Submitted to the Annals of Statistics

OPTIMAL ESTIMATION AND INFERENCE FOR MINIMIZER AND MINIMUM OF MULTIVARIATE ADDITIVE CONVEX FUNCTION

By Ran Chen

Massachusetts Institute of Technology

Contents

1. Introduction. Motivated by a wide range of applications, estimation and inference for the minimizer of nonparametric regression function has been a long standing problems in statistics (Kiefer and Wolfowitz, 1952; Blum, 1954; Chen, 1988). For fixed design, Belitser et al. (2012) establishes the minimax rate of convergence over a given smoothness class for estimating both the minimizer and minimum, Cai et al. (2022) establishes minimax rates for both estimation and inference for both minimizer and minimum under a non-asymptotic local minimax framework for univariate convex function. For sequential design, the minimax rate for estimation of minimizer has been established; see Chen et al. (1996); Polyak and Tsybakov (1990); Dippon (2003). Mokkadem and Pelletier (2007) introduces a companion for the Kiefer–Wolfowitz–Blum algorithm in sequential design for estimating both the minimizer and minimum.

Another related line of research is the stochastic continuum-armed bandits, which have been used to model online decision problems under uncertainty, with applications ranging from web advertising to adaptive routing. Stochastic continuum-armed bandits are in nature finding the maximizer (corresponding to the optimal action) of a nonparametric regression function through a sequence of actions. The objective is to minimize the expected total regret, which values a fine trade-off between exploration of new information and exploitation of historical information (Kleinberg, 2004; Auer et al., 2007; Kleinberg et al., 2019).

In the present paper, we consider optimal estimation and inference for the minimizer of *multivariate additive convex functions* under suitable non-asymptotic framework that can characterize the difficulty of the problem at individual functions. The local minimax framework, first proposed by Cai and Low (2015), has shown great power in estimation and inference for minimizer and minimum for univariate convex functions Cai et al. (2022). It is non-asymptotic and stricter than the classical minimax framework; measures the difficulty at individual function; makes it conceptually possible for establishing penalty-of-supper-efficiency type of results; and makes it possible to establish Uncertainty Principles. However, local minimax framework in general does not guarantee the existence of adaptive optimal procedures and could be too strict for measuring the difficulty of the problem. For the setting of multivariate additive convex functions, we show that local minimax framework fail to admit adaptive optimal procedures. However, the observation in this setting has a nice *projection representation*, from which we can get a sequence of independent sufficient statistics for each of its component. When the method class is restricted to that based on sufficient statistics from projection representation, which we call *separable methods*, adaptive optimality can be achieved under a suitable framework characterizing the difficulty for individual functions.

We consider both white noise model and nonparametric regression. We first focus on the white noise model, which is given by

$$\mathrm{d}Y(\mathbf{t}) = \mathbf{f}(\mathbf{t})\mathrm{d}\mathbf{t} + \varepsilon \mathrm{d}\mathbf{W}(\mathbf{t}), \mathbf{t} \in [0,1]^s, \tag{1.1}$$

where $\mathbf{W}(\mathbf{t})$ is a standard $(s,1)$-Brownian sheet on $[0,1]^s$, $\varepsilon > 0$ is the noise level. The drift function $\mathbf{f}$ is assume to be in $\mathcal{F}_s$, the collection of $s-$dimensional additive convex functions defined as follows. Function $\mathbf{f}$ is said to be an additive convex function if it can be written in the following form:

$$\mathbf{f}(\mathbf{t}) = f_0 + \sum_{i=1}^{s} f_i(t_i), \mathbf{t} = (t_1, t_2, \cdots, t_s) \in [0,1]^s, \tag{1.2}$$

where f_0 is a real number and for $1 \le i \le s$, f_i is in $\mathcal{F}$, the collection of univariate convex functions with unique minimizer, and f_i also satisfies $\int_0^1 f_i(t)\mathrm{d}t = 0$. Note that for any function $\mathbf{f}$ that can be written in the aforementioned decomposition, the decomposition is unique. And for $s = 1$, $\mathcal{F}_s = \mathcal{F}$. For clarity, we also write $Y_{\mathbf{f}}$ for Y under $\mathbf{f}$ to specify the true function. The goal is to optimally estimate the minimizer $Z(\mathbf{f}) = \arg\min_{\mathbf{t}\in[0,1]^s} \mathbf{f}(\mathbf{t})$ and also construct confidence hyper cube for $Z(\mathbf{f})$. Estimation and inference for the minimizer $Z(\mathbf{f})$ under nonparametric setting will be discussed later in section 4.

1.1. *Benchmarks Under Local Minimax Framework and Non-adaptivity.* The first step toward evaluating the performance of a procedure at individual convex functions in $\mathcal{F}_s$ is to define function-specific benchmarks for estimation and inference for minimizer. The first set of benchmarks we use is under local minimax framework (Cai and Low, 2015), which is used in estimation and inference for univariate convex functions (Cai et al., 2022).

For estimation, the hardness of the problem at an individual function is naturally captured by the expected squared distance. Further, under the local minimax framework, the benchmark is given by

$$R(\varepsilon; \mathbf{f}) = \sup_{\mathbf{g}\in\mathcal{F}_s} \inf_{\hat{Z}} \max_{\mathbf{h}\in\{\mathbf{f},\mathbf{g}\}} \mathbb{E}\left(\|\hat{Z} - Z(\mathbf{h})\|^2\right). \tag{1.3}$$

For any given $\mathbf{f} \in \mathcal{F}_s$, the benchmark $R(\varepsilon; \mathbf{f})$ quantifies the estimation accuracy at $\mathbf{f}$ of the minimizer $Z(\mathbf{f})$ against the hardest alternative of $\mathbf{f}$ within the function class $\mathcal{F}_s$.

We establish sharp minimax lower bound and upper bound for this benchmark. Interestingly, unlike univariate case, local minimax rate for multivariate case can not be adaptively achieved,

$$\inf_{\hat{Z}} \sup_{\mathbf{f}\in\mathcal{F}_s} \frac{\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z} - Z(\mathbf{f})\|^2\right)}{R(\varepsilon; \mathbf{f})} \ge C_1 s^{\frac{2}{3}}, \tag{1.4}$$

for an absolute positive constant C_1.

For confidence hyper cube with a pre-specified coverage, the hardness of the problem is naturally captured by the expected squared diameter. Let $\mathcal{I}_\alpha(\mathcal{S})$ be the collection of confidence hyper cubes for the minimizer $Z(\mathbf{f})$ with guaranteed coverage probability $1-\alpha$ for all $\mathbf{f} \in \mathcal{S}$. The benchmark under the local minimax framework, at $\mathbf{f}$, is given by the minimum expected squared diameter at $\mathbf{f}$ for all confidence hyper cube in $\mathcal{I}_\alpha(\{\mathbf{f},\mathbf{g}\})$ with the hardest alternative $\mathbf{g} \in \mathcal{F}_s$:

$$L_\alpha(\varepsilon;\mathbf{f}) = \sup_{\mathbf{g}\in\mathcal{F}_s} \inf_{CI\in\mathcal{I}_\alpha(\{\mathbf{f},\mathbf{g}\})} \max_{\mathbf{h}\in\{\mathbf{f},\mathbf{g}\}} \mathbb{E}_{\mathbf{h}}\left(diag\,(CI)^2\right), \tag{1.5}$$

where $diag\,(CI) = \sup\{\|\mathbf{t}_1-\mathbf{t}_2\| : \mathbf{t}_1,\mathbf{t}_2 \in CI\}$.

We establish minimax rate for this benchmark. Interestingly, unlike the univariate case, local minimax rate is not adaptively attainable.

$$\inf_{CI\in\mathcal{I}_\alpha(\mathcal{F}_s)} \sup_{\mathbf{f}\in\mathcal{F}_s} \frac{\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right)}{L_\alpha(\varepsilon;\mathbf{f})} \geq C_\alpha s^{\frac{2}{3}}, \tag{1.6}$$

where C_α is a positive constant depending on α.

The non-adaptivity results shows that the local minimax framework is too strict for multivariate convex functions, and calls for a better way of characterizing the difficulty of the problem.

1.2. *Separable Methods, Local Characterization, and Adaptive Procedures.* Interestingly, the observation $Y_{\mathbf{f}}$ admits a *projection representation*

$$\mathfrak{P}(Y_{\mathbf{f}}) = (\boldsymbol{\pi}_1(Y_{\mathbf{f}}),\cdots,\boldsymbol{\pi}_s(Y_{\mathbf{f}}),\mathbf{er}(Y_{\mathbf{f}}))$$

such that $\boldsymbol{\pi}_i(Y_{\mathbf{f}})$ is a sufficient statistic for f_i and all elements in $\mathfrak{P}(Y_{\mathbf{f}})$ are independent. So it is natural to focus on *separable methods*, which has i-th axis depending only on $\boldsymbol{\pi}_i(Y_{\mathbf{f}})$ for $i \in \{1,2\ldots,s\}$. Let $\mathcal{SE}$ be the collection of separable estimators and $\mathcal{SI}_\alpha(\mathcal{S})$ be the collection of separable confidence hyper cubes for the minimizer $Z(\mathbf{f})$ with guaranteed coverage probability $1-\alpha$ for all $\mathbf{f} \in \mathcal{S}$.

A natural way for characterizing the difficulty of the estimation problem (within separable estimators) individually at each function is to associate each function with a tag quantity $tag(\mathbf{f})$ such that there is a penalty for supper efficiency: if the risk at function $\mathbf{f}_0$ is much smaller than $tag(\mathbf{f}_0)$ then it has to pay a penalty at another function $\mathbf{f}_1$ by having a much larger risk than $tag(\mathbf{f}_1)$. We found a tag quantity $\beta(\mathbf{f})$ satisfying this property.

A natural way for characterizing the difficulty of the inference problem (within separable confidence hyper cubes) individually at each function is also to associate each function with a tag quantity $tag(\mathbf{f};\alpha)$ such that one of the following holds. 1, For all $\mathbf{f} \in \mathcal{F}_s$, for any $CI \in \mathcal{SI}_\alpha(\mathcal{F}_s)$, $\mathbb{E}_\mathbf{f}\left(diag\,(CI)^2\right) \geq C \cdot tag(\mathbf{f};\alpha)$ for some absolute positive constant C. 2, For any $CI \in \mathcal{SI}_\alpha(\mathcal{F}_s)$, $\sup_{\mathbf{f}\in\mathcal{F}_s} \frac{\mathbb{E}_\mathbf{f}\left(diag(CI)^2\right)}{tag(\mathbf{f};\alpha)}$ is lower bounded by an absolute positive constant. Note that the tag quantity satisfying the first condition serves as a hard lower bound for any function in $\mathcal{F}_s$, but it is not necessarily adaptively achievable. The tag quantity satisfying the second condition serves as a practical lower bound for adaptive optimal procedure, but the method can behave better at some functions. We found two tag functions $\Gamma_1(\mathbf{f};\alpha)$ and $\Gamma_2(\mathbf{f};\alpha)$ satisfying first and second property respectively, and $\frac{\Gamma_2(\mathbf{f};\alpha)}{\Gamma_1(\mathbf{f};\alpha)} \leq c \cdot (\log(s))^{\frac{1}{3}}$ for some absolute constant $c > 0$ for $\alpha \leq 0.2$.

With local characterization tag functions $\beta(\mathbf{f})$ and $\Gamma_2(\mathbf{f};\alpha)$, the next major step is to develop data-driven computationally efficient algorithms for the construction of adaptive estimator and adaptive confidence hypercube as well as establishing the optimality of these procedures at each $\mathbf{f} \in \mathcal{F}_s$.

The key idea behind the construction for each axis of the adaptive procedures is to first iteratively localize the minimizer by comparing the integrals over relevant subintervals together with a very carefully constructed stopping rule controlled by a user-specified parameter, and then add an additional estimation/inference procedure. For the final estimation/inference, is to carefully choose the control parameter of the axis-wise stopping rule and put together the output for each axis.

The resulting estimator $\hat{Z} \in \mathcal{SE}$ is shown to attain within a constant factor of the tag function $\beta\mathbf{f}$ simultaneously for all $\mathbf{f} \in \mathcal{F}_s$,

$$\mathbb{E}_\mathbf{f}\left(\|\hat{Z} - Z(\mathbf{f})\|^2\right) \leq C_2\beta(\mathbf{f}),$$

for some absolute positive constant C_2.

The resulting separable confidence hyper cube CI is shown to have guaranteed coverage probability $1-\alpha$ for all $\mathbf{f} \in \mathcal{F}_s$ while attain within a constant factor of the tag function $\Gamma_2(\mathbf{f};\alpha)$ for all $\mathbf{f} \in \mathcal{F}_s$. That is $CI \in \mathcal{SI}_\alpha(\mathcal{F}_s)$, and for all $\mathbf{f} \in \mathcal{F}_s$

$$\mathbb{E}_\mathbf{f}\left(diag\,(CI)^2\right) \leq C_3\Gamma_2(\mathbf{f};\alpha), \tag{1.7}$$

for some absolute positive constant C_3.

1.3. *Related Literature.* Leave this to paper

1.4. *Organization of the Paper.* In Section 2, we analysis local minimax risks, relating them to appropriate local modulus of continuity, in turn providing rate-sharp upper and lower bounds, and show that local minimax risks are not adaptively achievable. In Section 3, we introduce projection and separable methods; specify the tag quantities $\beta(\mathbf{f}), \Gamma_1(\mathbf{f}; \alpha), \Gamma_2(\mathbf{f}; \alpha)$ for local characterization and show their corresponding properties; introduce adaptive procedures and show their optimality. In Section 4, we consider the nonparametric regression model. We introduce the corresponding tag quantities, propose adaptive procedures and establish the optimality. In Section 5, we discuss some future directions. Proofs are given in Section 6.

1.5. *Notation.* We conclude this section with some notation that will be used in the rest of the paper. The cdf of the standard normal distribution is denoted by Φ. For $0 < \alpha < 1$, $z_\alpha = \Phi^{-1}(1-\alpha)$. For $\alpha = 0$, $z_\alpha = \infty$. We use $\|\cdot\|$ denote the L_2 norm for vectors, univariate functions and multivariate functions, depending on the setting. We use $\mathbb{1}\{A\}$ to be indicator function that takes 1 when event A happens and 0 otherwise. We use bold symbols to denote multivariate functions, e.g. $\mathbf{f}$, $\mathbf{g}$, $\mathbf{h}$. We use $f_1, \cdots, f_s$ to denote the component functions for $\mathbf{f}$ and f_0 for constant part for $\mathbf{f}$, similar convention for $\mathbf{g}$, $\mathbf{h}$. For a set $\mathcal{A} \subset [0,1]^s$, not necessarily hyper cube, denote its diameter $diag\,(\mathcal{A})$ to be the diameter of the smallest hypercube containing $\mathcal{A}$. Let $a \wedge b = \min\{a, b\}, a \vee b = \max\{a, b\}$ for real numbers a and b. We use $Z(\cdot)$ to denote the minimizer operator, for both $\mathbf{f} \in \mathcal{F}_s$ and $f \in \mathcal{F}$. Note that we use $\mathcal{I}_\alpha(\mathcal{S})$ to denote the collection of confidence hyper cubes for the minimizer with guaranteed coverage probability $1 - \alpha$ for all functions in $\mathcal{S}$. This can be generalized into univariate case when $\mathcal{S} \in \mathcal{F}$ and the hyper cube becomes interval.

2. Local Minimax Rates and Nonadaptivity. In this section, we first discuss the local minimax rates. We introduce the local modulus of continuity and use it to characterize the benchmarks for estimation and confidence hyper cubes introduced in Section 1.1, and provide rate-sharp bounds for the continuity modulus based on geometry properties of the functions. Then, we construct a set of functions such that for at least one of functions, there is a substantial gap between the risk/expected squared diameter and the aforementioned local minimax rates. These results show that the local minimax rates are not adaptively achievable for multivariate additive functions.

2.1. *Local Modulus of Continuity..* For any given function $\mathbf{f} \in \mathcal{F}_s$, we define the following local modulus of continuity for the minimizer.

$$\omega(\varepsilon; \mathbf{f}) = \sup\{\|Z(\mathbf{f}) - Z(\mathbf{g})\|^2 : \|\mathbf{f} - \mathbf{g}\|_2 \leq \varepsilon, g \in \mathcal{F}_s\}. \tag{2.1}$$

As in the case of minimizer and minimum operators for univariate convex functions or in the case of linear functionals, the local modulus $\omega(\varepsilon; \mathbf{f})$ clearly depends on $\mathbf{f}$ and can be regarded as an analogue of inverse Fisher Information in regular parametric model.

The following theorem characterizes the benchmarks for estimation and inference in terms of the corresponding local modulus of continuity.

THEOREM 2.1 (Sharp Lower Bounds). *Let* $0 < \alpha < 0.3$. *Then*

$$a\omega(\varepsilon; \mathbf{f}) \leq \quad R(\varepsilon; \mathbf{f}) \quad \leq A\omega(\varepsilon; \mathbf{f}), \tag{2.2}$$

$$b_\alpha \omega(\varepsilon/3; \mathbf{f}) \leq \quad L_\alpha(\varepsilon; \mathbf{f}) \quad \leq B_\alpha \omega(\varepsilon; \mathbf{f}), \tag{2.3}$$

where the constants a, A, b_α, B_α *can be taken as* $a = \Phi(-0.5) \approx 0.309$, $A = 3.1$, $b_\alpha = 0.6 - 2\alpha$, *and* $B_\alpha = 9(1 - 2\alpha)z_\alpha$

Theorem 2.1 shows that the benchmarks can be characterized in terms of continuity modulus of continuity. However, this continuity modulus is hard to compute. We now introduce two related geometry quantities to facilitate bounding the continuity modulus, which are also used in univariate case Cai et al. (2022). For $f \in \mathcal{F}$, $u \in \mathbb{R}$ and $\varepsilon > 0$, let $f_u(t) = \{f(t), u\}$, $M(f) = \min_{x \in [0,1]} f(x)$, and define

$$\rho_m(\varepsilon; f) = \sup\{u - \min\{f(x) : x \in [0, 1]\} : \|f - f_u\| \leq \varepsilon\}, \tag{2.4}$$

$$\rho_z(\varepsilon; f) = \sup\{|t - Z(f)| : f(t) \leq \rho_m(\varepsilon; f) + M(f), t \in [0, 1]\}. \tag{2.5}$$

The two quantities $\rho_m(\varepsilon; f)$ and $\rho_z(\varepsilon; f)$ can be intuitively perceived as filling water into the epigraph defined by the univariate convex function f until the "volume" (measured by $\|\cdot\|$) is equal to ε and then take the depth of the water and the width of the water surface.

With the geometric quantity $\rho_z(\varepsilon; f)$, we can establish a rate-sharp bound of modulus of continuity.

THEOREM 2.2 (Geometry Representation for Modulus of Continuity). *Let* $\rho_z(\varepsilon; f)$ *be defined in* (2.5) *for* $f \in \mathcal{F}$, *and let* $\mathbf{f} \in \mathcal{F}_s$. *Then*

$$\frac{1}{3} s^{-\frac{2}{3}} \sum_{i=1}^{s} \rho_z(\varepsilon; f_i)^2 \leq \omega(\varepsilon; \mathbf{f}) \leq \sum_{i=1}^{s} 9\rho_z(\varepsilon; f_i)^2. \tag{2.6}$$

And for any $\beta \le s$, *there exists* $\mathbf{f} \in \mathcal{F}_s$ *such that* $\sum_{i=1}^{s} \rho_z(\varepsilon; f_i)^2 = \beta$ *and*

$$\omega(\varepsilon; \mathbf{f}) \le 9 s^{-\frac{2}{3}} \sum_{i=1}^{s} \rho_z(\varepsilon; f_i)^2. \tag{2.7}$$

And for any $\beta \le s$, *and* $\delta_0 > 0$, *there exists* $\mathbf{f} \in \mathcal{F}_s$ *such that* $\sum_{i=1}^{s} \rho_z(\varepsilon; f_i)^2 = \beta$ *and*

$$\omega(\varepsilon; \mathbf{f}) \ge 9 \rho_z(\varepsilon; f_i)^2 - \delta_0. \tag{2.8}$$

Theorem 2.2 shows that the moduli of continuity varies within a multiple times of $s^{-\frac{2}{3}} \sum_{i=1}^{s} \rho_z(\varepsilon; f_i)^2$ and $\sum_{i=1}^{s} \rho_z(\varepsilon; f_i)^2$, with the order of both upper and lower bound attainable for some $\mathbf{f} \in \mathcal{F}_s$. Note that there is a difference between the order of the upper and lower bound, the difference of the order is $s^{-\frac{2}{3}}$. We will see later, this plays a role in the non-adaptivity.

2.2. *Non-adaptivity.* Section 2.1 provides sharp (up to constant) bounds for the benchmarks under local minimax frameworkin terms of local modulus of continuity and the geometric based quantity $\sum_{i=1}^{s} \rho_z(\varepsilon; f_i)^2$. In this section, we will show that the local minimax rates are not adaptively attainable.

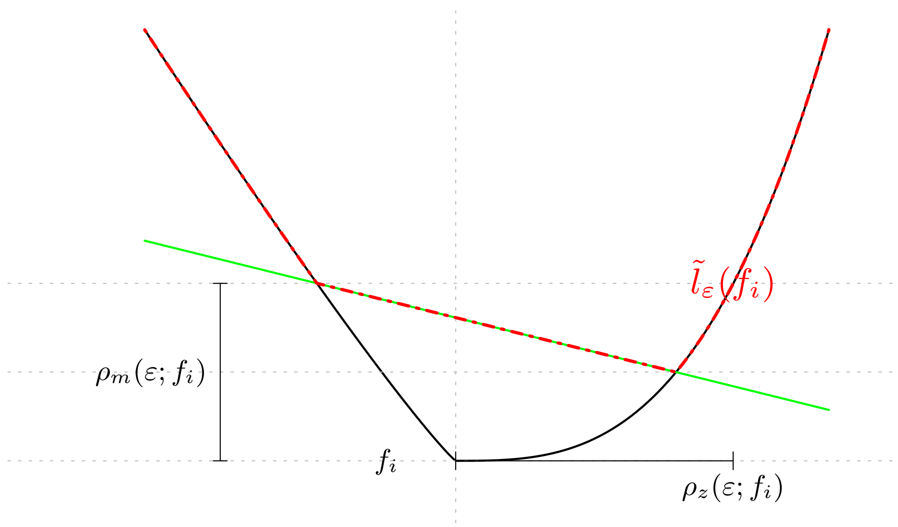
We start with constructing a set of functions associated to $\mathbf{f}$ for any $\mathbf{f} \in \mathcal{F}_s$. Recall that $\mathbf{f}$ can be written in $\mathbf{f}(\mathbf{t}) = f_0 + \sum_{i=1}^{s} f_i(t_i)$. The first step of construction is to define an alternative of f_i for all $i \in \{1, 2, \ldots, s\}$.

We will define two transformation mappings $\tilde{l}_\varepsilon(\cdot)$ and $l_\varepsilon(\cdot)$, with both from $\mathcal{F}$ to $\mathcal{F}$, and parametrized by ε. Let $x_{l,i}$ and $x_{r,i}$ be the left and right end points of the interval $\{x : f_i(x) \le M(f_i) + \rho_m(\varepsilon; f_i)\}$. Let $\tilde{x}_{l,i}$ and $\tilde{x}_{r,i}$ be the left and right end points of the interval $\{x : f(x) \le M(f_i) + \frac{1}{2}\rho_m(\varepsilon; f_i)\}$. Let $\mu = \frac{1}{2}\rho_m(\varepsilon; f_i) + M(f_i)$. Now we are ready to define $\tilde{l}_\varepsilon(f_i)$. For $0 \le 1 \le t$, let

$$\tilde{l}_\varepsilon(f_i)(t) = \begin{cases} \max\{f_i(t), \mu + \dfrac{-\rho_m(\varepsilon; f_i)}{2(\tilde{x}_{r,i} - x_{l,i})}(t - \tilde{x}_{r,i})\}, & \text{if } x_{r,i} = Z(f_i) + \rho_z(\varepsilon; f_i) \\ \max\{f_i(t), \mu + \dfrac{\rho_m(\varepsilon; f_i)}{2(x_{r,i} - \tilde{x}_{l,i})}(t - \tilde{x}_{l,i})\}, & \text{otherwise} \end{cases} \tag{2.9}$$

Figure 2.2 illustrates the construction of $\tilde{l}_\varepsilon(f_i)$, where the black line stands for f_i and red dot line stands for $\tilde{l}_\varepsilon(f_i)$.

Fig 1. *Illustration of Construction of* $\tilde{l}_\varepsilon(f_i)$



$l_\varepsilon(f_i)$ is defined by shifting $\tilde{l}_\varepsilon(f_i)$ to have integral 0:

$$l_\varepsilon(f_i)(t) = \tilde{l}_\varepsilon(f_i)(t) - \int_0^1 \tilde{l}_\varepsilon(f_i)(x)\mathrm{d}x. \tag{2.10}$$

$l_\varepsilon(f_i)$ will serve as an alternative function for i-th component in the construction of set of functions associated with $\mathbf{f}$ and parametrized by ε, $S_\varepsilon(\mathbf{f})$. $S_\varepsilon(\mathbf{f})$ is defined as the collection of functions with i-th components being f_i or $l_\varepsilon(f_i)$,

$$S_\varepsilon(\mathbf{f}) = \{\mathbf{f} + \sum_{i=1}^{s} \frac{a_i+1}{2}\left(l_\varepsilon(f_i) - f_i\right) : (a_1, a_2, \cdots, a_s) \in \{-1, 1\}^s\}. \tag{2.11}$$

The function sets $S_{\frac{1}{3}\varepsilon}(\mathbf{f})$ and $S_\varepsilon(\mathbf{f})$ defined in (2.11) with parameter $\frac{1}{3}\varepsilon$ and ε has an interesting property that the moduli of continuity of the functions in the sets are upper bounded by a multiple of the modulus of continuity of $\mathbf{f}$, as shown in Proposition 2.1.

Proposition 2.1 (Measurements of Function Set). *For* $S_{\frac{1}{3}\varepsilon}(\mathbf{f})$, $S_\varepsilon(\mathbf{f})$ *defined by* (2.11), *we have*

$$\sup_{\mathbf{h}\in S_{\frac{1}{3}\varepsilon}(\mathbf{f})} \omega(\varepsilon; \mathbf{h}) \le C_0\omega(\varepsilon; \mathbf{f}), \sup_{\mathbf{h}\in S_\varepsilon(\mathbf{f})} \omega(\varepsilon; \mathbf{h}) \le C_0\omega(\varepsilon; \mathbf{f}), \text{ for all } \mathbf{f} \in \mathcal{F}_s, \tag{2.12}$$

where C_0 is an absolute constant.

On the other hand, the lower bounds for the worst case in these function sets are of the same rate as the upper bound for continuity modulus in Theorem 2.2, as shown in Theorem 2.3.

THEOREM 2.3 (Lower Bounds for Function Set). *For $\mathbf{f} \in \mathcal{F}_s$, define $\beta(\mathbf{f}) = \sum_{i=1}^{s} \rho_z(\varepsilon; f_i)^2$. Let $0 < \alpha < 0.3$. Then*

$$\inf_{CI \in \mathcal{I}_\alpha(S_{\frac{1}{3}\varepsilon}(\mathbf{f}))} \sup_{\mathbf{h} \in S_{\frac{1}{3}\varepsilon}(\mathbf{f})} \mathbb{E}_{\mathbf{h}}(diag\,(CI)^2) \quad \geq c_\alpha \beta(\mathbf{f}), \tag{2.13}$$

$$\inf_{\hat{Z}} \sup_{\mathbf{h} \in S_\varepsilon(\mathbf{f})} \mathbb{E}_{\mathbf{h}} \left(\|\hat{Z} - Z(\mathbf{h})\|^2 \right) \quad \geq c_0 \beta(\mathbf{f}), \tag{2.14}$$

where c_0 is an absolute positive constant, and c_α is a positive constant depending on α.

From Proposition 2.1 and Theorem 2.3, naturally follows that the local minimax rates can not be adaptively achieved, as shown in Corollary 2.1.

COROLLARY 2.1 (Nonadaptivity).

$$\inf_{\hat{Z}} \sup_{\mathbf{f} \in \mathcal{F}_s} \frac{\mathbb{E}_{\mathbf{f}} \left(\|\hat{Z} - Z(\mathbf{f})\|^2 \right)}{R(\varepsilon; \mathbf{f})} \geq C_1 s^{\frac{2}{3}}, \tag{2.15}$$

where C_1 is an absolute positive constant. For $0 < \alpha < 0.3$,

$$\inf_{CI \in \mathcal{I}_\alpha(\mathcal{F}_s)} \sup_{\mathbf{f} \in \mathcal{F}_s} \frac{\mathbb{E}_{\mathbf{f}}(diag\,(CI)^2)}{L_\alpha(\varepsilon; \mathbf{f})} \geq C_\alpha s^{\frac{2}{3}}, \tag{2.16}$$

where C_α is a positive constant depending only on α.

3. Separable Methods, Local Characterization and Adaptive Optimal Procedures.. In Section 2, we show that the local minimax rates can not be adaptively achieved. In this section, we focus on a naturally arise wide class of methods that we call separable methods, which is based on a information-preserving projection representation of observation Y that we introduce in this section. We also introduce function-specific tag quantities $\beta(\mathbf{f})$, $\Gamma_1(\mathbf{f})$, $\Gamma_2(\mathbf{f})$ that can serve as benchmarks for estimation and inference in terms of can not be outperformed for any $\mathbf{f}$ or can not be outperformed uniformly. We develop adaptive optimal methods that achieve, up to a constant multiple, of the tag quantities.

3.1. *Projection Representation..* An interesting property of the observation $Y_{\mathbf{f}}$ (or Y) is that it admits a nice information-preserving *projection representation*, which maps Y to an $s+1-$tuple, where first s elements can roughly be considered as a of projection of the original stochastic process on each axis, and the last element is an $s-$dimensional stochastic process that can be considered as a remaining error.

DEFINITION 3.1 (Projection Representation). *For each $1 \le i \le s$, the $i-$th projection of Y, $\boldsymbol{\pi}_i(Y)$, is a univariate stochastic process that satisfies for $0 \le a_i < A_i \le 1$,*

$$\int_{[a_i,A_i]} \mathrm{d}\boldsymbol{\pi}_i(Y) = \int_{t_i\in[a_i,A_i],\mathbf{t}_{-i}\in[0,1]^{s-1}} \mathrm{d}Y - (A_i - a_i)\int_{[0,1]^s} \mathrm{d}Y, \tag{3.1}$$

where $\mathbf{t}_{-i} = \{t_1, \ldots, t_{i-1}, t_{i+1}, \ldots t_s\}$.

$\mathbf{er}(Y)$ is a stochastic process on $[0,1]^s$, such that for $\mathcal{A} = [a_1, A_1] \times [a_2, A_2] \times \cdots \times [a_s, A_s] \subset [0,1]^s$, we have

$$\int_{\mathcal{A}} \mathrm{d}\mathbf{er}(Y) = \int_{\mathcal{A}} \mathrm{d}Y - \sum_{i=1}^{s} \Pi_{j\neq i}(A_j - a_j)\int_{a_i}^{A_i} \mathrm{d}\boldsymbol{\pi}_i(Y). \tag{3.2}$$

The projection representation mapping $\mathfrak{P}(\cdot)$ of Y is

$$\mathfrak{P}(Y) = (\boldsymbol{\pi}_1(Y), \boldsymbol{\pi}_2(Y), \ldots, \boldsymbol{\pi}_s(Y), \mathbf{er}(Y)). \tag{3.3}$$

The reasons we call it a *projection representation* mapping are that $\mathfrak{P}(Y)$ preserves all information of Y, that $\mathfrak{P}(Y)$ has all of its elements, the projections and error, being mutually independent, and that its first s elements are sufficient statistics for corresponding function component f_i. More specifically, we have Proposition 3.1 summarizing the properties of projection representation.

PROPOSITION 3.1 (Property of Projection Representation). *Let $\mathfrak{P}(\cdot)$ be defined as in equation (3.3). Denote the class of stochastic process defined in (1.1) as $\mathfrak{Y}$. Then we have the followings.*

- *$\mathfrak{P}(\cdot)$ is a bijection from $\mathfrak{Y}$ to $\mathfrak{P}(\mathfrak{Y})$.*
- *$\mathfrak{P}(Y)$ has all elements being independent.*
- *$\boldsymbol{\pi}_i(Y)$ is a sufficient statistic for f_i, for $i \in \{1, 2, \ldots, s\}$.*

Also, it's easy to check that $\mathbf{er}(Y)$ only depends on f_0, thus not carrying information for $Z(\mathbf{f})$ by itself. Note that the minimizer $Z(\mathbf{f})$ can be written as $Z(\mathbf{f}) = (Z(f_1), Z(f_2), \ldots, Z(f_s))$, so its i-th element only depends on f_i. The information preserving representation $\mathfrak{P}(\cdot)$ plays the role of separating the relevant information of s axes into independent random variables.

3.2. *Separable Methods and Lower Bounds..* We introduce projection representation in Section 3.1, which gives independent random variables containing information of f_i separately, it's natural to focus on the class of methods that do estimation/inference for i-th coordinate, namely $Z(f_i)$, only based on the random variable containing information of f_i, namely $\boldsymbol{\pi}_i(Y)$. We call this class of methods *separable methods*, as defined in Definition 3.2.

DEFINITION 3.2 (Separable Methods). *Denote $\hat{Z}_i$ to be the ith element of $\hat{Z}$. The class of separable estimators is defined as*

$$\mathcal{SE} = \{\hat{Z} : \hat{Z}_i \textit{ only depends on } \boldsymbol{\pi}_i(Y)\}. \tag{3.4}$$

Denote CI_i of confidence hyper-cube CI to be the projection on its ith axis. The class of $1-\alpha$ separable confidence hyper-cube over $\mathcal{F}_s$ is defined as

$$\mathcal{SI}_\alpha(\mathcal{F}_s) = \{CI \in \mathcal{I}_\alpha(\mathcal{F}_s) : CI(Y)_i \textit{ only depends on } \boldsymbol{\pi}_i(Y)\}. \tag{3.5}$$

Now that focusing on separable methods, we need function-specific benchmarks to characterize the difficulty of the problems locally. As the local minimax framework suffers from non-adaptivity for the entire class of methods, we do not take it for separable methods for the concern of non-adaptivity. On the other hand, we resort to more fundamental properties that characterize good benchmarks for local characterization of problem difficulty.

Let a tag quantity associated to $\mathbf{f} \in \mathcal{F}_s$ be

$$\beta(\mathbf{f}) = \sum_{i=1}^{s} \rho_z(\varepsilon; f_i)^2. \tag{3.6}$$

Then we have that the estimator $\hat{Z}$ has to pay a price at some function if behaves too well at another function, in terms of risk defined by $\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z} - Z(\mathbf{f})\|^2\right)$, as shown in Theorem 3.1.

THEOREM 3.1 (Separable Estimator Lower Bound: Penalty for Supper Efficiency). *Let $\beta(\mathbf{f})$ be defined in (3.6). For any $\hat{Z} \in \mathcal{SE}$, for $0 < \gamma < \gamma_0$, if there is $\mathbf{f}_0 \in \mathcal{F}_s$ such that*

$$\mathbb{E}_{\mathbf{f}_0}\left(\|\hat{Z} - Z(\mathbf{f}_0)\|^2\right) \leq \gamma \beta(\mathbf{f}_0), \tag{3.7}$$

then there exists $\mathbf{f}_1 \in \mathcal{F}_s$ such that

$$\mathbb{E}_{\mathbf{f}_1}\left(\|\hat{Z} - Z(\mathbf{f}_1)\|^2\right) \geq c_2 \left(\log \frac{1}{\gamma}\right)^{\frac{2}{3}} \beta(\mathbf{f}_1), \tag{3.8}$$

where γ_0, c_2 are positive constants and $\gamma_0 < 1$.

Now we turn to the characterization of the difficulty for inference within separable methods. Theorem 3.2 shows the best separable confidence hyper cubes can achieve for each $\mathbf{f} \in \mathcal{F}_s$.

THEOREM 3.2 (Separable Confidence Hyper Cube Lower Bound, General). *For $0 < \alpha \leq 0.2$, suppose $CI \in \mathcal{SI}_\alpha(\mathcal{F}_s)$, then for $\mathbf{f} \in \mathcal{F}_s$, we have*

$$\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right) \geq 0.3 \inf_{\sum_{i=1}^s \alpha_i = 2\alpha; \alpha \geq \alpha_i \geq 0 \forall i} \sum_{i=1}^s \rho_z(z_{\alpha_i}\varepsilon; f_i)^2. \tag{3.9}$$

Now we extract two tag quantities

$$\Gamma_1(\mathbf{f}; \alpha) = \sum_{i=1}^s \rho_z(z_\alpha \varepsilon; f_i)^2, \qquad \Gamma_2(\mathbf{f}; \alpha) = \sum_{i=1}^s \rho_z(z_{\alpha/s}\varepsilon; f_i)^2. \tag{3.10}$$

We show in Theorem 3.3 that $\Gamma_1(\mathbf{f}; \alpha)$ can not be outperformed (up to a multiple of constant) at any $\mathbf{f} \in \mathcal{F}_s$, that $\Gamma_1(\mathbf{f}; \alpha)$ is not adaptively achievable, and that $\Gamma_2(\mathbf{f}; \alpha)$ can not be outperformed (up to a multiple of constant) uniformly.

THEOREM 3.3 (Separable Confidence Hyper Cube, Local Characterization). *Let $\Gamma_1(\mathbf{f}; \alpha)$ and $\Gamma_2(\mathbf{f}; \alpha)$ be defined in (3.10). Suppose $0 < \alpha \leq 0.2$, and $CI \in \mathcal{SI}_\alpha(\mathcal{F}_s)$. Then we have*

$$\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right) \geq 0.3 \cdot \Gamma_1(\mathbf{f}; \alpha), \text{ for all } \mathbf{f} \in \mathcal{F}_s, \tag{3.11}$$

$$\sup_{\mathbf{f} \in \mathcal{F}_s} \frac{\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right)}{\Gamma_1(\mathbf{f}; \alpha)} \geq \frac{1}{20}\left(\frac{z_{\alpha/s}}{z_\alpha}\right)^{\frac{2}{3}}, \tag{3.12}$$

and

$$\sup_{\mathbf{f} \in \mathcal{F}_s} \frac{\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right)}{\Gamma_2(\mathbf{f}; \alpha)} \geq \frac{1}{20}. \tag{3.13}$$

From Theorem 3.3, we can see optimal methods can be expected to achieve the rate $\Gamma_2(\mathbf{f}; \alpha)$, but not $\Gamma_1(\mathbf{f}; \alpha)$, and $\Gamma_1(\mathbf{f}; \alpha)$ serves as a floor for any separable method. With Proposition **??** from Cai et al. (2022), it is easy to see that $\frac{\Gamma_2(\mathbf{f};\alpha)}{\Gamma_1(\mathbf{f};\alpha)} \leq \left(\frac{2z_{\alpha/s}}{z_\alpha}\right)^{\frac{2}{3}}$. Note that $z_\zeta \sim \sqrt{\log \frac{1}{\zeta}}$ as $\zeta \to 0^+$, so $\frac{\Gamma_2(\mathbf{f};\alpha)}{\Gamma_1(\mathbf{f};\alpha)} \leq c \cdot \log s^{\frac{1}{3}}$ for an absolute constant $c > 0$ for $\alpha \leq 0.2$. Hence the difference between $\Gamma_1(\mathbf{f}; \alpha)$ and $\Gamma_2(\mathbf{f}; \alpha)$ is small.

3.3. *Adaptive Separable Procedures..* We now turn to the construction of data-driven and computationally efficient algorithms for estimation and confidence hyper cube for the minimizer $Z(\mathbf{f})$ under the white noise model. The construction of the procedures are given in this section. The procedures are shown to be adaptive to each individual function $\mathbf{f} \in \mathcal{F}_s$ in the sense that they simultaneously achieve, up to a universal constant, the corresponding tags $\beta(\mathbf{f})$, $\Gamma_2(\mathbf{f}; \alpha)$ for all $\mathbf{f} \in \mathcal{F}_s$, which is shown in Section 3.4.

Similar to the construction in Cai et al. (2022), we have three blocks: localization, stopping, and estimation/inference. But in multivariate setting, s can potentially grow to infinity and the projections $\boldsymbol{\pi}_i(Y)$ also has difference distribution with that in univariate case, our procedures are carefully tailored to accommodate for the new challenges.

3.3.1. *Sample Splitting.* For technical reasons, we split the data vector $V = (\boldsymbol{\pi}_1(Y), \boldsymbol{\pi}_2(Y), \ldots, \boldsymbol{\pi}_s(Y))$ into three independent pieces to ensure independence of the data used in the three steps.

Let $B_1^1(t), B_1^2(t), B_2^1(t), B_2^2(t), \ldots, B_s^1(t), B_s^2(t)$ be 2s independent standard Brownian motions that are also independent from Y. Let data vectors $V_l = (\mathbf{v}_1^l, \mathbf{v}_2^l, \ldots, \mathbf{v}_s^l)$, $V_r = (\mathbf{v}_1^r, \mathbf{v}_2^r, \ldots, \mathbf{v}_s^r)$ and $V_e = (\mathbf{v}_1^e, \mathbf{v}_2^e, \ldots, \mathbf{v}_s^e)$ be defined as follows.

(3.14)

$$\mathbf{v}_i^l(t) = \boldsymbol{\pi}_i(Y)(t) + \frac{\sqrt{2}}{2}\varepsilon\left(B_i^1(t) - t\int_0^1 B_i^1(x)\mathrm{d}x\right) + \frac{\sqrt{6}}{2}\varepsilon\left(B_i^2(t) - t\int_0^1 B_i^2(x)\mathrm{d}x\right),$$
$$\mathbf{v}_i^r(t) = \boldsymbol{\pi}_i(Y)(t) + \frac{\sqrt{2}}{2}\varepsilon\left(B_i^1(t) - t\int_0^1 B_i^1(x)\mathrm{d}x\right) - \frac{\sqrt{6}}{2}\varepsilon\left(B_i^2(t) - t\int_0^1 B_i^2(x)\mathrm{d}x\right),$$
$$\mathbf{v}_i^e(t) = \boldsymbol{\pi}_i(Y)(t) - \sqrt{2}\varepsilon\left(B_i^1(t) - t\int_0^1 B_i^1(x)\mathrm{d}x\right).$$

Then the concatenate vector of vectors V_l, V_r, V_e has all of its elements being independent, and for each axis $i \in \{1, 2 \ldots, s\}$, $\mathbf{v}_i^l(t), \mathbf{v}_i^r(t), \mathbf{v}_i^e(t)$ can be written as

$$\begin{aligned}
\mathrm{d}\mathbf{v}_i^l(t) &= f_i(t)\mathrm{d}t + \sqrt{3}\varepsilon\mathrm{d}\tilde{W}_i^l,\\
\mathrm{d}\mathbf{v}_i^r(t) &= f_i(t)\mathrm{d}t + \sqrt{3}\varepsilon\mathrm{d}\tilde{W}_i^r, \qquad (3.15)\\
\mathrm{d}\mathbf{v}_i^e(t) &= f_i(t)\mathrm{d}t + \sqrt{3}\varepsilon\mathrm{d}\tilde{W}_i^e,
\end{aligned}$$

where $\tilde{W}_i^l$, $\tilde{W}_i^r$, $\tilde{W}_i^e$ are independent standard Brownian Bridges.

3.3.2. *Localization.* We use V_l for localization step, and for each axis $k \in \{1, 2, \dots, s\}$, localization is based on $\mathbf{v}_k^l$.

We take an iterative localization procedure similar to Cai et al. (2022) on $\mathbf{v}_k^l$. For iterations (levels) $j = 0, 1, \dots,$ and possible location index at jth level $i = 0, 1, \dots, 2^j$, we denote the sub-interval length, sub-interval end points, and the index of the sub-interval containing the minimizer at level j to be

$$m_j = 2^{-j}, \; t_{j,i} = i \cdot m_j, \text{ and } i^*_{j,k} = \max\{i : Z(f_k) \in [t_{j,i-1}, t_{j,i}]\}. \tag{3.16}$$

For $j = 0, 1, \dots,$ and $i = 1, 2, \dots, 2^j$, define

$$X_{j,i,k} = \int_{t_{j,i-1}}^{t_{j,i}} \mathrm{d}\mathbf{v}_k^l(t),$$

where $\mathbf{v}_k^l$ is one of the three independent copies constructed above through sample splitting. For convenience, we define $X_{j,i,k} = +\infty$ for $j = 0, 1, \dots,$ and $i \in \mathbb{Z} \setminus \{1, 2, \dots, 2^j\}$.

Let $\hat{i}_{0,k} = 1$ and for $j = 1, 2, \dots,$ let

$$\hat{i}_{j,k} = \underset{2\hat{i}_{j-1}-2 \le i \le 2\hat{i}_{j-1}+1}{\arg\min} X_{j,i,k}.$$

Note that given the value of $\hat{i}_{j-1,k}$ at level $j-1$, in the next iteration the procedure halves the interval $[t_{\hat{i}_{j-1,k}-1}, t_{\hat{i}_{j-1,k}}]$ into two subintervals and selects the interval $[t_{\hat{i}_{j,k}-1}, t_{\hat{i}_{j,k}}]$ at level j from these and their immediate neighboring subintervals. So i only ranges over 4 possible values at level j.

3.3.3. *Stopping Rule.* For each axis, it is necessary to have a stopping rule to select a final subinterval constructed in the localization iterations and carry out the estimation/inference based on that. But unlike a unified stopping rule in Cai et al. (2022), we construct a series of stopping rules based on a user select parameter ζ, which we will specify later in the specific estimation/inference procedures. Again, for any $1 \le k \le s$, we focus on the stopping rules for k-th axis.

We use another independent copy $\mathbf{v}_k^r$ constructed in the sample splitting step to devise the stopping rules. For $j = 0, 1, \dots,$ and $i = 1, 2, \dots, 2^j$, let

$$\tilde{X}_{j,i,k} = \int_{t_{j,i-1}}^{t_{j,i}} \mathrm{d}\mathbf{v}_k^r(t).$$

Again, for convenience, we define $\tilde{X}_{j,i,k} = +\infty$ for $j = 0, 1, \dots,$ and $i \in \mathbb{Z} \setminus \{1, 2, \dots, 2^j\}$. Let the statistic $T_{j,k}$ be defined as

$$T_{j,k} = \min\{\tilde{X}_{j,\hat{i}_{j,k}+6,k} - \tilde{X}_{j,\hat{i}_{j,k}+5,k}, \; \tilde{X}_{j,\hat{i}_{j,k}-6,k} - \tilde{X}_{j,\hat{i}_{j,k}-5,k}\},$$

where we use the convention $+\infty - x = +\infty$ and $\min\{+\infty, x\} = x$, for any $-\infty \le x \le \infty$.

The stopping rule indexed by the parameter ζ is based on the value of $T_{j,k}$. Before we formally go into the stopping rule, it's helpful to look at the distribution of the elements defining $T_{j,k}$. Let $\sigma_j^2 = 6m_j\varepsilon^2$, some calculations show that when $\tilde{X}_{j,\hat{i}_{j,k}+6,k} - \tilde{X}_{j,\hat{i}_{j,k}+5,k} < \infty$, we have

(3.17)
$$\frac{\tilde{X}_{j,\hat{i}_{j,k}+6,k} - \tilde{X}_{j,\hat{i}_{j,k}+5,k}}{\sigma_j}\Bigg|\hat{i}_{j,k} \sim N\left(\frac{m_j\sqrt{m_j}}{\sqrt{6}\varepsilon} \times \frac{1}{m_j}\int_{t_{j,\hat{i}_j+5,k}}^{t_{j,\hat{i}_j+6,k}} \frac{f_k(t+m_j) - f_k(t)}{m_j}\mathrm{d}t, 1\right).$$

Note that the term

$$S_p(j,k) = \frac{1}{m_j}\int_{t_{j,\hat{i}_j+5,k}}^{t_{j,\hat{i}_j+6,k}} \frac{f_k(t+m_j) - f_k(t)}{m_j}\mathrm{d}t$$

can be interpreted as an average slope across the interval $[t_{j,\hat{i}_j+5,k}, t_{j,\hat{i}_j+6,k}]$ of the line determined by two points $(t, f(t))$ and $(t+m_j, f(t+m_j))$. Basic property of convex function shows that $S_p(j,k)$ is non-increasing with the increasing of j, and that $S_p(j,k) < 0$ implies $i^*_{j,k} \ge \hat{i}_j + 5$. These mean that a small number of $\frac{\tilde{X}_{j,\hat{i}_{j,k}+6,k} - \tilde{X}_{j,\hat{i}_{j,k}+5,k}}{\sigma_j}$ indicates localization procedure's choice of a far away sub-interval from the one minimizer lies in or a negligible signal which implies little or no gain in continuing the localization procedure.

Analogous results hold for $\frac{\tilde{X}_{j,\hat{i}_{j,k}-6,k} - \tilde{X}_{j,\hat{i}_{j,k}-5,k}}{\sigma_j}$.

Finally, the iteration stops at level $\hat{j}(\zeta, k)$, where

$$\hat{j}(\zeta,k) = \min\{j : \frac{T_{j,k}}{\sigma_j} \le z_\zeta\}. \tag{3.18}$$

The subinterval containing the minimizer $Z(f_k)$ is localized to be

$$[t_{\hat{j}(\zeta,k),\hat{i}_{\hat{j}(\zeta,k),k}-1}, t_{\hat{j}(\zeta,k),\hat{i}_{\hat{j}(\zeta,k),k}}].$$

3.3.4. *Estimation and Inference.* After obtaining, for each axis $k \in \{1, 2, \ldots, s\}$, a stopping step $\hat{j}(\zeta_k, k)$, an associated index at the stopping step $\hat{i}_{\hat{j}(\zeta_k,k),k}$, and a final interval $[t_{\hat{j}(\zeta_k,k),\hat{i}_{\hat{j}(\zeta_k,k),k}-1}, t_{\hat{j}(\zeta_k,k),\hat{i}_{\hat{j}(\zeta_k,k),k}}]$, all controlled by a parameter ζ_k, we use them to construct estimator and confidence hyper cube for the minimizer $Z(\mathbf{f})$.

For estimation, we set $\zeta_k = \zeta = \Phi(-2)$, for $k \in \{1, 2, \ldots, s\}$. The k-th axis of the estimator $\hat{Z}$ is given by the mid point of final interval:

$$\hat{Z}_k = \frac{t_{\hat{j}(\zeta,k),\hat{i}_{\hat{j}(\zeta,k),k}-1} + t_{\hat{j}(\zeta,k),\hat{i}_{\hat{j}(\zeta,k),k}}}{2}. \tag{3.19}$$

The final estimator $\hat{Z}$ is given by

$$\hat{Z} = (\hat{Z}_1, \hat{Z}_2, \ldots, \hat{Z}_s), \tag{3.20}$$

with $\hat{Z}_k$ defined in (3.19).

For inference, we set $\zeta_k = \zeta = \alpha/s$, for $k \in \{1, 2, \ldots, s\}$. The k-th axis CI_k of the hyper-cube CI is given by

$$CI_k = \left[2^{-\hat{j}(\zeta,k)+1}\left(\hat{i}_{\hat{j}(\zeta,k)-1,k} - 7\right), 2^{-\hat{j}(\zeta,k)+1}\left(\hat{i}_{\hat{j}(\zeta,k)-1,k} + 6\right)\right] \cap [0, 1]. \tag{3.21}$$

The confidence cube CI for the minimizer is give by

$$CI = CI_1 \times CI_2 \times \cdots \times CI_s, \tag{3.22}$$

where CI_k is defined in (3.21).

By the definition of our estimator $\hat{Z}$ and confidence hyper cube CI, their k-th axis only depend Y through $\boldsymbol{\pi}_k(Y)$. And further, both of them have mutually independent axes. So they are separable methods.

3.4. *Statistical Optimality..* In this section, we establish the optimality of the adaptive procedures constructed in Section 3.3. The results show that the date driven estimators and the confidence intervals achieves within a universal constant factor of their corresponding tags, $\beta(\mathbf{f})$, $\Gamma_2(\mathbf{f};\alpha)$, simultaneously for all $\mathbf{f} \in \mathcal{F}_s$. These results are non-asymptotic and function-specific, which are much stronger than the conventional minimax framework. We start with estimation.

THEOREM 3.4 (Estimation). *Let $\beta(\mathbf{f})$ be defined in (3.6). The estimator $\hat{Z}$ defined by (3.20) satisfies*

$$\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z} - Z(\mathbf{f})\|^2\right) \le C_2\beta(\mathbf{f}), \text{ for all } f \in \mathcal{F}_s, \tag{3.23}$$

where $C_2 > 0$ is an absolute constant.

The following holds for the confidence hyper cube CI.

THEOREM 3.5 (Confidence Hyper-cube). *Let* $\Gamma_2(\mathbf{f};\alpha)$ *be defined in* (3.10). *The confidence hyper cube CI defined by* (3.22) *is an* $1-\alpha$ *level confidence hyper cube for the minimizer* $Z(\mathbf{f})$. *For* $0<\alpha\le 0.3$, *its expected squared diameter satisfies*

$$\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right)\le C_3\Gamma_2(\mathbf{f};\alpha),$$

where C_3 *is an absolute constant.*

4. Nonparametric Regression. We have so far focused on the white noise model. The procedures and results presented in the previous sections can be extended to nonparametric regression, where we observe

$$y_{i_1,i_2,\ldots,i_s}=\mathbf{f}(i_1/n,i_2/n,\ldots,i_s/n)+\sigma z_{i_1,i_2,\ldots,i_s},0\le i_k\le n,\ \text{for } 1\le k\le s, \tag{4.1}$$

with $z_{i_1,i_2,\ldots,i_s}\overset{i.i.d}{\sim}N(0,1)$, $\mathbf{f}\in\mathcal{F}_s$. The noise level σ is assumed to be known. The tasks are the same as before: constructing optimal estimator and confidence interval for the minimizer within the class of separable methods, which we will introduce later, for $\mathbf{f}\in\mathcal{F}_s$. For simplicity of notation, we take $\mathbf{i}=(i_1,i_2,\ldots,i_s)$.

4.1. *Separable Methods, Local Characterization, and Lower Bounds.* Analogous to tag $\beta(\mathbf{f})$ enjoying non-supper-efficiency in Theorem 3.1 and tags $\Gamma_1(\mathbf{f};\alpha)$ and $\Gamma_2(\mathbf{f};\alpha)$ characterizing the difficulty of the inference problem in Theorem 3.3 for separable methods, we discuss similar benchmarks for the nonparametric regression model (4.1).

Analogous to the white noise model, the observation under nonparametric setting also admits a separable representation, as defined in Definition 4.1.

DEFINITION 4.1 (Projection Representation for Nonparametric Regression Model). *For* $k\in\{1,2,\ldots,s\}$, *the* k*-th projection of* $\{y_{\mathbf{i}}\}$, $\boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})$, *is an* $n+1$*-long random vector,*

$$\boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})=\left(\frac{\sum_{\mathbf{i}:i_k=1}y_{\mathbf{i}}}{(n+1)^{s-1}}-\frac{\sum_{\mathbf{i}}y_{\mathbf{i}}}{(n+1)^s},\frac{\sum_{\mathbf{i}:i_k=2}y_{\mathbf{i}}}{(n+1)^{s-1}}-\frac{\sum_{\mathbf{i}}y_{\mathbf{i}}}{(n+1)^s},,\ldots,\frac{\sum_{\mathbf{i}:i_k=s}y_{\mathbf{i}}}{(n+1)^{s-1}}-\frac{\sum_{\mathbf{i}}y_{\mathbf{i}}}{(n+1)^s}\right). \tag{4.2}$$

$\mathbf{er}(\{y_{\mathbf{i}}\})$ *is an* s*-dimension tensor with*

$$\mathbf{er}(\{y_{\mathbf{i}}\})_{i_1,i_2,\cdots,i_s}=y_{i_1,i_2,\cdots,i_s}-\sum_{k=1}^{s}\boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})_{i_k}, \tag{4.3}$$

for $0 \leq i_k \leq n$, $1 \leq k \leq s$.

The projection representation mapping $\mathfrak{P}(\cdot)$ *of observation* $\{y_{\mathbf{i}}\}$ *is given by*

$$\mathfrak{P}(\{y_{\mathbf{i}}\}) = (\boldsymbol{\pi}_1(\{y_{\mathbf{i}}\}), \boldsymbol{\pi}_2(\{y_{\mathbf{i}}\}), \boldsymbol{\pi}_s(\{y_{\mathbf{i}}\}), \mathbf{er}(\{y_{\mathbf{i}}\})). \tag{4.4}$$

Similar to white noise model, $\mathfrak{P}(\cdot)$ preserves the information of $\{y_{\mathbf{i}}\}$; has its $s+1$ elements being mutually independent; and separates the information for the s univariate component functions of $\mathbf{f}$ into its first s random variables, as shown in Proposition 4.1.

PROPOSITION 4.1 (Property of Projection Representation). *Let* $\mathfrak{P}(\cdot)$ *be define in equation* (4.4)*. Then we have*

- $\mathfrak{P}(\cdot)$ *is invertible,*
- $\mathfrak{P}(\{y_{\mathbf{i}}\})$ *has its* $s+1$ *elements being independent,*
- $\boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})$ *is sufficient statistic for* f_i.

Naturally, similar to white noise model, separable methods for nonparametric regression model are defined to be methods having its k-th coordinate depending only on $\boldsymbol{\pi}_k\{y_{\mathbf{i}}\}$, as in Definition 4.2.

DEFINITION 4.2 (Separable Methods). *Let* $\hat{Z}_k$ *be the* k*-th coordinate of* $\hat{Z}$*. The class of separable estimators is defined as*

$$\mathcal{SE} = \{\hat{Z} : \hat{Z}_k \textit{ only depends on } \boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})\}. \tag{4.5}$$

Let CI_k *be the projection of confidence hyper cube* CI *on* k*-th coordinate. The class of* $1-\alpha$ *separable confidence hyper cube over* $\mathcal{F}_s$ *is defined as*

$$\begin{aligned} \mathcal{SI}_\alpha(\mathcal{F}_s) = \{CI : &CI_k \textit{ only depends on } \boldsymbol{\pi}_k(\{y_{\mathbf{i}}\}) \textit{ for } k \in \{1, 2 \cdots, s\}, \\ &\inf_{\mathbf{f} \in \mathcal{F}_s} \mathbb{E}_{\mathbf{f}} \left(\mathbb{1}\{Z(\mathbf{f}) \in CI\}\right) \geq 1-\alpha\}. \end{aligned} \tag{4.6}$$

Similarly to white noise model, we first need to characterize the difficulty for the estimation/inference problem for each $\mathbf{f} \in \mathcal{F}_s$.

An additional complexity for the nonparametric regression is that two functions $\mathbf{f}$ and $\mathbf{g}$ can have same values on all grid points $\frac{\mathbf{i}}{n}$ while have different minimizers. We call this error caused by discretization *discretization error*:

$$\mathfrak{D}(\mathbf{f}; n) = \sup_{\mathbf{g} \in \mathcal{F}_s} \{\|Z(\mathbf{f}) - Z(\mathbf{g})\|^2 : \mathbf{f}(\frac{\mathbf{i}}{n}) = \mathbf{g}(\frac{\mathbf{i}}{n}) \text{ for all } \mathbf{i} \in \{0, 1, \dots, n\}^s\}. \tag{4.7}$$

It's apparent that $\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right) \geq (1-2\alpha)\mathfrak{D}(\mathbf{f};n)$ for $CI \in \mathcal{SI}_\alpha(\mathcal{F}_s)$. And $\sup_{\mathbf{f}} \frac{\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z}-Z(\mathbf{f})\|^2\right)}{\frac{1}{4}\mathfrak{D}(\mathbf{f};n)} \geq 1$.

Besides the discretization error, we associate each function with a tag quantities that measures the difficulty for estimation/inference at it in terms of the error caused by randomness of the observation.

For estimation, let the tag quantity $\psi(\mathbf{f})$ be defined as

$$\psi(\mathbf{f}) = \sum_{k=1}^{s} \rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)^2 \left(1 \wedge n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\right). \tag{4.8}$$

Then when if separable estimator $\hat{Z}$ behaves too well in terms of ψ but not discretization error, then the estimator has to pay a penalty at another function $\mathbf{f}_1$ in terms of ψ, as shown in Theorem 4.1.

THEOREM 4.1 (Estimation Lower Bound: Penalty for Supper Efficiency). *Let $\psi(\mathbf{f})$ be defined in (4.8). Suppose $\hat{Z} \in \mathcal{SE}$. Suppose $0 < \gamma < \gamma_1 < 1$, where γ_1 is an absolute positive constant. If $\mathbb{E}_{\mathbf{f}_0}\left(\|\hat{Z} - Z(\mathbf{f}_0)\|^2\right) \geq \frac{1}{4}\mathfrak{D}(\mathbf{f}_0;n)$, and*

$$\mathbb{E}_{\mathbf{f}_0}\left(\|\hat{Z} - Z(\mathbf{f}_0)\|^2\right) \leq \gamma\psi(\mathbf{f}_0), \tag{4.9}$$

for some $\mathbf{f}_0 \in \mathcal{F}_s$, then there exist $\mathbf{f}_1 \in \mathcal{F}_s$ such that

$$\mathbb{E}_{\mathbf{f}_1}\left(\|\hat{Z} - Z(\mathbf{f}_1)\|^2\right) \geq c_3 \left(\log(\frac{1}{\gamma})\right)^{\frac{2}{3}} \psi(\mathbf{f}_1), \tag{4.10}$$

where $c_3 > 0$ is an absolute constant.

Now we turn to characterizing the difficulty of inference with separable methods at individual functions $\mathbf{f} \in \mathcal{F}_s$. For simplicity of notation, we define

$$\varphi(\varepsilon; f) = \rho_z(\varepsilon; f)\left(1 \wedge \sqrt{n\rho_z(\varepsilon; f)}\right), \text{ for } f \in \mathcal{F}. \tag{4.11}$$

Theorem 4.2 shows the best separable hyper cubes can achieve for each $\mathbf{f} \in \mathcal{F}_s$.

THEOREM 4.2 (Confidence Hyper Cube Lower Bound). *Let $\varphi(\varepsilon; f)$ be defined in (4.11). For $0 < \alpha \leq 0.1$, suppose $CI \in \mathcal{SI}_\alpha(\mathcal{F}_s)$ defined in (4.6), then for $\mathbf{f} \in \mathcal{F}_s$, we have*

$$\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right) \geq (1-2\alpha)\mathfrak{D}(\mathbf{f};n), \tag{4.12}$$

and

$$\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right) \geq c_4 \inf_{\sum_{k=1}^{s}\alpha_k=2\alpha;\alpha\geq\alpha_i\geq 0,\forall k} \sum_{k=1}^{s} \varphi(z_{\alpha_i}\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2, \tag{4.13}$$

where $c_4 > 0$ is an absolute positive constant.

Now we extract two quantities

$$\Psi_1(\mathbf{f};\alpha) = \sum_{k=1}^{s} \varphi(z_\alpha \frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2, \Psi_2(\mathbf{f};\alpha) = \sum_{k=1}^{s} \varphi(z_{\alpha/s} \frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2, \tag{4.14}$$

where $\varphi(\cdot,\cdot)$ is defined in (4.11).

Theorem 4.3 shows that $\Psi_1(\mathbf{f};\alpha)$ can not be outperformed (up to a multiple of constant) at any $\mathbf{f} \in \mathcal{F}_s$, that $\Psi_1(\mathbf{f};\alpha)$ is not adaptively achievable, and that $\Psi_2(\mathbf{f};\alpha)$ can not be outperformed (up to a multiple of constant) uniformly.

THEOREM 4.3. *Let $\Psi_1(\mathbf{f};\alpha)$ and $\Psi_2(\mathbf{f};\alpha)$ be defined in* (4.14). *Suppose $0 < \alpha \leq 0.2$, and $CI \in \mathcal{SI}_\alpha(\mathcal{F}_s)$. Then we have*

$$\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right) \geq c_4 \cdot \Psi_1(\mathbf{f};\alpha),\ \textit{for all } \mathbf{f} \in \mathcal{F}_s, \tag{4.15}$$

$$\sup_{\mathbf{f}\in\mathcal{F}_s} \frac{\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right)}{\Psi_1(\mathbf{f};\alpha)} \geq c_5 \left(\frac{z_{\alpha/s}}{z_\alpha}\right)^2, \tag{4.16}$$

and

$$\sup_{\mathbf{f}\in\mathcal{F}_s} \frac{\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right)}{\Psi_2(\mathbf{f};\alpha)} \geq c_5, \tag{4.17}$$

where c_5 is a positive absolute constant, c_4 is the same as in Theorem 4.2.

Hence apart from discretization error, optimal method can be expected to achieve $\Gamma_2(\mathbf{f};\alpha)$ but not $\Gamma_1(\mathbf{f};\alpha)$. Also note that $\Psi_2(\mathbf{f};\alpha) \leq c_6 \log{(s)} \Psi_1(\mathbf{f};\alpha)$ for some absolute positive constant c_6 for $\alpha \leq 0.2$. So the difference between $\Psi_2(\mathbf{f};\alpha)$ and $\Psi_1(\mathbf{f};\alpha)$ is small.

4.2. *Separable Adaptive Procedures.* Similar to the white noise model, we split the data into three independent copies and then construct the estimators and confidence intervals for $Z(\mathbf{f})$ for $\mathbf{f} \in \mathcal{F}_s$ in three major steps: localization, stopping, and estimation/inference.

4.2.1. *Data Splitting.* Let $z_{k,i}^j \overset{i.i.d}{\sim} N(0,1)$, with $1 \le k \le s$, $1 \le i \le n$, $1 \le j \le 2$.

For each $1 \le k \le s$, we construct the following three sequences based on $\boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})$:

(4.18)

$$\nu_{k,i}^l = \boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})_i + \frac{\sigma}{(n+1)^{\frac{s-1}{2}}} \left\{ \frac{\sqrt{2}}{2} \left(z_{k,i}^1 - \frac{\sum_{l=0}^n z_{k,l}^1}{n+1} \right) + \frac{\sqrt{6}}{2} \left(z_{k,i}^2 - \frac{\sum_{l=0}^n z_{k,l}^2}{n+1} \right) \right\},$$
$$\nu_{k,i}^r = \boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})_i + \frac{\sigma}{(n+1)^{\frac{s-1}{2}}} \left\{ \frac{\sqrt{2}}{2} \left(z_{k,i}^1 - \frac{\sum_{l=0}^n z_{k,l}^1}{n+1} \right) - \frac{\sqrt{6}}{2} \left(z_{k,i}^2 - \frac{\sum_{l=0}^n z_{k,l}^2}{n+1} \right) \right\},$$
$$\nu_{k,i}^e = \boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})_i - \frac{\sigma}{(n+1)^{\frac{s-1}{2}}} \sqrt{2} \left(z_{k,i}^1 - \frac{\sum_{l=0}^n z_{k,l}^1}{n+1} \right),$$

for $i = 0, \cdots, n$. For convenience, let $\nu_{k,i}^l = \nu_{k,i}^r = \nu_{k,i}^e = \infty$ for $i \notin \{0, 1, \cdots, n\}$. It is easy to see that the three copies are independent, and the s collections of the three copies are also independent. For each k, we will use $\{\nu_{k,\cdot}^l\}$ for localization, $\{\nu_{k,\cdot}^r\}$ for stopping rule, and $\{\nu_{k,\cdot}^e\}$ for construction of the final estimation and inference procedures.

Let $J = \lfloor \log_2(n+1) \rfloor$. For $j = 0, 1, \cdots, J$, $i = 1, 2, \cdots, \lfloor \frac{n+1}{2^{J-j-1}} \rfloor$, the i-th block at level j consists of $\{\frac{(i-1)2^{J-j}}{n}, \frac{(i-1)2^{J-j}+1}{n}, \frac{i \cdot 2^{J-j}-1}{n}\}$. Denote the sum of observations in the i-th block at level j for the axis k, sequence u (u=l,r,e) as

$$\mathtt{Y}_{k,j,i}^u = \sum_{h=(i-1)2^{J-j}}^{i \cdot 2^{J-j}-1} \nu_{k,h}^u. \tag{4.19}$$

Again, let $\mathtt{Y}_{k,j,i}^u = +\infty$ when $i \notin \{1, 2, \cdots, \lfloor \frac{n+1}{2^{J-j-1}} \rfloor\}$ for $k \in \{1, 2, \cdots, s\}$, $u \in \{l, r, e\}$, $j \in \{0, 1, \cdots, J\}$.

4.2.2. *Localization.* For k-th axis, we use $\{\nu_{k,h}^l, h \in \{0, 1, \cdots, n\}\}$ to construct a localization procedure. Let $\hat{\mathtt{i}}_{k,0} = 1$, and for $j = 1, 2, \cdots, J$, let

$$\hat{\mathtt{i}}_{k,j} = \operatorname*{arg\,min}_{\max\{2\hat{\mathtt{i}}_{k,j-1}-2,1\} \le i \le \min\{2\hat{\mathtt{i}}_{k,j-1}+1, \lfloor \frac{n+1}{2^{J-j}} \rfloor\}} \mathtt{Y}_{k,j,i}^l. \tag{4.20}$$

This is similar to the localization step in the white noise model. In each iteration, the blocks at the previous level are split into two sub-blocks. The i-th block at level $j-1$ is split into two blocks, the $(2i-1)$-th block and the $2i$-th block, at level j. For a given $\hat{\mathtt{i}}_{k,j-1}$, $\hat{\mathtt{i}}_{k,j}$ is the sub-block with the smallest sum (i.e. $\mathtt{Y}_{k,j,i}^l$) among the two sub-blocks of $\hat{\mathtt{i}}_{k,j-1}$ and their immediate neighboring sub-blocks.

4.2.3. *Stopping Rule.* Similar to the stopping rule for the white noise model, define the statistic $\mathtt{T}_{k,j}$ as

$$\mathtt{T}_{k,j} = \min\{\mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}+6} - \mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}+5}, \mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}-6} - \mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}-5}\}.$$

Let $\tilde{\sigma}^2_{k,j} = 6 \times 2^{J-j} \times \frac{\sigma^2}{(n+1)^{s-1}}$. It is easy to see that when $\mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}+6} - \mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}+5} < \infty$,

$$\mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}+6} - \mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}+5} \Big| \hat{\mathtt{i}}_{k,j} \sim N\left(\sum_{h=(\hat{\mathtt{i}}_{k,j}+4)2^{J-j}}^{(\hat{\mathtt{i}}_{k,j}+5)2^{J-j}-1} \left(f_k(\frac{h+2^{J-j}}{n}) - f_k(\frac{h}{n})\right), \tilde{\sigma}^2_{k,j}\right). \tag{4.21}$$

Similar to white noise model, we define a series of stopping rules controlled by a parameter ζ.

Define

$$\check{\mathtt{j}}_k(\zeta) = \begin{cases} \min\{j : \mathtt{T}_{k,j} \le z_\zeta \tilde{\sigma}_{k,j}\} & \text{if } \{j : \mathtt{T}_{k,j} \le z_\zeta \tilde{\sigma}_{k,j}\} \cap \{0, 1, 2, \cdots, J\} \neq \emptyset \\ \infty & \text{otherwise} \end{cases}$$

and terminate the algorithm at level $\hat{\mathtt{j}}_k(\zeta) = \min\{J, \check{\mathtt{j}}_k(\zeta)\}$. So either either $\mathtt{T}_{k,j}$ triggers the stopping for some $0 \le j \le J$ or the algorithm reaches the highest possible level J.

With the localization strategy and the stopping rule, the final block, the $\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\zeta)}$-th block at level $\hat{\mathtt{j}}_k(\zeta)$ is given by

$$\{\frac{h}{n} : (\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\zeta)} - 1)2^{J-\hat{\mathtt{j}}_k(\zeta)} \le h \le \hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\zeta)} 2^{J-\hat{\mathtt{j}}_k(\zeta)} - 1\}.$$

4.2.4. *Estimation and Inference.* After we have, for each axis $k \in \{1, 2, \cdots, s\}$, our stopping step precursor $\check{\mathtt{j}}_k(\zeta)$, stopping step $\hat{\mathtt{j}}_k(\zeta)$, index associated with the stopping step $\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\zeta)}$, and the final block, we use them to construct estimator and confidence hyper cube for the minimizer of $\mathbf{f} \in \mathcal{F}_s$.

For estimation, let $\zeta = \Phi(-2)$. The k-th coordinate of $\hat{Z}$, $\hat{Z}_k$, is defined as

$$\hat{Z}_k = \begin{cases} -\frac{1}{2n} + \frac{1}{n}\left(2^{J-\hat{\mathtt{j}}_k(\zeta)} - 2^{J-\hat{\mathtt{j}}_k(\zeta)-1}\right), & \check{\mathtt{j}}_k(\zeta) < \infty \\ \frac{1}{n} \underset{\hat{\mathtt{i}}_{k,J}-2 \le i \le \hat{\mathtt{i}}_{k,J}+2}{\arg\min} \nu^e_{k,i-1} - \frac{1}{n}, & \check{\mathtt{j}}_k(\zeta) = \infty \end{cases}. \tag{4.22}$$

The final estimator $\hat{Z}$ is defined as

$$\hat{Z} = (\hat{Z}_1, \hat{Z}_2, \cdots, \hat{Z}_s), \tag{4.23}$$

where $\hat{Z}_k$ is defined in (4.22) for $k \in \{1, 2, \cdots, s\}$.

To construct the confidence hyper cube for $Z(\mathbf{f})$, for each axis $k \in \{1, \dots, s\}$, we set the parameter for stopping rule to be $\zeta_k = \alpha/2s$ and take a few adjacent blocks to the left and right of $\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\zeta_k)}$-th block at level $\hat{\mathtt{j}}_k(\zeta_k)$.

Let

$$L_k = \max\{0, \hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\alpha/2s)} - 7\}, U_k = \min\{\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\alpha/2s)} + 6, \lceil (n+1)2^{\hat{\mathtt{j}}_k(\alpha/2s)-J} \rceil\}.$$

When $\check{\mathtt{j}}_k(\alpha/2s) < \infty$, let

$$t_{k,lo} = \frac{2^{J-\hat{\mathtt{j}}_k(\alpha/2s)}}{n} L - \frac{1}{2n}, t_{k,hi} = \min\{\frac{2^{J-\hat{\mathtt{j}}_k(\alpha/2s)}}{n} - \frac{1}{2n}, 1\}. \tag{4.24}$$

When $\check{\mathtt{j}}_k(\alpha/2s) = \infty$, $t_{k,lo}$ and $t_{k,hi}$ are calculated by the following Algorithm 1.

As $\check{\mathtt{j}}_k(\alpha/2s) = \infty$ means that the $\mathtt{T}_{k,j}$ never triggers the stopping, which is a strong indicator that the signal is strong and discretization error could dominate. Algorithm 1 first specifies a range that the minimizer lies in with high probability (e.g. $1 - \alpha/2s$), and then shrinks the interval to locate the minimizer among the grid points within the original interval. After this step, the minimizer(s) among the grids are in the shrunk interval still with high probability (e.g. $1 - 3\alpha/4s$). Then in the case that shrunk interval detects only one grid-wise minimizer (i_m/n) and this minimizer does not indicates a discretizatino error larger or equal than $1/n$, we use a geometry property of convex functions to determine the final interval. Basically, the right most possible minimizer is or infinitely near the intersection of two lines : $y = f(i_m/n)$, and that joining $(\frac{i_m+1}{n}, f(\frac{i_m+1}{n}))$ with $(\frac{i_m+2}{n}, f(\frac{i_m+2}{n}))$. With observation $\nu^e_{k,i_m}, \nu^e_{k,i_m+1}, \nu^e_{k,i_m+2}$, we can infer the intersection of the aforementioned two lines and specify the right end point of the interval accordingly.

The k-th axis of confidence hyper cube CI_α is given by

$$CI_{k,\alpha} = [t_{k,lo}, t_{k,hi}]. \tag{4.25}$$

The $(1-\alpha)$-level confidence hyber cube CI_α is given by

$$CI_\alpha = CI_{1,\alpha} \times CI_{2,\alpha} \times \cdots CI_{s,\alpha}, \tag{4.26}$$

where $CI_{k,\alpha}$ is defined in (4.25).

By the definition of our estimator $\hat{Z}$ and confidence hyper cube CI_α, it's easy to check that they are separable methods.

Algorithm 1 Computing $t_{k,lo}$ and $t_{k,hi}$ when $\check{\mathtt{j}}_k(=)\infty$

$L_k \leftarrow \max\{0, \hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\alpha/2s)} - 8\}, U_k = \min\{n, \hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\alpha/2s)} + 6\}$, $\alpha_1 = \alpha/8s$, $\alpha_2 = \alpha/24s$

Generate $z^3_{k,0}, z^3_{k,2}, \cdots, , z^3_{k,n} \overset{i.i.e}{\sim} N(0,1)$

$$i_l \leftarrow \min\{\{U\} \cup \{i \in [L, U-1] : \nu^e_{k,i} - \nu^e_{k,i+1} + \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}} \left(z^3_{k,i} - z^3_{k,i+1} - 2z_{\alpha_1}\right) \le 0$$

$$i_r \leftarrow \max\{\{L-1\} \cup \{i \in [L, U-1] : \nu^e_{k,i} - \nu^e_{k,i+1} + \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}} \left(z^3_{k,i} - z^3_{k,i+1} + 2z_{\alpha_1}\right) \ge 0$$

if $i_l \le i_r$ **then**
  $t_{k,lo} = \max\{0, \frac{i_l - 1}{n}\}$, $t_{k,hi} = \max\{1, \frac{i_r+2}{n}\}$
end if
if $i_l = i_r + 1$ and $i_l \le n-2$ **then**
  if $\nu^e_{k,i_l+2} - \nu^e_{k,i_l+1} - \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}} \left(z^3_{k,i_l+2} - z^3_{k,i_l+1} - 2\sqrt{2}z_{\alpha_2}\right) > 0$ **then**

$$t_{hi} \leftarrow \left(\left(\frac{\nu^e_{k,i_l} - \nu^e_{k,i_l+1} - \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}} \left(z^3_{k,i_l} - z^3_{k,i_l+1} - 2\sqrt{2}z_{\alpha_2}\right)}{n\left(\nu^e_{k,i_l+2} - \nu^e_{k,i_l+1} - \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}} \left(z^3_{k,i_l+2} - z^3_{k,i_l+1} - 2\sqrt{2}z_{\alpha_2}\right)\right)} + \frac{1}{n}\right)_+ + \frac{i_l}{n}\right) \wedge \frac{i_l+1}{n}$$

  else
    $t_{hi} \leftarrow \frac{i_l}{n}$
  end if
end if
if $i_l = i_r + 1$ and $i_l \ge n-1$ **then**
  $t_{k,hi} = 1$
end if
if $i_l = i_r + 1$ and $i_l \ge 2$ **then**
  if $\nu^e_{k,i_l-2} - \nu^e_{k,i_l-1} - \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}} \left(z^3_{k,i_l-2} - z^3_{k,i_l-1} - 2\sqrt{2}z_{\alpha_2}\right) > 0$ **then**

$$t_{k,lo} \leftarrow \left(\left(-\frac{\nu^e_{k,i_l} - \nu^e_{k,i_l-1} - \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}} \left(z^3_{k,i_l} - z^3_{k,i_l-1} - 2\sqrt{2}z_{\alpha_2}\right)}{n\left(\nu^e_{k,i_l-2} - \nu^e_{k,i_l-1} - \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}} \left(z^3_{k,i_l-2} - z^3_{k,i_l-1} - 2\sqrt{2}z_{\alpha_2}\right)\right)}\right)_+ + \frac{i_l-1}{n}\right) \wedge \frac{i_l}{n}$$

  else
    $t_{k,lo} \leftarrow \frac{i_l}{n}$
  end if
end if
if $i_l = i_r + 1$ and $i_l \le 1$ **then**
  $t_{k,lo} = 0$
end if

4.3. *Statistical Optimality.* Now we establish the optimality of the adaptive procedures constructed in Section 4.2. The results show that our data-driven procedures are simultaneously optimal (up to constant factor) for all $f \in \mathcal{F}_s$, in terms of the tag quantities and discretization error introduced in Section 4.1. We start with the estimation.

THEOREM 4.4 (Estimation). *For $\psi(\mathbf{f})$ defined in (4.8), the estimator $\hat{Z}$ defined in (4.23) satisfies*

$$\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z} - Z(\mathbf{f})\|^2\right) \le C_4\psi(\mathbf{f}) + 2\mathfrak{D}(\mathbf{f}; n), \text{ for all } \mathbf{f} \in \mathcal{F}_s \tag{4.27}$$

where C_4 is an absolute positive constant.

The following holds for the separable confidence hyper cube CI_α.

THEOREM 4.5 (Inference). *For $0 < \alpha \le 0.3$, confidence cube CI_α defined in (4.26) is a $(1-\alpha)$-level confidence cube for the minimizer $Z(\mathbf{f})$ and its expected square diameter satisfies*

$$\mathbb{E}_{\mathbf{f}}\left(diag\,(CI_\alpha)^2\right) \le C_5\Psi(\mathbf{f}; \alpha) + 9\mathfrak{D}(\mathbf{f}; n), \text{ for all } \mathbf{f} \in \mathcal{F}_s \tag{4.28}$$

where C_5 is an absolute positive constant.

5. Discussion. In the present paper, we studied optimal estimation and inference for the minimizer of multivariate additive function in the white noise and nonparametric regression models within the class of separable methods under non-asymptotic benchmarks that characterize the difficulty of the statistical problem at individual functions. We have shown that local minimax framework (Cai and Low, 2015), unlike univariate case, does not fully captures the difficulty of estimation/inference problem in multivariate case for entire method class: local minimax rates are shown to be not adaptively achievable. We found an information-preserving representation of the observation, projection representation, and we focus on separable methods that are based on the representation. We turn to a definition-free framework that resorts to the fundamental link between benchmarks (tags) and the performance of the methods. These benchmarks are function-specific and can be easily transformed into rates of conventional minimax framework. This provides a way to characterize the difficulty of statistical problem locally in addition to local minimax framework, and also enlarge the meaning of minimax: we can add an variable denominator. It would be interesting to see how the local characterization discussed in paper works for problems where the difficulty for the statistical problem at different function varies or when we have different affordability for the price to pay at different function.

We also developed adaptively optimal procedures with respect to our benchmarks. Although some blocks of it looks similar to univariate case, no direct extension of the procedure of the univariate case can achieve the optimal rate for confidence hyper cube, it would have an additional multiplier of power function of dimension s.

The present work can be extended in different directions. We only consider multivariate additive functions, it would be interesting to investigate high-dimensional sparse additive functions with convexity constraints on each component function, and it would also be interesting to investigate general multivariate case. In our work, we focus on separable methods, it would be interesting to investigate the entire method class and see how they compare.

6. Proof.

6.1. *Notation.* Here we recollect or introduce notation that will be used later. We use $Z(f)$, $M(f)$ to denote the minimizer and minimum of function f, where f can be univariate or multivariate.

Recall that

(6.1)
$$\rho_m(\varepsilon; f) = \max\{\rho : \int_0^1 (\max\{\rho, f(t)\} - f(t))^2 \, \mathrm{d}t \le \varepsilon^2\} - M(f). \rho_z(\varepsilon; f) = \max\{|t - Z(f)| : f(t) \le \rho_m(\varepsilon;$$

for $f \in \mathcal{F}$.

6.2. *Proof of Theorem 2.1.* For the ease of notation, denote $\mathcal{D}$ to be $[0,1]^s$.

We start with lower bounds.

Let $\mathbf{f} \in \mathcal{F}_s$. Let $\mathbf{g} \in \mathcal{F}_s$, which we will specify later. Take $\theta \in \{-1, 1\}$ as parameter to be estimated, with $\mathbf{f}_1 = \mathbf{f}$ and $\mathbf{f}_{-1} = \mathbf{g}$.

For any estimator $\hat{Z}$ for estimating the minimizer, consider the projected estimator that projects $\hat{Z}$ to the line determined by $Z(\mathbf{f})$ and $Z(\mathbf{g})$:

$$\hat{Z}_p = Z(\mathbf{f}) + \langle \hat{Z} - Z(\mathbf{f}), \frac{Z(\mathbf{g}) - Z(\mathbf{f})}{\|Z(\mathbf{f}) - Z(\mathbf{g})\|} \rangle. \tag{6.2}$$

It's easy to see that

$$E_{\mathbf{f}} \left(\|\hat{Z}_p - Z(\mathbf{f})\|^2 \right) \le E_{\mathbf{f}} \left(\|\hat{Z} - Z(\mathbf{f})\|^2 \right)$$

and

$$E_{\mathbf{g}} \left(\|\hat{Z}_p - Z(\mathbf{g})\|^2 \right) \le E_{\mathbf{g}} \left(\|\hat{Z} - Z(\mathbf{g})\|^2 \right).$$

Therefore, we only need to consider the projected estimators $\hat{Z}_p$ for calculating $R(\varepsilon;\mathbf{f})$. Similarly, we only need to consider projected confidence hypercube CI_p is the smallest hypercube containing $\{Z(\mathbf{f})+\langle \mathbf{t}-Z(\mathbf{f}), \frac{Z(\mathbf{g})-Z(\mathbf{f})}{\|Z(\mathbf{g})-Z(\mathbf{f})\|}\rangle : \mathbf{t}\in CI\}$ for calculating $L_\alpha(\varepsilon;\mathbf{f})$, as projection does not weaken confidence level and projected hypercube has smaller hypercube-diameter.

Note that any projected estimator $\hat{Z}_p$ of the minimizer $Z(\mathbf{f}_\theta)$ gives an estimator of θ by

$$\hat{\theta} = \langle \frac{\hat{Z}_p - \frac{Z_p(\mathbf{f}_1)+Z_p(\mathbf{f}_{-1})}{2}}{\|\frac{Z_p(\mathbf{f}_1)-Z_p(\mathbf{f}_{-1})}{2}\|}, \frac{Z_p(\mathbf{f}_1)-Z_p(\mathbf{f}_{-1})}{\|Z_p(\mathbf{f}_1)-Z_p(\mathbf{f}_{-1})\|}\rangle,$$

and therefore $\mathbb{E}_\theta\|\hat{Z}_p - Z(\mathbf{f}_\theta)\|^2 = \|Z(\mathbf{f}_1)-Z(\mathbf{f}_{-1})\|^2\mathbb{E}_\theta\frac{|\hat{\theta}-\theta|}{2}$. Let $\mathbb{P}_\theta$ be the probability measure associated with the white noise model corresponding to $\mathbf{f}_\theta$. On the other hand, through calculating the Radon-Nikodym derivative $\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_{-1}}(Y)$, a sufficient statistic for θ is given by

$$W = \frac{\int_{\mathcal{D}}(\mathbf{f}_1(\mathbf{t})-\mathbf{f}_{-1}(\mathbf{t}))\mathrm{d}Y(\mathbf{t}) - \frac{1}{2}\int_{\mathcal{D}}(\mathbf{f}_1(\mathbf{t})^2-\mathbf{f}_{-1}(\mathbf{t})^2)\mathrm{d}\mathbf{t}}{\varepsilon\|\mathbf{f}_1-\mathbf{f}_{-1}\|}. \tag{6.3}$$

Then

$$W \sim N\left(\frac{\theta}{2}\cdot\frac{\|\mathbf{f}_1-\mathbf{f}_{-1}\|}{\varepsilon}, 1\right) \quad \text{under } \mathbb{P}_\theta.$$

Note that for any $\omega(\varepsilon;\mathbf{f}) > \delta > 0$, there exists $\mathbf{h}_\delta \in \mathcal{F}_s$ such that $\|\mathbf{f}-\mathbf{h}_\delta\| = \varepsilon$ and that $\|Z(\mathbf{f})-Z(\mathbf{h}_\delta)\|^2 \geq \omega(\varepsilon;\mathbf{f})-\delta$, we let $\mathbf{g}=\mathbf{h}_\delta$. Then we have $R(\varepsilon;f) \geq (\omega(\varepsilon;\mathbf{f})-\delta)\cdot r_1$, where r_1 is the minimax risk of the two-point problem based on an observation $X \sim N(\frac{\theta}{2},1)$,

$$r_1 = \inf_{\hat{\theta}}\max_{\theta=\pm 1}\mathbb{E}_\theta\frac{|\hat{\theta}-\theta|}{2}.$$

It is easy to see that $r_1 = \Phi(-0.5)$. Taking $\delta \to 0^+$, we have $R(\varepsilon;\mathbf{f}) \geq \Phi(-0.5)\omega(\varepsilon;\mathbf{f})$. So we have $a \geq \Phi(-0.5) \approx 0.309$.

Next, we show for $0<\alpha<0.3$ that $L_\alpha(\varepsilon;\mathbf{f}) \geq b_\alpha\omega(\varepsilon/3;\mathbf{f})$ where $b_\alpha = 0.6-2\alpha$.

We begin by recalling a lemma from Cai and Guo (2017) that can be easily modified to expected square diameter version.

Lemma 6.1 (Cai and Guo, 2017). *For any $CI \in \mathcal{I}_\alpha(\{\mathbf{f},\mathbf{g}\})$,*

$$\mathbb{E}_{\mathbf{f}} diag\,(CI)^2 \geq \|Z(\mathbf{f})-Z(\mathbf{g})\|^2(1-2\alpha-TV(P_{\mathbf{f}},P_{\mathbf{g}})),$$

where TV *denotes the total variation distance between the two distributions of the white noise models corresponding to* $\mathbf{f}$ *and* $\mathbf{g}$.

Again let $\mathbf{g} \in \mathcal{F}_s$. Then for $CI \in \mathcal{I}_\alpha(\{\mathbf{f}, \mathbf{g}\})$, by Lemma 6.1,

$$\mathbb{E}_{\mathbf{f}} diag\,(CI)^2 \geq |Z(\mathbf{f}) - Z(\mathbf{g})|^2 (1 - 2\alpha - \mathrm{TV}(P_{\mathbf{f}}, P_{\mathbf{g}})).$$

It is well known that $\mathrm{TV}(P_{\mathbf{f}}, P_{\mathbf{g}}) \leq \sqrt{\chi^2(P_{\mathbf{f}}, P_{\mathbf{g}})}$, where

$$\chi^2(P_{\mathbf{f}}, P_{\mathbf{g}}) = \int \left(\frac{dP_{\mathbf{f}}}{dP_{\mathbf{g}}}\right)^2 dP_{\mathbf{g}} - 1$$

is the χ^2 distance between $P_{\mathbf{f}}$ and $P_{\mathbf{g}}$. By Girsanov's theorem we can obtain the likelihood ratio

$$\frac{dP_{\mathbf{f}}}{dP_{\mathbf{g}}}(Y) = \exp\left(\int_{\mathcal{D}} \frac{\mathbf{f}(\mathbf{t}) - \mathbf{g}(\mathbf{t})}{\varepsilon^2} dY(\mathbf{t}) - \frac{1}{2}\int_{\mathcal{D}} \frac{\mathbf{f}(\mathbf{t})^2 - \mathbf{g}(\mathbf{t})^2}{\varepsilon^2} \mathrm{d}\mathbf{t}\right), \tag{6.4}$$

and hence

$$\begin{aligned}
\chi^2(P_{\mathbf{f}}, P_{\mathbf{g}}) &= \int_{\mathcal{D}} \exp\left(2\int_{\mathcal{D}} \frac{\mathbf{f}(\mathbf{t}) - \mathbf{g}(\mathbf{t})}{\varepsilon^2} dY(\mathbf{t}) - \int_{\mathcal{D}} \frac{\mathbf{f}(\mathbf{t})^2 - \mathbf{g}(\mathbf{t})^2}{\varepsilon^2} \mathrm{d}\mathbf{t}\right) \mathrm{d}P_{\mathbf{g}} - 1 \\
&= \exp\left(-\frac{\|\mathbf{f} - \mathbf{g}\|^2}{\varepsilon^2}\right) \mathbb{E} \exp\left(2\int_{\mathcal{D}} \frac{\mathbf{f}(\mathbf{t}) - \mathbf{g}(\mathbf{t})}{\varepsilon} \mathrm{d}\mathbf{W}(\mathbf{t})\right) - 1 \\
&= \exp\left(\frac{\|\mathbf{f} - \mathbf{g}\|^2}{\varepsilon^2}\right) - 1.
\end{aligned}$$

Using it to bound the total variation distance, we get

$$\mathbb{E}_{\mathbf{f}} diag\,(CI)^2 \geq \|Z(\mathbf{f}) - Z(\mathbf{g})\|^2 \left(1 - 2\alpha - \sqrt{\exp\left(\frac{\|\mathbf{f} - \mathbf{g}\|^2}{\varepsilon^2}\right) - 1}\right).$$

We continue by specifying $\mathbf{g}$. For any $\omega(\varepsilon/3; \mathbf{f}) > \delta > 0$, picking $\mathbf{g} = \mathbf{g}_\delta \in \mathcal{F}_s$ such that $\|\mathbf{f} - \mathbf{g}_\delta\| = \varepsilon/3$ and $|Z(\mathbf{f}) - Z(\mathbf{g}_\delta)| \geq \omega(\varepsilon/3; \mathbf{f}) - \delta$, we have $\mathbb{E}_{\mathbf{f}} diag\,(CI)^2 \geq (0.6 - 2\alpha)\,(\omega(\varepsilon/3; \mathbf{f}) - \delta)$. By taking $\delta \to 0^+$, we have

$$L_\alpha(\varepsilon; \mathbf{f}) \geq (0.6 - 2\alpha)\,\omega(\varepsilon/3; \mathbf{f}).$$

Now we turn to the upper bounds. We start with stating a property of $\omega(\varepsilon; \mathbf{f})$ in Proposition 6.1.

PROPOSITION 6.1. *Suppose* $\mathbf{f} \in \mathcal{F}_s$, $c \in (0, 1)$, *then we have*

$$\omega(\varepsilon; \mathbf{f}) \geq \omega(c\varepsilon; \mathbf{f}) \geq \frac{1}{9} \max\left\{(\frac{c}{2})^{\frac{2}{3}}, c\right\} \omega(\varepsilon; \mathbf{f}). \tag{6.5}$$

PROOF. The left hand side is apparent, we will prove the right hand side. Recalling Proposition 6.2, we have

$$\begin{aligned}\sup_{\sum_{i=1}^s b_i^2\le 1}\sum_{i=1}^s \rho_z(b_ic\varepsilon;f_i)^2 \le \omega(c\varepsilon;\mathbf{f}) \le 9\sup_{\sum_{i=1}^s b_i^2\le 1}\sum_{i=1}^s \rho_z(b_ic\varepsilon;f_i)^2,\\ \sup_{\sum_{i=1}^s b_i^2\le 1}\sum_{i=1}^s \rho_z(b_i\varepsilon;f_i)^2 \le \omega(\varepsilon;\mathbf{f}) \le 9\sup_{\sum_{i=1}^s b_i^2\le 1}\sum_{i=1}^s \rho_z(b_i\varepsilon;f_i)^2.\end{aligned} \tag{6.6}$$

Using Proposition **??** in Cai et al. (2022), namely

$$\max\left\{(\frac{q}{2})^{\frac{2}{3}},q\right\} \le \frac{\rho_z(q\varepsilon;f)}{\rho_z(\varepsilon;f)} \le 1,\ \text{for } q\in[0,1)$$

, we know $\rho_z(\epsilon;f)$ is a continuous function of $\epsilon\ge 0$ for $f\in\mathcal{F}$. So there exists $(\tilde b_1,\cdots,\tilde b_s)$ and $(\bar b_1,\cdots,\bar b_s)$ attaining the suprema:

(6.7)

$$\begin{aligned}&\tilde b_i\ge 0,\ \text{for } 1\le i\le s, \sum_{i=1}^s \tilde b_i^2 = 1, \sum_{i=1}^s \rho_z(\tilde b_ic\varepsilon;f_i)^2 = \sup_{\sum_{i=1}^s b_i^2\le 1}\sum_{i=1}^s \rho_z(b_ic\varepsilon;f_i)^2,\\ &\bar b_i\ge 0,\ \text{for } 1\le i\le s, \sum_{i=1}^s \bar b_i^2 = 1, \sum_{i=1}^s \rho_z(\bar b_i\varepsilon;f_i)^2 = \sup_{\sum_{i=1}^s b_i^2\le 1}\sum_{i=1}^s \rho_z(b_i\varepsilon;f_i)^2.\end{aligned}$$

Also we have

$$\sum_{i=1}^s \rho_z(\tilde b_ic\varepsilon;f_i)^2 \le \sum_{i=1}^s \rho_z(\tilde b_i\varepsilon;f_i)^2 \le \sum_{i=1}^s \rho_z(\bar b_i\varepsilon;f_i)^2, \tag{6.8}$$

and

$$\sum_{i=1}^s \rho_z(\tilde b_ic\varepsilon;f_i)^2 \ge \sum_{i=1}^s \rho_z(\bar b_ic\varepsilon;f_i)^2 \ge \sum_{i=1}^s \max\left\{(\frac{c}{2})^{\frac{2}{3}},c\right\}\rho_z(\bar b_i\varepsilon;f_i)^2. \tag{6.9}$$

Combining equations (6.6), (6.8), (6.9) we have

$$\omega(c\varepsilon;\mathbf{f}) \ge \frac{1}{9}\max\left\{(\frac{c}{2})^{\frac{2}{3}},c\right\}\omega(\varepsilon;\mathbf{f}). \tag{6.10}$$

□

Now we continue with the upper bounds.

Recalling W define in (6.3), let

$$\hat Z = sign(W)\cdot\frac{Z(\mathbf{f})-Z(\mathbf{g})}{2} + \frac{Z(\mathbf{f})+Z(\mathbf{g})}{2}. \tag{6.11}$$

Then

$$\mathbb{E}_{\mathbf{f}}(\|\hat{Z}-Z(\mathbf{f})\|^2) = \mathbb{E}_{\mathbf{g}}(\|\hat{Z}-Z(\mathbf{g})\|^2) = \|Z(\mathbf{f})-Z(\mathbf{g})\|^2\Phi(-\frac{\|\mathbf{f}-\mathbf{g}\|}{2\varepsilon}). \tag{6.12}$$

Therefore,

$$\begin{aligned} R(\varepsilon; f) &\le \sup_{f\in\mathcal{F}_s} \|Z(f)-Z(g)\|^2\Phi(-\frac{\|\mathbf{f}-\mathbf{g}\|}{2\varepsilon}) \\ &\le \sup_{c>0}\omega(c\varepsilon;\mathbf{f})\Phi(-\frac{c}{2}) \\ &\le \max\{0.5\omega(\varepsilon;\mathbf{f}), \sup_{c\ge 1}\omega(c\varepsilon;\mathbf{f})\Phi(-\frac{c}{2})\}. \end{aligned} \tag{6.13}$$

In addition

$$\sup_{c\ge 1}\omega(c\varepsilon;\mathbf{f})\Phi(-\frac{c}{2}) \le 9\sup_{c\ge 1}\min\{(2c)^{\frac{2}{3}}, c\}\Phi(-\frac{c}{2})\omega(\varepsilon;\mathbf{f}) \le 3.1\omega(\varepsilon;\mathbf{f}). \tag{6.14}$$

Take $A = 3.1$ gives the result.

For inference, let

$$CI_\alpha = \begin{cases} \{Z(\mathbf{g})\} & W < -z_\alpha + 0.5\frac{\|\mathbf{f}-\mathbf{g}\|}{\varepsilon} \\ \{Z(\mathbf{f})\} & W \ge (z_\alpha - \frac{\|\mathbf{f}-\mathbf{g}\|}{2\varepsilon}) \vee (-z_\alpha + \frac{\|\mathbf{f}-\mathbf{g}\|}{2\varepsilon}) \\ \{Z(\mathbf{f}) + (Z(\mathbf{g})-Z(\mathbf{f}))\cdot t : t\in[0,1]\} & \text{otherwise} \end{cases}.$$

Clearly, We have $P_{\mathbf{f}}(Z(\mathbf{f}) \notin CI_\alpha) \le \alpha, P_{\mathbf{g}}(Z(\mathbf{g}) \notin CI_\alpha) \le \alpha$. For the expected squared diameter, we have for $\theta \in \{-1, 1\}$,

$$\mathbb{E}_{\mathbf{f}_\theta} diag\,(CI_\alpha)^2 \le \|Z(\mathbf{f})-Z(\mathbf{g})\|^2\left(\Phi(z_\alpha - \frac{\|\mathbf{f}-\mathbf{g}\|}{\varepsilon}) - \alpha\right)_+. \tag{6.15}$$

Therefore

$$\begin{aligned} L_\alpha(\varepsilon;\mathbf{f}) &\le \sup_{\mathbf{g}\in\mathcal{F}_s}\|Z(\mathbf{f})-Z(\mathbf{g})\|^2\left(\Phi(z_\alpha-\frac{\|\mathbf{f}-\mathbf{g}\|}{\varepsilon})-\alpha\right)_+ \\ &\le \sup_{c>0}\omega(c\varepsilon;\mathbf{f})(\Phi(z_\alpha - c)-\alpha)_+ \\ &\le \max\{\omega(\varepsilon;\mathbf{f})(\Phi(z_\alpha)-\alpha)_+, \sup_{c>1}\omega(c\varepsilon;\mathbf{f})\,(\Phi(z_\alpha-c)-\alpha)_+\} \\ &\le \max\{\omega(\varepsilon;\mathbf{f})(1-2\alpha), \sup_{c>1}\omega(c\varepsilon;\mathbf{f})\,(\Phi(z_\alpha-c)-\alpha)_+\}. \end{aligned} \tag{6.16}$$

In addition note that for $0 < \alpha < 0.3$ we have $2z_\alpha > 1$, so we have

$$
\begin{aligned}
&\sup_{c>1} \omega(c\varepsilon; \mathbf{f}) \left(\Phi(z_\alpha - c) - \alpha\right)_+ \\
&\le 9\omega(\varepsilon; \mathbf{f}) \sup_{c>1} \min\{c, (2c)^{2/3}\}(\Phi(z_\alpha - c) - \alpha)_+ \\
&\le 9\omega(\varepsilon; \mathbf{f}) \max\left\{(1-2\alpha)\min\{z_\alpha, (2z_\alpha)^{2/3}\}\mathbb{1}\{z_\alpha \ge 1\}, (0.5-\alpha)\min\{2z_\alpha, (4z_\alpha)^{2/3}\}\right\} \\
&\le 9\omega(\varepsilon; \mathbf{f})(1-2\alpha)\min\{z_\alpha, (2z_\alpha)^{2/3}\}.
\end{aligned} \tag{6.17}
$$

Therefore, we have

$$
L_\alpha(\varepsilon; \mathbf{f}) \le 9(1-2\alpha)\min\{z_\alpha, (2z_\alpha)^{2/3}\}\omega(\varepsilon; \mathbf{f}). \tag{6.18}
$$

6.3. *Proof of Theorem 2.2.* We start with stating two propositions, which are proved later.

Proposition 6.2. *Let $\rho_z(\varepsilon; f)$ be defined in (2.5) for $f \in \mathcal{F}$, and let $\mathbf{f} \in \mathcal{F}_s$. Then*

$$
\sup_{\sum_{i=1}^s b_i^2 \le 1} \sum_{i=1}^s \rho_z(b_i\varepsilon; f_i)^2 \le \omega(\varepsilon; \mathbf{f}) \le \sup_{\sum_{i=1}^s b_i^2 \le 1} \sum_{i=1}^s 9\rho_z(b_i\varepsilon; f_i)^2, \tag{6.19}
$$

where b_i are non-negative.

Proposition 6.3. *Suppose $f_i \in \mathcal{F}$, for $i = 1, 2, \cdots, s$, then we have*

$$
\frac{1}{3}s^{-\frac{2}{3}} \sum_{i=1}^s \rho_z(\varepsilon; f_i)^2 \le \sup_{\sum_{i=1}^s b_i^2 \le 1} \sum_{i=1}^s \rho_z(b_i\varepsilon; f_i)^2 \le \sum_{i=1}^s \rho_z(\varepsilon; f_i)^2. \tag{6.20}
$$

And for any $\beta \le s$, exist $(f_1, \cdots, f_s)$ such that $\sum_{i=1}^s \rho_z(\varepsilon; f_i)^2 = \beta$ and

$$
\sup_{\sum_{i=1}^s b_i^2 \le 1} \sum_{i=1}^s \rho_z(b_i\varepsilon; f_i)^2 = s^{-\frac{2}{3}} \sum_{i=1}^s \rho_z(\varepsilon; f_i)^2. \tag{6.21}
$$

For $\beta \le s$, for any $\delta > 0$, there exist $(f_1, \cdots, f_s)$ such that $\sum_{i=1}^s \rho_z(\varepsilon; f_i)^2 = \beta$ and

$$
\sup_{\sum_{i=1}^s b_i^2 \le 1} \sum_{i=1}^s \rho_z(b_i\varepsilon; f_i)^2 \ge \sum_{i=1}^s \rho_z(\varepsilon; f_i)^2 - \delta. \tag{6.22}
$$

Inequality (6.20) in Proposition 6.3 and (6.19) in Proposition 6.2 implies Inequality 2.6 of Theorem 2.2.

Construct $\mathbf{f}(\mathbf{t}) = \sum_{i=1}^{s} \int_0^s f_i(x)\mathrm{d}x + \sum_{i=1}^{s} \left(f(t_i) - \int_0^s f_i(x)\mathrm{d}x\right)$ with f_i in Equation (6.21). Then together with the right hand side of Inequality (6.19) gives Inequality (2.7) of Theorem 2.2. Similar construct $\mathbf{f}$ with f_i in Inequality (6.22) with $\delta = \frac{\delta_0}{9}$ gives Inequality (2.8) in Theorem 2.2.

6.3.1. *Proof of Proposition 6.2.* Suppose $\mathbf{g} \in \mathcal{F}_s$, such that $\|\mathbf{g} - \mathbf{f}\| \leq \varepsilon$, $\mathbf{g}(\mathbf{t}) = g_0 + g_1(t_1) + g_2(t_2) + \cdots + g_s(t_s)$. Using the continuity of $\rho_z(\epsilon; f)$ with respect to ϵ implied by Proposition **??** in Cai et al. (2022), we know there exist $(\bar{b}_1, \bar{b}_2, \cdots, \bar{b}_s)$ such that

$$\bar{b}_i \geq 0, \text{ for } 1 \leq i \leq s, \sum_{i=1}^{s} \bar{b}_i^2 = 1, \sum_{i=1}^{s} \rho_z(\bar{b}_i\varepsilon; f_i)^2 = \sup_{\sum_{i=1}^{s} b_i^2 \leq 1} \sum_{i=1}^{s} \rho_z(b_i\varepsilon; f_i)^2. \tag{6.23}$$

We only need to prove

$$\sum_{i=1}^{s} \rho_z(\bar{b}_i\varepsilon; f_i)^2 \leq \omega(\varepsilon; \mathbf{f}) \leq \sum_{i=1}^{s} 9\rho_z(\bar{b}_i\varepsilon; f_i)^2. \tag{6.24}$$

We start with proving the upper bound.

Since $\|\mathbf{g} - \mathbf{f}\| \leq \varepsilon$, we have

$$\begin{aligned} \varepsilon^2 \geq \|\mathbf{f} - \mathbf{g}\|^2 &= \int_{\mathcal{D}} \left(f_0 - g_0 + \sum_{i=1}^{2} f_i(t_i) - g_i(t_i) \right)^2 \mathrm{d}\mathbf{t} \\ &= (f_0 - g_0)^2 + \sum_{i=1}^{s} \int_0^1 (f_i(t) - g_i(t))^2 \mathrm{d}t. \end{aligned} \tag{6.25}$$

Denote $\tilde{b}_i = \sqrt{\frac{\int_0^1 (f_i(t) - g_i(t))^2 \mathrm{d}t}{\varepsilon^2}}$ for $1 \leq i \leq s$, then we have $\sum_{i=1}^{s} \tilde{b}_i^2 = 1$.

Therefore, using Proposition **??** in Cai et al. (2022), we have

$$\|Z(\mathbf{f}) - Z(\mathbf{g})\|^2 = \sum_{i=1}^{s} |Z(f_i) - Z(g_i)|^2 \leq \sum_{i=1}^{s} 9\rho_z(\tilde{b}_i\varepsilon; f_i)^2 \leq \sum_{i=1}^{s} 9\rho_z(\bar{b}_i\varepsilon; f_i)^2. \tag{6.26}$$

For the lower bound, we construct a class of function $\mathbf{g}_\delta \in \mathcal{F}_s$, with $\frac{1}{2}\min_{1 \leq i \leq s} \rho_z(\bar{b}_i\varepsilon; f_i) > \delta > 0$. We construct the constant and components: $g_{\delta,i}$ for $0 \leq s$. Let $g_{\delta,0} = f_0$. For $1 \leq i \leq s$, suppose $x_{l,i}, x_{r,i}$ are left and right end points of the interval $\{x : f_i(x) \leq M(f_i) + \rho_m(\bar{b}_i\varepsilon; f_i)\}$. And without loss of generality, we assume $x_{r,i} = Z(f_i) + \rho_z(\bar{b}_i\varepsilon; f_i)$. Define univariate convex function $h_{\delta,i}$ as follow.

(6.27)
$$h_{\delta,i}(t) = \max\{f_i(t), f_i(x_{r,i}-\delta) - \frac{\rho_m(\bar{b}_i\varepsilon; f_i) + M(f_i) - f_i(x_{r,i}-\delta)}{x_{r,i}-\delta-x_{l,i}}(t-x_{r,i})\}.$$

Define univariate function $g_{\delta,i}$ as

$$g_{\delta,i}(t) = h_{\delta,i}(t) - \int_0^1 h_{\delta,i}(t)\mathrm{d}t. \tag{6.28}$$

Then we have $\int_0^1 g_{\delta,i}(t)\mathrm{d}t = 0$, so the definition defines a valid $\mathbf{g}_\delta \in \mathcal{F}_s$.

Further for $i = 1, 2\cdots, s$, we have

(6.29)
$$\int_0^1 \left(g_{\delta,i}(t) - f_i(t)\right)^2 \mathrm{d}t = \int_0^1 \left(h_{\delta,i}(t) - f_i(t)\right)^2 \mathrm{d}t - \left(\int_0^1 h_{\delta,i}(t)\mathrm{d}t\right)^2 \le \bar{b}_i^2\varepsilon^2,$$

and

$$|Z(g_{\delta,i}) - Z(f_i)| \ge \rho_z(\bar{b}_i\varepsilon; f_i) - \delta. \tag{6.30}$$

Therefore, we have

$$\|\mathbf{g}_\delta - \mathbf{f}\| \le \varepsilon^2, \|Z(\mathbf{g}_\delta) - Z(\mathbf{f})\|^2 \ge \sum_{i=1}^{s} \left(\rho_z(\bar{b}_i\varepsilon; f_i) - \delta\right)^2. \tag{6.31}$$

Let $\delta \to 0^+$, we have

$$\omega(\varepsilon; \mathbf{f}) \ge \sum_{i=1}^{s} \rho_z(\bar{b}_i\varepsilon; f_i)^2. \tag{6.32}$$

6.3.2. *Proof of Proposition 6.3.* We start with the right hand side and its almost-attainability.

Since $b_i \in [0, 1]$ for $1 \le i \le s$, we have $\rho_z(b_i\varepsilon; f_i) \le \rho_z(\varepsilon; f_i)$. The right hand side then apparently hold.

We first assume β in not an integer. Let $s_1 = \lfloor \beta - \delta \rfloor$, $s_2 = \beta - \lfloor \beta \rfloor$, $s_3 = s - \lceil \beta \rceil$.

Let $k_1, k_2, k_3 > 0$.

Now we start defining $f_i \in \mathcal{F}$ for $1 \le i \le s$.

If $s_1 \ge 1$, for $1 \le i \le s_1$, let

$$f_i(t) = k_1(t - \frac{1}{2}). \tag{6.33}$$

If $s_3 \geq 1$, for $n - s_3 + 1 \leq i \leq n$ let

$$f_i(t) = k_3(t - \frac{1}{2}). \tag{6.34}$$

Let

$$f_{s_1+1}(t) = k_2(t - \frac{1}{2}). \tag{6.35}$$

Suppose $0 < \delta < \frac{1}{2}s_2$.
If $s_3 \geq 1$, choose k_3 such that

$$\rho_z(\varepsilon; f_n) = \sqrt{\frac{\delta}{2s_3}}, \tag{6.36}$$

Define $s_4 = s_2 - \frac{\delta}{2}$ if $s_3 \geq 1$, otherwise $s_4 = s_2$. Choose k_2 such that

$$\rho_z(\varepsilon; f_{s_1+1}) = \sqrt{s_4}. \tag{6.37}$$

Now suppose b_{s_1+1} is the smallest $b \in [0, 1)$ such that

$$\rho_z(b\varepsilon; f_{s_1+1}) \geq \sqrt{s_4 - \frac{\delta}{2}}. \tag{6.38}$$

If $s_1 \geq 1$, choose k_1 such that

$$\rho_z(\sqrt{\frac{1-b^2}{s_1}}\varepsilon; f_1) = 1. \tag{6.39}$$

It's easy to verify that the above construction is legitimate and satisfy equation (6.22).

When $\beta = n$, choose large enough k such that $\rho_z(\frac{1}{\sqrt{s}}\varepsilon; k(t - 0.5)) = 1$, and let $f_i = k(t - 0.5)$ for $1 \leq k \leq s$.

When $\beta \leq n-1$ and is integer, for $\delta < 0.5$, let $s_1 = \beta - 1$, $s_3 = n - \beta$,$s_4 = 1 - \frac{\delta}{2}$. And choose k_3, k_2, k_1 as the case where β is not integer.

Now we proceed with the left hand side.

Recalling Proposition **??** in Cai et al. (2022), we have

$$\sup_{\sum_{i=1}^{s} b_i^2 \leq 1} \sum_{i=1}^{s} \rho_z(b_i\varepsilon; f_i)^2 \geq \sup_{\sum_{i=1}^{s} b_i^2 \leq 1} \sum_{i=1}^{s} (b_i^2/4)^{\frac{2}{3}} \rho_z(\varepsilon; f_i)^2 \geq \frac{1}{3} \left(\sum_{i=1}^{s} \rho_z(\varepsilon; f_i)^6 \right)^{\frac{1}{3}}, \tag{6.40}$$

The last inequality take $b_i = \sqrt{\frac{\rho_z(\varepsilon; f_i)^6}{\sum_{i=1}^{s} \rho_z(\varepsilon; f_i)^6}}$.

Cauchy-Schwarz inequality gives

$$\frac{1}{3}\left(\sum_{i=1}^{s}\rho_z(\varepsilon;f_i)^6\right)^{\frac{1}{3}} \geq \frac{1}{3}s^{-\frac{2}{3}}\sum_{i=1}^{s}\rho_z(\varepsilon;f_i)^2, \tag{6.41}$$

which concludes the left hand side.

For the attainability up to constant multiple, let $k>0$, which we will pick later. Let $f_i(t)=k(t-0.5)$ for $1\leq i\leq s$. Pick $k>0$ such that $\rho_z(\varepsilon;f_i)=\sqrt{\frac{\beta}{s}}$. Then we have that

$$\sup_{\sum_{i=1}^{s}b_i^2\leq 1}\sum_{i=1}^{s}\rho_z(b_i\varepsilon;f_i)^2 = \sup_{\sum_{i=1}^{s}b_i^2\leq 1}\sum_{i=1}^{s}b_i^{\frac{4}{3}}\rho_z(\varepsilon;f_i)^2 = \sup_{\sum_{i=1}^{s}b_i^2\leq 1}\sum_{i=1}^{s}b_i^{\frac{4}{3}}\frac{\beta}{s}. \tag{6.42}$$

Through basic calculation, we have $\sup_{\sum_{i=1}^{s}b_i^2\leq 1}\sum_{i=1}^{s}b_i^{\frac{4}{3}} = s^{\frac{1}{3}}$, which gives inequality (6.21).

6.4. *Proof of Proposition 2.1.* We begin with a lemma for $l_\varepsilon(f_i)$, which we will prove later.

Lemma 6.2. *For $q\geq 0$, we have*

$$\rho_z(q\varepsilon;l_\varepsilon(f_i)) \leq 2(8\sqrt{3}q)^{\frac{2}{3}}\rho_z(\varepsilon;f_i). \tag{6.43}$$

With lemma 6.2 and Proposition **??** in Cai et al. (2022) we have that for $q\geq 0$, $b\in(0,1)$,

$$\rho_z(q\varepsilon;l_\varepsilon(f_i)) \leq 2(8\sqrt{3}q)^{\frac{2}{3}}\rho_z(\varepsilon;f_i) \leq 2(8\sqrt{3}q)^{\frac{2}{3}}(\frac{2}{b})^{\frac{2}{3}}\rho_z(b\varepsilon;f_i). \tag{6.44}$$

Also we have the following holds apparently

$$\rho_z(q\varepsilon;f_i) \leq 2(8\sqrt{3}q)^{\frac{2}{3}}(\frac{2}{b})^{\frac{2}{3}}\rho_z(b\varepsilon;f_i) \tag{6.45}$$

for $q\geq 0$, $b\in(0,1)$.

Using Proposition 6.2, we have

(6.46)

$$\begin{aligned}\sup_{\mathbf{h}\in S_{\frac{\varepsilon}{3}}(\mathbf{f})}\omega(\varepsilon;\mathbf{h}) &\le \sup_{\mathbf{h}\in S_{\frac{\varepsilon}{3}}(\mathbf{f})} 9 \sup_{\sum_{i=1}^s b_i^2\le 1}\sum_{i=1}^s \rho_z(b_i\frac{\varepsilon}{3};h_i)^2\\ &\le \sum_{i=1}^s 9\times 4(16\sqrt{3})^{\frac{4}{3}}\rho_z(\frac{\varepsilon}{3};f_i)^2 \le 9\times 4(16\sqrt{3})^{\frac{4}{3}}\omega(\varepsilon;\mathbf{f}),\\ \sup_{\mathbf{h}\in S_{\varepsilon}(\mathbf{f})}\omega(\varepsilon;\mathbf{h}) &\le \sup_{\mathbf{h}\in S_{\varepsilon}(\mathbf{f})} 9 \sup_{\sum_{i=1}^s b_i^2\le 1}\sum_{i=1}^s \rho_z(b_i\varepsilon;h_i)^2\\ &\le \sum_{i=1}^s 9\times 4(16\sqrt{3})^{\frac{4}{3}}\rho_z(\varepsilon;f_i)^2 \le 9\times 4(16\sqrt{3})^{\frac{4}{3}}\omega(\varepsilon;\mathbf{f}).\end{aligned}$$

6.4.1. *Proof of Lemma 6.2.* When $q=0$, the statement holds trivially. We consider the case that $q>0$ later. For $i\in\{1,2,\cdots,s\}$, recall that

$$l_\varepsilon(f_i)=\tilde{l}_\varepsilon(f_i)-\int_0^1\tilde{l}_\varepsilon(f_i)(t)\mathrm{d}t. \tag{6.47}$$

So we have for any $q>0$, we have $\rho_z(q\varepsilon;l_\varepsilon(f_i))=\rho_z(q\varepsilon;\tilde{l}_\varepsilon(f_i))$. We only need to prove the statement for $\tilde{l}_\varepsilon(f_i)$.

Suppose $x_{l,i}$, $x_{r,i}$ are left and right end points of the interval $\{x: f_i(x)\le M(f_i)+\rho_m(\varepsilon;f_i)\}$. Without loss of generality, assume $x_{r,i}=Z(f_i)+\rho_m(\varepsilon;f_i)$.

Define a reference function g_i as

$$g_i(t)=\max\{f_i(t),M(h)+\rho_m(\varepsilon;f_i)\}. \tag{6.48}$$

Since $g_i(t)=f_i(t)=\tilde{l}_\varepsilon(f_i)(t)$ for $t\in[0,1]/(x_{l,i},x_{r,i})$, suppose $\|g_i-\tilde{l}_\varepsilon(f_i)\|=\kappa\varepsilon$, then $\rho_z(\kappa\varepsilon;\tilde{l}_\varepsilon(f_i))\le 2\rho_z(\varepsilon;f_i)$, $\rho_m(\kappa\varepsilon;\tilde{l}_\varepsilon(f_i))=\frac{1}{2}\rho_m(\varepsilon;f_i)$. It's easy to see

(6.49)

$$\frac{1}{2}\varepsilon^2\ge\frac{1}{6}\rho_z(\varepsilon;f_i)\rho_m(\varepsilon;f_i)^2\ge\|g_i-\tilde{l}_\varepsilon(f_i)\|^2\ge\frac{1}{3}\frac{\rho_z(\varepsilon;f_i)}{2}\frac{\rho_m(\varepsilon;f_i)^2}{4}\ge\frac{\varepsilon^2}{48}.$$

So we know $\kappa\in[\frac{1}{4\sqrt{3}},\sqrt{\frac{1}{2}}]$, and we have that

$$\rho_z(\frac{1}{4\sqrt{3}}\varepsilon;\tilde{l}_\varepsilon(f_i))\le\rho_z(\kappa\varepsilon;\tilde{l}_\varepsilon(f_i))\le 2\rho_z(\varepsilon;f_i). \tag{6.50}$$

Therefore, for $q\ge\frac{1}{4\sqrt{3}}$, using Proposition **??**, we have

$$\begin{aligned}\rho_z(q\varepsilon;\tilde{l}_\varepsilon(f_i)) &\le \min\{(8\sqrt{3}q)^{\frac{2}{3}},4\sqrt{3}q\}\rho_z(\frac{1}{4\sqrt{3}}\varepsilon;\tilde{l}_\varepsilon(f_i))\\ &\le 2\min\{(8\sqrt{3}q)^{\frac{2}{3}},4\sqrt{3}q\}\rho_z(\varepsilon;f_i).\end{aligned} \tag{6.51}$$

On the other hand, for $q \leq \frac{1}{4\sqrt{3}}$, given the piece-wise linearity of $\tilde{l}_\varepsilon(f_i)$ in $[x_{l,i}, x_{r,i}]$, we have

(6.52)
$$\rho_z(q\varepsilon; \tilde{l}_\varepsilon(f_i)) = \left(\frac{q}{\kappa}\right)^{\frac{2}{3}} \rho_z(\kappa\varepsilon; \tilde{l}_\varepsilon(f_i)) \leq 2\left(\frac{q}{\kappa}\right)^{\frac{2}{3}} \rho_z(\varepsilon; f_i) \leq 2(4\sqrt{3}q)^{\frac{2}{3}} \rho_z(\varepsilon; f_i).$$

6.5. *Proof of Theorem 2.3.* Recall that elements in function set $S_\varepsilon(\mathbf{f})$ and $S_{\frac{\varepsilon}{3}}(\mathbf{f})$ can be represented by $s-$dimensional vector $\mathbf{a} = (a_1, a_2, \cdots, a_s)$, with $a_i = 1$ corresponding to $h_i = l_\varepsilon(f_f)$ (or $h_i = l_{\frac{\varepsilon}{3}}(f_f)$) and $a_i = -1$ corresponding to $h_i = f_i$. We inter-changeably use $\mathbf{a}$ for $\mathbf{h}$. We use $\mathbf{a}_{-j}$ to denote $(a_1, \cdots, a_{j-1}, a_{j+1}, \cdots, a_s)$.

With start with the lower bound of confidence hyper-cube. For any $CI \in S_{\frac{\varepsilon}{3}}(\mathcal{I}_{\frac{\varepsilon}{3}}(\mathbf{f}))$, for any given data denotes the coordinates of CI are expressed as $CI = [u_1, U_1] \times [u_2, U_2] \times \cdots \times [u_s, U_s]$, then we know that $[u_i, U_j]$ is a $1-\alpha$ confidence interval for $Z(h_i)$ and we have

(6.53)
$$\begin{aligned}
&\sup_{\mathbf{h} \in S_{\frac{\varepsilon}{3}}(\mathbf{f})} \mathbb{E}_{\mathbf{h}}(diag\,(CI)^2) \\
&\geq 2^{-s} \sum_{\mathbf{a} \in \{-1,1\}^s} \mathbb{E}_{\mathbf{a}} \left(\sum_{i=1}^{s} (u_i - U_i)^2 \right) \\
&\geq \sum_{i=1}^{s} 2^{-s+1} \sum_{\mathbf{a}_{-i} \in \{-1,1\}^{s-1}} \frac{1}{2} \left(\mathbb{E}_{\mathbf{a}_{-i}, a_i=-1}(u_i - U_i)^2 + \mathbb{E}_{\mathbf{a}_{-i}, a_i=1}(u_i - U_i)^2 \right) \\
&\geq \sum_{i=1}^{s} \frac{1}{2}(0.6 - 2\alpha)\omega(\frac{\varepsilon}{3}; f_i).
\end{aligned}$$

The last inequality is due to applying Theorem 2.1 to the setting when $s = 1$.

For any i, bound $\omega(\frac{\varepsilon}{3}; f_i)$ with Proposition 6.1 and Proposition 6.2 applying to the setting $s = 1$, we have $\frac{1}{2}(0.6 - 2\alpha)\omega(\frac{\varepsilon}{3}; f_i) \geq c_\alpha \rho_z(\varepsilon; f_i)^2$ for constant c_α that depends on α only, thus conclude the proof for confidence hyper-cube.

For the lower bound for estimation, similarly for any $\hat{Z} = (\hat{Z}_1, Z_2, \cdots, Z_s)$,

we have

$$\begin{aligned} &\sup_{\mathbf{h}\in S_\varepsilon(\mathbf{f})} \mathbb{E}_{\mathbf{h}}\left(\|\hat{Z}-Z(\mathbf{h})\|^2\right) \\ &\geq 2^{-s+1}\sum_{i=1}^{s}\sum_{\mathbf{a}_{-i}\in\{-1,1\}^{s-1}}\frac{1}{2}\left(\mathbb{E}_{\mathbf{a}_{-i},a_i=1}\left(\hat{Z}_i-Z\left(l_\varepsilon(f_i)\right)\right)^2+\mathbb{E}_{\mathbf{a}_{-i},a_i=-1}\left(Z_i-Z(f_i)\right)^2\right) \\ &\geq \frac{1}{2}\sum_{i=1}^{s}R(\varepsilon;f_i). \end{aligned} \tag{6.54}$$

$R(\varepsilon; f_i)$ in last inequality is using the definition $R(\varepsilon;\cdot)$ in the setting that $s=1$.

This combine with Theorem 2.1 and Theorem 2.2 applied to the setting $s=1$ concludes the proof for estimation.

6.6. *Proof of Corollary 2.1.* It suffices to prove the following lemma.

Lemma 6.3. *For $0<\alpha<0.3$, we have*

$$\sup_{\mathbf{f}\in\mathcal{F}_s}\inf_{CI\in\mathcal{I}_\alpha(S_{\frac{1}{3}\varepsilon}(\mathbf{f}))}\sup_{\mathbf{h}\in S_\varepsilon(\mathbf{f})}\frac{\mathbb{E}_{\mathbf{h}}(diag\,(CI)^2)}{L_\alpha(\varepsilon;\mathbf{h})} \geq C_\alpha s^{\frac{2}{3}}, \tag{6.55}$$

$$\sup_{f\in\mathcal{F}_s}\inf_{\hat{Z}}\sup_{\mathbf{h}\in S_\varepsilon(\mathbf{f})}\frac{\mathbb{E}_{\mathbf{h}}\left(\|\hat{Z}-Z(\mathbf{h})\|^2\right)}{R(\varepsilon;\mathbf{h})} \geq C_1 s^{\frac{2}{3}}. \tag{6.56}$$

where C_α is a positive constant depending on α and C_1 is a positive constant.

Proof. Recalling $\beta(\mathbf{f})$ defined in Theorem 2.3, using Theorem 2.2, we have

$$\sup_{\mathbf{f}\in\mathcal{F}_s}\frac{\beta(\mathbf{f})}{\omega(\varepsilon/3;\mathbf{f})} \geq \sup_{\mathbf{f}\in\mathcal{F}_s}\frac{\beta(\mathbf{f})}{\omega(\varepsilon;\mathbf{f})} \geq s^{\frac{2}{3}}/9. \tag{6.57}$$

Then Theorem 2.3 Proposition 2.1 combined with Theorem 2.1 gives

$$\begin{aligned}
&\sup_{\mathbf{f}\in\mathcal{F}_s} \inf_{CI\in\mathcal{I}_\alpha(S_{\frac{1}{3}\varepsilon}(\mathbf{f}))} \sup_{\mathbf{h}\in S_\varepsilon(\mathbf{f})} \frac{\mathbb{E}_{\mathbf{h}}(diag\,(CI)^2)}{L_\alpha(\varepsilon;\mathbf{h})} \\
&\geq \sup_{\mathbf{f}\in\mathcal{F}_s} \inf_{CI\in\mathcal{I}_\alpha(S_{\frac{1}{3}\varepsilon}(\mathbf{f}))} \sup_{\mathbf{h}\in S_\varepsilon(\mathbf{f})} \frac{\mathbb{E}_{\mathbf{h}}(diag\,(CI)^2)}{b_\alpha\omega(\varepsilon/3;\mathbf{h})} \\
&\geq \sup_{\mathbf{f}\in\mathcal{F}_s} \inf_{CI\in\mathcal{I}_\alpha(S_{\frac{1}{3}\varepsilon}(\mathbf{f}))} \sup_{\mathbf{h}\in S_\varepsilon(\mathbf{f})} \frac{\mathbb{E}_{\mathbf{h}}(diag\,(CI)^2)}{b_\alpha C_0\omega(\varepsilon;\mathbf{f})} \\
&\geq \frac{c_\alpha\beta(\mathbf{f})}{b_\alpha C_0\omega(\varepsilon;\mathbf{f})} \geq C_\alpha s^{\frac{2}{3}}.
\end{aligned} \tag{6.58}$$

Similarly We have

$$\sup_{f\in\mathcal{F}_s} \inf_{\hat{Z}} \sup_{\mathbf{h}\in S_\varepsilon(\mathbf{f})} \frac{\mathbb{E}_{\mathbf{h}}\left(\|\hat{Z}-Z(\mathbf{h})\|^2\right)}{R(\varepsilon;\mathbf{h})} \geq C_1 s^{\frac{2}{3}}. \tag{6.59}$$

□

6.7. *Proof of Proposition 3.1.* We start with the first item.

Suppose $\mathfrak{P}(Y^1) = \mathfrak{P}(Y^2)$ for $Y^1, Y^2 \in \mathfrak{Y}$. Then for $\mathcal{A} = [a_1, A_1] \times [a_2, A_2] \times \cdots \times [a_s, A_s] \subset [0,1]^s$, we have

$$\begin{aligned}
\int_{\mathcal{A}} \mathrm{d}Y^1 &= \int_{\mathcal{A}} \mathrm{d}\mathbf{er}(Y^1) + \sum_{i=1}^{s} \Pi_{j\neq i}(A_j - a_j) \int_{a_i}^{A_i} \mathrm{d}\boldsymbol{\pi}_i(Y^1) \\
&= \int_{\mathcal{A}} \mathrm{d}\mathbf{er}(Y^2) + \sum_{i=1}^{s} \Pi_{j\neq i}(A_j - a_j) \int_{a_i}^{A_i} \mathrm{d}\boldsymbol{\pi}_i(Y^2) \\
&= \int_{\mathcal{A}} \mathrm{d}Y^2.
\end{aligned} \tag{6.60}$$

Therefore, using Dynkin's $\pi-\lambda$ theorem, $Y^1 = Y^2$.

Now we continue with the second item.

Again, from Dynkin's $\pi-\lambda$ theorem, we only need to prove that for any $[a_1, A_1], [a_2, A_2], \cdots, [a_s, A_s] \subset [0,1]$ and $\mathfrak{B} = [b_1, B_1]\times[b_2, B_2]\times\cdots\times[b_s, B_s]$, the following variables are independent:

$$\int_{[a_1,A_1]} \mathrm{d}\boldsymbol{\pi}_1(Y), \int_{[a_2,A_2]} \mathrm{d}\boldsymbol{\pi}_2(Y), \cdots, \int_{[a_s,A_s]} \mathrm{d}\boldsymbol{\pi}_s(Y), \int_{[b_1,B_1]\times[b_2,B_2]\times[b_s,B_s]} \mathrm{d}\mathbf{er}(Y).$$

Note that $\boldsymbol{\pi}_i(Y)[A_i] - \boldsymbol{\pi}_i(Y)[a_i] = \int_{[a_1,A_1]} \mathrm{d}\boldsymbol{\pi}_i(Y)$, but we use integral form whenever possible to ease understanding as we have stochastic processes of different dimensions.

From the definition 3.1 of $\boldsymbol{\pi}_i(Y)$ and $\mathbf{er}(Y)$, we know that

$$(\int_{[a_1,A_1]} \mathrm{d}\boldsymbol{\pi}_1(Y), \int_{[a_2,A_2]} \mathrm{d}\boldsymbol{\pi}_2(Y), \cdots, \int_{[a_s,A_s]} \mathrm{d}\boldsymbol{\pi}_s(Y), \int_{\mathfrak{B}} \mathrm{d}\mathbf{er}(Y))$$

is joint normal random vector. To prove independence we only need to prove the correlations are zero.

For $1 \le i < j \le s$, we have

$$\begin{aligned}
&COV(\int_{[a_i,A_i]} \mathrm{d}\boldsymbol{\pi}_i(Y), \int_{[a_j,A_j]} \mathrm{d}\boldsymbol{\pi}_j(Y)) \\
&= \mathbb{E}\Bigg(\left(\int_{t_i \in [a_i,A_i], \mathbf{t}_{-i} \in [0,1]^{s-1}} \mathrm{d}W - (A_i - a_i) \int_{[0,1]^s} \mathrm{d}W \right) \cdot \\
&\quad \left(\int_{t_j \in [a_j,A_j], \mathbf{t}_{-j} \in [0,1]^{s-1}} \mathrm{d}W - (A_j - a_j) \int_{[0,1]^s} \mathrm{d}W \right) \Bigg) \\
&= 0.
\end{aligned} \tag{6.61}$$

For $1 \le i \le s$, suppose $\mathcal{A}_i = \{\mathbf{t} : t_i \in [a_i, A_i], \mathbf{t}_{-i} \in [0,1]^{s-1}\}$, and $V(\cdot)$ denotes the volume (length when one dimensional, area when two dimensional, etc.), we have

$$\begin{aligned}
&COV(\int_{[a_i,A_i]} \mathrm{d}\boldsymbol{\pi}_i(Y), \int_{\mathfrak{B}} \mathrm{d}Y) \\
&= \mathbb{E}\Bigg(\left(\int_{t_i \in [a_i,A_i], \mathbf{t}_{-i} \in [0,1]^{s-1}} \mathrm{d}W - (A_i - a_i) \int_{[0,1]^s} \mathrm{d}W \right) \cdot \\
&\quad \left(\int_{\mathfrak{B}} \mathrm{d}W - \sum_{j=1}^{s} \Pi_{k \ne j}(B_k - b_k) \int_{t_j \in [b_j,B_j], \mathbf{t}_{-j} \in [0,1]^{s-1}} \mathrm{d}W + s\Pi_{k=1}^{s}(B_k - b_k) \int_{[0,1]^s} \mathrm{d}W \right) \Bigg) \\
&= V(\mathcal{A}_i \cap \mathfrak{B}) - (A_i - a_i)V(\mathfrak{B}) - \sum_{j \ne i} \Pi_{k \ne j}(B_k - b_k)(B_j - b_j)(A_i - a_i) \\
&\quad - V([a_i, A_i] \cap [b_i, B_i])\Pi_{j \ne i}(B_j - b_j) + s(A_i - a_i)\Pi_{i=1}^{s}(B_i - b_i) + 0 \\
&= 0.
\end{aligned} \tag{6.62}$$

Therefore, we prove the independence.

Now we continue with the sufficiency property. Recalling the Radon-

Nikodym derivative calculated in (6.4), we have that for $\mathbf{f}, \mathbf{g} \in \mathcal{F}_s$

$$\frac{dP_{\mathbf{f}}}{dP_{\mathbf{g}}}(Y) = \exp\left(\int_{[0,1]^s} \frac{\mathbf{f}(\mathbf{t}) - \mathbf{g}(\mathbf{t})}{\varepsilon^2} dY(\mathbf{t}) - \frac{1}{2}\int_{[0,1]^s} \frac{\mathbf{f}(\mathbf{t})^2 - \mathbf{g}(\mathbf{t})^2}{\varepsilon^2} \mathrm{d}\mathbf{t}\right)$$
$$= \exp\left(\frac{1}{\varepsilon^2}\sum_{i=1}^s \int_0^1 (f_i(t) - g_i(t))\mathrm{d}\boldsymbol{\pi}_i(Y) - \frac{1}{2\varepsilon^2}\int_{[0,1]^s} \left(\mathbf{f}(\mathbf{t})^2 - \mathbf{g}(\mathbf{t})^2\right)\mathrm{d}\mathbf{t}\right). \tag{6.63}$$

Hence we concludes the proof.

6.8. *Proof of Theorem 3.1.* Suppose additive form of $\mathbf{f}_0$ is $\mathbf{f}_0 = f_{0,0} + \sum_{i=1}^s f_{0,i}(t_i)$.

Suppose $h_0(\gamma) > 1$ be a function depending on γ. Separate indexes $1 \le i \le s$ into two groups:

$$\begin{aligned} Ind_1 &= \{1 \le i \le s : \mathbb{E}_{\mathbf{f}_0}\left(\|\hat{Z}_i - Z(\mathbf{f}_0)_i\|^2\right) \le h_0(\gamma)\gamma\rho_z(\varepsilon; f_{0,i})^2\}, \\ Ind_2 &= \{1 \le i \le s : \mathbb{E}_{\mathbf{f}_0}\left(\|\hat{Z}_i - Z(\mathbf{f}_0)_i\|^2\right) > h_0(\gamma)\gamma\rho_z(\varepsilon; f_{0,i})^2\}. \end{aligned} \tag{6.64}$$

Then we have

$$\sum_{i \in Ind_2} \rho_z(\varepsilon; f_{0,i})^2 \le \frac{1}{h_0(\gamma)}\sum_{i=1}^s \rho_z(\varepsilon; f_{0,i})^2. \tag{6.65}$$

Note that for $\mathbf{g}, \mathbf{h} \in \mathcal{F}_s$, Suppose the distribution of $\boldsymbol{\pi}_i(Y_{\mathbf{g}})$ and $\boldsymbol{\pi}_i(Y_{\mathbf{h}})$ are $\mathbb{Q}_{\mathbf{g}}$, $\mathbb{Q}_{\mathbf{h}}$, then we have

$$\frac{\mathrm{d}\mathbb{Q}_{\mathbf{g}}}{\mathrm{d}\mathbb{Q}_{\mathbf{h}}}(\boldsymbol{\pi}_i(Y)) = \exp\left(\int_0^1 \frac{g_i(t) - h_i(t)}{\varepsilon^2}\mathrm{d}\boldsymbol{\pi}_i(Y) - \frac{1}{2\varepsilon^2}\int_0^1 \left(h_i(t)^2 - g_i(t)^2\right)\mathrm{d}t\right). \tag{6.66}$$

Hence if we let $\theta = 1$ for g_i and $\theta = -1$ for h_i, a sufficient statistic for θ is

$$W_i = \int_0^1 \frac{g_i(t) - h_i(t)}{\|g_i - h_i\|\varepsilon}\mathrm{d}\boldsymbol{\pi}_i(Y) - \frac{\int_0^1 g_i(t)^2 - h_i(t)^2 \mathrm{d}t}{2\|g_i - h_i\|\varepsilon} \sim N(\frac{\theta}{2}\cdot\frac{\|g_i - h_i\|}{\varepsilon}, 1), \tag{6.67}$$

under $\mathbb{P}_\theta$.

Also, note that for any $g, h \in \mathcal{F}$, if we let $\tilde{g}(t) = g(t) - \int_0^1 g(x)\mathrm{d}x$ and $\tilde{h}(t) = h(t) - \int_0^1 h(x)\mathrm{d}x$, then we have $\|g - h\| \ge \|\tilde{g} - \tilde{h}\|$.

Now, using the construction of the alternative function (g_2) in the proof of Lemma A.5 in Cai et al. (2021), we have that when $h_0(\gamma)\gamma < \sqrt{0.0063}$, for $i \in Ind_1$, there exists $f_{2,i} \in \mathcal{F}$ such that

$$|Z(f_{2,i}) - Z(f_{0,i})| \ge \rho_z(\varepsilon; f_i), \tag{6.68}$$

and for any $\tau > 0$

$$\rho_z(\tau; f_{2,i}) \leq \frac{2^{\frac{7}{3}}}{3^{\frac{1}{3}}}\left(\frac{\tau}{z_{2h_0(\gamma)\gamma}\varepsilon}\right)^{\frac{2}{3}}|Z(f_{2,i}) - Z(f_{0,i})|, \tag{6.69}$$

also for any $\mathbf{f} \in \mathcal{F}_s$ such that $f_i(t) = f_{2,i}(t) - \int_0^1 f_{2,i}(x)\mathrm{d}x$ the following holds

$$\mathbb{E}_{\mathbf{f}}\|\hat{Z}_i - Z(\mathbf{g})_i\|^2 \geq \frac{1}{8}|Z(f_{0,i} - Z(f_{2,i})|^2. \tag{6.70}$$

The last statement uses similar argument as in Inequality (B.19) in Cai et al. (2021).

Therefore, let

$$\mathbf{f}_1 = f_{0,0} + \sum_{i \in Ind_1}\left(f_{2,i}(t_i) - \int_0^1 f_{2,i}(x)\mathrm{d}x\right) + \sum_{i \in Ind_2} f_{0,i}(t_i), \tag{6.71}$$

we have

$$\mathbb{E}_{\mathbf{f}_1}\left(\|Z(\mathbf{f}_1) - \hat{Z}\|^2\right) \geq \sum_{i \in Ind_1}\frac{1}{8}|Z(f_{0,i}) - Z(f_{1,i})|^2 \geq \sum_{i \in Ind_1}\frac{1}{8}z_{2h_0(\gamma)\gamma}^{\frac{4}{3}}\frac{3^{\frac{2}{3}}}{2^{\frac{14}{3}}}\rho_z(\varepsilon; f_{1,i})^2, \tag{6.72}$$

where $f_{1,i} = f_{2,i}(t_i) - \int_0^1 f_{2,i}(x)\mathrm{d}x$.

We also according to (6.65) and (6.68), we have

$$\sum_{i \in Ind_1}|Z(f_{0,i}) - Z(f_{1,i})|^2 \geq (h_0(\gamma) - 1)\sum_{i \in Ind_2}\rho_z(\varepsilon; f_{1,i})^2 \tag{6.73}$$

Therefore, let $h_0(\gamma) = 1 + \frac{1}{8}z_{2\gamma}^{\frac{4}{3}}\frac{3^{\frac{2}{3}}}{2^{\frac{14}{3}}}$, we have

$$\mathbb{E}_{\mathbf{f}_1}\left(\|Z(\mathbf{f}_1) - \hat{Z}\|^2\right) \geq \frac{1}{16}z_{2h_0(\gamma)\gamma}^{\frac{4}{3}}\frac{3^{\frac{2}{3}}}{2^{\frac{14}{3}}}\beta(\mathbf{f}_1). \tag{6.74}$$

Note that $h_0(\gamma)\gamma \sim \gamma(\sqrt{\log\frac{1}{\gamma}})^{\frac{4}{3}}$ when $\gamma \to 0^+$, so $h_0(\gamma)\gamma \to 0^+$ as $\gamma \to 0^+$. Hence $h_0(\gamma)\gamma < \sqrt{0.0063}$ for small enough γ, the arguments go through.

According to Inequality (6.74), and that $z_{2h_0(\gamma)\gamma} \sim \sqrt{\log\frac{1}{\gamma}}$, we have concludes Inequality (3.8).

6.9. *Proof of Theorem 3.2.* Since $CI \in \mathcal{SI}_\alpha(\mathcal{F}_s)$, let CI_i be the confidence interval indicated by the projection on ith axis of the hyper cube CI: $CI_i(\boldsymbol{\pi}_i(Y)) = CI(Y)_i$. Since the distribution of $\boldsymbol{\pi}_i(Y_{\mathbf{f}})$ only depends on f_i, we have

$$\begin{aligned} \alpha &\geq \sup_{\mathbf{f}\in\mathcal{F}_s} P(Z(\mathbf{f}) \in CI(Y_{\mathbf{f}})) \\ &= P(Z(f_i) \in CI_i(\boldsymbol{\pi}_i(Y_{\mathbf{f}})),\ \text{for } 1 \leq i \leq s) \\ &= \Pi_{i=1}^s P(Z(f_i) \in CI_i(\boldsymbol{\pi}_i(Y_{\mathbf{f}}))). \end{aligned} \tag{6.75}$$

Denote $\alpha_i = P(Z(f_i) \notin CI_i(\boldsymbol{\pi}_i(Y_{\mathbf{f}})))$, then we have

$$1 - \alpha \leq \Pi_{i=1}^s (1 - \alpha_i). \tag{6.76}$$

It's clear the $\alpha_i \leq \alpha$ for all $1 \leq i \leq s$.

We introduce a basic property as in Lemma 6.4, the proof of which is at a later part.

Lemma 6.4. *For $\sum_{i=1}^s \alpha_i < 1$ and $\frac{1}{2} \geq \alpha_i \geq 0$, we have*

$$\Pi_{i=1}^s (1 - \alpha_i) \leq 1 - \frac{1}{2}\sum_{i=1}^s \alpha_i.$$

We only need to verify that $\sum_{i=1}^s \alpha_i < 1$ to invoke this lemma and get $\sum_{i=1}^s \alpha_i \leq 2\alpha$.

We argue by contradiction. If $\sum_{i=1}^s \alpha_i \geq 1$, recalling that $(1-\frac{1}{n})^n$ increases with n increasing, we have

$$1 - \alpha \leq \Pi_{i=1}^s (1 - \alpha_i) \leq (1 - \frac{\sum_{j=1}^s \alpha_j}{s})^s \leq \lim_{s\to\infty} (1 - \frac{1}{s})^s = \frac{1}{e}. \tag{6.77}$$

This is in contradiction with $\alpha \leq 0.1$. So we have $\sum_{i=1}^s \alpha_i < 1$ and thus $\sum_{i=1}^s \alpha_i \leq 2\alpha$.

Note that we have

$$\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right) \geq \sum_{i=1}^s \mathbb{E}_{\mathbf{f}}\left(|CI_i(\boldsymbol{\pi}_i(Y))|^2\right). \tag{6.78}$$

So we are left with bounding $\mathbb{E}_{\mathbf{f}}\left(|CI_i(\boldsymbol{\pi}_i(Y_{\mathbf{f}}))|^2\right)$ for CI_i being $1 - \alpha_i$ confidence interval for $Z(f_i)$ based on $\boldsymbol{\pi}_i(Y)$. We summarize it into the following lemma, which immediately concludes the proof.

LEMMA 6.5. *For $0 < \alpha < 1$ and $CI \in \mathcal{SI}_\alpha(\mathcal{F}_s)$, denote $CI_i(\boldsymbol{\pi}_i(Y)) = CI(Y)$, if for some $0 \le \alpha_i \le 0.2$,*

$$P_{\mathbf{f}}(Z(f_i) \in CI_i(\boldsymbol{\pi}_i(Y))) \ge 1 - \alpha_i, \tag{6.79}$$

then

$$\mathbb{E}_{\mathbf{f}}\left(|CI_i(\boldsymbol{\pi}_i(Y_{\mathbf{f}}))|^2\right) \ge 0.3\rho_z(z_{\alpha_i}\varepsilon; f_i)^2, \tag{6.80}$$

for all $\mathbf{f} \in \mathcal{F}_s$.

PROOF. When $\alpha_i = 0$, $\mathbb{E}_{\mathbf{f}}\left(|CI_i(\boldsymbol{\pi}_i(Y_{\mathbf{f}}))|^2\right) = 1 = \rho_z(z_{\alpha_i}\varepsilon; f_i)^2 \ge 0.3\rho_z(z_{\alpha_i}\varepsilon; f_i)^2$. Now we only consider $\alpha_i > 0$.

Now for f_i, for any δ such that $\min\{\frac{1}{20}\varepsilon, \frac{1}{10}\rho_z(\frac{1}{10}\varepsilon; f_i)\} > \delta > 0$, we will construct $g_i \in \mathcal{F}$ such that $\int_0^1 g_i(t)\mathrm{d}t = 0$ and the followings hold. If we let $\mathbf{g}(\mathbf{t}) = f_0 + \sum_{j\neq i} f_j(t_j) + g_i(t_i)$, then for any CI_i satisfying $P(Z(g_i) \in CI_i(\boldsymbol{\pi}_i(Y_{\mathbf{g}}))) \ge 1 - \alpha_i, P(Z(f_i) \in CI_i(\boldsymbol{\pi}_i(Y_{\mathbf{f}}))) \ge 1 - \alpha_i$, we have

$$\mathbb{E}_{\mathbf{f}}\left(|CI_i(\boldsymbol{\pi}_i(Y))|^2\right) \ge \left(\frac{1}{2} - \alpha_i\right)\left(\rho_z(z_{\alpha_i}\varepsilon; f_i) - \delta\right)^2 \ge 0.3\left(\rho_z(z_{\alpha_i}\varepsilon; f_i) - \delta\right)^2. \tag{6.81}$$

Let x_l and x_r be the left and right end points of the interval $\{x : f_i(x) \le M(f_i) + \rho_m(z_{\alpha_i}\varepsilon; f_i)\}$. Without loss of generality, we assume $x_r = Z(f_i) + \rho_z(z_{\alpha_i}\varepsilon; f_i)$.

Let $H = \max\{f_i(x_r - \delta), M(f_i) + \rho_m(z_{\alpha_i}\varepsilon; f_i) - \delta\}$. Define the $\tilde{g}_i \in \mathcal{F}$ as

$$\tilde{g}_i(t) = \max\{f_i(t), H + \frac{H - M(f_i) - \rho_m(z_{\alpha_i}\varepsilon; f_i)}{x_r - \delta - x_l}(t - x_r + \delta)\}.$$

Then we know that $z_{\alpha_i}\varepsilon \ge \|\tilde{g}_i - f_i\| \ge z_{\alpha_i}\varepsilon - 2\delta$ and that $|Z(\tilde{g}_i) - Z(f_i)| \ge \rho_z(z_{\alpha_i}\varepsilon; f_i) - \delta$.

Let $g_i \in \mathcal{F}$ be defined as $g_i(t) = \tilde{g}_i(t) - \int_0^1 \tilde{g}_i(x)\mathrm{d}x$. Then we have $\|g_i - f_i\| \le z_{\alpha_i}\varepsilon$, and $Z(g_i) \ge x_r - \delta$. Then recalling the Radon-Nikodym derivative in Equation (6.66) and the sufficient statistics (6.67), using the Neyman-Pearson lemma we have

$$\begin{aligned} &P_{\mathbf{f}}([Z(f_i), Z(g_i)] \subset CI_i(\boldsymbol{\pi}_i(Y))) \\ &\ge 1 - \alpha_i - P_{\mathbf{f}}(g_i \notin CI_i(\boldsymbol{\pi}_i(Y))) \\ &\ge 1 - \alpha_i - \Phi(-z_{\alpha_i} + \frac{\|g_i - f_i\|}{\varepsilon}) \ge 0.5 - \alpha_i. \end{aligned} \tag{6.82}$$

Therefore we have Inequality (6.81), and letting $\delta \to 0^+$ we concludes the statement. □

6.9.1. *Proof of Lemma 6.4.* For s =2, $(1-\alpha_1)(1-\alpha_2) = 1-\frac{1}{2}(\alpha_1+\alpha_2)-(\frac{1}{4}-(\frac{1}{2}-\alpha_1)(\frac{1}{2}-\alpha_2)) \leq 1-\frac{1}{2}(\alpha_1+\alpha_2)$.

Suppose this holds for $s \leq k$, when $s = k+1$, we have

$$\begin{aligned}\Pi_{i=1}^{k+1}(1-\alpha_i) &\leq (1-\alpha_{k+1})(1-\frac{1}{2}\sum_{i=1}^{k}\alpha_i) \\ &= 1-\frac{1}{2}\sum_{i=1}^{k+1}\alpha_i - \frac{1}{2}\alpha_{k+1}(1-\sum_{i=1}^{k}\alpha_i) \\ &\leq 1-\frac{1}{2}\sum_{i=1}^{k+1}\alpha_i.\end{aligned} \tag{6.83}$$

6.10. *Proof of Theorem 3.3.* Suppose $CI \in \mathcal{SI}_\alpha(\mathcal{F}_s)$.

Suppose

$$\alpha_i = \sup_{\mathbf{g}\in\mathcal{F}_s} P(Z(g_i) \notin CI(Y)_i). \tag{6.84}$$

Suppose the permutation $p_1 = (p_1(1), p_1(2), \cdots, p_1(s))$ satisfies

$$\alpha_{p_1(i)} \geq \alpha_{p_1(j)}, \tag{6.85}$$

for $i \leq j$.

Then by lemma 6.4, we have that $\sum_i \alpha_i \leq 2\alpha$, which gives that for all $i \geq \lfloor \frac{s}{2} \rfloor + 1$, we have $\alpha_{p_1(i)} \leq \min\{\alpha, \frac{4\alpha}{s}\}$.

Let

$$\tau = \min\{\alpha, \frac{4\alpha}{s}\}.$$

For any $\mathbf{f} \in \mathcal{F}_s$, suppose the permutation $p_2 = (p_2(1), p_2(2), \cdots, p_2(s))$ satisfies

$$\rho_z(z_\tau\varepsilon; f_{p_2(i)}) \leq \rho_z(z_\tau\varepsilon; f_{p_2(j)}), \tag{6.86}$$

for $i \leq j$.

Then we have that

$$\begin{aligned}\sum_{i\geq\lfloor\frac{s}{2}\rfloor+1}\rho_z(z_\tau\varepsilon; f_{p_2(i)})^2 &\geq \frac{1}{2}\sum_{i=1}^{s}\rho_z(z_\tau\varepsilon; f_{p_2(i)})^2 \\ &\geq \frac{1}{2}\left(\frac{\min\{z_{\frac{4\alpha}{s}}, z_\alpha\}}{z_{\alpha/s}}\right)^{\frac{4}{3}}\Gamma_2(\mathbf{f};\alpha) > \frac{1}{2}2^{-\frac{4}{3}}\Gamma_2(\mathbf{f};\alpha),\end{aligned} \tag{6.87}$$

for $\alpha \le 0.2$.

Now we define $\mathbf{f}_1(\mathbf{t})$ for $\mathbf{t} = (t_1, t_2, \ldots, t_s) \in [0,1]^s$. Let

$$\mathbf{f}_1(\mathbf{t}) = f_0 + \sum_{i=1}^{s} f_{p_2(i)}(t_{p_1(i)})$$

,

Then with lemma 6.5 we have

(6.88)
$$\mathbb{E}_{\mathbf{f}_1}\left(diag\,(CI)^2\right) \ge 0.3 \sum_{i \ge \lfloor \frac{s}{2} \rfloor + 1} \rho_z(z_\tau \varepsilon; f_{p_2(i)})^2 > \frac{1}{20}\Gamma_2(\mathbf{f};\alpha) = \frac{1}{20}\Gamma_2(\mathbf{f}_1;\alpha).$$

Note that the choice of $\mathbf{f}$ is arbitrary, when we further let

$$f_i = H \cdot |x - 0.5|, \tag{6.89}$$

for an H such that $\rho_z(z_{\alpha/s}\varepsilon; f_i) < \frac{1}{2}$. Then we have $\Gamma_2(\mathbf{f}_1;\alpha) = \left(\frac{z_{\alpha/s}}{z_\alpha}\right)^{\frac{2}{3}} \Gamma_1(\mathbf{f}_1;\alpha)$. Combining with Inequality (6.88), we concludes the proof.

6.11. *Proof of Theorem 3.4.* Since we have

$$\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z} - Z(\mathbf{f})\|^2\right) = \sum_{k=1}^{s} \mathbb{E}_{\mathbf{f}}\left(|\hat{Z}_k - Z(f_k)|^2\right), \tag{6.90}$$

we only need to prove that there is an absolute constant $C_2 > 0$ such that for $1 \le k \le s$,

$$\mathbb{E}_{\mathbf{f}}\left(|\hat{Z}_k - Z(f_k)|^2\right) \le C_2 \rho_z(\varepsilon; f_k)^2. \tag{6.91}$$

Now we focus on any given $k \in \{1, \ldots, s\}$.

Note that for each level $j \ge 1$, the localization and stopping rule only based on the following random variables $\{\tilde{X}_{j,i,k} - \tilde{X}_{j,i-1,k} : i = 2, \ldots, 2^j\} \cup \{X_{j,i,k} - X_{j,i-1,k} : i = 2, \ldots, 2^j\}$.

If we construct two stochastic process $\tilde{\mathbf{v}}^l$ and $\tilde{\mathbf{v}}^r$ in the following way

$$\begin{aligned} \mathrm{d}\tilde{\mathbf{v}}^l(t) &= f_k(t)\mathrm{d}t + \sqrt{3}\varepsilon \mathrm{d}W^l, \\ \mathrm{d}\tilde{\mathbf{v}}^r(t) &= f_k(t)\mathrm{d}t + \sqrt{3}\varepsilon \mathrm{d}W^r, \end{aligned} \tag{6.92}$$

where W^l and W^r are independent Brownian Motion, and also define $O_{j,i,k}, \tilde{O}_{j,i,k}$ in the same way as $X_{j,i,k}, \tilde{X}_{j,i,k}$ with $\mathbf{v}^l$ and $\mathbf{v}^r$ replaced by $\tilde{\mathbf{v}}^l$ and $\tilde{\mathbf{v}}^r$, then we know that the distribution under $\mathbf{f}$ of the infinite dimension object $Ds(X,k)$ that concatenate the following vectors with $j = 1, 2, \ldots$:

$$(\tilde{X}_{j,2,k} - \tilde{X}_{j,1,k}, \tilde{X}_{j,3,k} - \tilde{X}_{j,2,k}, \ldots, \tilde{X}_{j,2^j,k} - \tilde{X}_{j,2^j-1,k}, \\ X_{j,2,k} - X_{j,1,k}, X_{j,3,k} - X_{j,2,k}, \ldots, X_{j,2^j,k} - X_{j,2^j-1,k}) \tag{6.93}$$

is the same with that having $O_{j,i,k}, \tilde{O}_{j,i,k}$ in the place of $X_{j,i,k}, \tilde{X}_{j,i,k}$, which we call $Ds(O,k)$.

Also note that the localization procedure, stopping procedure and construction of each axis of the estimator goes in parallel with the univariate estimator in Cai et al. (2022), and that the distribution of random variables playing a role in the entire estimation procedure (i.e. $Ds(X,k)$) is the same with that of $Ds(O,k)$.

Hence bounding $E_{\mathbf{f}}\left(|\hat{Z}_k - Z(f_k)|^2\right)$ here is the same with bounding $\mathbb{E}_{f_k}\left(|\tilde{Z} - Z(f_k)|^2\right)$ with $\tilde{Z}$ being the estimator of the minimizer of the univariate function in the setting of Cai et al. (2022).

Resort to the proof of that of Theorem **??** Cai et al. (2022) with the quantities bounding $|\tilde{Z} - Z(f_k)|$ there being replaced by the square of it, we have

$$\mathbb{E}_{\mathbf{f}}\left(|\hat{Z}_k - Z(f_k)|^2\right) \le \mathbb{E}_{f_k}\left(|\tilde{Z} - Z(f_k)|^2\right) \le C_2 \rho_z(\varepsilon; f_k)^2, \tag{6.94}$$

for an absolute constant C_2.

6.12. *Proof of Theorem 3.5.* We prove the theorem by proving two propositions.

PROPOSITION 6.4 (Coverage). *The confidence hyper cube CI defined by (3.22) is an $1-\alpha$ level confidence cube for minimizer.*

PROPOSITION 6.5 (Expected Squared Diameter). *For $\alpha \le 0.3$, and confidence hypercube CI defined by (3.22), we have*

$$\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right) \le C_3 \sum_{k=1}^{s} \rho_z(z_{\alpha/s}; f_k)^2, \tag{6.95}$$

where C_3 is an absolute constant.

6.12.1. *Proof of Proposition 6.4.* According to our construction, we know that

$$P_{\mathbf{f}}(Z(\mathbf{f}) \in CI) = \Pi_{k=1}^{s} P_{\mathbf{f}}(Z(f_k) \in CI_k) \ge \Pi_{k=1}^{s} \inf_{\mathbf{f}\in\mathcal{F}_s} P_{\mathbf{f}}(Z(f_k) \in CI_k). \tag{6.96}$$

So it suffices to prove that $\inf_{\mathbf{f}\in\mathcal{F}_s} P_{\mathbf{f}}(Z(f_k)\in CI_k)\geq 1-\frac{\alpha}{s}$.

Denote $\grave{j}_k=\min\{j:|\hat{i}_{j,k}-i^*_{j,k}|\geq 7\}$.Then we have for any $\mathbf{f}\in\mathcal{F}_s$,

(6.97)

$$P_{\mathbf{f}}(Z(f_k)\notin CI_k)=P_{\mathbf{f}}(\grave{j}_k<\hat{j}(\alpha/s,k))=\sum_{j=3}^{\infty}\mathbb{E}_{\mathbf{f}}(\mathbb{E}_{\mathbf{f}}(\mathbb{1}\{j<\hat{j}(\alpha/s,k)\}|\mathbf{v}_k^l)\mathbb{1}\{\grave{j}_k=j\})$$

$$\leq\sum_{j=3}^{\infty}\mathbb{E}_f(\alpha/s\mathbb{1}\{\grave{j}_k=j\})\leq\alpha/s.$$

The first inequality is due to the distribution in (3.17) and that for the $\frac{\tilde{X}_{j,\hat{i}_{j,k}-6,k}-\tilde{X}_{j,\hat{i}_{j,k}-5,k}}{\sigma_j}$, as well as the facts that $\hat{i}_{j,k}$ only depends on $\mathbf{v}_k^l$, that $\mathbf{v}_k^l$ and $\mathbf{v}_k^r$ are independent, and that $j=\grave{j}_k$ implies $S_p(j,k)\leq 0$ or that for the left side is non-positive.

This concludes the proof.

6.12.2. *Proof of Proposition 6.5.* Since we have

$$\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right)=\sum_{k=1}^{s}\mathbb{E}_{\mathbf{f}}\left(\|CI_k\|^2\right), \tag{6.98}$$

it suffice to prove that there exists an absolute constant $C_3>0$ such that for any $k\in\{1,2,\ldots,s\}$, the following holds

$$\mathbb{E}_{\mathbf{f}}\left(\|CI_k\|^2\right)\leq C_3\rho_z(z_{\alpha/s}\varepsilon;f_k)^2. \tag{6.99}$$

Now we recollect and introduce some notation that indicate the levels at which the localization procedure picks a interval far away from the right one.

$$\begin{aligned}\tilde{j}_k&=\min\{j:|\hat{i}_{j,k}-i^*_{j,k}|\geq 2\},\\ \acute{j}_k&=\min\{j:|\hat{i}_{j,k}-i^*_{j,k}|\geq 5\},\\ \grave{j}_k&=\min\{j:|\hat{i}_{j,k}-i^*_{j,k}|\geq 7\}.\end{aligned} \tag{6.100}$$

It's clear that for any $j\geq\tilde{j}_k$ we have

$$|\hat{i}_{j,k}-i^*_{j,k}|\geq 2. \tag{6.101}$$

We also introduce a quantity as follow.

$$j_k^*=\min\{j:m_j\leq\frac{\rho_z(\varepsilon;f_k)}{4}\}. \tag{6.102}$$

We have

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{f}}(\|CI_k\|^2) \\
&\leq 169\sum_{j=3}^{\infty}\mathbb{E}_{\mathbf{f}}(2^{-2j}\mathbb{1}\{\hat{j}(\alpha/s,k)=j\}) \\
&\leq 169\sum_{j=3}^{\infty}\mathbb{E}_{\mathbf{f}}(2^{-2j}\mathbb{1}\{\hat{j}(\alpha/s,k)=j,\acute{j}_k\leq j\}) + 169\sum_{j=3}^{\infty}\mathbb{E}_{\mathbf{f}}(2^{-2j}\mathbb{1}\{\hat{j}(\alpha/s,k)=j,\acute{j}_k> j\}) \\
&\leq 169\sum_{j=3}^{\infty}\mathbb{E}_{\mathbf{f}}(2^{-2\acute{j}_k}\mathbb{1}\{\hat{j}(\alpha/s,k)=j,\acute{j}_k\leq j\}) + 169\sum_{j=3}^{\infty}\mathbb{E}_{\mathbf{f}}(2^{-2j}\mathbb{1}\{\hat{j}(\alpha/s,k)=j,\acute{j}_k> j\}) \\
&\leq 169\mathbb{E}_{\mathbf{f}}(2^{-2\acute{j}_k}) + 169\sum_{j=3}^{\infty}\mathbb{E}_{\mathbf{f}}(2^{-2j}\mathbb{1}\{\hat{j}(\alpha/s,k)=j,\acute{j}_k> j\}).
\end{aligned} \tag{6.103}
$$

We will bound the two terms separately, now we start with the first term.

Note that we have $\grave{j}_k \geq \acute{j}_k \geq \tilde{j}_k$ and that $\tilde{j}_k = j$ implies one of the following happens:

$$
\begin{aligned}
&\{X_{j,i^*_{j,k}+1,k}\geq X_{j,i^*_{j,k}+2,k}\},\{X_{j,i^*_{j,k}+1,k}\geq X_{j,i^*_{j,k}+3,k}\},\{X_{j,i^*_{j,k}+1,k}\geq X_{j,i^*_{j,k}+4,k}\},\\
&\{X_{j,i^*_{j,k}-1,k}\geq X_{j,i^*_{j,k}-2,k}\},\{X_{j,i^*_{j,k}-1,k}\geq X_{j,i^*_{j,k}-3,k}\},\{X_{j,i^*_{j,k}-1,k}\geq X_{j,i^*_{j,k}-4,k}\}.
\end{aligned} \tag{6.104}
$$

Also we have for $j \geq j^*_k + 3$, $m_j > \rho_z(\varepsilon; f_k)$.

So we have

(6.105)

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{f}}(2^{-2\acute{j}_k}) \\
&\leq \mathbb{E}_{\mathbf{f}}(2^{-2\tilde{j}_k}) \leq 2^{-2j_k^*+6} + \sum_{j=3}^{j_k^*-4} 2^{-2j}\mathbb{E}_{\mathbf{f}}(\mathbb{1}\{\tilde{j}_k = j\}) \\
&\leq 4\rho_z(\varepsilon; f_k)^2 + \sum_{j=3}^{j_k^*-4} 2^{-2j} \times 2 \times \Bigg(\Phi(-\frac{\rho_m(\varepsilon; f_k)}{\rho_z(\varepsilon; f_k)} \frac{(2^{j_k^*-3-j}\rho_z(\varepsilon; f_k))^{\frac{3}{2}}}{\sqrt{3}\varepsilon}) + \\
&\qquad \Phi(-2\frac{\rho_m(\varepsilon; f_k)}{\rho_z(\varepsilon; f_k)} \frac{(2^{j_k^*-3-j}\rho_z(\varepsilon; f_k))^{\frac{3}{2}}}{\sqrt{3}\varepsilon}) + \Phi(-3\frac{\rho_m(\varepsilon; f_k)}{\rho_z(\varepsilon; f_k)} \frac{(2^{j_k^*-3-j}\rho_z(\varepsilon; f_k))^{\frac{3}{2}}}{\sqrt{3}\varepsilon}) \Bigg) \\
&\leq 4\rho_z(\varepsilon; f_k)^2 + \sum_{j=3}^{j_k^*-4} 2^{-2j} \times 2 \times \Bigg(\Phi(-2^{\frac{3(j_k^*-3-j)}{2}} \frac{1}{\sqrt{2}} \frac{1}{\sqrt{3}}) + \\
&\qquad \Phi(-2 \times 2^{\frac{3(j_k^*-3-j)}{2}} \frac{1}{\sqrt{2}} \frac{1}{\sqrt{3}}) + \Phi(-3 \times 2^{\frac{3(j_k^*-3-j)}{2}} \frac{1}{\sqrt{2}} \frac{1}{\sqrt{3}}) \Bigg) \\
&\leq 4\rho_z(\varepsilon; f_k)^2 + 32\rho_z(\varepsilon; f_k)^2 \Big(\frac{\Phi(-\frac{2}{\sqrt{3}})}{1 - 8\sqrt{2}\exp(-\frac{7}{2} \cdot \frac{4}{3})} + \frac{\Phi(-\frac{4}{\sqrt{3}})}{1 - 8\sqrt{2}\exp(-\frac{7}{2} \cdot \frac{16}{3})} \\
&\qquad + \frac{\Phi(-2\sqrt{3})}{1 - 8\sqrt{2}\exp(-\frac{7}{2} \cdot 12)} \Big) \\
&\leq 4\rho_z(\varepsilon; f_k)^2 + 4.5\rho_z(\varepsilon; f_k)^2 = 8.5\rho_z(\varepsilon; f_k)^2.
\end{aligned}
$$

Now we turn to the second term in Inequality (6.103). We first define three quantities.

Let the average of f_k over $[t_{j,i-1}, t_{j,i}]$ to be

$$\bar{f}_{j,i,k} = 2^j \int_{2^{-j}\times(i-1)}^{2^{-j}\times i} f_k(t)dt.$$

For $i > 2^j$ or $i \leq 0$, define $\bar{f}_{j,i,k} = +\infty$. And suppose $\infty - a = \infty$ for $a \in [-\infty, \infty]$, and $\min\{\infty, a\} = a$ for $a \in [-\infty, \infty]$.

Let the minimum of the difference of the two neighboring intervals be

$$\Xi_{j,k} = \min\{\bar{f}_{j,i^*_{j,k}+2,k} - \bar{f}_{j,i^*_{j,k}+1,k}, \bar{f}_{j,i^*_{j,k}-2,k} - \bar{f}_{j,i^*_{j,k}-1,k}\}. \tag{6.106}$$

Let $j(\zeta, k)$ be the level j such that the signal part in $T_{j,k}$ is relatively small, specifically defined as follow.

$$j(\zeta, k) = \min\{j : \Xi_{j,k} \cdot 2^{-\frac{j}{2}} \frac{1}{\sqrt{6}\varepsilon} \leq z_\zeta + 1\}. \tag{6.107}$$

Note that $j(\zeta, k)$ is a determined quantity depending only on ζ and f_k. Recall that $\hat{j}(\alpha/s, k)$ is the stopping level, which is a random variable.

Also note that for $j \leq j(\alpha/s, k) - 1$ we have

$$\Xi_{j,k} \cdot 2^{-\frac{j}{2}} \frac{1}{\sqrt{6}\varepsilon} \geq 2^{\frac{3(j(\alpha/s,k)-1-j)}{2}} (z_{\alpha/s} + 1) \tag{6.108}$$

With these quantities, we have

(6.109)
$$\begin{aligned}
&\sum_{j=3}^{\infty} 2^{-2j} \mathbb{E}_{\mathbf{f}}(\mathbb{1}\{\hat{j}(\alpha/s, k) = j, \acute{j}_k > j\}) \\
&\leq 2^{-2j(\alpha/s,k)+1} + \sum_{j=3}^{j(\alpha/s,k)-1} 2^{-2j} \Phi(-(z_{\alpha/s} + 1) \times 2^{\frac{3}{2}(j-j(\alpha/s,k)+1)} + z_{\alpha/s}) \\
&\leq 2^{-2j(\alpha/s,k)+1} + 2^{-2j(\alpha/s,k)+2} \Phi(-1) \frac{1}{1 - \Phi(-2\sqrt{2})/\Phi(-1)} \\
&< 3 \cdot 2^{-2j(\alpha/s,k)}.
\end{aligned}$$

Now we introduce a lemma.

Lemma 6.6. *For $j(\zeta, k)$ defined in (6.107), with $\zeta \leq 0.3$ we have*

$$\left(\frac{6\sqrt{2}(z_\zeta + 1)}{z_\zeta}\right)^{\frac{2}{3}} \rho_z(z_\zeta \varepsilon; f_k) \geq 2^{-j(\zeta,k)}. \tag{6.110}$$

Proof. Without loss of generality, we assume

$$\bar{f}_{j,i_{j(\zeta,k),k}+2,k} - \bar{f}_{j,i_{j(\zeta,k),k}+1,k} = \Xi_{j(\zeta,k)}.$$

Let $\mu_k = \min\{f_k(\max\{t_{j(\zeta,k),i^*_{j(\zeta,k),k}-2}, 0\}), f_k(t_{j(\zeta,k),i^*_{j(\zeta,k),k}+1})\}$. Let the $g_{lo} \in \mathcal{F}$ be defined as $g_{lo}(t) = \max\{f_k(t), \mu_k\}$.

For simplicity of notation, let $j_0 = j(\zeta, k)$, $i^* = i^*_{j(\zeta,k),k}$.

Therefore,

(6.111)
$$\begin{aligned}
\|g_{lo} - f_k\|^2 &\leq (\mu_k - M(f_k))^2 \cdot 3 \cdot 2^{-j_0} \\
&\leq (f_k(t_{j_0,i^*+1}) - f_k(t_{j_0,i^*}) + f_k(t_{j_0,i^*}) - M(f_k))^2 \cdot 3 \cdot 2^{-j_0} \\
&\leq (\bar{f}_{j,i^*+2} - \bar{f}_{j,i^*+1})^2 \cdot 3 \cdot 2^{-j_0} \\
&\leq ((z_\zeta + 1) \cdot 2^{\frac{j_0}{2}} \sqrt{6}\varepsilon)^2 \cdot 3 \cdot 2^{-j_0} \\
&= 6(z_\zeta + 1)^2 \times 3\varepsilon^2.
\end{aligned}$$

Therefore,

$$2^{-j_0} \leq \rho_z(3\sqrt{2}(z_\zeta + 1)\varepsilon; f_k) \leq (\frac{6\sqrt{2}(z_\zeta+1)}{z_\zeta})^{\frac{2}{3}} \rho_z(z_\zeta \varepsilon; f_k). \tag{6.112}$$

The last inequality is due to Proposition **??** in Cai et al. (2022) and that $z_\zeta \geq z_{0.3} = 0.524$

□

Lemma 6.6 combined with Inequality (6.109), and note that $\alpha/s \leq 0.3$ we have

$$\sum_{j=3}^{\infty} 2^{-2j} \mathbb{E}_{\mathbf{f}}(\mathbb{1}\{\hat{j}(\alpha/s, k) = j, \acute{j}_k > j\}) < 136 \rho_z(z_{\alpha/s}\varepsilon; f_k)^2. \tag{6.113}$$

Also note that for $\alpha \leq 0.3$, we have $\rho_z(\varepsilon; f_k) < 2.6\rho_z(z_{\alpha/s}; f_k)$.

Therefore both terms in Inequality 6.103 are bounded by multiple times $\rho_z(z_{\alpha/s}; f_k)^2$. We conclude the proof.

6.12.3. *Proof of Proposition 4.1.* The idea of the proof is very similar to that for white noise model.

Invertibility follows from definition. Independence follows from the observation that the concatenation of the elements is this $s+1$ tuple $\mathfrak{P}(\{y_{\mathbf{i}}\})$ follows a joint normal distribution and that covariance of of elements from different places of the tuple is 0. The sufficiency rises from factorization of the probability.

6.13. *Proof of Theorem 4.1.* For simplicity of notation, we also use $\mathfrak{D}(\mathbf{f}; n)$ to denote the discretization error when $\mathbf{f}$ is univariate (i.e. setting $s = 1$).

Suppose $\mathbf{f}_0(\mathbf{t}) = f_0 + \sum_{k=1}^{s} f_k(t_k)$.

Let $h_0(\gamma) > 1$ be a function of γ that we will specify later. Define several index sets.

(6.114)

$$Ind_0 = \{k \in \{1, 2, \ldots, s\} : \frac{1}{4}\mathfrak{D}(f_k; n) \geq h_0(\gamma)\gamma\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)^2 \left(1 \wedge n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\right)\}$$

$$Ind_1 = \{k \in \{1, 2, \ldots, s\}/Ind_0 :$$

$$\mathbb{E}_{\mathbf{f}_0}\left(|\hat{Z}_k - Z(\mathbf{f}_0)_k|^2\right) \geq \gamma h_0(\gamma)\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)^2 \left(1 \wedge n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\right)\}$$

$$Ind_2 = \{1, 2, \ldots, s\}/(Ind_0 \cup Ind_1)$$

Then we have Ind_0, Ind_1, Ind_2 are a partition of $\{1,2,\ldots,s\}$ and the followings.

(6.115)

$$\sum_{k\in Ind_0\cup Ind_1} \rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2\left(1\wedge n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\right)$$
$$\leq \frac{2}{h_0(\gamma)}\sum_{k\in\{1,2,\ldots,s\}} \rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2\left(1\wedge n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\right).$$
$$\sum_{k\in Ind_0\cup Ind_1} \mathbb{E}_{\mathbf{f}_0}\left(|\hat{Z}_k - Z(\mathbf{f}_0)_k|^2\right)$$
$$\leq \gamma\sum_{k\in\{1,2,\ldots,s\}} \rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2\left(1\wedge n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\right).$$

For all k $\in Ind_2$, we have

$$\mathbb{E}_{\mathbf{f}_0}\left(|\hat{Z}_k - Z(\mathbf{f}_0)_k|^2\right) < \gamma h_0(\gamma)\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2\left(1\wedge n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\right) \text{ and}$$
$$\frac{1}{4}\mathfrak{D}(f_k;n) < h_0(\gamma)\gamma\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2\left(1\wedge n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\right)$$

For $f\in\mathcal{F}$, if $\rho_z(\varepsilon;f)\geq \frac{2}{n}$, then we know that

$$\sup_{g\in\mathcal{F}:g(i/n)=f(i/n)\text{ for }i\in\{1,2,\ldots,n\}} \rho_z(\varepsilon;g)\leq 6\rho_z(\varepsilon;f). \tag{6.116}$$

For $\mathbf{f}\in$, if $\rho_z(\varepsilon;f)\geq\frac{4}{n}$, we have

$$\frac{1}{3}\rho_z(\varepsilon;f)\leq \inf_{g\in\mathcal{F}:g(i/n)=f(i/n)\text{ for }i\in\{1,2,\ldots,n\}}\rho_z(\varepsilon;g) \tag{6.117}$$

Suppose $h_0(\gamma)\gamma<0.1$. For $f\in\mathcal{F}$, if

$$\rho_z(\varepsilon;f)<\frac{2}{n}$$

and

$$\mathfrak{D}(f;n)<4h_0(\gamma)\gamma\rho_z(\varepsilon;f)^2\left(1\wedge n\rho_z(\varepsilon;f)\right),$$

with some basic calculation we know

(6.118)

$$\tilde{C}_1\left(h_0(\gamma)\gamma\right)\rho_z(\varepsilon;f)\leq \inf_{g\in\mathcal{F}:g(i/n)=f(i/n)\text{ for }i\in\{1,2,\ldots,n\}}\rho_z(\varepsilon;g)$$
$$\leq \sup_{g\in\mathcal{F}:g(i/n)=f(i/n)\text{ for }i\in\{1,2,\ldots,n\}}\rho_z(\varepsilon;g)\leq \tilde{C}_2\left(h_0(\gamma)\gamma\right)\rho_z(\varepsilon;f),$$

where $\tilde{C}_1(h_0(\gamma)\gamma)$ and $\tilde{C}_1(h_0(\gamma)\gamma)$ are two constants depending on $h_0(\gamma)\gamma$, and

$$\lim_{\tau\to 0^+}\tilde{C}_1(\tau)=\lim_{\tau\to 0^+}\tilde{C}_2(\tau)=1.$$

Choose $\tau_0<0.1$ such that $0.5\le\tilde{C}_1(\tau_0)\le\tilde{C}_2(\tau_0)\le 2$.

Therefore, when $h_0(\gamma)\gamma<\min\{0.1,\tau_0\}$ we have that for $k\in Ind_2$,

(6.119)
$$\begin{aligned}&\frac{1}{3}\sup_{g\in\mathcal{F},g(i/n)=f_k(i/n)\text{ for }i\in\{1,2,\dots,n\}}\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};g)\\&\le\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\le 6\inf_{g\in\mathcal{F},g(i/n)=f_k(i/n)\text{ for }i\in\{1,2,\dots,n\}}\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};g).\end{aligned}$$

Define a linear operator $\mathfrak{L}_n(\cdot)$ on $\mathcal{F}$ such that $\mathfrak{L}_n(f)$ is the piecewise linear function joining $(\frac{i}{n},f(\frac{i}{n}))$.

Next, for $k\in Ind_2$, we will construct $f_{1,k}$ for kth component for $\mathbf{f}_1$. Note that $\boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})$ under $\mathbf{f}$ has the same distribution with the following vector $\nu=(\nu_0,\dots,\nu_n)$:

$$\nu_i=f_k(\frac{i}{n})-\frac{\sum_{j=0}^n f_k(\frac{j}{n})}{n+1}+\frac{\sigma}{(n+1)^{\frac{s-1}{2}}}\left(z_i-\frac{\sum_{j=0}^n z_j}{n+1}\right).\tag{6.120}$$

We will use ν_i under f_k as a proxy of $\boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})$ under $\mathbf{f}$.

For any $g\in\mathcal{F}$, suppose $\tilde{f}_\theta=g$ for $\theta=-1$ and $\tilde{f}_\theta=f_k$ for $\theta=1$, denote $\bar{f}_k=\frac{\sum_{j=0}^n f_k(\frac{j}{n})}{n+1}$, $\bar{g}=\frac{\sum_{j=0}^n g(\frac{j}{n})}{n+1}$, and $l_n(f,g)=\sqrt{\frac{\sum_{j=1}^n(f(\frac{i}{n})-g(\frac{i}{n}))^2}{n+1}}$, then a sufficient statistic for θ is

(6.121)
$$\begin{aligned}W_0&=\frac{(n+1)^{\frac{s-2}{2}}}{2l_n(f_k-\bar{f}_k,g-\bar{g})\sigma}\sum_{i=0}^n\left(2\nu_i-f_k(\frac{i}{n})-g(\frac{i}{n})+\bar{f}_k+\bar{g}\right)\left(f_k(\frac{i}{n})-\bar{f}_k-g(\frac{i}{n})+\bar{g}\right)\\&\sim N\left(\frac{\theta}{2}\frac{l_n(f_k-\bar{f}_k,g-\bar{g})(n+1)^{\frac{s}{2}}}{\sigma},1\right),\end{aligned}$$

under P_θ.

Also note that we have

$$l_n(f_k-\bar{f}_k,g-\bar{g})\le l_n(f_k,g).\tag{6.122}$$

We still use the construction of g_2 in Lemma A.5 in Cai et al. (2021) based on $\mathfrak{L}_n(f_k)$, we will specify μ there in two cases. In the following proof we also use the g_1 constructed there. Note that $\mathfrak{L}_n(f_k)$ can have two minimizers, but the constructions are still valid.

When $|Z(g_1) - Z(\mathfrak{L}_n(f_k)| \geq \frac{1}{2n}$, then we have

$$l_n(\mathfrak{L}_n(f_k), g_2) \leq \sqrt{\frac{n}{n+1}}\sqrt{6}\|\mathfrak{L}_n(f_k) - g_2\|. \tag{6.123}$$

When $|Z(g_1) - Z(\mathfrak{L}_n(f_k))| < \frac{1}{2n}$, we have

$$l_n(\mathfrak{L}_n(f_k), g_2) \leq \sqrt{\frac{n}{n+1}}\sqrt{6}\sqrt{\frac{1}{2n|Z(g_1) - Z(\mathfrak{L}_n(f_k))|}}\|\mathfrak{L}_n(f_k) - g_2\| \tag{6.124}$$

For notation simplicity, we denote $\eta(f) = \rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f)(1 \wedge \sqrt{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f)})$.

We now specify μ for g_2 differently in two cases, $\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}\sqrt{\frac{n+1}{6n}}; \mathfrak{L}_n(f_k)) \geq \frac{1}{2n}$ and $\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}\sqrt{\frac{n+1}{6n}}; \mathfrak{L}_n(f_k)) < \frac{1}{2n}$.

For $|Z(g_1) - Z(\mathfrak{L}_n(f_k)| \geq \frac{1}{2n}$, specify ε in the proof of Lemma A.5 in Cai et al. (2021) to be $\frac{\sigma}{(n+1)^{\frac{s}{2}}}\sqrt{\frac{n+1}{6n}}$, and specify μ the same way there to construct g_2. We only need to make sure $\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}\sqrt{\frac{n+1}{6n}}; \mathfrak{L}_n(f_k)) \geq 2\left(\sqrt{\mathfrak{D}(f_k; n)} + \sqrt{h_0(\gamma)\gamma}\eta(f)\right)$ for the arguments to go through, so that exists an absolute constant $\tilde{C}_3 > 0$, such that

$$\begin{aligned} |Z(g_2) - Z(f_k)| &\geq \rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}\sqrt{\frac{n+1}{6n}}; \mathfrak{L}_n(f_k)) - \sqrt{\mathfrak{D}(f_k, n)}, \\ \rho_z(\xi; g_2) &\leq \tilde{C}_3\left(\frac{\xi}{z_{2h_0(\gamma)\gamma\frac{\sigma}{(n+1)^{\frac{s}{2}}}}}\right)^{\frac{2}{3}}|Z(g_2) - Z(\mathfrak{L}_n(f_k))|. \end{aligned} \tag{6.125}$$

We only need to choose γ and h_0 such that $h_0(\gamma)\gamma \leq \min\{\tau_0, \frac{1}{18^2}, \sqrt{0.0063} \cdot (\frac{1}{2\sqrt{6}})^{\frac{2}{3}}\}$.

Further choose $\tau_1 < \min\{\tau_0, \frac{1}{18^2, \sqrt{0.0063}} \cdot (\frac{1}{2\sqrt{6}})^{\frac{2}{3}}\}\}$ such that for $h_0(\gamma)\gamma < \tau_1$,

$$\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}\sqrt{\frac{n+1}{6n}}; \mathfrak{L}_n(f_k)) - \sqrt{\mathfrak{D}(f_k, n)} \geq \frac{1}{7}\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}\sqrt{\frac{n+1}{6n}}; f_k). \tag{6.126}$$

According to $l_n(\mathfrak{L}_n(f_k), g_2) \leq \sqrt{\frac{n}{n+1}}\sqrt{6}\|\mathfrak{L}_n(f_k) - g_2\|$ and Inequality (B.19) in Cai et al. (2021), we have

$$\mathbb{E}_{g_2}(\|\hat{Z}_k - Z(g_2)\|^2) \geq \frac{1}{8}|Z(g_2) - f_k|^2. \tag{6.127}$$

We take $g_2 - \int_0^1 g_2$ to be the kth component for $\mathbf{f}_1$.

For $\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}\sqrt{\frac{n+1}{6n}}; \mathfrak{L}_n(f_k)) < \frac{1}{2n}$, first consider the same construction as in the case $\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}\sqrt{\frac{n+1}{6n}}; \mathfrak{L}_n(f_k)) \geq \frac{1}{2n}$, If $|Z(g_2) - \mathfrak{L}_n(f_k)| \geq \frac{1}{2n}$. then we have

$$l_n(\mathfrak{L}_n(f_k), g_2) \leq \sqrt{\frac{n}{n+1}}\sqrt{6}\|\mathfrak{L}_n(f_k) - g_2\|, \tag{6.128}$$

and same arguments goes.

Otherwise, $|Z(g_2) - \mathfrak{L}_n(f_k)| < \frac{1}{2n}$, denote $\rho_M = |Z(g_2) - \mathfrak{L}_n(f_k)|$, specify ε in the proof of Lemma A.5 in Cai et al. (2021) to be a series of numbers

$$\varepsilon(\rho) = \frac{\sigma}{(n+1)^{\frac{s}{2}}}\sqrt{\frac{n+1}{6n}}\sqrt{2\rho n},$$

and construct g_2 (depending on $\rho \leq \rho_M$) the same way there. Since $\rho_M < \frac{1}{2n}$, we know that the piece wise linear function g_2 and piece wise linear function $\mathfrak{L}_n(f_k)$ has intersections within the $\frac{1}{n}$ long interval around $Z(\mathfrak{L}_n(f_k))$ on each side (if exists).

Then we have that

(6.129)

$$\|g_2 - \mathfrak{L}_n(f_k)\| = \varepsilon(\rho) \times z_{2h_0(\gamma)\gamma}$$

$$|Z(g_2) - Z(\mathfrak{L}_n(f_k))| \geq \rho_z(\frac{\varepsilon(\rho) \times z_{2h_0(\gamma)\gamma}}{\sqrt{5}}; \mathfrak{L}_n(f_k)) = \rho_z(\varepsilon(\frac{1}{2n}); \mathfrak{L}_n(f_k)) \left(\frac{\sqrt{2n\rho}z_{2h_0(\gamma)\gamma}}{\sqrt{5}}\right)^{\frac{2}{3}}$$

$$|Z(g_2) - Z(\mathfrak{L}_n(f_k))| \leq \rho_z(\varepsilon(\rho) \times z_{2h_0(\gamma)\gamma}; \mathfrak{L}_n(f_k)) = \rho_z(\varepsilon(\frac{1}{2n}); \mathfrak{L}_n(f_k)) \left(\sqrt{2n\rho}z_{2h_0(\gamma)\gamma}\right)^{\frac{2}{3}}$$

$$\begin{aligned}\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; g_2) &\leq \tilde{C}_4 \left(\frac{\sqrt{3}}{\sqrt{(n+1)\rho} \times z_{2h_0(\gamma)\gamma}}\right)^{\frac{2}{3}} |Z(g_2) - Z(\mathfrak{L}_n(f_k))| \\ &\leq \tilde{C}_4 \times 3^{\frac{1}{3}} \rho_z(\varepsilon(\frac{1}{2n}); \mathfrak{L}_n(f_k))\end{aligned}$$

$$\frac{l_n(g_2, \mathfrak{L}_n(f_k))}{\frac{\sigma}{(n+1)^{\frac{s}{2}}}} \leq z_{2h_0(\gamma)\gamma}\sqrt{\frac{\rho}{|Z(g_2) - Z(\mathfrak{L}_n(f_k))|}} \leq z_{2h_0(\gamma)\gamma}^{\frac{2}{3}} \frac{(2n\rho)^{\frac{1}{3}}}{\sqrt{2n\rho_z(\varepsilon(\frac{1}{2n}); \mathfrak{L}_n(f_k))}} 5^{\frac{1}{3}},$$

where $\tilde{C}_4 > 0$ is an absolute positive constant.

Choose ρ such that $2n\rho = \left(2n\rho_z(\varepsilon(\frac{1}{2n}); \mathfrak{L}_n(f_k))\right)^{\frac{3}{2}}$. This choice gives a valid $\rho \le \rho_M$ as $\|g_2 - \mathfrak{L}_n(f_k)\| \le \frac{\sigma}{(n+1)^{\frac{s}{2}}}\sqrt{\frac{n+1}{6n}} \times z_{2h_0(\gamma)\gamma}$.

An immediate consequence of this choice of ρ is that

(6.130)

$$|Z(g_2) - Z(\mathfrak{L}_n(f_k))| \ge \rho_z(\varepsilon(\frac{1}{2n}); \mathfrak{L}_n(f_k))\sqrt{2n\rho_z(\varepsilon(\frac{1}{2n}); \mathfrak{L}_n(f_k))}\frac{z_{2h_0(\gamma)\gamma}^{\frac{2}{3}}}{5^{\frac{1}{3}}}$$
$$|Z(g_2) - Z(\mathfrak{L}_n(f_k))| \ge \tilde{C}_5\rho_z(\varepsilon(\frac{1}{2n}); g_2)\sqrt{2n\rho_z(\varepsilon(\frac{1}{2n}); g_2)}z_{2h_0(\gamma)\gamma}{}^{\frac{2}{3}},$$

where $\tilde{C}_5 > 0$ is an absolute positive constant.

Choosing $h_0(\gamma)\gamma \le \min\{\tau_1, \tau_2\}$ such that $z_{2h_0(\gamma)\gamma}^{\frac{2}{3}} 5^{\frac{1}{3}} \le z_{2h_0(\gamma)\gamma}$, using the arguments in Inequality (B.19) in the proof of Lemma A.5 in Cai et al. (2021), we have

$$\mathbb{E}_{g_2}\left(|\hat{Z}_k - Z(g_2)|^2\right) \ge \frac{1}{8}|Z(g_2) - Z(f_k)|^2. \tag{6.131}$$

Also note that $\frac{1}{4}\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f) \le \rho_z(\varepsilon(\frac{1}{2n}); f) \le \rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f)$ for any $f \in \mathcal{F}$.

Combining two case, we have that for $h_0(\gamma)\gamma \le \min\{\tau_1, \tau_2\}$ there exist $g_2 \in \mathcal{F}$, such that

(6.132)

$$|Z(g_2) - Z(f_k)| \ge \tilde{C}_6\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\sqrt{\left(1 \wedge n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\right)}$$
$$|Z(g_2) - Z(f_k)| \ge \tilde{C}_7 z_{2h_0(\gamma)\gamma}^{\frac{2}{3}}\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; g_2)\sqrt{\left(1 \wedge n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; g_2)\right)}$$
$$\mathbb{E}_{g_2}\left(|\hat{Z}_k - Z(g_2)|^2\right) \ge \frac{1}{8}|Z(g_2) - Z(f_k)|^2,$$

where $\tilde{C}_6, \tilde{C}_7$ are positive constants. Let $f_{1,k} = g_2 - \int_0^1 g_2$.

Now we turn back to the components in $Ind_0 \cup Ind_1$. For $k \in Ind_0 \cup Ind_1$, let $f_{1,k} = f_k$.

Then we have that

(6.133)
$$\mathbb{E}_{\mathbf{f}_1}(\|\hat{Z}-Z(\mathbf{f}_1)\|^2) \geq \sum_{k\in Ind_0\cup Ind_1} \mathbb{E}_{\mathbf{f}_0}\left(\|\hat{Z}_k - Z(f_k)\|^2\right) + \sum_{k\in Ind_2} \frac{1}{8}|Z(f_{1,k})-Z(f_k)|^2,$$

$$\sum_{k\in Ind_2} |Z(f_{1,k})-Z(f_k)|^2 \geq \sum_{k\in Ind_2} \tilde{C}_6^2 \rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)^2 \left(1\wedge n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\right)$$

$$\sum_{k\in Ind_2} |Z(f_{1,k})-Z(f_k)|^2 \geq \sum_{k\in Ind_2} \tilde{C}_7^2 z_{2h_0(\gamma)\gamma}^{\frac{4}{3}} \rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_{1,k})^2 \left(1\wedge n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_{1,k})\right).$$

Recalling the Inequality (6.115), we have

$$\begin{aligned}
&\sum_k \rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_{1,k})^2 \left(1\wedge n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_{1,k})\right) \\
&\leq \left(\frac{2}{h_0(\gamma)-2}\frac{1}{\tilde{C}_6^2} + z_{2h_0(\gamma)\gamma}^{-\frac{4}{3}}\frac{1}{\tilde{C}_7^2}\right) \sum_{k\in Ind_2} |Z(f_{1,k})-Z(f_k)|^2 \leq \\
&\leq \sum 8\left(\frac{2}{h_0(\gamma)-2}\frac{1}{\tilde{C}_6^2} + z_{2h_0(\gamma)\gamma}^{-\frac{4}{3}}\frac{1}{\tilde{C}_7^2}\right) \mathbb{E}_{\mathbf{f}_1}(\|\hat{Z}-Z(\mathbf{f}_1)\|^2).
\end{aligned} \tag{6.134}$$

Set $h_0(\gamma) = 2 + z_{2\gamma}^{\frac{4}{3}}$. Then we know $h_0(\gamma)\gamma \sim \gamma\left(\log(\frac{1}{\gamma})\right)^{\frac{2}{3}}$ and $h_0(\gamma)\gamma \to 0^+$ as $\gamma \to 0^+$. So we there is $\gamma_1 > 0$ such that all the requirements for $h_0(\gamma)\gamma$ are satisfied. Also note that $z_{2h_0(\gamma)\gamma} \sim \sqrt{\log(\frac{1}{\gamma})}$ as $\gamma \to 0^+$. Therefore, we have the statement.

6.14. *Proof of Theorem 4.2.* The first assertion is almost trivial. For any $\frac{1}{2}\mathfrak{D}(\mathbf{f}; n) > \delta > 0$, let $\mathbf{g}_1, \mathbf{g}_2 \in \mathcal{F}_s$, such that $\mathbf{g}_1(\frac{\mathbf{i}}{n}) = \mathbf{g}_2(\frac{\mathbf{i}}{n}) = \mathbf{f}(\frac{\mathbf{i}}{n})$ for all $\mathbf{i} \in \{0, 1, \ldots, n\}^s$ and $\|Z(\mathbf{g}_1) - Z(\mathbf{g}_2)\|^2 \geq \mathfrak{D}(\mathbf{f}; n) - \delta$.

Then we have $\{y_{\mathbf{i}}\}$ has same distribution under $\mathbf{g}_1$, $\mathbf{g}_2$ and $\mathbf{f}$. Then we have

(6.135)
$$\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right) \geq \mathbb{E}_{\mathbf{f}}\left(\mathbb{1}\{Z(\mathbf{g}_1), Z(\mathbf{g}_2) \in CI\}\right)(\mathfrak{D}(\mathbf{f}; n) - \delta) \geq (1-2\alpha)(\mathfrak{D}(\mathbf{f}; n) - \delta).$$

Let $\delta \to 0^+$ gives the first statement.

Now we turn to the second statement.

Suppose $\alpha_k = \inf_{\mathbf{f}\in\mathcal{F}_s} P_{\mathbf{f}}\left(Z(\mathbf{f})_k \in CI_k\right)$, then the fact that CI is a separable confidence hyper cube implies that

$$(1-\alpha) \geq \Pi_{k=1}^s (1-\alpha_k). \tag{6.136}$$

Then by Lemma 6.4 and the arguments following that, we have

$$\sum_{k=1}^{s} \alpha_k \leq 2\alpha. \tag{6.137}$$

Apparently $0 \leq \alpha_k \leq \alpha$

Note that $\varphi(z_{\alpha_k}\frac{\sigma}{(n+1)^{\frac{s}{2}}}, f_k)$ increases with α_k and that CI_k only depends on $\boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})$ (the distribution of which only depends on f_k, thus we use f_k to indicate its distribution), we are only left with proving the following lemma

LEMMA 6.7. *Suppose $\alpha_k \leq 0.2$. For $1-\alpha_k$ level confidence interval CI_k of $Z(f_k)$ based on $\boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})$*

$$\mathbb{E}_{f_k}\left(|CI_k|^2\right) \geq c_4\varphi(z_{\alpha_k}\frac{\sigma}{(n+1)^{\frac{s}{2}}}, f_k)^2 \tag{6.138}$$

holds for all $f_k \in \mathcal{F}$ that $\int_0^1 f_k = 0$.

PROOF. When $\alpha_k = 0$, it apparently holds. Now we consider the case that $\alpha_k > 0$.

Now, given the distribution of $\boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})$, which is discussed in the proof of Theorem 4.1, we will rephrase this statement and make it a bit stronger into the following lemma.

LEMMA 6.8. *Suppose $f \in \mathcal{F}$ such that $\int_0^1 f = 0$. Observation $\nu = (\nu_0, \dots, \nu_n)$ has its each element being $\nu_i = f(\frac{i}{n}) + \varepsilon(z_i - \frac{\sum_i^s z_i}{n+1})$, where $z_i \overset{i.i.d}{\sim} N(0,1)$. Suppose zeta ≤ 0.2. Then $1-\zeta$ confidence interval CI for $Z(f)$ satisfies*

$$\mathbb{E}_f(CI^2) \geq c_4\varphi(z_\zeta\frac{\varepsilon}{\sqrt{n}}; f)^2. \tag{6.139}$$

PROOF. For any $g \in \mathcal{F}$, the distribution of ν under g is the same as that under $g - \int_0^1 g(t)\mathrm{d}t$, and $Z(g) = Z(g - \int_0^1 g(t)\mathrm{d}t)$. Therefore, we can extend the function class $\mathcal{F} \cap \{g : \int_0^1 g(t)\mathrm{d}t = 0\}$ to $\mathcal{F}$. If for this larger class (i.e. $\mathcal{F}$), the lemma holds, then the original lemma holds.

For any $g \in \mathcal{F}$, suppose $\tilde{f}_\theta = g$ for $\theta = -1$ and $\tilde{f}_\theta = f$ for $\theta = 1$, denote $\bar{f} = \frac{\sum_{j=0}^n f(\frac{j}{n})}{n+1}$, $\bar{g} = \frac{\sum_{j=0}^n g(\frac{j}{n})}{n+1}$, and $l_n(f,g) = \sqrt{\frac{\sum_{j=1}^n (f(\frac{i}{n}) - g(\frac{i}{n}))^2}{n+1}}$, then a sufficient statistic for θ is

(6.140)
$$W_0 = \frac{\frac{1}{\sqrt{n+1}}}{2l_n(f-\bar{f}, g-\bar{g})\varepsilon} \sum_{i=0}^{n} \left(2\nu_i - f(\frac{i}{n}) - g(\frac{i}{n}) + \bar{f} + \bar{g}\right)\left(f(\frac{i}{n}) - \bar{f} - g(\frac{i}{n}) + \bar{g}\right)$$
$$\sim N\left(\frac{\theta}{2}\frac{l_n(f-\bar{f}, g-\bar{g})(n+1)^{\frac{1}{2}}}{\varepsilon}, 1\right),$$

under P_θ.

Also note that

$$l_n(f-\bar{f}, g-\bar{g}) \le l_n(f,g). \tag{6.141}$$

We discuss two cases: $\rho_z((z_\zeta)\frac{\varepsilon}{\sqrt{6(n+1)}}; f) \ge \frac{1}{2n}$ and $\rho_z(z_\zeta\frac{\varepsilon}{\sqrt{6(n+1)}}; f) < \frac{1}{2n}$.

Define a function g_μ depending on μ, $g_\mu(t) = \max\{f(t), \mu + M(f)\}$.

When $\rho_z(z_\zeta\frac{\varepsilon}{\sqrt{6(n+1)}}; f) \ge \frac{1}{2n}$, let $\mu = \rho_m(z_\zeta\frac{\varepsilon}{\sqrt{6(n+1)}}; f)$ then we have

$$l_n(f, g_\mu)\sqrt{n+1} \le \sqrt{6n}\frac{z_\zeta\varepsilon}{\sqrt{6(n+1)}} < \varepsilon z_\zeta. \tag{6.142}$$

Then we have

(6.143)
$$\mathbb{E}_f(|CI|^2) \ge (0.5-\zeta)\rho_z(z_\zeta\frac{\varepsilon}{\sqrt{6(n+1)}}; f)^2 \ge \tilde{c}_4\rho_z(z_\zeta\frac{\varepsilon}{\sqrt{n}}; f)^2(1\wedge\sqrt{n\rho_z(z_\zeta\frac{\varepsilon}{\sqrt{n}}; f)})^2,$$

for an absolute positive constant $\tilde{c}_4$.

When $\rho_z(z_\zeta\frac{\varepsilon}{\sqrt{6(n+1)}}; f) < \frac{1}{2n}$.

Let

(6.144)
$$\mu = \min\left\{f\left(Z(f) - \rho_z(z_\zeta\frac{\varepsilon}{\sqrt{6(n+1)}}; f)\sqrt{2n\rho_z(z_\zeta\frac{\varepsilon}{\sqrt{6(n+1)}}; f)}\right),\right.$$
$$\left. f\left(Z(f) + \rho_z(z_\zeta\frac{\varepsilon}{\sqrt{6(n+1)}}; f)\sqrt{2n\rho_z(z_\zeta\frac{\varepsilon}{\sqrt{6(n+1)}}; f)}\right)\right\},$$

Then we have

$$
\begin{aligned}
& l_n(f,g_\mu)\sqrt{n+1} \\
& \le \sqrt{\frac{6}{\rho_z(z_\zeta\frac{\varepsilon}{\sqrt{6(n+1)}};f)}}\|f-g_\mu\| \\
& \le \sqrt{\frac{6}{\rho_z(z_\zeta\frac{\varepsilon}{\sqrt{6(n+1)}};f)}}\sqrt{n\rho_z(z_\zeta\frac{\varepsilon}{\sqrt{6(n+1)}};f)z_\zeta\frac{\varepsilon}{\sqrt{6(n+1)}}} \\
& \le z_\zeta\varepsilon.
\end{aligned} \tag{6.145}
$$

Therefore, we have

(6.146)

$$
\begin{aligned}
\mathbb{E}_f(|CI|^2) &\ge (0.5-\zeta)\rho_z(z_\zeta\frac{\varepsilon}{\sqrt{6(n+1)}};f)^2(1\wedge\sqrt{2n\rho_z(z_\zeta\frac{\varepsilon}{\sqrt{6(n+1)}};f)})^2 \\
&\ge \tilde{c}_5\rho_z(z_\zeta\frac{\varepsilon}{\sqrt{n}};f)^2(1\wedge\sqrt{n\rho_z(z_\zeta\frac{\varepsilon}{\sqrt{n}};f)})^2,
\end{aligned}
$$

for an absolute positive constant $\tilde{c}_5$.

Taking $c_4=\min\{\tilde{c}_4,\tilde{c}_5\}$ gives the statement of the lemma.

□

□

6.15. *Proof of Theorem 4.3.* The first statement follows naturally from Theorem 4.2.

For the second statement, Let $f_0=0$, $f_k=H|t-0.5|$ for all $k\in\{1,2,\cdots,s\}$, where we specify H to be large enough that that $\rho_z(z_{\alpha/s}\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\le\frac{1}{n}$.

Suppose $\alpha_k=P(Z(f_k)\notin CI_k)$. Apparently $\alpha_k\le\alpha$.

Without loss of generality, we can assume $\alpha_1\le\alpha_2\le\cdots,\alpha_s$. Then $\sum_{k=1}^s\alpha_k\le 2\alpha$, according to Lemma 6.4 and arguments following that.

Then we know that for $k\le\lfloor s/2\rfloor$, $\alpha_k\le\min\{\alpha,4\alpha/s\}$. Let $\tau=\ min\{\alpha,4\alpha/s\}$.

By Lemma 6.7, we have

(6.147)

$$
\mathbb{E}_{\mathbf{f}}\left(diag\,(CI)^2\right)\ge c_4\lfloor s/2\rfloor\varphi(z_\tau\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2=c_4\lfloor s/2\rfloor\left(\frac{z_\tau}{z_{\alpha/s}}\right)^2\varphi(z_{\alpha/s}\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2\ge 0.1c_4\Psi_2(\mathbf{f};\alpha).
$$

On the other hand

$$
\Psi_2(\mathbf{f};\alpha)=\left(\frac{z_{\alpha/s}}{z_\alpha}\right)^2 Psi_1(\mathbf{f};\alpha). \tag{6.148}
$$

This concludes the theorem.

6.16. *Proof of Theorem 4.4.* We have

$$\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z}-Z(\mathbf{f})\|^2\right) \leq \sum_{k=1}^{s} \mathbb{E}_{\mathbf{f}}\left(\|\hat{Z}_k - Z(f_k)\|^2\right). \tag{6.149}$$

Note that we have

$$\rho_z((z_\zeta+1)\frac{\sqrt{6}\sigma}{\sqrt{n}(n+1)^{\frac{s-1}{2}}}; f_k) \leq \left(3\times 2\sqrt{3}\right)^{\frac{2}{3}} \rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k) \tag{6.150}$$

for $\zeta \geq \Phi(-2)$, so it is sufficient to prove the following

$$\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z}_k - Z(f_k)\|^2\right) \leq \check{C}_2 \rho_z(; f_k) + 2\mathfrak{D}(f_k; n), \tag{6.151}$$

for an absolute constant $\check{C}_2 > 0$.

Now we proceed with this.

First let

(6.152)

$$\xi_k(\zeta) = \sup\Bigg\{\xi : \min\Bigg\{\sqrt{\xi}\left[f_k(Z(f_k)+\xi) - M(f_k)\right], \\ \sqrt{\xi}\left[f_k\left(Z(f_k)-\xi\right) - M(f_k)\right]\Bigg\} \times \frac{\sqrt{n}}{\sqrt{6}\sigma/(n+1)^{\frac{s-1}{2}}} \leq z_\zeta + 1\Bigg\},$$

Then let

$$\mathtt{j}_k(\zeta) = \max\{j : \frac{2^{J-j}}{n} > \xi_k(\zeta)\}. \tag{6.153}$$

We further introduce the following quantities.

$$\begin{aligned} \mathtt{i}^*_{k,j} &= \max\{i : Z(f_k) \in \left[\frac{2^{J-j}\cdot(i-1)}{n} - \frac{1}{2n}, \frac{2^{J-j}\cdot i}{n} - \frac{1}{2n}\right]\} \\ \tilde{\mathtt{j}}_k &= \min\left(\{j : |\hat{\mathtt{i}}_{k,j} - \mathtt{i}^*_{k,j}| \geq 2\} \cup \infty\right), \\ \acute{\mathtt{j}}_k &= \min\left(\{j : |\hat{\mathtt{i}}_{k,j} - \mathtt{i}^*_{k,j}| \geq 5\} \cup \infty\right), \\ \grave{\mathtt{j}}_k &= \min\left(\{j : |\hat{\mathtt{i}}_{k,j} - \mathtt{i}^*_{k,j}| \geq 7\} \cup \infty\right). \end{aligned} \tag{6.154}$$

Then we immediately have the following facts that we summarize into a lemma.

LEMMA 6.9. *For $j \le \min\{J, \mathtt{j}_k(\zeta)\}$, we have*

$$\frac{1}{\tilde{\sigma}_{k,j}} \sum_{h=(\mathtt{i}^*_{k,j}+1)2^{J-j}}^{(\mathtt{i}^*_{k,j}+2)2^{J-j}-1} \left(f_k(\frac{h+2^{J-j}}{n}) - f_k(\frac{h}{n}) \right) \ge 2^{\frac{3}{2}(\mathtt{j}_k(\zeta)-j)} \left(z_\zeta + 1\right), \tag{6.155}$$

and

$$\frac{1}{\tilde{\sigma}_{k,j}} \sum_{h=(\mathtt{i}^*_{k,j}-2)2^{J-j}}^{(\mathtt{i}^*_{k,j}-1)2^{J-j}-1} \left(f_k(\frac{h+2^{J-j}}{n}) - f_k(\frac{h}{n}) \right) \ge 2^{\frac{3}{2}(\mathtt{j}_k(\zeta)-j)} \left(z_\zeta + 1\right). \tag{6.156}$$

Let $j = \tilde{\mathtt{j}}_k$, then one of the following happens

$$\begin{aligned} &\mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}+2} \le \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}+1}, \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}+3} \le \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}+1}, \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}+4} \le \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}+1} \\ &\mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}-2} \le \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}-1}, \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}-3} \le \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}-1}, \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}-4} \le \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}-1}. \end{aligned} \tag{6.157}$$

Now we will state three lemmas, the proofs of which are left to latter parts.

LEMMA 6.10. *Suppose $\zeta \le 0.5$.*

$$\mathbb{E}_{\mathbf{f}} \left(2^{-2\tilde{\mathtt{j}}_k} \mathbb{1}\{\tilde{\mathtt{j}}_k \le J\} \right) \le \check{C}_0 2^{-2\mathtt{j}_k(\zeta)} \left(1 \wedge 2^{J-\mathtt{j}_k(\zeta)}\right), \tag{6.158}$$

where $\check{C}_0 = \max\{\sup_{x\ge 1} 2x^2\Phi(-x), 2\}$.

LEMMA 6.11. *Suppose $\zeta \le 0.5$.*

$$\mathbb{E}_{\mathbf{f}} \left(2^{-2\check{\mathtt{j}}_k(\zeta)} \mathbb{1}\{\check{\mathtt{j}}_k(\zeta) < \infty\} \mathbb{1}\{\tilde{\mathtt{j}}_k > \check{\mathtt{j}}_k(\zeta)\} \right) \le \check{C}_0 2^{-2\mathtt{j}_k(\zeta)} \left(1 \wedge 2^{J-\mathtt{j}_k(\zeta)}\right), \tag{6.159}$$

where $\check{C}_0 = \max\{\sup_{x\ge 1} 2x^2\Phi(-x), 2\}$.

REMARK 6.1. Note that the left hand side of Inequality (**??**) does not depend on ζ, but we state this more general lemma.

LEMMA 6.12. *Suppose $\zeta \le 0.5$.*

$$\begin{aligned} &\mathbb{E}_{\mathbf{f}} \left(|\hat{Z}_k - Z(f_k)|^2 \mathbb{1}\{\check{\mathtt{j}}_k(\zeta) = \infty, \tilde{\mathtt{j}}_k > J\} \right) \\ &\le 64 \cdot 2^{-2\mathtt{j}_k(\zeta)} \left(1 \wedge 2^{J-\mathtt{j}_k(\zeta)}\right) + 2\mathfrak{D}(f_k; n). \end{aligned} \tag{6.160}$$

With these lemmas, we have that

$$\mathbb{E}_{\mathbf{f}}\left(|\hat{Z}_k - Z(f_k)|^2\right) \le \check{C}_1 \cdot 2^{-2\mathtt{j}_k(\zeta)}\left(1 \wedge 2^{J-\mathtt{j}_k(\zeta)}\right) + 2\mathfrak{D}(f_k;n), \tag{6.161}$$

where $\check{C}_1 = 64 + 2\check{C}_0$.

Now we introduce the flowing lemma about $\xi_k(\zeta)$ and $\mathtt{j}_k(\zeta)$, which immediately concludes the proof.

Lemma 6.13.

$$2\rho_z((z_\zeta+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k) \ge \xi_k(\zeta) \ge \frac{1}{2}\rho_z((z_\zeta+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k). \tag{6.162}$$

$$\frac{n+2}{2} \le 2^J \le n+1, 2^{-\mathtt{j}_k(\zeta)} \le \frac{2n}{2^J}\xi_k(\zeta) \le 8\rho_z((z_\zeta+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k). \tag{6.163}$$

6.16.1. *Proof of Lemma 6.10.* A basic property of normal tail bound is that $\frac{\Phi-2\sqrt{2}x}{\Phi(-x)}$ decreases with $x>0$ increasing.

$$\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(2^{-2\tilde{\mathtt{j}}_k}\mathbb{1}\{\tilde{\mathtt{j}}_k \le J\}\right)\\
&\le \sum_{j=1}^{J} 2^{-2\mathtt{j}_k(\zeta)} \cdot 2^{-2j+2\mathtt{j}_k(\zeta)}\left(\Phi(-2^{\frac{3}{2}(\mathtt{j}_k(\zeta)-j)}(z_\zeta+1))\mathbb{1}\{j \le \mathtt{j}_k(\zeta)\} + \mathbb{1}\{j > \mathtt{j}_k(\zeta)\}\right)\\
&\le \mathbb{1}\{J \le \mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)} \cdot 2^{-2J+2\mathtt{j}_k(\zeta)}\Phi(-2^{\frac{3}{2}(\mathtt{j}_k(\zeta)-J)}(z_\zeta+1))\frac{1}{1-4\frac{\Phi(-2\sqrt{2})}{\Phi(-1)}}\\
&\quad + \mathbb{1}\{J > \mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)}\left(\frac{1}{1-4\frac{\Phi(-2\sqrt{2})}{\Phi(-1)}} + \frac{1}{3}\right)\\
&\le \mathbb{1}\{J \le \mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)} \cdot 2^{J-\mathtt{j}_k(\zeta)}\sup_{x\ge 1} 2x^2\Phi(-x) + 2\cdot\mathbb{1}\{J > \mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)}
\end{aligned} \tag{6.164}$$

Let $\check{C}_0 = \max\{\sup_{x\ge1} 2x^2\Phi(-x), 2\}$, then we have the lemma.

6.16.2. *Proof of Lemma 6.11.* By our stopping rule, apparently $\check{\mathtt{j}}_k(\zeta) \ge 1$.

(6.165)

$$\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(2^{-2\check{\mathtt{j}}_k(\zeta)}\mathbb{1}\{\check{\mathtt{j}}_k(\zeta)<\infty\}\mathbb{1}\{\tilde{\mathtt{j}}_k>\check{\mathtt{j}}_k(\zeta)\}\right)\\
&=\sum_{j=1}^{J}2^{-2j}\mathbb{E}_{\mathbf{f}}\left(\mathbb{E}_{\mathbf{f}}\left(\mathbb{1}\{\tilde{\mathtt{j}}_k>\check{\mathtt{j}}_k(\zeta)=j\}\big|\nu^l_{k,i}\right)\right)\\
&\le\sum_{j=1}^{J}2^{-2\mathtt{j}_k(\zeta)}\cdot 2^{-2j+2\mathtt{j}_k(\zeta)}\left(\Phi(-2^{\frac{3}{2}(\mathtt{j}_k(\zeta)-j)}(z_\zeta+1))\mathbb{1}\{j\le\mathtt{j}_k(\zeta)\}+\mathbb{1}\{j>\mathtt{j}_k(\zeta)\}\right)\\
&\le\mathbb{1}\{J\le\mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)}\cdot 2^{-2J+2\mathtt{j}_k(\zeta)}\Phi(-2^{\frac{3}{2}(\mathtt{j}_k(\zeta)-J)}(z_\zeta+1))\frac{1}{1-4\frac{\Phi(-2\sqrt{2})}{\Phi(-1)}}\\
&\quad+\mathbb{1}\{J>\mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)}\left(\frac{1}{1-4\frac{\Phi(-2\sqrt{2})}{\Phi(-1)}}+\frac{1}{3}\right)\\
&\le\mathbb{1}\{J\le\mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)}\cdot 2^{J-\mathtt{j}_k(\zeta)}\sup_{x\ge 1}2x^2\Phi(-x)+2\mathbb{1}\{J>\mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)}
\end{aligned}$$

Let $\check{C}_0=\max\{\sup_{x\ge1}2x^2\Phi(-x),2\}$, then we have the lemma.

6.16.3. *Proof of Lemma 6.12.* Note that $\check{\mathtt{j}}_k(\zeta)=\infty,\tilde{\mathtt{j}}_k>J$ means that

$$\{i:f_k(\frac{i}{n})=\min_{l\in\{0,1,\cdots,n\}}\}\subset\{\hat{\mathtt{i}}_{k,J}-3,\hat{\mathtt{i}}_{k,J}-2,\hat{\mathtt{i}}_{k,J}-1,\hat{\mathtt{i}}_{k,J},\hat{\mathtt{i}}_{k,J}+1\},\tag{6.166}$$

and that

$$Z(f_k)\le[\frac{\hat{\mathtt{i}}_{k,J}-3}{n},\frac{\hat{\mathtt{i}}_{k,J}+1}{n}].\tag{6.167}$$

When $\mathtt{j}_k(\zeta)\le J$, then we have $2^{-\mathtt{j}_k(\zeta)}\ge 2^{-J}\ge\frac{1}{n+1}$.

(6.168)

$$\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(|\hat{Z}_k-Z(f_k)|^2\mathbb{1}\{\check{\mathtt{j}}_k(\zeta)=\infty,\tilde{\mathtt{j}}_k>J\}\right)\le\frac{16}{n^2}\\
&\le 16\left(\frac{n+1}{n}\right)^2 2^{-2\mathtt{j}_k(\zeta)}\left(1\wedge 2^{J-\mathtt{j}_k(\zeta)}\right)\le 64\cdot 2^{-2\mathtt{j}_k(\zeta)}\left(1\wedge 2^{J-\mathtt{j}_k(\zeta)}\right).
\end{aligned}$$

When $\mathtt{j}_k(\zeta)\ge J+1$, denote $i_m=\arg\min_{i:f_k(\frac{i}{n})=\min_{l\in\{0,1,\cdots,n\}}}|\frac{i}{n}-\hat{Z}_k|$, the index of the position at which f_k is minimized while being closest to the estimator. Note that this is deterministic when f_k has unique minimizer among grid points but is a random variable when f_k has two minimizers among grid points.

Then according to Lemma 6.9 we know that

$$\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(|\hat{Z}_k - Z(f_k)|^2 \mathbb{1}\{\check{\mathrm{j}}_k(\zeta) = \infty, \tilde{\mathrm{j}}_k > J\}\right) \\
&\le 2\mathbb{E}_{\mathbf{f}}\left(|\hat{Z}_k - \frac{i_m}{n}|^2\right) + 2\mathfrak{D}(f_k; n) \\
&\le 2 \times \frac{16}{n^2} \times 4\Phi(-2^{\frac{3}{2}(\mathrm{j}_k(\zeta)-J)}(z_\zeta + 1)) + 2\mathfrak{D}(f_k; n) \\
&\le 128\left(\frac{n+1}{n}\right)^2 2^{-2J}\Phi(-2^{\frac{3}{2}(\mathrm{j}_k(\zeta)-J)}) + 2\mathfrak{D}(f_k; n) \\
&\le 128\left(\frac{n+1}{n}\right)^2 2^{-2\mathrm{j}_k(\zeta)} \cdot 2^{J-\mathrm{j}_k(\zeta)} \cdot 2^3\Phi(-\sqrt{8}) \\
&< 10 \cdot 2^{-2\mathrm{j}_k(\zeta)} \cdot 2^{J-\mathrm{j}_k(\zeta)} + 2\mathfrak{D}(f_k; n)
\end{aligned} \tag{6.169}$$

Hence we concludes the proof.

6.16.4. *Proof of Lemma 6.13.* Denote

$$\Delta_{1,k} = \frac{1}{2}\rho_z((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k),$$

and

$$\Delta_{2,k} = \min\{f_k(Z(f_k) + \Delta_{1,k}), f_k(Z(f_k) - \Delta_{1,k})\} - M(f_k).$$

Then we have that

$$\begin{aligned}
&\Delta_{1,k}\Delta_{2,k}^2 \\
&\le \|f_k - \max\{f_k, M(f_k) + \rho_m((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\}\|^2 \\
&= \left((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}\right)^2.
\end{aligned} \tag{6.170}$$

Denote

$$\Delta_{3,k} = 2\rho_z((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k),$$

and

$$\Delta_{4,k} = \min\{f_k(Z(f_k) + \Delta_{1,k}), f_k(Z(f_k) - \Delta_{1,k})\} - M(f_k).$$

Clearly that

$$\Delta_{4,k} \geq \rho_m((z_\zeta+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k).$$

Then we have that

$$\begin{aligned}
&\Delta_{3,k}\Delta_{4,k}^2 \\
&\geq \|f_k - \max\{f_k, M(f_k)+\rho_m((z_\zeta+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\}\|^2 \\
&= \left((z_\zeta+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}\right)^2.
\end{aligned} \tag{6.171}$$

REMEMBER TO ADD BACK THE FOLLOWINGS

6.17. *Proof of Theorem 4.5* . Note that the axes of the hyper cube CI_α are independence from each other, so the following two propositions are sufficient to give the statement of the theorem.

Proposition 6.6. *For $CI_{k,\alpha}$ defined in (4.25)*

$$\mathbb{E}_{\mathbf{f}}\left(\mathbb{1}\{Z(f_k)\notin CI_{k,\alpha}\}\right) \leq \alpha/s, \tag{6.172}$$

for all $\mathbf{f}\in\mathcal{F}_s$

Proposition 6.7. *For $CI_{k,\alpha}$ defined in (4.25)*

$$\mathbb{E}_{\mathbf{f}}\left(|t_{k,hi}-t_{l,lo}|^2\right) \leq C_5\rho_z(z_{\alpha/s}\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)^2\left(1\wedge n\rho_z(z_{\alpha/s}\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\right)+9\mathfrak{D}(f_k;n), \tag{6.173}$$

for all $\mathbf{f}\in\mathcal{F}_s$, for an absolute positive constant C_5.

Before we continue with the proofs of the proposition, recall the quantities we defined in Equation (6.154) and (6.153).

And we further introduce the following quantities that will be used frequently

$$i_{m,l}=\min\{i: f(\frac{i}{n})=\min_{h\in\{0,1,\cdots,n\}} f(\frac{h}{n})\}, i_{m,r}=\max\{i: f(\frac{i}{n})=\min_{h\in\{0,1,\cdots,n\}} f(\frac{h}{n})\}. \tag{6.174}$$

On the event $\{\check{\mathrm{j}}_k(\alpha/2s)=\infty\}$, we define a "bad" event. Let the event that first shrinking step misses the target be

$$B_1=\{i_l\geq i_{m,l}+1\}\cup\{i_r\leq i_{m,2}-2\}. \tag{6.175}$$

We will define more "bad" events in the proofs of the propositions, usually denoted by B_h for $h = 2, 3, 4, \cdots$.

On the event $\{\check{\mathtt{j}}_k(\alpha/2s) = \infty\}$, from our definition, it is clear that $i_l \leq i_r + 1$.

6.17.1. *Proof of Proposition 6.6.* The event that $\{Z(f_k) \notin CI_{k,\alpha}\}$ can be partitioned into the followings

$$\begin{aligned}\{Z(f_k) \notin CI_{k,\alpha}\} \subset &\{\grave{\mathtt{j}}_k \leq \hat{\mathtt{j}}_k(\alpha/2s) - 1\} \\ &\cup \left(\{\grave{\mathtt{j}}_k \geq \hat{\mathtt{j}}_k(\alpha/2s), \check{\mathtt{j}}_k(\alpha/2s) = \infty\} \cap B_1\right) \\ &\cup \left(\left(\{\grave{\mathtt{j}}_k \geq \hat{\mathtt{j}}_k(\alpha/2s), \check{\mathtt{j}}_k(\alpha/2s) = \infty\} \cap B_1^c\right) \cap \{Z(f_k) \notin CI_{k,\alpha}\}\right).\end{aligned} \tag{6.176}$$

We will bound them separately.

$$\mathbb{E}_{\mathbf{f}}\left(\mathbb{1}\{\grave{\mathtt{j}}_k \leq \hat{\mathtt{j}}_k(\alpha/2s) - 1\}\right) \leq \mathbb{E}_{\mathbf{f}}\left(\left(\mathbb{1}\{\mathtt{T}_{k,\grave{\mathtt{j}}_k} \geq \tilde{\sigma}_{k,\grave{\mathtt{j}}_k}(z_{\alpha/2s})\}\middle|\nu^l_{k,\cdot}\right)\right) \leq \alpha/2s. \tag{6.177}$$

On event $\{\grave{\mathtt{j}}_k \geq \hat{\mathtt{j}}_k(\alpha/2s), \check{\mathtt{j}}_k(\alpha/2s) = \infty\}$, we know that $L_k \leq i_{m,l} \leq i_{m,r} \leq U_k$. Therefore, we have

$$\begin{aligned}&\mathbb{E}_{\mathbf{f}}\left(\{\grave{\mathtt{j}}_k \geq \hat{\mathtt{j}}_k(\alpha/2s), \check{\mathtt{j}}_k(\alpha/2s) = \infty\} \cap B_1\right) \\ &\leq P(\nu^e_{k,i_{m,l}} - \nu^e_{k,i_{m,l}+1} + \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}}\left(z^3_{k,i_{m,l}} - z^3_{k,i_{m,l}+1}\right) > 2\sqrt{3}\frac{\sigma}{(n+1)^{\frac{s-1}{2}}} z_{\alpha_1}) \\ &\quad + P(\nu^e_{k,i_{m,r}-1} - \nu^e_{k,i_{m,r}} + \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}}\left(z^3_{k,i_{m,r}-1} - z^3_{k,i_{m,r}}\right) < -\frac{2\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}} z_{\alpha_1}) \\ &\leq 2\alpha_1 \leq \alpha/4s.\end{aligned} \tag{6.178}$$

On the event $\{\grave{\mathtt{j}}_k \geq \hat{\mathtt{j}}_k(\alpha/2s), \check{\mathtt{j}}_k(\alpha/2s) = \infty\} \cap B_1^c$, we know that only when $i_l = i_r + 1 \leq n - 1$, $t_{k,hi} < \min\{\frac{i_{m,r+1}}{n}, 1\}$ could happen, and only when $i_l = i_r + 1 \geq 1$, $t_{k,lo} > \max\{\frac{i_{m,l-1}}{n}, 0\}$ could happen. And note that $i_{m,r} \leq i_l = i_r + 1 \leq i_{m,l}$ indicates that $i_{m,l} = i_{m,r}$, which we denote as i_m. So in the following we only consider f_k with unique minimizer on grids. Also we have in these cases $i_l = i_m$. We have that

$$\begin{aligned}&P_{\mathbf{f}}\left(\left(\{\grave{\mathtt{j}}_k \geq \hat{\mathtt{j}}_k(\alpha/2s), \check{\mathtt{j}}_k(\alpha/2s) = \infty\} \cap B_1^c\right) \cap \{Z(f_k) \notin CI_{k,\alpha}\}\right) \\ &\leq \mathbb{E}_{\mathbf{f}}\left(\mathbb{1}\{i_m = i_l = i_r + 1 \leq n-1, t_{k,hi} < Z(f_k)\}\right) \\ &\quad + \mathbb{E}_{\mathbf{f}}\left(\mathbb{1}\{i_m = i_l = i_r + 1 \geq 1, t_{k,lo} > Z(f_k)\}\right).\end{aligned} \tag{6.179}$$

The arguments bounding the two terms are similar, so we only show that for the first one.

Use $t_{k,r}$ to denote the intersection between the two lines

$$l_1 : y = f(\frac{i_m}{n}), l_2 : y(t) = f(\frac{i_m+1}{n}) + \frac{f(\frac{i_m+2}{n}) - f(\frac{i_m+1}{n})}{1/n}(t - \frac{i_m+1}{n}). \tag{6.180}$$

It is clear that $Z(f_k) \leq t_{k,r}$.

Basic calculation shows that

$$t_{k,r} = \frac{f_k(\frac{i_m}{n}) - f_k(\frac{i_m+1}{n})}{n(f_k(\frac{i_m+2}{n}) - f_k(\frac{i_m+1}{n}))} + \frac{i_m+1}{n}. \tag{6.181}$$

It is easy to check that the distribution of

$$\begin{aligned}&\Bigg(\nu^e_{k,i_m} - \nu^e_{k,i_m+1} - \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}}\left(z^3_{k,i_m} - z^3_{k,i_m+1} - 2\sqrt{2}z_{\alpha_2}\right),\\ &\nu^e_{k,i_m+2} - \nu^e_{k,i_m+1} - \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}}\left(z^3_{k,i_m+2} - z^3_{k,i_m+1} - 2\sqrt{2}z_{\alpha_2}\right)\Bigg)\end{aligned} \tag{6.182}$$

is the same with the following

$$\begin{aligned}&\Bigg(f_k(\frac{i_m}{n}) + \frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}}\cdot\eta_0 - f_k(\frac{i_m+1}{n}) - \frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}}\cdot\eta_1 + \frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}}\cdot 2z_{\alpha_2},\\ &f_k(\frac{i_m+2}{n}) + \frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}}\cdot\eta_2 - f_k(\frac{i_m+1}{n}) - \frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}}\cdot\eta_1 + \frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}}\cdot 2z_{\alpha_2}\Bigg),\end{aligned} \tag{6.183}$$

where $\eta_0, \eta_1, \eta_2 \overset{i.i.d}{\sim} N(0,1)$ and also independent from i_l, i_r.

Note that under the event

$$\{\eta_0 \geq -z_{\alpha_2}, \eta_1 \leq z_{\alpha_2}, \eta_2 \geq -z_{\alpha_2}\},$$

we have $t_{k,hi} \geq t_{k,r}$. Hence we have that

$$\begin{aligned}&\mathbb{E}_{\mathbf{f}}\left(\mathbb{1}\{i_m = i_l = i_r + 1 \leq n-1, t_{k,hi} < Z(f_k)\}\right)\\ \leq &P(\eta_0 < -z_{\alpha_2}) + P(\eta_1 > z_{\alpha_2}) + P(\eta_2 < -z_{\alpha_2}) \leq 3\alpha_2 = \frac{\alpha}{8s}.\end{aligned} \tag{6.184}$$

Similar arguments show that

$$\mathbb{E}_{\mathbf{f}}\left(\mathbb{1}\{i_m = i_l = i_r + 1 \geq 1, t_{k,lo} > Z(f_k)\}\right) \leq 3\alpha_2 = \frac{\alpha}{8s}.$$

Therefore we have

$$P_{\mathbf{f}}(Z(f_k) \notin CI_k) \leq \alpha/2s + 2\alpha_1 + 6\alpha_2 = \alpha/s. \tag{6.185}$$

6.17.2. *Proof of Proposition 6.7.*

$$\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(|CI_k|^2\right)\\
&\leq 169\mathbb{E}_{\mathbf{f}}\left(\frac{2^{2J-2\hat{\mathrm{j}}_k(\alpha/2s)}}{n^2}\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s)<\infty,\check{\mathrm{j}}_k(\alpha/2s)<\tilde{\mathrm{j}}_k\}\right)\\
&+169\mathbb{E}_{\mathbf{f}}\left(\frac{2^{2J-2\tilde{\mathrm{j}}_k}}{n^2}\mathbb{1}\{\tilde{\mathrm{j}}_k\leq\hat{\mathrm{j}}_k(\alpha/2s)\}\right)\\
&+\mathbb{E}_{\mathbf{f}}\left(|CI_k|^2\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s)=\infty,\tilde{\mathrm{j}}_k>J\}\right)
\end{aligned} \tag{6.186}$$

Recall Lemma 6.10, 6.11 and 6.13, we have first two terms being bounded by multiple times $\rho_z((z_{\alpha/2s}+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)$, specifically,

$$\begin{aligned}
&E_{\mathbf{f}}\left(|CI_k|^2\right)\\
&\leq \check{C}_3\rho_z((z_{\alpha/2s}+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)^2\left(1\wedge n\rho_z((z_{\alpha/2s}+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right)\\
&\quad+\mathbb{E}_{\mathbf{f}}\left(|CI_k|^2\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s)=\infty,\tilde{\mathrm{j}}_k>J\}\right),
\end{aligned} \tag{6.187}$$

where $\check{C}_3>0$ is an absolute constant.

Note that $\frac{z_{\alpha/2s}+1}{z_{\alpha/s}}<4$, and invoke Proposition **??** in Cai et al. (2022), it suffices to bound the remaining term.

We proceed to bound the remaining term. Note that

$$\begin{aligned}
&\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\leq\left(2\frac{z_{\alpha/8s}}{z_{\alpha/s}}\cdot 4\sqrt{3}\sqrt{\frac{n+1}{n}}\right)^{\frac{2}{3}}\rho_z(z_{\alpha/s}\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k),\\
&\frac{n+1}{n}\leq 2,\qquad\qquad \frac{z_{\alpha/8s}}{z_{\alpha/s}}<4 \text{ for } \alpha\leq 0.3.
\end{aligned} \tag{6.188}$$

So it is sufficient to have the following lemma for concluding the proof.

Lemma 6.14.

$$\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(|CI_k|^2\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s)=\infty,\tilde{\mathrm{j}}_k>J\}\right)\leq\\
&\check{C}_4\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)^2\left(1\wedge n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right)+9\mathfrak{D}(f_k;n)
\end{aligned} \tag{6.189}$$

where $\check{C}_4>169$ is an absolute constant.

PROOF. When

$$\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k) \geq \frac{1}{n}, \tag{6.190}$$

lemma 6.14 holds.

Now we consider the case that

$$\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k) < \frac{1}{n}. \tag{6.191}$$

Note that this means that for $i \geq i_{m,r}$,

$$\begin{aligned} f_k(\frac{i+1}{n}) - f_k(\frac{i}{n}) &\geq \frac{1}{n}\frac{\rho_m(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)}{\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)} \\ &\geq \frac{1}{\sqrt{2}}z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}}\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right)^{-\frac{3}{2}}. \end{aligned} \tag{6.192}$$

and similarly for $i \leq i_{m,l}$, we have

$$f_k(\frac{i-1}{n}) - f_k(\frac{i}{n}) \geq \frac{1}{\sqrt{2}}z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}}\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right)^{-\frac{3}{2}}. \tag{6.193}$$

Note that on the event $\{\check{\mathrm{j}}_k(\alpha/2s) = \infty, \tilde{\mathrm{j}}_k > J\}$, we have that $L_k \leq i_{m,l} \leq i_{m,r} \leq U_k$. We define a "bad" event

$$B_2 = \{i_l \leq i_{m,l} - 1\} \cup \{i_r \geq i_{m,r}\}. \tag{6.194}$$

Then we know that

$$\begin{aligned} &P_{\mathbf{f}}(B_2 \cap \{\check{\mathrm{j}}_k(\alpha/2s) = \infty, \tilde{\mathrm{j}}_k > J\}) \\ &\leq 13\Phi\left(-\sqrt{2}z_{\alpha/8s}\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right)^{-\frac{3}{2}} + z_{\alpha_1}\right). \end{aligned} \tag{6.195}$$

On the other hand, for the bad event B_1 defined in (6.175), we have

$$\begin{aligned} &P_{\mathbf{f}}(B_1 \cap \{\check{\mathrm{j}}_k(\alpha/2s) = \infty, \tilde{\mathrm{j}}_k > J\}) \\ &\leq \Phi\left(-\sqrt{2}z_{\alpha/8s}\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right)^{-\frac{3}{2}} - z_{\alpha_1}\right). \end{aligned} \tag{6.196}$$

Hence we have

(6.197)
$$
\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(|CI_k|^2\mathbb{1}\{B_1\cup B_2\}\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s)=\infty,\tilde{\mathrm{j}}_k>J\}\right)\\
&\leq \frac{169}{n^2}\times 14\Phi\left(-(\sqrt{2}-1)z_{\alpha/8s}\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right)^{-\frac{3}{2}}\right)\\
&\leq \check{C}_5\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)^2\left(1\wedge n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right),
\end{aligned}
$$

where $\check{C}_5=169\times 14\times\sup_{x>1}x^2\Phi(-(\sqrt{2}-1)x)$.

On the remaining event

$$(B_1\cup B_2)^c\cap\{\check{\mathrm{j}}_k(\alpha/2s)=\infty,\tilde{\mathrm{j}}_k>J\},$$

we have that

$$i_l=i_{m,l},i_r=i_{m,r}-1.$$

Now we have two cases. Case 1: $i_{m,l}=i_{m,r}-1$, or $i_{m,l}=i_{m,r}=1$ or $i_{m,l}=i_{m,r}=n-1$. Case 2: $i_{m,l}=i_{m,r}$ and $i_{m,l}\neq 1$ and $i_{m,l}\neq n-1$.

For the case 1 , we have $\mathfrak{D}(f_k;n)\geq\frac{1}{n^2}$, so we have

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(|CI_k|^2\mathbb{1}\{(B_1\cup B_2)^c\}\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s)=\infty,\tilde{\mathrm{j}}_k>J\}\right)\\
&\leq\frac{9}{n^2}\leq 9\mathfrak{D}(f_k;n).
\end{aligned}
\tag{6.198}
$$

Combining with Inequality (6.197), we have lemma 6.14.

For the case 2, denote $i_m=i_{m,l}=i_{m,r}$, we have

(6.199)
$$
\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(|CI_k|^2\mathbb{1}\{(B_1\cup B_2)^c\}\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s)=\infty,\tilde{\mathrm{j}}_k>J\}\right)\\
&\leq\mathbb{E}_{\mathbf{f}}\left(2(t_{k,hi}-i_m)^2\mathbb{1}\{(B_1\cup B_2)^c\}\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s)=\infty,\tilde{\mathrm{j}}_k>J,i_m\leq n-2\}\right)\\
&\quad+\mathbb{E}_{\mathbf{f}}\left(2(t_{k,lo}-i_m)^2\mathbb{1}\{(B_1\cup B_2)^c\}\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s)=\infty,\tilde{\mathrm{j}}_k>J,i_m\geq 2\}\right).
\end{aligned}
$$

The arguments for bounding the two terms are almost identical (flipping everything around i_m), we only bound the first and second share the same bound.

Recall $t_{k,r}$ defined in Equation (6.181), for simplicity of notation, denote

$$D=(B_1\cup B_2)^c\cap\{\check{\mathrm{j}}_k(\alpha/2s)=\infty,\tilde{\mathrm{j}}_k>J,i_m\leq n-2\}$$

we have

$$\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(2(t_{k,hi}-i_m)^2\mathbb{1}\{D\}\right)\\
&\le \mathbb{E}_{\mathbf{f}}\left(\left(4(t_{k,hi}-t_{k,r})_+^2+4(t_{k,r}-\frac{i_m}{n})^2\right)\mathbb{1}\{D\}\right)\\
&\le 4\mathfrak{D}(f_k;n)+4\mathbb{E}_{\mathbf{f}}\left((t_{k,hi}-t_{k,r})_+^2\mathbb{1}\{D\}\right).
\end{aligned} \tag{6.200}$$

Recall the joint distribution of the quantities in the numerator and denominator of $t_{k,hi}$ under $(B_1\cup B_2)^c\cap\{\check{\mathrm{j}}_k(\alpha/2s)=\infty,\tilde{\mathrm{j}}_k>J,i_m\le n-2\}$, as explained in Equation (6.183), denote $\varepsilon=\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}}$, when further under the event $t_{k,hi}>\frac{i_m}{n}$ (the only one we need to consider), $t_{k,hi}-t_{k,r}$ is no larger than

(6.201)
$$\begin{aligned}
&t_{k,hi}-t_{k,r}\le\\
&\frac{\varepsilon\eta_0\left(f_k(\frac{i_m+2}{n})-f_k(\frac{i_m+1}{n})\right)+\varepsilon\eta_1\left(f_k(\frac{i_m}{n})-f_k(\frac{i_m+2}{n})\right)+\varepsilon\eta_2\left(f_k(\frac{i_m+1}{n})-f_k(\frac{i_m}{n})\right)}{n\left(f_k(\frac{i_m+2}{n})-f_k(\frac{i_m+1}{n})+\varepsilon\eta_2-\varepsilon\eta_1+2\varepsilon z_{\alpha_2}\right)\left(f_k(\frac{i_m+2}{n})-f_k(\frac{i_m+1}{n})\right)}\\
&+\frac{2z_{\alpha_2}\varepsilon\left(f_k(\frac{i_m+2}{n})-f_k(\frac{i_m}{n})\right)}{n\left(f_k(\frac{i_m+2}{n})-f_k(\frac{i_m+1}{n})+\varepsilon\eta_2-\varepsilon\eta_1+2\varepsilon z_{\alpha_2}\right)\left(f_k(\frac{i_m+2}{n})-f_k(\frac{i_m+1}{n})\right)}.
\end{aligned}$$

The reason it is not an equation is due to the possibility of upper truncation if $t_{k,hi}$ by $\frac{i_m+1}{n}$

Now we consider a "good" event

(6.202)
$$A=\{\eta_1\le\frac{f_k(\frac{i_m+2}{n})-f_k(\frac{i_m+1}{n})}{6\varepsilon}+\frac{1}{2}\varepsilon z_{\alpha_2},\eta_2\ge-\frac{f_k(\frac{i_m+2}{n})-f_k(\frac{i_m+1}{n})}{6\varepsilon}-\frac{1}{2}\varepsilon z_{\alpha_2}\}.$$

Under this good event A, we have

(6.203)
$$f_k(\frac{i_m+2}{n})-f_k(\frac{i_m+1}{n})+\varepsilon\eta_2-\varepsilon\eta_1+2\varepsilon z_{\alpha_2}\ge\frac{2}{3}\left(f_k(\frac{i_m+2}{n})-f_k(\frac{i_m+1}{n})\right)+\varepsilon z_{\alpha_2}.$$

Then we have that

(6.204)
$$
\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left((t_{k,hi}-t_{k,r})_+^2\mathbb{1}\{D\cap A\}\right)\\
&\le 4\frac{1}{n^2}\left(\frac{\varepsilon}{\frac{2}{3}\left(f_k(\frac{i_m+2}{n})-f_k(\frac{i_m+1}{n})\right)+\varepsilon z_{\alpha_2}}\right)^2\left(1+4+1+16z_{\alpha_2}^2\right)\\
&\le 4\frac{1}{n^2}\left(\frac{1}{\frac{2}{3}\cdot 2z_{\alpha/8s}\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right)^{-\frac{3}{2}}+z_{\alpha/24s}}\right)^2\left(6+16z_{\alpha/24s}^2\right)\\
&\le \rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)^2\cdot\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right)\left(13.5+36\left(\frac{z_{\alpha/24s}}{z_{\alpha/8s}}\right)^2\right).
\end{aligned}
$$

The second inequality is due to Inequality (6.192).

Also note that $\frac{z_{\alpha/24s}}{z_{\alpha/8s}}<2$ for $\alpha<1$. Hence we have that

(6.205)
$$
\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left((t_{k,hi}-t_{k,r})_+^2\mathbb{1}\{D\cap A\}\right)\\
&<86\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)^2\cdot\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right).
\end{aligned}
$$

For event $A^c\cap D$, we have

$$
P(A^c\cap D)\le 2\Phi\left(-\frac{z_{\alpha/8s}}{3}\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right)^{-\frac{3}{2}}\right). \quad (6.206)
$$

Therefore we have

(6.207)
$$
\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left((t_{k,hi}-t_{k,r})_+^2\mathbb{1}\{D\cap A^c\}\right)\\
&\le 18\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)^2\cdot\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right).
\end{aligned}
$$

Adding the event A^c and A and going back to Inequality (6.200), we have the first term in (6.199) bounded. Using similar arguments, the second term can be bounded by the same bound. So we have

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(|CI_k|^2 \mathbb{1}\{(B_1 \cup B_2)^c\} \mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s) = \infty, \tilde{\mathrm{j}}_k > J\}\right) \\
&\leq 8D(f_k; n) + 832\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)^2 \cdot \left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\right).
\end{aligned} \tag{6.208}
$$

This concludes case 2, thus the proof of the lemma.

$\square$

References.

Auer, P., Ortner, R., and Szepesvári, C. (2007). Improved rates for the stochastic continuum-armed bandit problem. In Bshouty, N. H. and Gentile, C., editors, *Learning Theory*, pages 454–468, Berlin, Heidelberg. Springer Berlin Heidelberg.

Belitser, E., Ghosal, S., van Zanten, H., et al. (2012). Optimal two-stage procedures for estimating location and size of the maximum of a multivariate regression function. *The Annals of Statistics*, 40(6):2850–2876.

Blum, J. R. (1954). Multidimensional stochastic approximation methods. *The Annals of Mathematical Statistics*, 25(4):737–744.

Cai, T. T., Chen, R., and Zhu, Y. (2021). Supplement to "Estimation and Inference for Minimizer and Minimum of Convex Functions: Optimality, Adaptivity, and Uncertainty Principles".

Cai, T. T., Chen, R., and Zhu, Y. (2022). Estimation and inference for minimizer and minimum of convex functions: Adaptivity, optimality, and uncertainty principles. *The Annals of Statistics, submitted*.

Cai, T. T. and Guo, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *The Annals of Statistics*, 45(2):615–646.

Cai, T. T. and Low, M. G. (2015). A framework for estimation of convex functions. *Statistica Sinica*, pages 423–456.

Chen, H. (1988). Lower rate of convergence for locating a maximum of a function. *The Annals of Statistics*, 16(3):1330–1334.

Chen, H., Huang, M.-N. L., and Huang, W.-J. (1996). Estimation of the location of the maximum of a regression function using extreme order statistics. *Journal of Multivariate Analysis*, 57(2):191–214.

Dippon, J. (2003). Accelerated randomized stochastic optimization. *The Annals of Statistics*, 31(4):1260–1281.

Kiefer, J. and Wolfowitz, J. (1952). Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(3):462–466.

Kleinberg, R. (2004). Nearly tight bounds for the continuum-armed bandit problem. In *Proceedings of the 17th International Conference on Neural Information Processing Systems*, NIPS'04, pages 697—704, Cambridge, MA, USA. MIT Press.

Kleinberg, R., Slivkins, A., and Upfal, E. (2019). Bandits and experts in metric spaces. *J. ACM*, 66(4).

Mokkadem, A. and Pelletier, M. (2007). A companion for the Kiefer–Wolfowitz–Blum stochastic approximation algorithm. *The Annals of Statistics*, 35(4):1749–1772.

Polyak, B. T. and Tsybakov, A. B. (1990). Optimal order of accuracy of search algorithms in stochastic optimization. *Problemy Peredachi Informatsii*, 26(2):45–53.

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
USA
ran1chen@mit.edu