2} \log(np) \log^2(n)) \), but the rate in (106) is

\[
\Omega(n^{-\eta} + n^{-\delta + c} + n^{c \log_2(1 - \lambda)}) \geq \Omega(n^{-\eta}) \geq \Omega(n^{-1/8})
\]

(109)

- (Case 2) \( 0 < \lambda \leq 1/2 \). Then it holds

\[
1 - \log_2(1 - \lambda) \leq 1 - \log_2(1/2) = 2
\]

(110)

and hence \( \phi(\lambda) \geq \log_2(1 - \lambda)/2 \). So our rate in (107) is \( O(n^{\log_2(1 - \lambda)/2} \log(np) \log^2(n)) \), but the rate in (106) is

\[
\Omega(n^{-\eta} + n^{-\delta + c} + n^{c \log_2(1 - \lambda)}) \geq \Omega(n^{c \log_2(1 - \lambda)}) \geq \Omega(n^{1/2 \log_2(1 - \lambda)})
\]

(111)

D Auxiliary results

**Lemma D.1** (Bernstein’s inequality) Let \( Z_1, \ldots, Z_n \) be i.i.d. random variables satisfying \( |\mathbb{E}((Z_1 - \mathbb{E}(Z_1))^k)| \leq (1/2)k!\gamma^{-k} \) for some constants \( \gamma, b > 0 \) and for all \( k \geq 2 \). Then for any \( t > 0 \),

\[
\mathbb{P}\left( \left| \frac{1}{n} \sum_{i=1}^{n} Z_i - \mathbb{E}(Z_i) \right| > t \right) \leq 2 \exp\left( -\frac{n}{4} \left( \frac{t^2}{\gamma^2} \wedge \frac{t}{b} \right) \right)
\]

**Lemma D.2** (Binomial tail bound) Let \( Z_1, \ldots, Z_n \) be i.i.d. random variables with \( \mathbb{P}(Z_i = 1) = \alpha \) and \( \mathbb{P}(Z_i = 0) = 1 - \alpha \). Then for any \( t \in (0, 1) \),

\[
\mathbb{P}\left( \left| \frac{1}{n} \sum_{i=1}^{n} Z_i \right| > t \right) \leq \exp\left( -n \left[ t \log \left( \frac{t}{\alpha} \right) + (1 - t) \log \left( \frac{1 - t}{1 - \alpha} \right) \right] \right)
\]

**Lemma D.3** For any \( t \in (0, 3/4) \), \( \log(1 - t) > - t^2 \).

**Proof.** For \( t \in (0, 3/4) \),

\[
\log(1 - t) + t^2 = \frac{t^2}{2} - \sum_{k=3}^{\infty} \frac{t^k}{k!} \geq \frac{t^2}{2} - \frac{1}{6} \sum_{k=3}^{\infty} t^k = \frac{t^2}{2} - \frac{t^3}{6(1 - t)} > 0.
\]

□
Lemma D.4 Suppose $Z$ is a random variable satisfying $\mathbb{E}(e^{\lambda Z}) \leq e^{\lambda^2 \sigma^2 / 2}$ for all $\lambda \in \mathbb{R}$, where $\sigma > 0$ is a constant; then

$$\mathbb{E}(|Z|^k) \leq 9\sigma^k k!$$

Proof. By Chernoff inequality it holds $\mathbb{P}(|Z| > t) \leq 2 \exp(-t^2/(2\sigma^2))$ for all $t > 0$. As a result,

$$\mathbb{E}(|Z|^k)/(k!\sigma^k) \leq \mathbb{E}(e^{\lambda |Z|/\sigma}) = \int_0^\infty e^t \mathbb{P}(|Z| > t) \, dt$$

$$\leq \int_0^\infty 2 \exp\left(\frac{t^2}{2}\right) \, dt = 2\sqrt{e} \int_0^\infty \exp(-t^2/2) \, dt$$

$$\leq 2\sqrt{e} \int_{-\infty}^\infty \exp(-t^2/2) \, dt = 2\sqrt{2\pi e} \leq 9$$

□

Lemma D.5 For any integer $k \geq 2$ it holds $\frac{1}{k^2} - \frac{4}{(k+1)^2} \leq \frac{1}{(k+1)^2}$.

Proof. For any $k \geq 2$ it holds

$$\frac{(2k+1)(k+1)}{2k^2} = \left(1 + \frac{1}{2k}\right)(1 + \frac{1}{k}) \leq \left(1 + \frac{1}{4}\right)(1 + \frac{1}{2}) < 2$$

Multiplying $2/(k+1)^2$ in the display above, we have

$$\frac{2k+1}{2k(k+1)^2} < \frac{4}{(k+1)^2}$$

The proof is complete by noting that $\frac{2k+1}{2k(k+1)^2} = \frac{1}{k^2} - \frac{1}{(k+1)^2}$. □

Lemma D.6 Let $[a, b]$ be a sub-interval of $[0, 1]$, and $c \in (a, b)$. Let $h$ be a function on $[a, b]$ such that $h$ is differentiable on $(a, c)$ and $(c, b)$, but can be discontinuous at $c$. Denote $\Delta h(c) := \lim_{t \to c+} h(t) - \lim_{t \to c-} h(t)$. Suppose

$$\Delta h(c) > 4 \max\left\{\int_a^c |h'(t)| \, dt, \int_c^b |h'(t)| \, dt\right\}$$  \hspace{1cm} (112)

Then it holds

$$\inf_{w \in \mathbb{R}} \int_a^b (h(t) - w)^2 \, dt \geq \min\{c-a, b-c\}(\Delta h(c))^2/16$$

Proof. We assume that $h$ is not continuous at $c$, since otherwise, the conclusion holds true trivially. We use the notation $h(c+) := \lim_{t \to c+} h(t)$ and $h(c-) := \lim_{t \to c-} h(t)$. Without loss of generality, assume $h(c+) > h(c-)$. For $w \geq (1/2)(h(c+) + h(c-))$, it holds $w - h(c-) \geq (1/2)\Delta h(c)$. By (112), we know that for any $t \in (a, c)$,

$$|h(t) - h(c-)| \leq \int_a^c |h'(\tau)| \, d\tau \leq \frac{1}{4} \Delta h(c)$$

Hence for all $t \in (a, c)$,

$$w - h(t) = w - h(c-) + h(c-) - h(t) \geq \frac{1}{2} \Delta h(c) - \frac{1}{4} \Delta h(c) = \frac{1}{4} \Delta h(c)$$

As a result,

$$\int_a^b (h(t) - w)^2 \, dt \geq \int_a^c (h(t) - w)^2 \, dt \geq (c-a)(\Delta h(c))^2/16$$  \hspace{1cm} (113)

For $w < (1/2)(h(c+) + h(c-))$, similarly, we can prove

$$\int_a^b (h(t) - w)^2 \, dt \geq (b-c)(\Delta h(c))^2/16$$  \hspace{1cm} (114)

The proof is complete by combining (113) and (114). □