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

By Ran Chen

Massachusetts Institute of Technology

In this paper, we consider optimal estimation and inference for the minimizer and minimum of *multivariate additive convex functions* under suitable non-asymptotic framework that can characterize the difficulty of the problem at individual functions. We provide sharp minimax lower bounds for both the estimation accuracy and expected volume (length) of confidence hypercube (interval) for the minimizer and minimum. We provide statistically optimal and computationally efficient algorithm for these four tasks.

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. (2023a) 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.

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 $1 \leq i \leq s$, 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 (1.2), 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 minimum $M(\mathbf{f}) = \min_{\mathbf{t}\in[0,1]^s} \mathbf{f}(\mathbf{t})$ and also construct confidence hyper cube for $Z(\mathbf{f})$ and confidence interval for $M(\mathbf{f})$. Estimation and inference for the minimizer $Z(\mathbf{f})$ and minimum $M(\mathbf{f})$ under nonparametric setting will be discussed later in section 4.

1.1. *Non-asymptotic Function-specific Benchmarks.* 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. For estimation and inference of minimum and estimation of minimizer, we investigate it under local minimax framework (Cai and Low, 2015), which is also used in estimation and inference for univariate convex functions by Cai et al. (2023a). For inference of minimizer, the same two-point local minimax framework is not as appropriate and we take a non-asymptotic function-specific benchmark that measures exactly the best behavior that any method can achieve.

For estimation of the minimizer, 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

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

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

For estimation of the minimum, the hardness of the problem at an individual function $\mathbf{f}$ is naturally captured by the expected squared error. Further, under the local minimax framework, it is given by

$$R_m(\varepsilon;\mathbf{f}) = \sup_{\mathbf{g}\in\mathcal{F}_s} \inf_{\hat{M}} \max_{\mathbf{h}\in\{\mathbf{f},\mathbf{g}\}} \mathbb{E}_{\mathbf{h}}\left(\|\hat{M} - M(\mathbf{h})\|^2\right). \tag{1.4}$$

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

For estimation problems, we show that the benchmarks are valid good benchmarks in the sense that if it is significantly out performed at function $\mathbf{f} \in \mathcal{F}_s$, then a penalty need to be paid at another function $\mathbf{f}_1 \in \mathcal{F}_s$. We establish sharp minimax rates for these benchmarks and construct procedures attain the minimax rates, up to a constant factor depending on dimension s, simultaneously for all $\mathbf{f} \in \mathcal{F}_s$.

For confidence hyper cube of the minimizer with a pre-specified coverage, the hardness of the problem is naturally captured by the expected volume. Let $\mathcal{I}_{z,\alpha}(\mathcal{S})$ 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 a non-asymptotic function-specific framework, at $\mathbf{f}$, is given by the minimum expected volume at $\mathbf{f}$ for all confidence hyper cube in $\mathcal{I}_{z,\alpha}(\mathcal{F}_s)$:

$$L_{\alpha,z}(\varepsilon;\mathbf{f}) = \inf_{CI_{z,\alpha}\in\mathcal{I}_{z,\alpha}(\mathcal{F}_s)} \mathbb{E}_{\mathbf{f}}\left(V(CI_{z,\alpha})\right), \tag{1.5}$$

where $V(CI_{z,\alpha})$ is the volume of the confidence hyper cubes. Unlike local minimax framework, which measures the best a confidence hyper cube with the pre-specified probability coverage at $\mathbf{f}$ and a hardest $\mathbf{g} \in \mathcal{F}_s$ can achieve, this benchmark takes hyper cubes in $\mathcal{I}_{z,\alpha}(\mathcal{F}_s)$ (i.e. it has pre-specified probability coverage for all $\mathbf{g} \in \mathcal{F}_s$). It is easy to see that this benchmark depends on $\mathbf{f}$ and is the best that any method can achieve at $\mathbf{f}$.

For confidence interval of the minimum with a pre-specified coverage, the hardness of the problem is naturally captured by the expected length. Let $\mathcal{I}_{m,\alpha}(\mathcal{S})$ be the collection of confidence intervals for the minimum $M(\mathbf{f})$ with guaranteed coverage probability $1-\alpha$ for all $\mathbf{f} \in \mathcal{S}$. Under the local minimax framework, the benchmark is given by

$$L_{\alpha,m}(\varepsilon;\mathbf{f}) = \sup_{\mathbf{g}\in\mathcal{F}_s} \inf_{CI_{m,\alpha}\in\mathcal{I}_{m,\alpha}(\{\mathbf{f},\mathbf{g}\})} \mathbb{E}_{\mathbf{f}}\left(|CI_{m,\alpha}|\right), \tag{1.6}$$

1.2. *Projection Representation and Optimal Procedures.* Another major step in our analysis is developing data-driven and computationally efficient algorithms for the construction of estimators and confidence interval (hyper cube) as well as establishing the optimality of these procedures at each $f \in \mathcal{F}$.

An interesting observation is that $Y_{\mathbf{f}}$ 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. Also $Y_{\mathbf{f}}$ can be fully recovered from $\mathfrak{P}(Y_{\mathbf{f}})$. The estimators and confidence interval (hyper cube) are constructed based on this observation by doing estimation and inference on each component and carefully join them together.

The key idea behind the construction for each component of the optimal 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. The final estimation/inference is to carefully choose the control parameter of the component-wise stopping rule and put together the output for each axis.

The resulting estimators, $\hat{Z}$ for $Z(\mathbf{f})$ and $\hat{M}$ for $M(\mathbf{f})$, are shown to attain within a dimension-dependent constant of the benchmarks $R_z(\varepsilon; \mathbf{f})$ $R_m(\varepsilon; \mathbf{f})$ simultaneously for all $\mathbf{f} \in \mathcal{F}_s$,

$$\mathbb{E}_{\mathbf{f}} \left(\|\hat{Z} - Z(\mathbf{f})\|^2 \right) \leq C_{z,s} R_z(\varepsilon; \mathbf{f}), \tag{1.7}$$

$$\mathbb{E}_{\mathbf{f}} \left(\|\hat{M} - M(\mathbf{f})\|^2 \right) \leq C_{m,s} R_m(\varepsilon; \mathbf{f}), \tag{1.8}$$

for constants $C_{z,s}$ and $C_{m,s}$ depending on dimension s only.

The resulting confidence interval (hyper cube), $CI_{z,\alpha}$ for $Z(\mathbf{f})$ and $CI_{m,\alpha}$ for $M(\mathbf{f})$, are shown to have the pre-specified coverage $(1 - \alpha)$ while having expected length (volume) being adaptive to $\mathbf{f}$ and attaining within a coverage-dimension-dependent constant of the benchmarks $L_{\alpha,z}(\varepsilon; \mathbf{f}), L_{\alpha,m}(\varepsilon; \mathbf{f})$ for all $\mathbf{f} \in \mathcal{F}_s$. That is,

$$\mathbb{E}_{\mathbf{f}} \left(V(CI_{z,\alpha}) \right) \leq C_{z,s,\alpha} L_{\alpha,z}(\varepsilon; \mathbf{f}), \tag{1.9}$$

$$\mathbb{E}_{\mathbf{f}} \left(|CI_{m,\alpha}| \right) \leq C_{m,s,\alpha} L_{\alpha,m}(\varepsilon; \mathbf{f}), \tag{1.10}$$

where $C_{z,s,\alpha}$ and $C_{m,s,\alpha}$ are constants depending on dimension s and α only.

1.3. *Organization of the Paper.* In Section 2, we analyze local minimax risks, relating them to appropriate local modulus of continuity, in turn providing rate-sharp upper and lower bounds. We also provide lower bound for the benchmark for inference of the minimizer in Section 2. In Section 3, we introduce projection representation of the observation, provide computationally efficient adaptive procedures and show their optimality. In Section 4, we consider the nonparametric regression model. We introduce the corresponding benchmarks, propose adaptive procedures and establish the optimality. Proofs are given in appendix Section 6.

1.4. *Notation.* We conclude this section with some notation that will be used in the section. 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 $\|\cdot\|$ to denote the L_2 norm for vectors, univariate functions and multivariate functions, depending on the setting. We use $\mathbb{1}\{A\}$ to denote 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 $\mathbf{g}$, $\mathbf{h}$. 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, and $M(\cdot)$ to denote the minimum operator, for both $\mathbf{f} \in \mathcal{F}_s$ and $f \in \mathcal{F}$. Note that we use $\mathcal{I}_{z,\alpha}(\mathcal{S})$ 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 $\mathcal{S} \subset \mathcal{F}$ and the hyper cube becomes interval.

We use $\mathcal{I}_{m,\alpha}(\mathcal{S})$ to denote the collection of confidence intervals for the minimum with guaranteed coverage probability $1-\alpha$ for all functions in $\mathcal{S}$. This can be generalized into univariate case when $\mathcal{S} \subset \mathcal{F}$.

2. Local Minimax Rates and Lower Bounds.

In this section, we discuss the local minimax rates and the lower bound for inference of the minimizer. We introduce the local moduli of continuity and use it to characterize the benchmarks for estimation of minimizer and estimation and inference of minimum introduced in Section 1.1. We provide rate-sharp bounds for the continuity moduli based on geometry properties of the functions. As we use a different benchmark for inference of minimizer, we provide lower bound of it in this section.

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

$$\omega_z(\varepsilon;\mathbf{f}) = \sup\{\|Z(\mathbf{f})-Z(\mathbf{g})\|^2 : \|\mathbf{f}-\mathbf{g}\|_2 \le \varepsilon, \mathbf{g}\in\mathcal{F}_s\} \tag{2.1}$$
$$\omega_m(\varepsilon;\mathbf{f}) = \sup\{\|M(\mathbf{f})-M(\mathbf{g})\|^2 : \|\mathbf{f}-\mathbf{g}\|_2 \le \varepsilon, \mathbf{f}\in\mathcal{F}_s\}, \tag{2.2}$$
$$\tilde{\omega}_m(\varepsilon;\mathbf{f}) = \sup\{\|M(\mathbf{f})-M(\mathbf{g})\| : \|\mathbf{f}-\mathbf{g}\|_2 \le \varepsilon, \mathbf{f}\in\mathcal{F}_s\}. \tag{2.3}$$

As in the case of linear functionals or in the case of minimizer and minimum operators for univariate convex functions, the local moduli $\omega_z(\varepsilon;\mathbf{f})$, $\omega_m(\varepsilon;\mathbf{f})$, $\tilde{\omega}_m(\varepsilon;\mathbf{f})$ 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 moduli of continuity.

THEOREM 2.1 (Sharp Lower Bounds). *Let $R_z(\varepsilon;\mathbf{f})$ be defined in (1.3), $R_m(\varepsilon;\mathbf{f})$ be defined in (1.4), and $L_{\alpha,m}(\varepsilon;\mathbf{f})$ be defined in (1.6). Let $0<\alpha\le 0.1$. Then*

$$a\omega_z(\varepsilon;\mathbf{f}) \le R_z(\varepsilon;\mathbf{f}) \le A\omega_z(\varepsilon;\mathbf{f}), \tag{2.4}$$
$$a\omega_m(\varepsilon;\mathbf{f}) \le R_m(\varepsilon;\mathbf{f}) \le A\omega_m(\varepsilon;\mathbf{f}) \tag{2.5}$$
$$b_\alpha\tilde{\omega}_m(\varepsilon;\mathbf{f}) \le L_{\alpha,m}(\varepsilon;\mathbf{f}) \le B_\alpha\tilde{\omega}_m(\varepsilon;\mathbf{f}) \tag{2.6}$$

where the constants a, A, b_α, B_α can be taken as $a=0.1$, $A=3.1$, $b_\alpha = 0.6-\alpha$, and $B_\alpha = 2(1-2\alpha)z_\alpha$.

Theorem 2.1 shows that the benchmarks can be characterized in terms of continuity moduli of continuity. However, this continuity moduli is hard to compute. Now we related it to geometric quantities of $\mathbf{f}$. We first introduce two geometric quantities for univariate convex function $f\in\mathcal{F}$, which are also used by Cai et al. (2023a). For $f\in\mathcal{F}$, $u\in\mathbb{R}$ and $\varepsilon>0$, let $f_u(t) = \max\{f(t),u\}$, $M(f)=\min_{x\in[0,1]} f(x)$, and define

$$\rho_m(\varepsilon;f) = \sup\{u-\min\{f(x): x\in[0,1]\} : \|f-f_u\|\le\varepsilon\}, \tag{2.7}$$
$$\rho_z(\varepsilon;f) = \sup\{|t-Z(f)| : f(t)\le \rho_m(\varepsilon;f)+M(f), t\in[0,1]\}. \tag{2.8}$$

With the geometric quantity $\rho_z(\varepsilon;f)$ for univariate convex function $f\in\mathcal{F}$, we can establish a rate-sharp bound of modulus of continuity for the minimizer for multivariate additive convex function $\mathbf{f}\in\mathcal{F}_s$.

THEOREM 2.2 (Geometry Representation for Modulus of Continuity for Minimzer). *Let $\rho_z(\varepsilon;f)$ be defined in (2.8) for $f\in\mathcal{F}$, and let $\mathbf{f}\in\mathcal{F}_s$. Let*

$\omega_z(\varepsilon; \mathbf{f})$ *be defined in* (2.1)*. Then*

$$\frac{1}{3} s^{-\frac{2}{3}} \sum_{i=1}^{s} \rho_z(\varepsilon; f_i)^2 \leq \omega_z(\varepsilon; \mathbf{f}) \leq \sum_{i=1}^{s} 9\rho_z(\varepsilon; f_i)^2. \tag{2.9}$$

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

$$\omega_z(\varepsilon; \mathbf{f}) \leq 9s^{-\frac{2}{3}} \sum_{i=1}^{s} \rho_z(\varepsilon; f_i)^2. \tag{2.10}$$

And for any $\beta \leq s$*, 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*

$$\omega_z(\varepsilon; \mathbf{f}) \geq \rho_z(\varepsilon; f_i)^2 - \delta_0. \tag{2.11}$$

Theorem 2.2 shows that the modulus of continuity for minimizer varies within an absolute constant multiple times of

$$s^{-\frac{2}{3}} \sum_{i=1}^{s} \rho_z(\varepsilon; f_i)^2 \text{ and } \sum_{i=1}^{s} \rho_z(\varepsilon; f_i)^2,$$

with the order of both upper and lower bound attainable for some $\mathbf{f} \in \mathcal{F}_s$.

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

Theorem 2.3 (Geometry Representation for Modulus of Continuity for Minimum). *Let* $\rho_z(\varepsilon; \mathbf{f})$ *be defined in* (2.8) *and* $\rho_m(\varepsilon; \mathbf{f})$ *be defined in* (2.7) *for* $f \in \mathcal{F}$*. Let* $\omega_m(\varepsilon; \mathbf{f})$ *be defined in* (2.2) *and* $\tilde{\omega}_m(\varepsilon; \mathbf{f})$ *be defined in* (2.3) *for* $\mathbf{f} \in \mathcal{F}_s$*. Then*

$$\frac{1}{1 + \sum_{i=1}^{s}(1 \wedge 2\rho_z(\varepsilon; f_i))} \sum_{i=1}^{s} \rho_m(\varepsilon; f_i)^2 \leq \omega_m(\varepsilon; \mathbf{f}) \leq 9(1 + \frac{1}{s}) \sum_{i=1}^{s} \rho_m(\varepsilon; f_i)^2, \tag{2.12}$$

$$\sqrt{\frac{1}{1 + \sum_{i=1}^{s}(1 \wedge 2\rho_z(\varepsilon; f_i))} \sum_{i=1}^{s} \rho_m(\varepsilon; f_i)^2} \leq \tilde{\omega}_m(\varepsilon; \mathbf{f}) \leq \sqrt{9(1 + \frac{1}{s}) \sum_{i=1}^{s} \rho_m(\varepsilon; f_i)^2}. \tag{2.13}$$

Theorem 2.3 shows that the modulus of continuity for minimum $\omega_m(\varepsilon; \mathbf{f})$ is of the order $\sum_{k=1}^{s} \rho_m(\varepsilon; f_k)^2$ and $\tilde{\omega}_m(\varepsilon; \mathbf{f})$ is of the order $\sqrt{\sum_{k=1}^{s} \rho_m(\varepsilon; f_k)^2}$.

Now we have done establishing the local minimax rates for three tasks, we turn to establishing the lower bound for the benchmark of inference of the minimizer.

THEOREM 2.4 (Lower Bound for Expected Volume of Confidence Hyper Cube for Minimizer). *Let $L_{\alpha,z}(\varepsilon;\mathbf{f})$ be defined in (1.5) for $\mathbf{f}\in\mathcal{F}_s$ and $\rho_z(\varepsilon;f)$ be defined in (2.8) for $f\in\mathcal{F}$. Then we have*

$$L_{\alpha,z}(\varepsilon;\mathbf{f})\geq C_{\alpha,s}\Pi_{i=1}^{s}\rho_z(\varepsilon;f_i), \tag{2.14}$$

where $C_{\alpha,s}$ is a positive constant depending on α and s.

2.2. *Penalty for Super-efficiency.* We have shown that the estimation benchmarks $R_z(\varepsilon;\mathbf{f})$ and $R_m(\varepsilon;\mathbf{f})$ can be characterized by intrinsic geometric quantities of $\mathbf{f}$. Now we show that these benchmarks can not be essentially uniformly out performed. That is, if the benchmark is significantly out performed at function $\mathbf{f}\in\mathcal{F}_s$, then it needs to pay a penalty at another function $\mathbf{f}_1\in\mathcal{F}_s$. These benchmarks, similar to that in the univariate case, play a role analogous to the information lower bound in the classic statistic.

THEOREM 2.5 (Penalty for Supper-Efficiency). *For any estimator of the minimizer $\hat{Z}$, if $\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z}-Z(\mathbf{f})\|^2\right)\leq\gamma R_z(\varepsilon;\mathbf{f})$ for $\mathbf{f}\in\mathcal{F}_s$ and $\gamma<\gamma_0$, where γ_0 is a positive constant, then there exists $\mathbf{f}_1\in\mathcal{F}_s$ such that*

$$\mathbb{E}_{\mathbf{f}_1}\left(\|\hat{Z}-Z(\mathbf{f}_1)\|^2\right)\geq c_{z,s}(\log\frac{1}{\gamma})^{\frac{2}{3}}R_z(\varepsilon;\mathbf{f}_1), \tag{2.15}$$

where $c_{z,s}$ is a constant depending on s only.

Similarly, for any estimator of the minimum $\hat{M}$, if $\mathbb{E}_{\mathbf{f}}(|\hat{M}-M(\mathbf{f})|^2)\leq\gamma R_m(\varepsilon;\mathbf{f})$ for $\mathbf{f}\in\mathcal{F}_s$ and $\gamma<\gamma_0/s$, where γ_0 is a positive constant, then there exists $\mathbf{f}_1\in\mathcal{F}_s$ such that

$$\mathbb{E}_{\mathbf{f}_1}\left(|\hat{M}-M(\mathbf{f}_1)|^2\right)\geq c_{m,s}(\log\frac{1}{\gamma})^{\frac{2}{3}}R_m(\varepsilon;\mathbf{f}_1), \tag{2.16}$$

where $c_{m,s}$ is a constant depending on s only.

3. Projection Representation and Adaptive Optimal Procedures.. We now turn to the construction of data-driven and computationally efficient algorithms for estimation and inference of minimizer and minimum for white noise model. Our construction is based on an information-preserving representation of the observation $Y_{\mathbf{f}}$, which we call *Projection Representation.* We show that our procedures achieve, up to a universal constant depending on dimension s and confidence level $1-\alpha$, the corresponding benchmarks $R_z(\varepsilon;\mathbf{f})$, $R_m(\varepsilon;\mathbf{f})$, $L_{\alpha,z}(\varepsilon;\mathbf{f})$, $L_{\alpha,m}(\varepsilon;\mathbf{f})$, simultaneously for all $\mathbf{f}\in\mathcal{F}_s$.

3.1. *Projection Representation..* The construction of the procedures is based on an interesting property of the observation $Y_{\mathbf{f}}$ (or Y) that Y 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 projection of the original stochastic process on each coordinate, 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 $1 \leq i \leq s$, the $i-$th projection of Y, $\boldsymbol{\pi}_i(Y)$, is a univariate stochastic process that satisfies for $0 \leq a_i < A_i \leq 1$,*

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

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 component function 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. Instead, it carries part of the information of $M(\mathbf{f})$. Note that the minimizer $Z(\mathbf{f})$ can be written as $Z(\mathbf{f}) =$

$(Z(f_1), Z(f_2), \dots, Z(f_s))$, so its i-th element only depends on f_i. Similarly $M(\mathbf{f})$ can be written as $M(\mathbf{f}) = f_0 + \sum_{k=1}^{s} M(f_k)$, so each component in $\mathfrak{P}(Y_{\mathbf{f}})$ serves as a sufficient statistics for each of the adding components of $M(\mathbf{f})$. The information preserving representation $\mathfrak{P}(\cdot)$ plays the role of separating the relevant information of s coordinates into independent random variables.

3.2. *Adaptive Procedures..* Now we are ready to introduce the construction of data-driven and computationally efficient algorithms for estimation and confidence interval (hyper cube) for the minimum $M(\mathbf{f})$ and the minimizer $Z(\mathbf{f})$ under the white noise model in this section. The procedures constructed in this section are shown in Section 3.3 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 depending on dimension s and confidence level $1-\alpha$, the corresponding benchmarks, simultaneously for all $\mathbf{f} \in \mathcal{F}_s$.

Similar to the construction in Cai et al. (2023a), we have three blocks: localization, stopping, and estimation/inference. But since $\boldsymbol{\pi}_i(Y)$ has different distribution with that in the univariate case, and we also need to account for the dimension, our procedures are carefully tailored to accommodate for the new challenges.

3.2.1. *Sample Splitting.* For technical reasons, we split the first s coordinates of the projection representation (i.e. $\mathfrak{P}(Y)$), $V = (\boldsymbol{\pi}_1(Y), \boldsymbol{\pi}_2(Y), \dots, \boldsymbol{\pi}_s(Y))$, into three independent pieces to ensure independence of the data used in the three steps.

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

(3.4)
$$\mathbf{v}_i^l(t) = \boldsymbol{\pi}_i(Y)(t) + \frac{\sqrt{2}}{2}\varepsilon\left(B_i^1(t) - t\int_0^1 B_i^1(x)\mathrm{d}x\right) + \frac{\sqrt{6}}{2}\varepsilon\left(B_i^2(t) - t\int_0^1 B_i^2(x)\mathrm{d}x\right),$$
$$\mathbf{v}_i^r(t) = \boldsymbol{\pi}_i(Y)(t) + \frac{\sqrt{2}}{2}\varepsilon\left(B_i^1(t) - t\int_0^1 B_i^1(x)\mathrm{d}x\right) - \frac{\sqrt{6}}{2}\varepsilon\left(B_i^2(t) - t\int_0^1 B_i^2(x)\mathrm{d}x\right),$$
$$\mathbf{v}_i^e(t) = \boldsymbol{\pi}_i(Y)(t) - \sqrt{2}\varepsilon\left(B_i^1(t) - t\int_0^1 B_i^1(x)\mathrm{d}x\right).$$

Then the concatenate vector of vectors V_l, V_r, V_e has all of its 3s elements being independent, and for each axis $i \in \{1, 2 \dots, s\}$, $\mathbf{v}_i^l(t), \mathbf{v}_i^r(t), \mathbf{v}_i^e(t)$ can be written as

$$\begin{aligned} \mathrm{d}\mathbf{v}_i^l(t) &= f_i(t)\mathrm{d}t + \sqrt{3}\varepsilon \mathrm{d}\tilde{W}_i^l, \\ \mathrm{d}\mathbf{v}_i^r(t) &= f_i(t)\mathrm{d}t + \sqrt{3}\varepsilon \mathrm{d}\tilde{W}_i^r, \\ \mathrm{d}\mathbf{v}_i^e(t) &= f_i(t)\mathrm{d}t + \sqrt{3}\varepsilon \mathrm{d}\tilde{W}_i^e, \end{aligned} \tag{3.5}$$

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

3.2.2. *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 that in Cai et al. (2023a) 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

$$m_j = 2^{-j},\ t_{j,i} = i \cdot m_j, \text{ and } i_{j,k}^* = \max\{i : Z(f_k) \in [t_{j,i-1}, t_{j,i}]\}. \tag{3.6}$$

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.

3.2.3. *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 univariate case, we construct a series of stopping rules based on a user select parameter $\zeta > 0$, 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 $\zeta > 0$ 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

(3.7)
$$\frac{\tilde{X}_{j,\hat{i}_{j,k}+6,k} - \tilde{X}_{j,\hat{i}_{j,k}+5,k}}{\sigma_j}\Bigg|\hat{i}_{j,k} \sim N\left(\frac{m_j\sqrt{m_j}}{\sqrt{6}\varepsilon} \times \frac{1}{m_j}\int_{t_{j,\hat{i}_j+5,k}}^{t_{j,\hat{i}_j+6,k}} \frac{f_k(t+m_j) - f_k(t)}{m_j}\mathrm{d}t, 1\right).$$

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 either 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

$$\hat{j}(\zeta,k) = \min\{j : \frac{T_{j,k}}{\sigma_j} \le z_\zeta\}. \tag{3.8}$$

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

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

3.2.4. *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 $\zeta_k > 0$, we use them to construct estimator and confidence interval (hyper cube) for the minimum $M(\mathbf{f})$ and the minimizer $Z(\mathbf{f})$.

For estimation of the minimizer, 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:

$$\hat{Z}_k = \frac{t_{\hat{j}(\zeta,k),\hat{i}_{\hat{j}(\zeta,k),k}-1} + t_{\hat{j}(\zeta,k),\hat{i}_{\hat{j}(\zeta,k),k}}}{2}. \tag{3.9}$$

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

$$\hat{Z} = (\hat{Z}_1, \hat{Z}_2, \ldots, \hat{Z}_s), \tag{3.10}$$

with $\hat{Z}_k$ defined in (3.9).

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

$$CI_k = \left[2^{-\hat{j}(\zeta,k)+1}\left(\hat{i}_{\hat{j}(\zeta,k)-1,k} - 7\right), 2^{-\hat{j}(\zeta,k)+1}\left(\hat{i}_{\hat{j}(\zeta,k)-1,k} + 6\right)\right] \cap [0, 1]. \tag{3.11}$$

The confidence cube CI for the minimizer is give by

$$CI_{z,\alpha} = CI_1 \times CI_2 \times \cdots \times CI_s, \tag{3.12}$$

where CI_k is defined in (3.11).

For estimation and inference of the minimum, let

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

for $1 \le i \le 2^j$, and $+\infty$ for $i \notin \{1, 2, \cdots, 2^j\}$.

For estimation of the minimum $M(\mathbf{f})$, let $\zeta_k = \zeta = \Phi(-2)$ for $k = 1, 2 \cdots, s$. Let the *final index* for estimator construction for k-th coordinate be

$$\begin{aligned} i_{F,k} = \hat{i}_{\hat{j}(\zeta,k)-1,k} + 2\Bigg(& \mathbb{1}\{\tilde{X}_{\hat{j}(\zeta,k),\hat{i}_{\hat{j}(\zeta,k)}+6,k} - \tilde{X}_{\hat{j}(\zeta,k),\hat{i}_{\hat{j}(\zeta,k)}+5,k} \le 2\sigma_{\hat{j}(\zeta,k)}\} \\ & - \mathbb{1}\{\tilde{X}_{\hat{j}(\zeta,k),\hat{i}_{\hat{j}(\zeta,k)}-6,k} - \tilde{X}_{\hat{j}(\zeta,k),\hat{i}_{\hat{j}(\zeta,k)}-5,k} \le 2\sigma_{\hat{j}(\zeta,k)}\}\Bigg). \end{aligned} \tag{3.13}$$

The estimator of the minimum is given by

$$\hat{M} = Y(1,1,\cdots,1) - Y(0,0,\cdots,0) + \sum_{k=1}^{s} 2^{\hat{j}(\zeta,k)} \bar{X}_{\hat{j}(\zeta,k),i_{F,k},k}. \tag{3.14}$$

For inference of the minimum, let $\zeta_k = \zeta = \alpha/4s$ for $k = 1, 2\cdots, s$. Define an intermediate estimator of the minimum by

$$\hat{\mathbf{f}}_1 = Y(1,1,\cdots,1) - Y(0,0,\cdots,0) + \sum_{k=1}^{s} 2^{\hat{j}(\zeta,k)+3} \min_{16(\hat{i}_{\hat{j}(\zeta,k),k-1}-7)<i\leq 16(\hat{i}_{\hat{j}(\zeta,k),k-1}+6)} \bar{X}_{\hat{j}(\zeta,k)+2,i,k}. \tag{3.15}$$

Let U_n be the cumulative distribution function of $\tilde{u} = \max\{u_1,\cdots,u_n\}$, where

$$u_1,\cdots,u_n \overset{i.i.d}{\sim} N(0,1),$$

and define

$$S_{n,\beta} = U_n^{-1}(1-\beta). \tag{3.16}$$

In other words, $S_{n,\beta}$ is the $(1-\beta)$ quantile of the distribution of the maximum of n *i.i.d.* standard normal variables.

Let

$$\begin{aligned} \mathbf{f}_{hi} &= \hat{\mathbf{f}}_1 + S_{208,\alpha/8s} \times \sqrt{3}\varepsilon \sum_{k=1}^{s} 2^{\frac{\hat{j}(\zeta,k)+3}{2}} + z_{\alpha/8}\sqrt{3}\varepsilon s \\ \mathbf{f}_{lo} &= \hat{\mathbf{f}}_1 - z_{\alpha/4}\sqrt{3}\varepsilon \sqrt{1 + \sum_{k=1}^{s} 2^{\hat{j}(\zeta,k)+3}} - \sum_{k=1}^{s} z_{\alpha/4s}\sqrt{3}\cdot 2\varepsilon \cdot 2^{\frac{\hat{j}(\zeta,k)+3}{2}}. \end{aligned} \tag{3.17}$$

Then the $(1-\alpha)$ level confidence interval for $M(\mathbf{f})$ is

$$CI_{m,\alpha} = [\mathbf{f}_{lo}, \mathbf{f}_{hi}]. \tag{3.18}$$

3.3. *Statistical Optimality..* In this section, we establish the optimality of the adaptive procedures constructed in Section 3.2. The results show that the date driven estimators and the confidence interval (hyper cube) achieve within a universal constant factor depending on s and α only of their corresponding benchmarks 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 of the minimizer.

THEOREM 3.1 (Estimation for Miminizer). *The estimator $\hat{Z}$ defined by (3.10) satisfies*

$$\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z}-Z(\mathbf{f})\|^2\right) \le C_{z,s} R_z(\varepsilon;\mathbf{f}),\ \textit{for all } \mathbf{f}\in\mathcal{F}_s, \tag{3.19}$$

where $C_{z,s} > 0$ is a constant depending on dimension s.

The following holds for the confidence hyper cube $CI_{z,\alpha}$.

THEOREM 3.2 (Confidence Hyper-cube for Minimizer). *For $0 < \alpha \le 0.3$, the confidence hyper cube $CI_{z,\alpha}$ defined by (3.12) is a $1-\alpha$ level confidence hyper cube for the minimizer $Z(\mathbf{f})$. Its expected volume satisfies*

$$\mathbb{E}_{\mathbf{f}}\left(V(CI)\right) \le C_{z,s,\alpha} L_{\alpha,z}(\varepsilon;\mathbf{f}),$$

where $C_{z,s,\alpha}$ is a positive constant depending on s and α.

THEOREM 3.3 (Estimation for Minimum). *The estimation $\hat{M}$ defined in (3.14) satisfies*

$$E\left((\hat{M}-M(\mathbf{f}))^2\right) \le C_{m,s} R_m(\varepsilon;\mathbf{f}), \tag{3.20}$$

where $C_{m,s}$ is a positive constant depending on dimension s.

THEOREM 3.4 (Confidence Interval for Minimum). *For $0 < \alpha \le 0.3$, the confidence interval defined by (3.18) is a $1-\alpha$ level confidence interval for the minimum $M(\mathbf{f})$ satisfying*

$$\mathbb{E}\left(|CI_{m,\alpha}|\right) \le C_{m,s,\alpha} L_{\alpha,m}(\varepsilon;\mathbf{f}), \tag{3.21}$$

where $C_{m,s,\alpha}$ is a positive constant depending on α and s.

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 estimators and confidence interval (hyper cube) for the minimizer $Z(\mathbf{f})$ and the minimum $M(\mathbf{f})$, for $\mathbf{f}\in\mathcal{F}_s$. For simplicity of notation, we take $\mathbf{i} = (i_1, i_2, \ldots, i_s)$. To avoid trivial case, we suppose $n \ge 2$.

4.1. *Local Minimax Rates, Discretization Error and Separable Representation.* Analogous to the benchmarks for the white noise model defined in Equations (1.3), (1.4), (1.6), we define similar benchmarks for the nonparametric regression model (4.1) with $n+1$ equally spaced observations. Denote by $\mathcal{I}_{m,\alpha,n}(\mathfrak{F})$ the collection of $(1-\alpha)$ level confidence intervals for $M(f)$ on a function class $\mathfrak{F}$ under the regression model (4.1) and let

$$\begin{aligned}
\tilde{\mathtt{R}}_{z,n}(\sigma;\mathbf{f}) &= \sup_{\mathbf{g}\in\mathcal{F}_s} \inf_{\hat{Z}} \max_{h\in\{\mathbf{f},\mathbf{g}\}} \mathbb{E}_h\|\hat{Z}-Z(h)\|^2, \\
\tilde{\mathtt{R}}_{m,n}(\sigma;\mathbf{f}) &= \sup_{\mathbf{g}\in\mathcal{F}_s} \inf_{\hat{M}} \max_{h\in\{\mathbf{f},\mathbf{g}\}} \mathbb{E}_h(\hat{M}-M(h))^2, \\
\tilde{\mathtt{L}}_{m,\alpha,n}(\sigma;\mathbf{f}) &= \sup_{\mathbf{g}\in\mathcal{F}_s} \inf_{CI_{m,\alpha}\in\mathcal{I}_{m,\alpha,n}(\{\mathbf{f},\mathbf{g}\})} \mathbb{E}_{\mathbf{f}}|CI_{m,\alpha}|.
\end{aligned} \tag{4.2}$$

For confidence hyper cube for minimizer, denote $\mathcal{I}_{z,\alpha,n}(\mathfrak{F})$ the collection of $(1-\alpha)$ level confidence hyper cube on a function class $\mathfrak{F}$ under the regression model (4.1) and let

$$\tilde{\mathtt{L}}_{z,\alpha,n}(\sigma;\mathbf{f}) = \inf_{CI\in\mathcal{I}_{m,\alpha,n}(\mathcal{F}_s)} \mathbb{E}_{\mathbf{f}} V(CI). \tag{4.3}$$

It is clear that the expected volume for confidence hyper cube of the minimizer can not be smaller than $\tilde{\mathtt{L}}_{z,\alpha,n}(\sigma;\mathbf{f})$, which is also function-specific, i.e. depending on $\mathbf{f}$.

Compared with white noise model, in addition to the difference in the probability structure caused by discrete observations, estimation and inference for both $Z(\mathbf{f})$ and $M(\mathbf{f})$ incur additional discretization errors, even in the noiseless case. See the appendix Section 6.12 for further discussion.

4.1.1. *Separable Representation.* 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 $k \in \{1,2,\dots,s\}$, the k-th projection of $\{y_{\mathbf{i}}\}$, $\boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})$, is an $n+1$-long random vector,*

$$\boldsymbol{\pi}_k(\{y_{\mathbf{i}}\}) = \left(\frac{\sum_{\mathbf{i}:i_k=1} y_{\mathbf{i}}}{(n+1)^{s-1}} - \frac{\sum_{\mathbf{i}} y_{\mathbf{i}}}{(n+1)^s}, \frac{\sum_{\mathbf{i}:i_k=2} y_{\mathbf{i}}}{(n+1)^{s-1}} - \frac{\sum_{\mathbf{i}} y_{\mathbf{i}}}{(n+1)^s}, \dots, \frac{\sum_{\mathbf{i}:i_k=s} y_{\mathbf{i}}}{(n+1)^{s-1}} - \frac{\sum_{\mathbf{i}} y_{\mathbf{i}}}{(n+1)^s}\right). \tag{4.4}$$

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

$$\mathbf{er}(\{y_{\mathbf{i}}\})_{i_1,i_2,\cdots,i_s} = y_{i_1,i_2,\cdots,i_s} - \sum_{k=1}^{s} \boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})_{i_k}, \tag{4.5}$$

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

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

$$\mathfrak{P}(\{y_{\mathbf{i}}\}) = (\boldsymbol{\pi}_1(\{y_{\mathbf{i}}\}), \boldsymbol{\pi}_2(\{y_{\mathbf{i}}\}), \boldsymbol{\pi}_s(\{y_{\mathbf{i}}\}), \mathbf{er}(\{y_{\mathbf{i}}\})). \tag{4.6}$$

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* (4.6)*. 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_k*.*

4.2. *Optimal Procedures.* Similar to the white noise model, we split the data into three independent copies and then construct the estimators and confidence interval (hyper cube) for $Z(\mathbf{f})$ and $M(\mathbf{f})$ 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 $\boldsymbol{\pi}_k(\{y_{\mathbf{i}}\})$:

(4.7)

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

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

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

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 sequences for each axis k are independent, and the s collections of the three sequences 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 $j = 0, 1, \cdots, J$, $i = 1, 2, \cdots, \lfloor \frac{n+1}{2^{J-j}} \rfloor$, 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

$$\mathtt{Y}^u_{k,j,i} = \sum_{h=(i-1)2^{J-j}}^{i \cdot 2^{J-j}-1} \nu^u_{k,h}. \tag{4.8}$$

Again, let $\mathtt{Y}^u_{k,j,i} = +\infty$ when $i \notin \{1, 2, \cdots, \lfloor \frac{n+1}{2^{J-j}} \rfloor\}$ 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 $\{\nu^l_{k,h}, h \in \{0, 1, \cdots, n\}\}$ to construct a localization procedure. Let $\hat{\mathtt{i}}_{k,0} = 1$, and for $j = 1, 2, \cdots, J$, let

$$\hat{\mathtt{i}}_{k,j} = \underset{\max\{2\hat{\mathtt{i}}_{k,j-1}-2,1\} \le i \le \min\{2\hat{\mathtt{i}}_{k,j-1}+1, \lfloor \frac{n+1}{2^{J-j}} \rfloor\}}{\arg\min} \mathtt{Y}^l_{k,j,i}. \tag{4.9}$$

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. $\mathtt{Y}^l_{k,j,i}$) 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, for axis k, define the statistic $\mathtt{T}_{k,j}$ based on the sequence $Y^r_{k,\cdot,\cdot}$ 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 $\mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}+6} - \mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}+5} < \infty$,

$$\mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}+6} - \mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}+5} \Big| \hat{\mathtt{i}}_{k,j} \sim N\left(\sum_{h=(\hat{\mathtt{i}}_{k,j}+4)2^{J-j}}^{(\hat{\mathtt{i}}_{k,j}+5)2^{J-j}-1} \left(f_k(\frac{h+2^{J-j}}{n}) - f_k(\frac{h}{n}) \right), \tilde{\sigma}^2_{k,j} \right). \tag{4.10}$$

Similar to white noise model, we define a series of stopping rules controlled by a parameter $\zeta > 0$.

Define a *stopping step precursor* $\check{\mathtt{j}}_k(\zeta)$ as

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

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

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

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 $\mathbf{f} \in \mathcal{F}_s$, as well as estimator and confidence interval for the minimum of $\mathbf{f} \in \mathcal{F}_s$.

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

$$\hat{Z}_k = \begin{cases} -\frac{1}{2n} + \frac{1}{n}\left(2^{J-\hat{\mathtt{j}}_k(\zeta)} - 2^{J-\hat{\mathtt{j}}_k(\zeta)-1}\right), & \check{\mathtt{j}}_k(\zeta) < \infty \\ \frac{1}{n} \underset{\hat{\mathtt{i}}_{k,J}-2 \leq i \leq \hat{\mathtt{i}}_{k,J}+2}{\arg\min} \quad \nu^e_{k,i-1} - \frac{1}{n}, & \check{\mathtt{j}}_k(\zeta) = \infty \end{cases}. \tag{4.11}$$

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

$$\hat{Z} = (\hat{Z}_1, \hat{Z}_2, \cdots, \hat{Z}_s), \tag{4.12}$$

where $\hat{Z}_k$ is defined in (4.11) for $k \in \{1, 2, \cdots, s\}$.

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

Let

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

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

$$t_{k,lo} = \frac{2^{J-\hat{\mathtt{j}}_k(\alpha/2s)}}{n} L_k - \frac{1}{2n}, t_{k,hi} = \min\{\frac{2^{J-\hat{\mathtt{j}}_k(\alpha/2s)}}{n} U_k - \frac{1}{2n}, 1\}. \tag{4.13}$$

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

The key ideas of Algorithm 1 are as follows.

$\check{\mathtt{j}}_k(\alpha/2s) = \infty$ means that $\mathtt{T}_{k,j}$ 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 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 $1/n$ (i.e. $i_m = 1$ or $i_m = n-1$), we use a geometry property of convex functions to determine the final interval. Basically, the right most possible minimizer is or is infinitely near to the intersection of two lines : $y = f(i_m/n)$, and the line 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 $CI_{z,\alpha}$ is given by

$$CI_{k,\alpha} = [t_{k,lo}, t_{k,hi}]. \tag{4.14}$$

The $(1-\alpha)$-level confidence hyper cube $CI_{z,\alpha}$ is given by

$$CI_{z,\alpha} = CI_{1,\alpha} \times CI_{2,\alpha} \times \cdots \times CI_{s,\alpha}, \tag{4.15}$$

where $CI_{k,\alpha}$ is defined in (4.14).

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

$L_k \leftarrow \max\{0, 2\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\alpha/2s)-1} - 15\}, U_k = \min\{n, 2\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\alpha/2s)-1} + 12\}$, $\alpha_1 = \alpha/8s$, $\alpha_2 = \alpha/24s$

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

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

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

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

Now we turn to the construction of the estimator and confidence interval for the minimum.

We start with estimation for the minimum $M(\mathbf{f})$. Let $\zeta = \Phi(-2)$. For axis

k, let

$$\Delta_k = \mathbb{1}\{\mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}+6} - \mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}+5} \le z_\zeta \tilde{\sigma}^2_{k,\hat{\mathtt{j}}_k(\zeta)}\} - \mathbb{1}\{\mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}-6} - \mathtt{Y}^r_{k,j,\hat{\mathtt{i}}_{k,j}-5} \le z_\zeta \tilde{\sigma}^2_{k,\hat{\mathtt{j}}_k(\zeta)}\}.$$

The estimator for $M(\mathbf{f})$ is given as follows.

We define s intermediate estimators $\hat{M}_k$ as

$$\hat{M}_k = \begin{cases} 2^{\hat{\mathtt{j}}_k(\zeta)-J} \mathtt{Y}^e_{k,\hat{\mathtt{j}}_k(\zeta),\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\zeta)}+2\Delta_k}, & \check{\mathtt{j}}_k(\zeta) < \infty \\ \min\limits_{\hat{\mathtt{i}}_{k,J}-2 \le i \le \hat{\mathtt{i}}_{k,J}+2} \nu^e_{k,i-1}, & \check{\mathtt{j}}_k(\zeta) = \infty \end{cases}. \tag{4.16}$$

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

$$\hat{M} = \frac{1}{(n+1)^s} \sum_{\mathbf{i} \in \{0,1,2,\cdots,n\}^s} \mathbf{er}(\{y_\mathbf{i}\}) + \sum_{k=1}^s \hat{M}_k. \tag{4.17}$$

Now we continue with the confidence interval for the minimum $M(\mathbf{f})$. Let $\zeta_k = \zeta = \alpha/4s$.

Define the step number that will be used for constructing the interval as

$$j_{F,k} = \begin{cases} \check{\mathtt{j}}_k(\zeta) + 3, & \text{for } \check{\mathtt{j}}_k(\zeta) \le J \\ \infty, & \text{for } \check{\mathtt{j}}_k(\zeta) = \infty \end{cases} \tag{4.18}$$

Basically, we go three steps forward from the step that the test statistic $\mathtt{T}_{k,j}$ triggers the stopping rule.

Define

$$\begin{aligned} I_{k,lo} &= 2^{(j_{F,k}\wedge J)-\hat{\mathtt{j}}_k(\zeta)+1} \times \left(\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\zeta)-1} - 7\right), \\ I_{k,hi} &= 2^{(j_{F,k}\wedge J)-\hat{\mathtt{j}}_k(\zeta)+1} \times \left(\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\zeta)-1} + 6\right) + 1 \end{aligned} \tag{4.19}$$

We first define 3 sets of s intermediate estimators $\{\tilde{M}_{k,md} : 1 \le k \le s\}, \{\tilde{M}_{k,hi} : 1 \le k \le s\}, \{\tilde{M}_{k,lo} : 1 \le k \le s\}$ as

$$\tilde{M}_{k,md} = \min_{I_{k,lo} \le i \le I_{k,hi}} \mathtt{Y}^e_{k,(j_{F,k}\wedge J),i} \times 2^{(j_{F,k}\wedge J)-J}, \tag{4.20}$$

$$\tilde{M}_{k,hi} = \tilde{M}_{k,md} + S_{210,\alpha/8s} \times \sqrt{3} \frac{\sigma}{(n+1)^{\frac{s-1}{2}}} \times 2^{\frac{(j_{F,k}\wedge J)-J}{2}} \tag{4.21}$$

and

(4.22)

$$\tilde{M}_{k,lo} = \tilde{M}_{k,md} - \frac{3\sigma(z_{\alpha/4s}+1)}{(n+1)^{\frac{s-1}{2}}} \times 2^{\frac{j_{F,k}-J}{2}} - S_{210,\alpha/8s} \times \sqrt{3} \frac{\sigma}{(n+1)^{\frac{s-1}{2}}} \times 2^{\frac{(j_{F,k}\wedge J)-J}{2}} \text{ for } j_{F,k} \le J.$$

Let $\tilde{M}_{k,lo}$ be computed by Algorithm 2 when $j_{F,k} > J$. Algorithm 2 is based on the geometric property of the convex function f that for any $1 \le i \le n-2$,

$$\inf_{t\in[\frac{i}{n},\frac{i+1}{n}]} f(t) \ge \inf_{t\in[\frac{i}{n},\frac{i+1}{n}]} \max\left\{ \frac{f_k(\frac{i+2}{n}) - f_k(\frac{i+1}{n})}{1/n}(t - \frac{i+1}{n}) + f_k(\frac{i+1}{n}), \frac{f_k(\frac{i}{n}) - f_k(\frac{i-1}{n})}{1/n}(t - \frac{i}{n}) + f_k(\frac{i}{n}) \right\}.$$

Algorithm 2 Computing $\tilde{M}_{k,lo}$ when $j_{F,k} > J$

$k_l \leftarrow \max\{0, I_{k,lo} - 1\}$, $k_r \leftarrow \min\{n-1, I_{k,hi} - 2\}$,$H \leftarrow S_{k_r-k_l+4,\frac{\alpha}{24s}}\sqrt{3}\frac{\sigma}{(n+1)^{\frac{s-1}{2}}} + z_{\frac{\alpha}{48s}}\frac{\sqrt{3}\sigma}{(n+1)^{\frac{s}{2}}}$

if $k_l = 0$ **then**

$v_{r,0}(t) \leftarrow \frac{\nu^e_{k,2}-\nu^e_{k,1}+2H}{1/n}(t-1/n) + \nu^e_{k,1} - H$,$h(0) \leftarrow \min_{t\in[0,1/n]} v_{r,0}(t)$

end if

if $k_r = n-1$ **then**

$v_{l,n-1}(t) \leftarrow \frac{\nu^e_{k,n-1}-\nu^e_{k,n-2}-2H}{1/n}(t-\frac{n-1}{n})+\nu^e_{k,n-1}-H$, $h(n-1) = \min_{t\in[\frac{n-1}{n},1]} v_{l,n-1}(t)$

end if

for $i = (k_l \vee 1), \cdots, (k_r \wedge n-2)$ **do**

Define two linear functions:

$$v_{l,i}(t) = \frac{\nu^e_{k,i} - \nu^e_{k,i-1} - 2H}{1/n}(t - \frac{i}{n}) + \nu^e_{k,i} - H,$$

$$v_{r,i} = \frac{\nu^e_{k,i+2} - \nu^e_{k,i+1} + 2H}{1/n}(t - \frac{i+1}{n}) + \nu^e_{k,i+1} - H$$

$h(i) = \min_{t\in[\frac{i}{n},\frac{i+1}{n}]} \max\{v_{l,i}(t), v_{r,i}(t)\}$

end for

$\tilde{M}_{k,lo} \leftarrow \min\{h(i) : k_l \le i \le k_r\} \wedge \tilde{M}_{k,hi}$

Let

(4.23)
$$\tilde{M}_{hi} = \frac{1}{(n+1)^s} \sum_{\mathbf{i}\in\{0,1,2,\cdots,n\}^s} \mathbf{er}(\{y_{\mathbf{i}}\}) + \sum_{k=1}^{s} \tilde{M}_{k,hi} + z_{\alpha/8} \cdot 2\sqrt{3}\frac{\sigma}{(n+1)^{\frac{s}{2}}} s,$$

(4.24)
$$\tilde{M}_{lo} = \frac{1}{(n+1)^s} \sum_{\mathbf{i}\in\{0,1,2,\cdots,n\}^s} \mathbf{er}(\{y_{\mathbf{i}}\}) + \sum_{k=1}^{s} \tilde{M}_{k,lo} - z_{\alpha/8} \cdot 2\sqrt{3}\frac{\sigma}{(n+1)^{\frac{s}{2}}} s.$$

The confidence interval for the minimum $M(\mathbf{f})$ is given by

$$CI_{m,\alpha} = [\tilde{M}_{lo}, \tilde{M}_{hi}]. \tag{4.25}$$

4.3. *Statistical Optimality.* Now we establish the optimality of the adaptive procedures constructed in Section 4.2. The results show that our procedures are simultaneously optimal (up to a constant depending on dimension and confidence level) for $\mathbf{f} \in \mathcal{F}_s$ in terms our benchmarks introduced in (4.2) and (4.3).

We begin with the estimator of the minimizer.

THEOREM 4.1 (Estimation for Minimizer). *The estimator $\hat{Z}$ defined in (4.12) satisfies*

$$\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z} - Z(\mathbf{f})\|^2\right) \le Q_{z,s}\tilde{\mathtt{R}}_{z,n}(\sigma;\mathbf{f}), \text{ for all } \mathbf{f} \in \mathcal{F}_s \tag{4.26}$$

where $Q_{z,s}$ is a positive constant depending on s.

For the confidence hyper cube $CI_{z,\alpha}$ of $Z(\mathbf{f})$, we have the following result.

THEOREM 4.2 (Inference for Minimizer). *For $0 < \alpha \le 0.3$, confidence cube $CI_{z,\alpha}$ defined in (4.15) is a $(1-\alpha)$-level confidence cube for the minimizer $Z(\mathbf{f})$ and its expected volume satisfies*

$$\mathbb{E}_{\mathbf{f}}\left(V(CI_{z,\alpha})\right) \le Q_{z,s,\alpha}\tilde{\mathtt{L}}_{z,\alpha,n}(\sigma;\mathbf{f}), \text{ for all } \mathbf{f} \in \mathcal{F}_s \tag{4.27}$$

where $Q_{z,s,\alpha}$ is a positive constant depending on s and α only.

Similarly, the estimator and confidence interval for the minimizer $M(\mathbf{f})$ also achieve within a constant depending on s and α of the corresponding benchmark simultaneously for all $\mathbf{f} \in \mathcal{F}_s$.

THEOREM 4.3 (Estimation for Minimum). *The estimator $\hat{M}$ defined in (4.17) satisfies*

$$\mathbb{E}\left(\hat{M} - M(\mathbf{f})\right)^2 \le Q_{m,s}\tilde{\mathtt{R}}_{m,n}(\sigma;\mathbf{f}) \tag{4.28}$$

where $Q_{m,s}$ is a positive constant depending on s.

THEOREM 4.4 (Inference for Minimum). *For $0 < \alpha \le 0.3$, the confidence interval $CI_{m,\alpha}$ defined in (4.25) is a $(1-\alpha)$ level confidence interval for minimum $M(\mathbf{f})$ and its expected length satisfies*

$$\mathbb{E}(|CI_{m,\alpha}|) \le Q_{m,s,\alpha}\tilde{\mathtt{L}}_{m,\alpha,n}(\sigma;\mathbf{f}), \tag{4.29}$$

where $Q_{m,s,\alpha}$ is a positive constant depending on dimension s and α.

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

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

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 minimizer. 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 $R_z(\varepsilon; \mathbf{f})$. 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 $L_{\alpha,z}(\varepsilon; \mathbf{f})$, 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 $\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_{-1}}(Y)$ by Girsanov's theorem,

$$\frac{dP_{\mathbf{f}}}{dP_{\mathbf{g}}}(Y) = \exp\left(\int_{\mathcal{D}} \frac{\mathbf{f}(\mathbf{t}) - \mathbf{g}(\mathbf{t})}{\varepsilon^2} dY(\mathbf{t}) - \frac{1}{2}\int_{\mathcal{D}} \frac{\mathbf{f}(\mathbf{t})^2 - \mathbf{g}(\mathbf{t})^2}{\varepsilon^2}\mathrm{d}\mathbf{t}\right), \tag{6.3}$$

a sufficient statistic for θ is given by

$$W = \frac{\int_{\mathcal{D}}(\mathbf{f}_1(\mathbf{t}) - \mathbf{f}_{-1}(\mathbf{t}))\mathrm{d}Y(\mathbf{t}) - \frac{1}{2}\int_{\mathcal{D}}(\mathbf{f}_1(\mathbf{t})^2 - \mathbf{f}_{-1}(\mathbf{t})^2)\mathrm{d}\mathbf{t}}{\varepsilon\|\mathbf{f}_1 - \mathbf{f}_{-1}\|}. \tag{6.4}$$

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 $\omega_z(\varepsilon;\mathbf{f}) > \delta > 0$, there exists $\mathbf{h}_\delta \in \mathcal{F}_s$ such that $\|\mathbf{f} - \mathbf{h}_\delta\| = \varepsilon$ and that $\|Z(\mathbf{f}) - Z(\mathbf{h}_\delta)\|^2 \geq \omega_z(\varepsilon;\mathbf{f}) - \delta$, we let $\mathbf{g} = \mathbf{h}_\delta$. Then we have $R_z(\varepsilon; f) \geq (\omega_z(\varepsilon;\mathbf{f}) - \delta)\cdot r_2$, where r_2 is the minimax risk of the two-point problem based on an observation $X \sim N(\frac{\theta}{2}, 1)$,

$$r_2 = \inf_{\hat{\theta}} \max_{\theta=\pm 1} \mathbb{E}_\theta \frac{|\hat{\theta} - \theta|^2}{4}.$$

Elementary calculation shows that $r_2 \geq 0.1$. Taking $\delta \to 0^+$, we have $R_z(\varepsilon;\mathbf{f}) \geq 0.1\omega_z(\varepsilon;\mathbf{f})$. So we have $a \geq 0.1$.

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

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

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

PROOF. The left hand side is apparent, we will prove the right hand side. Using Proposition 6.3, we have

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

Using Proposition 2.1 by Cai et al. (2023a), namely

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

, we know $\rho_z(\epsilon; f)$ is a continuous function of $\epsilon \geq 0$ 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)

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

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

Also we have

$$\sum_{i=1}^s \rho_z(\tilde{b}_i c\varepsilon; f_i)^2 \leq \sum_{i=1}^s \rho_z(\tilde{b}_i \varepsilon; f_i)^2 \leq \sum_{i=1}^s \rho_z(\bar{b}_i \varepsilon; f_i)^2, \tag{6.8}$$

and

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

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

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

□

Now we continue with the upper bounds.
Recalling W define in (6.4), 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,

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

In addition

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

Take $A = 3.1$ gives the result.

Now we turn to the minimum and start with estimation. We start with the lower bound.

Recall that W defined in (6.4) is a sufficient statistics for θ.

Then similarly to the proof of that for minimizer we have that

$$
R_m(\varepsilon; \mathbf{f}) \geq a\omega_m(\varepsilon; \mathbf{f}). \tag{6.15}
$$

For the upper bound. We start with a proposition.

Proposition 6.2. *For $c > 1$, we have*

$$
\omega_m(c\varepsilon; \mathbf{f}) \leq c^2 \omega_m(\varepsilon; \mathbf{f}), \tilde{\omega}_m(c\varepsilon; \mathbf{f}) \leq c\tilde{\omega}_m(\varepsilon; \mathbf{f}). \tag{6.16}
$$

Proof. Suppose $\mathbf{g}$ satisfies $\|\mathbf{g} - \mathbf{f}\|_2 \leq c\varepsilon$. Then calculation show that

$$
|g_0 - f_0|^2 + \sum_{i=1}^{s} \|g_i - f_i\|^2 \leq c^2 \varepsilon^2, \tag{6.17}
$$

Let $h_i(t) = \frac{1}{c} g_i(t) + \frac{c-1}{c} f_i(t)$. Let $\mathbf{h}(\mathbf{t}) = \frac{1}{c} g_0 + \frac{c-1}{c} f_0 + \sum_{i=1}^{s} h_i(t_i)$ Then we have that

$$
\|\mathbf{h} - \mathbf{f}\|^2 \leq \varepsilon^2, \tag{6.18}
$$

and that

$$
|M(\mathbf{h}) - M(\mathbf{f})| = \frac{1}{c} |M(\mathbf{g}) - M(\mathbf{f})|. \tag{6.19}
$$

This gives the statement of the proposition.

□

Recalling W define in (6.4), let

$$\hat{M} = sign(W) \cdot \frac{M(\mathbf{f}) - M(\mathbf{g})}{2} + \frac{M(\mathbf{f}) + M(\mathbf{g})}{2}. \tag{6.20}$$

Then

$$\mathbb{E}_{\mathbf{f}}(\|\hat{M} - M(\mathbf{f})\|^2) = \mathbb{E}_{\mathbf{g}}(\|\hat{M} - M(\mathbf{g})\|^2) = \|M(\mathbf{f}) - M(\mathbf{g})\ |^2 \Phi(-\frac{\|\mathbf{f} - \mathbf{g}\|}{2\varepsilon}). \tag{6.21}$$

With Proposition 6.2 we have that

$$\begin{aligned} R_m(\varepsilon; \mathbf{f}) &\leq \sup_{c>0} \omega_m(c\varepsilon; \mathbf{f}) \Phi(-\frac{c}{2}) \leq \max\{0.5\omega_m(\varepsilon; \mathbf{f}), \sup_{c\geq 1} \omega_m(c\varepsilon; \mathbf{f}) \Phi(-\frac{c}{2})\} \\ &\leq \omega_m(\varepsilon; \mathbf{f}) \max\{0.5, \sup_{c\geq 1} c^2 \Phi(-\frac{c}{2})\} \leq \omega_m(\varepsilon; \mathbf{f}). \end{aligned} \tag{6.22}$$

For the inference of the minimum, we again start with the lower bound.

$$\begin{aligned} L_{\alpha,m}(\varepsilon; \mathbf{f}) &\geq \sup_{\mathbf{g} \in \mathcal{F}_s} \inf_{CI_{m,\alpha} \in \mathcal{I}_{m,\alpha}(\mathbf{f},\mathbf{g})} \mathbb{P}_{\mathbf{f}}(\{M(\mathbf{g}), M(\mathbf{f})\} \in CI_{m,\alpha}) |M(\mathbf{f}) - M(\mathbf{g})| \\ &\geq \sup_{\mathbf{g} \in \mathcal{F}_s, \|\mathbf{g} - \mathbf{f}\| \leq \varepsilon} (1 - \alpha - \mathbb{P}_{\mathbf{f}}(M(\mathbf{g}) \notin \mathcal{I}_{m,\alpha}(\mathbf{f}, \mathbf{g}))) \tilde{\omega}_m(\varepsilon; \mathbf{f}) \\ &\geq (1 - \alpha - \Phi(-z_\alpha + 1)) \tilde{\omega}_m(\varepsilon; \mathbf{f}) \geq (0.6 - \alpha) \tilde{\omega}_m(\varepsilon; \mathbf{f}). \end{aligned} \tag{6.23}$$

The second to last inequality is due to Neyman-Pearson inequality.

For the upper bound, we recollect our sufficient statistics (6.4) and associated notation, let

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

Clearly, we have $P_{\mathbf{f}}(M(\mathbf{f}) \notin CI_\alpha) \leq \alpha, P_{\mathbf{g}}(M(\mathbf{g}) \notin CI_\alpha) \leq \alpha$. For the expected squared length, we have for $\theta \in \{-1, 1\}$,

$$\mathbb{E}_{\mathbf{f}_\theta}(|CI_{m,\alpha}|) \leq \|M(\mathbf{f}) - M(\mathbf{g})\| \left(\Phi(z_\alpha - \frac{\|\mathbf{f} - \mathbf{g}\|}{\varepsilon}) - \alpha\right)_+ \tag{6.24}$$

$$\begin{aligned} \mathbb{E}_{\mathbf{f}_\theta}(|CI_{m,\alpha}|) &\leq \max\{\tilde{\omega}_m(\varepsilon; \mathbf{f})(1 - 2\alpha), \sup_{c>1} \tilde{\omega}_m(c\varepsilon; \mathbf{f})(\Phi(z_\alpha - c) - \alpha)_+\} \\ &\leq \tilde{\omega}_m(\varepsilon; \mathbf{f}) \max\{(1 - 2\alpha), \sup_{c>1} c(\Phi(z_\alpha - c) - \alpha)_+\} \\ &\leq \tilde{\omega}_m(\varepsilon; \mathbf{f})(1 - 2\alpha) \times 2z_\alpha. \end{aligned} \tag{6.25}$$

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

PROPOSITION 6.3. *Let $\rho_z(\varepsilon; f)$ be defined in (2.8) for $f \in \mathcal{F}$, and let $\mathbf{f} \in \mathcal{F}_s$. Then*

$$\sup_{\sum_{i=1}^s b_i^2 \leq 1} \sum_{i=1}^s \rho_z(b_i\varepsilon; f_i)^2 \leq \omega_z(\varepsilon; \mathbf{f}) \leq \sup_{\sum_{i=1}^s b_i^2 \leq 1} \sum_{i=1}^s 9\rho_z(b_i\varepsilon; f_i)^2, \tag{6.26}$$

where b_i are non-negative.

PROPOSITION 6.4. *Suppose $f_i \in \mathcal{F}$, for $i = 1, 2, \cdots, s$, then we have*

$$\frac{1}{3} s^{-\frac{2}{3}} \sum_{i=1}^s \rho_z(\varepsilon; f_i)^2 \leq \sup_{\sum_{i=1}^s b_i^2 \leq 1} \sum_{i=1}^s \rho_z(b_i\varepsilon; f_i)^2 \leq \sum_{i=1}^s \rho_z(\varepsilon; f_i)^2. \tag{6.27}$$

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

$$\sup_{\sum_{i=1}^s b_i^2 \leq 1} \sum_{i=1}^s \rho_z(b_i\varepsilon; f_i)^2 = s^{-\frac{2}{3}} \sum_{i=1}^s \rho_z(\varepsilon; f_i)^2. \tag{6.28}$$

For $\beta \leq s$, 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

$$\sup_{\sum_{i=1}^s b_i^2 \leq 1} \sum_{i=1}^s \rho_z(b_i\varepsilon; f_i)^2 \geq \sum_{i=1}^s \rho_z(\varepsilon; f_i)^2 - \delta. \tag{6.29}$$

Inequality (6.27) in Proposition 6.4 and (6.26) in Proposition 6.3 implies Inequality 2.9 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 (6.28). Then together with the right hand side of Inequality (6.26) gives Inequality (2.10) of Theorem 2.2. Similar construct $\mathbf{f}$ with f_i in Inequality (6.29) with $\delta_0 = \delta$ gives Inequality (2.11) in Theorem 2.2.

6.3.1. *Proof of Proposition 6.3.* 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 2.1 by Cai et al. (2023a), we know there exist $(\bar{b}_1, \bar{b}_2, \cdots, \bar{b}_s)$ such that

$$\bar{b}_i \geq 0, \text{ for } 1 \leq i \leq s, \sum_{i=1}^s \bar{b}_i^2 = 1, \sum_{i=1}^s \rho_z(\bar{b}_i\varepsilon; f_i)^2 = \sup_{\sum_{i=1}^s b_i^2 \leq 1} \sum_{i=1}^s \rho_z(b_i\varepsilon; f_i)^2. \tag{6.30}$$

We only need to prove

$$\sum_{i=1}^{s}\rho_z(\bar{b}_i\varepsilon;f_i)^2 \le \omega_z(\varepsilon;\mathbf{f}) \le \sum_{i=1}^{s}9\rho_z(\bar{b}_i\varepsilon;f_i)^2. \tag{6.31}$$

We start with proving the upper bound.
Since $\|\mathbf{g}-\mathbf{f}\|\le\varepsilon$, we have

$$\begin{aligned}\varepsilon^2 \ge \|\mathbf{f}-\mathbf{g}\|^2 &= \int_{\mathcal{D}}\left(f_0-g_0+\sum_{i=1}^{2}f_i(t_i)-g_i(t_i)\right)^2\mathrm{d}\mathbf{t}\\ &= (f_0-g_0)^2+\sum_{i=1}^{s}\int_0^1(f_i(t)-g_i(t))^2\mathrm{d}t.\end{aligned}\tag{6.32}$$

Denote $\tilde{b}_i=\sqrt{\frac{\int_0^1(f_i(t)-g_i(t))^2\mathrm{d}t}{\varepsilon^2}}$ for $1\le i\le s$, then we have $\sum_{i=1}^{s}\tilde{b}_i^2=1$. Therefore, using Proposition 2.2 by Cai et al. (2023a), we have

$$\|Z(\mathbf{f})-Z(\mathbf{g})\|^2=\sum_{i=1}^{s}|Z(f_i)-Z(g_i)|^2\le\sum_{i=1}^{s}9\rho_z(\tilde{b}_i\varepsilon;f_i)^2\le\sum_{i=1}^{s}9\rho_z(\bar{b}_i\varepsilon;f_i)^2. \tag{6.33}$$

For the lower bound, we construct a class of function $\mathbf{g}_\delta\in\mathcal{F}_s$, with $\frac{1}{2}\min_{1\le i\le s}\rho_z(\bar{b}_i\varepsilon;f_i)>\delta>0$. We construct the constant and components: $g_{\delta,i}$ for $0\le s$. Let $g_{\delta,0}=f_0$. For $1\le i\le s$, suppose $x_{l,i},x_{r,i}$ are left and right end points of the interval $\{x:f_i(x)\le M(f_i)+\rho_m(\bar{b}_i\varepsilon;f_i)\}$. 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.

$$h_{\delta,i}(t)=\max\{f_i(t),f_i(x_{r,i}-\delta)-\frac{\rho_m(\bar{b}_i\varepsilon;f_i)+M(f_i)-f_i(x_{r,i}-\delta)}{x_{r,i}-\delta-x_{l,i}}(t-x_{r,i})\}. \tag{6.34}$$

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

$$g_{\delta,i}(t)=h_{\delta,i}(t)-\int_0^1h_{\delta,i}(t)\mathrm{d}t. \tag{6.35}$$

Then we have $\int_0^1g_{\delta,i}(t)\mathrm{d}t=0$, so the definition defines a valid $\mathbf{g}_\delta\in\mathcal{F}_s$. Further for $i=1,2\cdots,s$, we have

$$\int_0^1(g_{\delta,i}(t)-f_i(t))^2\,\mathrm{d}t=\int_0^1(h_{\delta,i}(t)-f_i(t))^2\,\mathrm{d}t-\left(\int_0^1h_{\delta,i}(t)\mathrm{d}t\right)^2\le\bar{b}_i^2\varepsilon^2, \tag{6.36}$$

and

$$|Z(g_{\delta,i}) - Z(f_i)| \geq \rho_z(\bar{b}_i\varepsilon; f_i) - \delta. \tag{6.37}$$

Therefore, we have

$$\|\mathbf{g}_\delta - \mathbf{f}\|^2 \leq \varepsilon^2, \|Z(\mathbf{g}_\delta) - Z(\mathbf{f})\|^2 \geq \sum_{i=1}^{s} \left(\rho_z(\bar{b}_i\varepsilon; f_i) - \delta\right)^2. \tag{6.38}$$

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

$$\omega_z(\varepsilon; \mathbf{f}) \geq \sum_{i=1}^{s} \rho_z(\bar{b}_i\varepsilon; f_i)^2. \tag{6.39}$$

6.3.2. *Proof of Proposition 6.4.* We start with the right hand side and its almost-attainability.

Since $b_i \in [0,1]$ for $1 \leq i \leq s$, we have $\rho_z(b_i\varepsilon; f_i) \leq \rho_z(\varepsilon; f_i)$. 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 $1 \leq i \leq s$.

If $s_1 \geq 1$, for $1 \leq i \leq s_1$, let

$$f_i(t) = k_1(t - \frac{1}{2}). \tag{6.40}$$

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

$$f_i(t) = k_3(t - \frac{1}{2}). \tag{6.41}$$

Let

$$f_{s_1+1}(t) = k_2(t - \frac{1}{2}). \tag{6.42}$$

Suppose $0 < \delta < \frac{1}{2}s_2$.

If $s_3 \geq 1$, choose k_3 such that

$$\rho_z(\varepsilon; f_n) = \sqrt{\frac{\delta}{2s_3}}, \tag{6.43}$$

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

$$\rho_z(\varepsilon; f_{s_1+1}) = \sqrt{s_4}. \tag{6.44}$$

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

$$\rho_z(b\varepsilon; f_{s_1+1}) \geq \sqrt{s_4 - \frac{\delta}{2}}. \tag{6.45}$$

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

$$\rho_z(\sqrt{\frac{1-b^2}{s_1}}\varepsilon; f_1) = 1. \tag{6.46}$$

It's easy to verify that the above construction is legitimate and satisfy equation (6.29).

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.

Proposition 2.1 by Cai et al. (2023a), we have

$$\sup_{\sum_{i=1}^s b_i^2 \leq 1} \sum_{i=1}^s \rho_z(b_i\varepsilon; f_i)^2 \geq \sup_{\sum_{i=1}^s b_i^2 \leq 1} \sum_{i=1}^s (b_i^2/4)^{\frac{2}{3}} \rho_z(\varepsilon; f_i)^2 \geq \frac{1}{3}\left(\sum_{i=1}^s \rho_z(\varepsilon; f_i)^6\right)^{\frac{1}{3}}, \tag{6.47}$$

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

$$\frac{1}{3}\left(\sum_{i=1}^s \rho_z(\varepsilon; f_i)^6\right)^{\frac{1}{3}} \geq \frac{1}{3} s^{-\frac{2}{3}} \sum_{i=1}^s \rho_z(\varepsilon; f_i)^2, \tag{6.48}$$

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

$$\sup_{\sum_{i=1}^s b_i^2 \leq 1} \sum_{i=1}^s \rho_z(b_i\varepsilon; f_i)^2 = \sup_{\sum_{i=1}^s b_i^2 \leq 1} \sum_{i=1}^s b_i^{\frac{4}{3}} \rho_z(\varepsilon; f_i)^2 = \sup_{\sum_{i=1}^s b_i^2 \leq 1} \sum_{i=1}^s b_i^{\frac{4}{3}} \frac{\beta}{s}. \tag{6.49}$$

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 (6.28).

6.4. *Proof of Theorem 2.3.* We start with the upper bound. Suppose $\|\mathbf{g}-\mathbf{f}\| \le \varepsilon$. Suppose $\mathbf{g}(\mathbf{t}) = g_0 + \sum_{i=1}^s g_i(t_i)$, where $\int_0^1 g_i(t)\mathrm{d}t = 0$. Calculation show that $\|\mathbf{g}-\mathbf{f}\| \le \varepsilon$ implies

$$|g_0 - f_0|^2 + \sum_{i=1}^s \|g_i - f_i\|^2 \le \varepsilon^2. \tag{6.50}$$

Suppose $\varepsilon_i = \|g_i - f_i\|$. Then we have that

(6.51)

$$\begin{aligned}
&|M(\mathbf{g}) - M(\mathbf{f})|^2 \le (|g_0 - f_0| + \sum_{i=1}^s |M(g_i) - M(f_i)|)^2 \le (|g_0 - f_0| + \sum_{i=1}^s 3\rho_m(\varepsilon_i; f_i))^2 \\
&\le (|g_0 - f_0| + \sum_{i=1}^s 3(\frac{\varepsilon_i}{\varepsilon})^{\frac{4}{3}}\rho_m(\varepsilon; f_i))^2 \\
&\le \left(\varepsilon^2 + \sum_{i=1}^s \rho_m(\varepsilon; f_i)^2\right)\left(\left(\frac{|g_0 - f_0|}{\varepsilon}\right)^2 + \sum_{i=1}^s 9(\frac{\varepsilon_i}{\varepsilon})^{\frac{8}{3}}\right) \\
&\le \left(\sum_{i=1}^s \rho_m(\varepsilon; f_i)^2\right)\frac{9(s+1)}{s},
\end{aligned}$$

where the second Inequality is due to Proposition 2.1.

Now that we have the upper bound, we turn to the lower bound. Let

$$\varepsilon_i = \frac{\rho_m(\varepsilon; f_i)}{\sqrt{\sum_{j=1}^s \rho_m(\varepsilon; f_j)^2}}\sqrt{\frac{1}{1 + \sum_{i=1}^s (1 \wedge 2\rho_z(\varepsilon; f_i))}}\varepsilon. \tag{6.52}$$

Suppose $\delta > 0$ is small enough quantity, which will be set going to 0 later. We construct components of an alternative function. Without loss of generality we assume $t_{i,l}$, $t_{i,r}$ are the left and right end points of the interval $\{t : f_i(t) \le M(f_i) + \rho_m(\varepsilon_i; f_i)\}$ and that $t_{i,r} = Z(f_i) + \rho_z(\varepsilon_i; f_i)$. Suppose $g_{i,\delta}(t) = \max\{f_i(t), f_i(t_l) + \frac{-\delta}{t_{i,r} - t_{i,l}}(t - t_{i,l})\}$, and let $\mathbf{h}_\delta(\mathbf{t}) = f_0 + \sum_{i=1}^s g_i(t_i)$. Then we have for small enough $\delta > 0$,

$$\begin{aligned}
\|\mathbf{h}_\delta - \mathbf{f}\|^2 &\le (\sum_{i=1}^s \int_0^1 g_i(t)\mathrm{d}t)^2 + \sum_{i=1}^s \varepsilon_i^2 - \sum_{i=1}^s (\int_0^1 g_i(t)\mathrm{d}t)^2 \\
&\le \sum_{i=1}^s \varepsilon_i^2 (1 + \sum_{i=1}^s (1 \wedge 2\rho_z(\varepsilon_i; f_i))) \le \varepsilon^2.
\end{aligned} \tag{6.53}$$

We also have

$$\lim_{\delta\to 0^+}(M(\mathbf{h}_\delta)-M(\mathbf{f}))\geq \sum_{i=1}^s \rho_m(\varepsilon_i;f_i)\geq \sum_{i=1}^s \rho_m(\varepsilon;f_i)\frac{\varepsilon_i}{\varepsilon}$$
$$\geq\sqrt{\sum_{i=1}^s \rho_m(\varepsilon;f_i)^2}\sqrt{\frac{1}{1+\sum_{i=1}^s(1\wedge 2\rho_z(\varepsilon;f_i))}}. \tag{6.54}$$

This gives the lower bound.

6.5. *Proof of Theorem 2.4.*

$$\begin{aligned}&\inf_{CI_{z,\alpha}\in\mathcal{I}_{z,\alpha}(\mathcal{F}_s)}\mathbb{E}_{\mathbf{f}}\left(V(CI_{z,\alpha})\right)\\ &\geq \sup_{\mathbf{g}\in\mathcal{F}_s}\inf_{CI_{z,\alpha}\in\mathcal{I}_{z,\alpha}(\mathbf{f},\mathbf{g})}\mathbb{E}_{\mathbf{f}}\left(V(CI_{z,\alpha})\right)\\ &\geq \sup_{\mathbf{g}\in\mathcal{F}_s}\inf_{CI_{z,\alpha}\in\mathcal{I}_{z,\alpha}(\mathbf{f},\mathbf{g})}\mathbb{E}_{\mathbf{f}}\left(\mathbb{1}\{\{Z(\mathbf{f}),Z(\mathbf{g})\}\subset CI_{z,\alpha}\}\right)\sup_{\mathbf{g}\in\mathcal{F}_s}\Pi_{i=1}^s|Z(g_i)-Z(f_i)|\\ &\geq \sup_{\mathbf{g}\in\mathcal{F}_s}\left(1-\alpha-\Phi(-z_\alpha+\frac{\|\mathbf{f}-\mathbf{g}\|}{\varepsilon})\right)\sup_{\mathbf{g}\in\mathcal{F}_s}\Pi_{i=1}^s|Z(g_i)-Z(f_i)|\end{aligned} \tag{6.55}$$

Let $g_{i,\delta}$ be constructed as follows. Without loss of generality, we assume $t_{i,r}=Z(f_i)+\rho_z(\varepsilon/\sqrt{s};f_i)$ satisfies $f_i(t_{i,r})\leq\rho_m(\varepsilon/\sqrt{s};f_i)+M(f_i)$ and $t_{i,l}$ is the left end point of $\{t: f_i(t)\leq\rho_m(\varepsilon/\sqrt{s};f_i)+M(f_i)\}$. Let

$$g_{i,\delta}(t)=\max\{f_i(t),M(f_i)+\rho_m(\varepsilon/\sqrt{s};f_i)+\frac{-\delta}{t_{i,r}-t_{i,l}}(t-t_{i,l})\}. \tag{6.56}$$

Define

$$\mathbf{g}_\delta(\mathbf{t})=f_0+\sum_{i=1}^s g_{i,\delta}(t_i)-\sum_{i=1}^s\int_0^1 g_{i,\delta}(t)\mathrm{d}t.$$

It's clear that

$$\|\mathbf{g}_\delta-\mathbf{f}\|\leq\varepsilon.$$

It is obvious that $Z(g_{\delta,i})=Z(g_{i,\delta})$.

$$\lim_{\delta\to 0^+}\Pi_{i=1}^s|Z(g_{\delta,i})-Z(f_i)|\geq\Pi_{i=1}^s\rho_z(\varepsilon/\sqrt{s};f_i)\geq(\frac{1}{2\sqrt{s}})^{\frac{2s}{3}}\Pi_{i=1}^s\rho_z(\varepsilon;f_i). \tag{6.57}$$

Going back to Inequality (6.55) we have that

$$\inf_{CI_{z,\alpha}\in\mathcal{I}_{z,\alpha}(\mathcal{F}_s)}\mathbb{E}_{\mathbf{f}}\left(V(CI_{z,\alpha})\right)\geq(0.6-\alpha)(\frac{1}{2\sqrt{s}})^{\frac{2s}{3}}\Pi_{i=1}^s\rho_z(\varepsilon;f_i). \tag{6.58}$$

6.6. *Proof of Theorem 2.5.* We prove the theorem by proving the following two propositions.

Proposition 6.5. *For any estimator of the minimizer, $\hat{Z}$, if*

$$\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z}-Z(\mathbf{f})\|^2\right) \leq \gamma R_z(\varepsilon;\mathbf{f})$$

for $\mathbf{f}\in\mathcal{F}_s$ and $\gamma<\gamma_0$, where γ_0 is a positive constant, then there exists $\mathbf{f}_1\in\mathcal{F}_s$ such that

$$\mathbb{E}_{\mathbf{f}_1}\left(\|\hat{Z}-Z(\mathbf{f}_1)\|^2\right) \geq c_{z,s}(\log\frac{1}{\gamma})^{\frac{2}{3}}R_z(\varepsilon;\mathbf{f}_1), \tag{6.59}$$

where $c_{z,s}$ is a constant depending on s only.

Proposition 6.6. *For any estimator of the minimum, $\hat{M}$, if*

$$\mathbb{E}_{\mathbf{f}}(|\hat{M}-M(\mathbf{f})|^2) \leq \gamma R_m(\varepsilon;\mathbf{f})$$

for $\mathbf{f}\in\mathcal{F}_s$ and $\gamma<\gamma_0/s$, where γ_0 is a positive constant, then there exists $\mathbf{f}_1\in\mathcal{F}_s$ such that

$$\mathbb{E}_{\mathbf{f}_1}\left(|\hat{M}-M(\mathbf{f}_1)|^2\right) \geq c_{m,s}(\log\frac{1}{\gamma})^{\frac{2}{3}}R_m(\varepsilon;\mathbf{f}_1), \tag{6.60}$$

where $c_{m,s}$ is a constant depending on s only.

6.6.1. *Proof of Proposition 6.5.* Let $\sigma=\frac{\Phi^{-1}(1-6\cdot 9\cdot 2\gamma)\varepsilon}{\sqrt{5}}$. Let $F(\gamma)=(\sigma/\varepsilon)^2$.

Then for $\gamma\leq 0.0024558/54$, we have $\sigma\geq\sqrt{\frac{4}{3}}\varepsilon$.

Suppose $(w_1,w_2,\cdots,w_s)$ achieves

$$\sup_{\sum_{j=1}^s w_j^2\leq 1, w_j\geq 0}\sum_i^s \rho_z(w_i\varepsilon;f_i)^2. \tag{6.61}$$

The compactness of $\{(w_1,w_2,\cdots,w_s):\sum_{j=1}^s w_j^2\leq 1, w_j\geq 0\}$ and the continuity of $\sum_i^s\rho_z(w_i\varepsilon;f_i)^2$ implies that supremum is attainable. So $(w_1,w_2,\cdots,w_s)$ is well defined. Also, it's easy to see that $\sum_{j=1}^s w_j^2=1$.

Denote set B as

$$B=\{(b_1,b_2,\cdots,b_s):\sum_{i=1}^s b_i\leq 1, b_i\geq\max\{\frac{w_i}{\sqrt{F(\gamma)}},\sqrt{1/4s}\}\}. \tag{6.62}$$

It's clear that B is not null set, and

$$(\sqrt{\frac{w_1^2}{F(\gamma)}+\frac{1}{4s}},\sqrt{\frac{w_2^2}{F(\gamma)}+\frac{1}{4s}},\cdots,\sqrt{\frac{w_s^2}{F(\gamma)}+\frac{1}{4s}})\in B. \tag{6.63}$$

Let $(b_1,b_2,\cdots,b_s)$ achieves

$$\sup_{(b_1,b_2,\cdots,b_s)\in B}\left(\sum_{k=1}^{s}\rho_z(b_k\sqrt{F(\gamma)}\varepsilon;f_k)^2\right)^3\Big/\left(\sum_{i=1}^{s}\frac{\rho_z(b_i\sqrt{F(\gamma)}\varepsilon;f_i)^4}{\rho_m(b_i\sqrt{F(\gamma)}\varepsilon;f_i)^4}\right). \tag{6.64}$$

Then it is clear that

$$\begin{aligned}&\left(\sum_{k=1}^{s}\rho_z(b_k\sqrt{F(\gamma)}\varepsilon;f_k)^2\right)^3\Big/\left(\sum_{i=1}^{s}\frac{\rho_z(b_i\sqrt{F(\gamma)}\varepsilon;f_i)^4}{\rho_m(b_i\sqrt{F(\gamma)}\varepsilon;f_i)^4}\right)\\ &\geq\min_{1\leq k\leq s}\left(\rho_z(b_k\sqrt{F(\gamma)}\varepsilon;f_k)^2\right)^3\Big/\left(\frac{\rho_z(b_k\sqrt{F(\gamma)}\varepsilon;f_k)^4}{\rho_m(b_k\sqrt{F(\gamma)}\varepsilon;f_k)^4}\right)\\ &\geq\min_{1\leq k\leq s}\left(\frac{1}{2}b_k^2F(\gamma)\varepsilon^2\right)^2\geq\left(\frac{F(\gamma)}{8s}\varepsilon^2\right)^2,\end{aligned} \tag{6.65}$$

and that

$$\sum_{k=1}^{s}\rho_z(b_k\sqrt{F(\gamma)}\varepsilon;f_k)^2\geq\sum_{k=1}^{s}\rho_z(w_k\varepsilon;f_k)^2\geq\frac{1}{9}\omega_z(\varepsilon;\mathbf{f}), \tag{6.66}$$

where the very last inequality comes from Proposition 6.3.

For each $1\leq k\leq s$, we construct $\tilde{f}_k$.

Let x_l,x_r be the left and right end points of the interval $\{x:f_k(x)\leq M(f_k)+\rho_m(b_k\sigma;f_k)\}$. Without loss of generality, suppose $f_k\left(Z(f_k)+\rho_z(b_k\sigma;f_k)\right)\leq M(f_k)+\rho_m(b_k\sigma;f_k)$.

Let $g_{2,k}(t)=\max\{f_k(t),f_k(x_r)+\frac{M(f_k)+2\rho_m(b_k\sigma;f_k)-f_k(x_r)}{x_l-x_r}(t-x_r)\}$.

Calculation similar to that in Lemma C.8 in Cai et al. (2023b) shows that

$$\begin{aligned}&\|g_{2,k}-f_k\|\leq\sqrt{5}b_k\sqrt{F(\gamma)}\varepsilon\\ &\rho_z(\eta;g_{2,k})\leq(\frac{16}{3})^{\frac{1}{3}}(\frac{\eta}{\sqrt{b_k^2\sigma^2/3}})^{\frac{2}{3}}\rho_z(b_k\sigma;f_k).\end{aligned} \tag{6.67}$$

Let

$$\mathbf{g}(\mathbf{t})=f_0+\sum_{k=1}^{s}\left(g_{2,k}(t_k)-\int_0^1g_{2,k}(t)\mathrm{d}t\right). \tag{6.68}$$

Then we know that

$$\|\mathbf{g}-\mathbf{f}\| \leq \Phi^{-1}(1-6\cdot 9\cdot 2\gamma)\varepsilon, \tag{6.69}$$

that

$$\|Z(\mathbf{g})-Z(\mathbf{f})\|^2 = \sum_{k=1}^{s} \rho_z(b_k\sigma; f_k)^2 \geq \frac{1}{9}\omega_z(\varepsilon;\mathbf{f}), \tag{6.70}$$

and that

$$\begin{aligned}\omega_z(\varepsilon;\mathbf{g}) &\leq 9 \sup_{\sum_{j=1}^{s} d_j^2\leq 1, d_j\geq 0} \sum_{k=1}^{s} \rho_z(d_j\varepsilon; g_{2,k})^2 \\ &\leq 9 \sup_{\sum_{j=1}^{s} d_j^2\leq 1, d_j\geq 0} \sum_{k=1}^{s} (\frac{16}{3})^{\frac{2}{3}} (\frac{d_k\varepsilon}{\sqrt{b_k^2\sigma^2/3}})^{\frac{4}{3}} \rho_z(b_k\sigma; f_k)^2.\end{aligned} \tag{6.71}$$

Taking derivative of

$$\sum_{k=1}^{s} (\frac{d_k}{\sqrt{b_k^2}})^{\frac{4}{3}} \rho_z(b_k\sigma; f_k)^2 \tag{6.72}$$

with respect to

$$(d_1^2, d_2^2, \cdots, d_s^2), \tag{6.73}$$

we have

$$\left(\frac{2}{3}(d_1^2)^{-\frac{1}{3}} b_1^{-\frac{4}{3}} \rho_z(b_1\sigma; f_1)^2, \cdots, \frac{2}{3}(d_s^2)^{-\frac{1}{3}} b_s^{-\frac{4}{3}} \rho_z(b_s\sigma; f_s)^2\right). \tag{6.74}$$

Note that the constraint for $d_1^2, d_2^2, \cdots, d_s^2$ is

$$\sum_{k=1}^{s} d_k^2 = 1, d_j^2 \geq 0 \text{ for } 1\leq j\leq s. \tag{6.75}$$

Therefore, we have that

$$\sum_{k=1}^{s}(\frac{d_k}{\sqrt{b_k^2}})^{\frac{4}{3}}\rho_z(b_k\sigma;f_k)^2 \le \sum_{k=1}^{s}\left(\frac{\rho_z(b_k\sigma;f_k)^6/b_k^4\sum_{j=1}^{s}\left(\rho_z(b_j\sigma;f_j)^6/b_j^4\right)}{b_k^2}\right)^{\frac{2}{3}}\rho_z(b_k\sigma;f_k)^2 \le \left(\sum_{j=1}^{s}\rho_z(b_j\sigma;f_j)^6/b_j^4\right)^{\frac{1}{3}} \le \left(\sum_{j=1}^{s}\sigma^4 4\cdot\frac{\rho_z(b_j\sigma;f_j)^4}{\rho_m(b_j\sigma;f_j)^4}\right)^{\frac{1}{3}}. \tag{6.76}$$

Using Inequality (6.65) and going back to Inequality (6.71), we have that

$$\begin{aligned}\omega_z(\varepsilon;\mathbf{g}) &\le 9\cdot(16\cdot 3)^{\frac{2}{3}}(\frac{\varepsilon}{\sigma})^{\frac{4}{3}}\cdot\sigma^{\frac{4}{3}}\cdot 4^{\frac{1}{3}}(\frac{8s}{F(\gamma)\varepsilon^2})^{\frac{2}{3}}\sum_{k=1}^{s}\rho_z(b_k\sigma;f_k)^2\\ &= 9\cdot(16\cdot 3)^{\frac{2}{3}}\cdot 2^{\frac{8}{3}}(\frac{s}{F(\gamma)})^{\frac{2}{3}}\|Z(\mathbf{f})-Z(\mathbf{g})\|^2.\end{aligned} \tag{6.77}$$

Recall that when we let $\mathbf{f}_\theta = \mathbf{f}$ for $\theta = 1$ and $\mathbf{f}_\theta = \mathbf{g}$ for $\theta = -1$, a sufficient statistic would be W defined in (6.4).

Note that we have

$$\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z}-Z(\mathbf{f})\|^2\right) \le \gamma R_z(\varepsilon;\mathbf{f}) \le 6\gamma\omega_z(\varepsilon;\mathbf{f}), \tag{6.78}$$

where the last Inequality comes from Theorem 2.1.

Denote event $D = \{\|\hat{Z}-Z(\mathbf{f})\| \ge \frac{1}{18}\omega_z(\varepsilon;\mathbf{f})\}$. Then

$$P_{\mathbf{f}}(D) \le \frac{6\gamma\omega_z(\varepsilon;\mathbf{f})}{\frac{1}{18}\omega_z(\varepsilon;\mathbf{f})} = 108\gamma \le 0.00491163. \tag{6.79}$$

So we have that

$$P_{\mathbf{g}}(D) \le \frac{1}{2}. \tag{6.80}$$

Hence we have that

(6.81)
$$\begin{aligned}\mathbb{E}_{\mathbf{g}}\left(\|\hat{Z}-Z(\mathbf{g})\|^2\right) &\geq \mathbb{E}_{\mathbf{g}}\left(\left(\|Z(\mathbf{f})-Z(\mathbf{g})\|-\frac{1}{18}\omega_z(\varepsilon;\mathbf{f})\right)_+^2 \mathbb{1}\{D^c\}\right)\\ &\geq \mathbb{E}_{\mathbf{g}}\left(\frac{1}{4}\|Z(\mathbf{f})-Z(\mathbf{g})\|^2\mathbb{1}\{D^c\}\right) \geq \frac{1}{8}\|Z(\mathbf{f})-Z(\mathbf{g})\|^2\\ &\geq \frac{1}{8}\frac{1}{9}(16\cdot 3)^{-\frac{2}{3}}\cdot 2^{-\frac{8}{3}}\frac{F(\gamma)^{\frac{2}{3}}}{s^{\frac{2}{3}}}\omega_z(\varepsilon;\mathbf{g})\\ &\geq \frac{1}{8}\frac{1}{9}(16\cdot 3)^{-\frac{2}{3}}\cdot 2^{-\frac{8}{3}}\frac{1}{6}R_z(\varepsilon;\mathbf{g})\frac{F(\gamma)^{\frac{2}{3}}}{s^{\frac{2}{3}}}.\end{aligned}$$

Note that $F(\gamma)=z_{108\gamma}^2/5$, so $F(\gamma)\sim\log(\frac{1}{\gamma})$, so we have

$$\mathbb{E}_{\mathbf{g}}\left(\|\hat{Z}-Z(\mathbf{g})\|^2\right) \geq c_z\cdot s^{-\frac{2}{3}}\log(\frac{1}{\gamma})^{\frac{2}{3}}R_z(\varepsilon;\mathbf{g}). \tag{6.82}$$

for some constant $c_z>0$.

Letting $c_{z,s}=c_z\cdot s^{-\frac{2}{3}}$ and $\mathbf{f}_1=\mathbf{g}$ gives the statement of the Proposition.

6.6.2. *Proof of Proposition 6.6.* Take $\sigma=\Phi^{-1}(1-108(s+1)^2\gamma/s)\varepsilon$.

Suppose $\gamma\leq 0.158655s/108(s+1)^2$. Then we know that $\sigma>1$

Take the construction of $\mathbf{h}_\delta$ in the Proof of Theorem 2.3 withe noise level being σ. Then we know that

(6.83)
$$\begin{aligned}&\|\mathbf{h}_\delta-\mathbf{f}\|\leq\sigma,\\ &\lim_{\delta\to 0^+}\|M(\mathbf{f})-M(\mathbf{h}_\delta)\|^2 \geq \frac{\sum_{k=1}^s\rho_m(\sigma;f_k)^2}{1+s} \geq \left(\frac{\sigma}{\varepsilon}\right)^{\frac{4}{3}}\frac{\sum_{k=1}^s\rho_m(\varepsilon;f_k)^2}{1+s}\\ &\qquad\geq \Phi^{-1}(1-2(s+1)\gamma)^{\frac{4}{3}}\frac{\sum_{k=1}^s\rho_m(\varepsilon;h_{\delta,k})^2}{1+s}\\ &\qquad\geq \Phi^{-1}(1-2(s+1)\gamma)^{\frac{4}{3}}\frac{s}{9(s+1)^2}\omega_m(\varepsilon;\mathbf{h}_\delta)\\ &\qquad\geq \Phi^{-1}(1-2(s+1)\gamma)^{\frac{4}{3}}\frac{s}{9(s+1)^2}\frac{1}{6}R_m(\varepsilon;\mathbf{h}_\delta).\end{aligned}$$

Note that $\frac{\sigma}{\varepsilon}>1$. Hence, there exists $\delta_0>0$, such that for $\delta_0>\delta>0$, we have

$$\|M(\mathbf{f})-M(\mathbf{h}_\delta)\|^2 \geq \frac{s}{9(s+1)^2}\omega_m(\varepsilon;\mathbf{f}) \geq \frac{s}{54(s+1)^2}R_m(\varepsilon;\mathbf{f}). \tag{6.84}$$

Denote event

$$D = \{\|\hat{M} - M(\mathbf{f})\|^2 \geq \frac{s}{108(s+1)^2} R_m(\varepsilon; \mathbf{f})\}. \tag{6.85}$$

Then we know that

$$P_{\mathbf{f}}(D) \leq \gamma \cdot \frac{108(s+1)^2}{s}. \tag{6.86}$$

So

$$P_{\mathbf{h}_\delta}(D) \leq \frac{1}{2}. \tag{6.87}$$

Therefore, we have that

$$\begin{aligned} \mathbb{E}_{\mathbf{h}_\delta}\left(\|\hat{M} - M(\mathbf{h}_\delta)\|^2\right) &\geq \mathbb{E}_{\mathbf{h}_\delta}\left((1 - \frac{1}{\sqrt{2}})^2 \|M(\mathbf{f}) - M(\mathbf{h}_\delta)\|^2 \mathbb{1}\{D^c\}\right) \\ &\geq \frac{3 - 2\sqrt{2}}{4} \|M(\mathbf{f}) - M(\mathbf{h}_\delta)\|^2. \end{aligned} \tag{6.88}$$

From Inequality (6.83), we know that there exists $0 < \delta_1 < \delta_0$, such that for $\delta < \delta_1$, we have

$$\|M(\mathbf{f}) - M(\mathbf{h}_\delta)\|^2 \geq \Phi^{-1}(1 - 2(s+1)\gamma)^{\frac{4}{3}} \frac{s}{55(s+1)^2} R_m(\varepsilon; \mathbf{h}_\delta). \tag{6.89}$$

Hence,

$$\mathbb{E}_{\mathbf{h}_\delta}\left(\|\hat{M} - M(\mathbf{h}_\delta)\|^2\right) \geq \frac{3 - 2\sqrt{2}}{4} \Phi^{-1}(1 - 2(s+1)\gamma)^{\frac{4}{3}} \frac{s}{55(s+1)^2} R_m(\varepsilon; \mathbf{h}_\delta). \tag{6.90}$$

Note that $\Phi^{-1}(1-2(s+1)\gamma)^{\frac{4}{3}} \sim \log(\frac{1}{s\gamma})^{\frac{2}{3}}$ as $\gamma \to 0^+$ and that $\log(\frac{1}{s\gamma})^{\frac{2}{3}} \geq (\log(\frac{1}{\gamma})/\log(s))^{\frac{2}{3}}$ for $\gamma < \frac{1}{3s}$, so we have the statement by taking $\mathbf{f}_1 = \mathbf{h}_\delta$.

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

$$\begin{aligned} \int_{\mathcal{A}} \mathrm{d}Y^1 &= \int_{\mathcal{A}} \mathrm{d}\mathbf{er}(Y^1) + \sum_{i=1}^{s} \Pi_{j\neq i}(A_j - a_j) \int_{a_i}^{A_i} \mathrm{d}\boldsymbol{\pi}_i(Y^1) \\ &= \int_{\mathcal{A}} \mathrm{d}\mathbf{er}(Y^2) + \sum_{i=1}^{s} \Pi_{j\neq i}(A_j - a_j) \int_{a_i}^{A_i} \mathrm{d}\boldsymbol{\pi}_i(Y^2) \\ &= \int_{\mathcal{A}} \mathrm{d}Y^2. \end{aligned} \tag{6.91}$$

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

$$[a_1,A_1],[a_2,A_2],\cdots,[a_s,A_s]\subset[0,1] \text{ and } \mathfrak{B}=[b_1,B_1]\times[b_2,B_2]\times\cdots\times[b_s,B_s],$$

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

$$\begin{aligned}
&COV(\int_{[a_i,A_i]}\mathrm{d}\boldsymbol{\pi}_i(Y),\int_{[a_j,A_j]}\mathrm{d}\boldsymbol{\pi}_j(Y))\\
&=\mathbb{E}\Bigg(\left(\int_{t_i\in[a_i,A_i],\mathbf{t}_{-i}\in[0,1]^{s-1}}\mathrm{d}W-(A_i-a_i)\int_{[0,1]^s}\mathrm{d}W\right)\cdot\\
&\quad\left(\int_{t_j\in[a_j,A_j],\mathbf{t}_{-j}\in[0,1]^{s-1}}\mathrm{d}W-(A_j-a_j)\int_{[0,1]^s}\mathrm{d}W\right)\Bigg)\\
&=0.
\end{aligned}\tag{6.92}$$

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

(6.93)

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

Therefore, we prove the independence.

Now we continue with the sufficiency property. Recalling the Radon-Nikodym derivative calculated in (6.3), we have that for $\mathbf{f},\mathbf{g}\in\mathcal{F}_s$

(6.94)

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

Hence we concludes the proof.

6.8. *Proof of Theorem 3.1.* Recalling Theorem 2.1 and Theorem 2.2, we know that it suffices to prove that

$$
\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z} - Z(\mathbf{f})\|^2\right) \le C_2\sum_{k=1}^{s}\rho_z(\varepsilon;\mathbf{f})^2, \tag{6.95}
$$

for an absolute constant $C_2 > 0$.

Since we have

$$
\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z} - Z(\mathbf{f})\|^2\right) = \sum_{k=1}^{s}\mathbb{E}_{\mathbf{f}}\left(|\hat{Z}_k - Z(f_k)|^2\right), \tag{6.96}
$$

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

$$\mathbb{E}_{\mathbf{f}}\left(|\hat{Z}_k - Z(f_k)|^2\right) \le C_2 \rho_z(\varepsilon; f_k)^2. \tag{6.97}$$

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

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

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

$$\begin{aligned} \mathrm{d}\tilde{\mathbf{v}}^l(t) &= f_k(t)\mathrm{d}t + \sqrt{3}\varepsilon \mathrm{d}W^l, \\ \mathrm{d}\tilde{\mathbf{v}}^r(t) &= f_k(t)\mathrm{d}t + \sqrt{3}\varepsilon \mathrm{d}W^r, \end{aligned} \tag{6.98}$$

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 $j = 1, 2, \dots$:

$$\begin{aligned} (&\tilde{X}_{j,2,k} - \tilde{X}_{j,1,k}, \tilde{X}_{j,3,k} - \tilde{X}_{j,2,k}, \dots, \tilde{X}_{j,2^j,k} - \tilde{X}_{j,2^j-1,k}, \\ &X_{j,2,k} - X_{j,1,k}, X_{j,3,k} - X_{j,2,k}, \dots, X_{j,2^j,k} - X_{j,2^j-1,k}) \end{aligned} \tag{6.99}$$

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 by Cai et al. (2023a), 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 univariate case as in Cai et al. (2023a).

Resort to the proof of that of Theorem 3.1 in Cai et al. (2023a) with the quantities bounding $|\tilde{Z} - Z(f_k)|$ there being replaced by the square of it, we have

$$\mathbb{E}_{\mathbf{f}}\left(|\hat{Z}_k - Z(f_k)|^2\right) \le \mathbb{E}_{f_k}\left(|\tilde{Z} - Z(f_k)|^2\right) \le C_2 \rho_z(\varepsilon; f_k)^2, \tag{6.100}$$

for an absolute constant C_2.

6.9. *Proof of Theorem 3.2.* Recalling the lower bound of $L_{\alpha,z}(\varepsilon;\mathbf{f})$ established in Theorem 2.4 and Proposition 2.1 in Cai et al. (2023a), it suffices to prove the following two two propositions.

PROPOSITION 6.7 (Coverage). *The confidence hyper cube $CI_{z,\alpha}$ defined by (3.12) is an $1-\alpha$ level confidence cube for minimizer.*

PROPOSITION 6.8 (Expected Volume). *For $\alpha \le 0.3$, and confidence hyper cube $CI_{z,\alpha}$ defined by (3.12), we have*

$$\mathbb{E}_{\mathbf{f}}\left(V(CI_{z,\alpha})\right) \le C_3^{\frac{s}{2}} \sum_{k=1}^{s} \rho_z(z_{\alpha/s}\varepsilon; f_k), \tag{6.101}$$

where C_3 is an absolute positive constant.

Note that $\rho_z(z_{\alpha/s}\varepsilon; f_k) \le (2z_{\alpha/s})^{\frac{2}{3}}\rho_z(\varepsilon; f_k)$, so these two propositions lead to the theorem.

6.9.1. *Proof of Proposition 6.7.* By the definition of confidence hyper cube $CI_{z,\alpha}$ in (3.12), its k-th coordinate CI_k only depend Y through $\boldsymbol{\pi}_k(Y)$. So it has mutually independent coordinates. Hence we have

$$P_{\mathbf{f}}(Z(\mathbf{f}) \in CI_{z,\alpha}) = \Pi_{k=1}^{s} P_{\mathbf{f}}(Z(f_k) \in CI_k) \ge \Pi_{k=1}^{s} \inf_{\mathbf{f}\in\mathcal{F}_s} P_{\mathbf{f}}(Z(f_k) \in CI_k). \tag{6.102}$$

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

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

$$\begin{aligned} &P_{\mathbf{f}}(Z(f_k) \notin CI_k) = P_{\mathbf{f}}(\grave{j}_k < \hat{j}(\alpha/s,k)) = \sum_{j=3}^{\infty} \mathbb{E}_{\mathbf{f}}(\mathbb{E}_{\mathbf{f}}(\mathbb{1}\{j < \hat{j}(\alpha/s,k)\}|\mathbf{v}_k^l)\mathbb{1}\{\grave{j}_k = j\}) \\ &\le \sum_{j=3}^{\infty} \mathbb{E}_f(\alpha/s\mathbb{1}\{\grave{j}_k = j\}) \le \alpha/s. \end{aligned} \tag{6.103}$$

The first inequality is due to the distribution in (3.7) 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 $S_p(j,k) \le 0$ or that for the left side is non-positive.

This concludes the proof.

6.9.2. *Proof of Proposition 6.8.* Note that the coordinates of the confidence hyper cube are independent, so we have

$$\mathbb{E}_{\mathbf{f}}\left(V(CI_{z,\alpha})\right) = \Pi_{k=1}^{s}\mathbb{E}_{\mathbf{f}}\left(\|CI_k\|\right), \tag{6.104}$$

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

$$\mathbb{E}_{\mathbf{f}}\left(\|CI_k\|^2\right) \le C_3 \rho_z(z_{\alpha/s}\varepsilon; f_k)^2. \tag{6.105}$$

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.

$$\begin{aligned} \tilde{j}_k &= \min\{j : |\hat{i}_{j,k} - i^*_{j,k}| \ge 2\}, \\ \acute{j}_k &= \min\{j : |\hat{i}_{j,k} - i^*_{j,k}| \ge 5\}, \\ \grave{j}_k &= \min\{j : |\hat{i}_{j,k} - i^*_{j,k}| \ge 7\}. \end{aligned} \tag{6.106}$$

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

$$|\hat{i}_{j,k} - i^*_{j,k}| \ge 2. \tag{6.107}$$

We also introduce a quantity as follow.

$$j^*_k = \min\{j : m_j \le \frac{\rho_z(\varepsilon; f_k)}{4}\}. \tag{6.108}$$

We have

(6.109)

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

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:

(6.110)

$$\{X_{j,i^*_{j,k}+1,k} \geq X_{j,i^*_{j,k}+2,k}\}, \{X_{j,i^*_{j,k}+1,k} \geq X_{j,i^*_{j,k}+3,k}\}, \{X_{j,i^*_{j,k}+1,k} \geq X_{j,i^*_{j,k}+4,k}\},$$
$$\{X_{j,i^*_{j,k}-1,k} \geq X_{j,i^*_{j,k}-2,k}\}, \{X_{j,i^*_{j,k}-1,k} \geq X_{j,i^*_{j,k}-3,k}\}, \{X_{j,i^*_{j,k}-1,k} \geq X_{j,i^*_{j,k}-4,k}\}.$$

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

So we have

(6.111)

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

Now we turn to the second term in Inequality (6.109). 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 $i \le 0$, 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

$$\Xi_{j,k} = \min\{\bar{f}_{j,i^*_{j,k}+2,k} - \bar{f}_{j,i^*_{j,k}+1,k}, \bar{f}_{j,i^*_{j,k}-2,k} - \bar{f}_{j,i^*_{j,k}-1,k}\}. \tag{6.112}$$

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

$$j(\zeta, k) = \min\{j : \Xi_{j,k} \cdot 2^{-\frac{j}{2}} \frac{1}{\sqrt{6}\varepsilon} \le z_\zeta + 1\}. \tag{6.113}$$

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 $j \le j(\alpha/s, k) - 1$ we have

$$\Xi_{j,k} \cdot 2^{-\frac{j}{2}} \frac{1}{\sqrt{6}\varepsilon} \ge 2^{\frac{3(j(\alpha/s,k)-1-j)}{2}} (z_{\alpha/s} + 1) \tag{6.114}$$

With these quantities, we have

$$\begin{aligned} &\sum_{j=3}^{\infty} 2^{-2j} \mathbb{E}_{\mathbf{f}}(\mathbb{1}\{\hat{j}(\alpha/s, k) = j, \acute{j}_k > j\}) \\ &\le 2^{-2j(\alpha/s,k)+1} + \sum_{j=3}^{j(\alpha/s,k)-1} 2^{-2j} \Phi(-(z_{\alpha/s} + 1) \times 2^{\frac{3}{2}(j - j(\alpha/s,k)+1)} + z_{\alpha/s}) \\ &\le 2^{-2j(\alpha/s,k)+1} + 2^{-2j(\alpha/s,k)+2} \Phi(-1) \frac{1}{1 - \Phi(-2\sqrt{2})/\Phi(-1)} \\ &< 3 \cdot 2^{-2j(\alpha/s,k)}. \end{aligned} \tag{6.115}$$

Now we introduce a lemma.

Lemma 6.1. *For $j(\zeta, k)$ defined in (6.113), with $\zeta \le 0.3$ we have*

$$(\frac{6\sqrt{2}(z_\zeta + 1)}{z_\zeta})^{\frac{2}{3}} \rho_z(z_\zeta \varepsilon; f_k) \ge 2^{-j(\zeta,k)}. \tag{6.116}$$

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,

$$\begin{aligned}
\|g_{lo} - f_k\|^2 &\leq (\mu_k - M(f_k))^2 \cdot 3 \cdot 2^{-j_0} \\
&\leq (f_k(t_{j_0,i^*+1}) - f_k(t_{j_0,i^*}) + f_k(t_{j_0,i^*}) - M(f_k))^2 \cdot 3 \cdot 2^{-j_0} \\
&\leq (\bar{f}_{j,i^*+2} - \bar{f}_{j,i^*+1})^2 \cdot 3 \cdot 2^{-j_0} \\
&\leq ((z_\zeta + 1) \cdot 2^{\frac{j_0}{2}} \sqrt{6}\varepsilon)^2 \cdot 3 \cdot 2^{-j_0} \\
&= 6(z_\zeta + 1)^2 \times 3\varepsilon^2.
\end{aligned} \tag{6.117}$$

Therefore,

$$2^{-j_0} \leq \rho_z(3\sqrt{2}(z_\zeta + 1)\varepsilon; f_k) \leq (\frac{6\sqrt{2}(z_\zeta + 1)}{z_\zeta})^{\frac{2}{3}} \rho_z(z_\zeta \varepsilon; f_k). \tag{6.118}$$

The last inequality is due to Proposition 2.1 in Cai et al. (2023a) and that $z_\zeta \geq z_{0.3} = 0.524$

□

Lemma 6.1 combined with Inequality (6.115), and note that $\alpha/s \leq 0.3$ we have

$$\sum_{j=3}^{\infty} 2^{-2j} \mathbb{E}_{\mathbf{f}}(\mathbb{1}\{\hat{j}(\alpha/s, k) = j, \acute{j}_k > j\}) < 136 \rho_z(z_{\alpha/s}\varepsilon; f_k)^2. \tag{6.119}$$

Also note that for $\alpha \leq 0.3$, we have $\rho_z(\varepsilon; f_k) < 2.6 \rho_z(z_{\alpha/s}\varepsilon; f_k)$.
Therefore both terms in Inequality 6.109 are bounded by multiple times $\rho_z(z_{\alpha/s}\varepsilon; f_k)^2$. We conclude the proof.

6.10. *Proof of Theorem 3.3.* Recalling Theorem 2.1 and Theorem 2.2, it suffice to prove

$$E\left((\hat{M} - M(\mathbf{f}))^2\right) \leq C_m \left(\sum_{k=1}^{s} \rho_m(\varepsilon; f_k)\right)^2, \tag{6.120}$$

for an absolute positive constant C_m.
We proceed to prove this.
Recall that $\zeta = \Phi(-2)$.

Note that $Y(1,1,\cdots,1)-Y(0,0,\cdots,0)$, $2^{\hat{j}(\zeta,k)}\bar{X}_{\hat{j}(\zeta,k),i_{F,k},k}$ for $k=1,2,\cdots,s$ are independent. Therefore,

(6.121)
$$\mathbb{E}\left((\hat{M}-M(\mathbf{f}))^2\right) \leq \left(\sqrt{\mathbb{E}(Y(1,1,\cdots,1)-Y(0,0,\cdots,0)-f_0)^2}+\sum_{k=1}^{s}\sqrt{\mathbb{E}\left(2^{\hat{j}(\zeta,k)}\bar{X}_{\hat{j}(\zeta,k),i_{F,k},k}-M(f_k)\right)^2}\right)^2.$$

Recollect the notation

$$\bar{f}_{j,i,k}=2^j\int_{2^{-j}(i-1)}^{2^{-j}\cdot i}f_k(t)\mathrm{d}t. \tag{6.122}$$

Recall that the location procedure, the stopping rule and the definition of $i_{F,k}$ parallel those of univariate case introduced in Cai et al. (2023a), so we have that $\bar{f}_{\hat{j}(\zeta,k),i_{F,k},k}$ has the same distribution with that of $\hat{f}$ in the proof of Theorem 3.3 with f_k being the true function.

Hence we have that

$$\mathbb{E}\left(\bar{f}_{\hat{j}(\zeta,k),i_{F,k},k}-M(f_k)\right)^2 \leq \tilde{C}_m\rho_m(\varepsilon;f_k)^2 \tag{6.123}$$

for all $k\in\{1,2,\cdots,s\}$, where $\tilde{C}_m$ is a positive absolute constant.

Also note that

$$\bar{X}_{\hat{j}(\zeta,k),i_{F,k},k}|(\hat{j}(\zeta,k),i_{F,k})\sim N(\bar{f}_{\hat{j}(\zeta,k),i_{F,k},k},(1-2^{-\hat{j}(\zeta,k)})2^{-\hat{j}(\zeta,k)}\times 3\varepsilon^2).$$

So we have that

(6.124)
$$\begin{aligned}&\mathbb{E}\left(2^{\hat{j}(\zeta,k)}\bar{X}_{\hat{j}(\zeta,k),i_{F,k},k}-M(f_k)\right)^2\\&=\mathbb{E}\left(2^{\hat{j}(\zeta,k)}\bar{X}_{\hat{j}(\zeta,k),i_{F,k},k}-\bar{f}_{\hat{j}(\zeta,k),i_{F,k},k}\right)^2+\mathbb{E}\left(\bar{f}_{\hat{j}(\zeta,k),i_{F,k},k}-M(f_k)\right)^2\\&\leq\mathbb{E}((1-2^{-\hat{j}(\zeta,k)})2^{\hat{j}(\zeta,k)}\times 3\varepsilon^2)+\tilde{C}_m\rho_m(\varepsilon;f_k)^2.\end{aligned}$$

Now we will bound $\mathbb{E}(2^{\hat{j}(\zeta,k)}\times 3\varepsilon^2)$. Note that $\zeta=\Phi(-2)<0.3$, so we have that

(6.125)

$$\mathbb{E}_{\mathbf{f}}(2^{\hat{j}(\zeta,k)}) \leq \sum_{j=1}^{j(\zeta,k)+3} \mathbb{E}_{\mathbf{f}}(\hat{j}(\zeta,k)=j) \times 2^j + \sum_{j=j(\zeta,k)+4}^{\infty} \mathbb{E}_{\mathbf{f}}(\hat{j}(\zeta,k)=j) \times 2^j$$
$$\leq 2^{j(\zeta,k)+4} + \sum_{j=j(\zeta,k)+4}^{\infty} 2^j \Phi(-z_\zeta + \frac{z_\zeta+1}{64})^{j-j(\zeta,k)-4}$$
$$\leq 2^{j(\zeta,k)+4} + 2^{j(\zeta,k)+4} \cdot \frac{1}{1-0.03} \leq \frac{4}{\rho_z(z_\zeta\varepsilon; f_k)} \times 33.$$

The last inequality is due to Lemma 6.3.

Going back to Inequality (6.121) we have that

$$\mathbb{E}\left((\hat{M}-M(\mathbf{f}))^2\right) \leq \left(\varepsilon + \sum_{k=1}^{s}\sqrt{132\frac{3\varepsilon^2}{\rho_z(z_\zeta\varepsilon; f_k)} + \tilde{C}_m\rho_m(\varepsilon; f_k)^2}\right)^2$$
$$\leq \left(\varepsilon + \sum_{k=1}^{s}\sqrt{800+\tilde{C}_m} \times \rho_m(z_\zeta\varepsilon; f_k)\right)^2 \tag{6.126}$$
$$\leq C_m\left(\sum_{k=1}^{s}\rho_m(\varepsilon; f_k)\right)^2.$$

6.11. *Proof of Theorem 3.4.* Recalling the lower bound for $L_{\alpha,m}(\varepsilon;\mathbf{f})$ established in Theorem 2.1 and Theorem 2.2, it suffices to prove the following propositions.

PROPOSITION 6.9 (Coverage). *The confidence interval $CI_{m,\alpha}$ defined by (3.18) is an $1-\alpha$ level confidence cube for minimum.*

PROPOSITION 6.10 (Expected Length). *For $\alpha \leq 0.3$, and confidence interval $CI_{m,\alpha}$ defined by (3.18), we have*

$$\mathbb{E}_{\mathbf{f}}(|CI_{m,\alpha}|) \leq \tilde{C}_{m,s,\alpha}\sum_{k=1}^{s}\rho_m(\varepsilon; f_k), \tag{6.127}$$

where $\tilde{C}_{m,s,\alpha}$ is an absolute positive constant depending on s and α.

6.11.1. *Proof of Proposition 6.9.* Recall that $\zeta = \alpha/4s$. Let the event A_1 be

(6.128)
$$A_1 = \Big\{ Z(f_k) \in [2^{-\hat{j}(\zeta,k)+1} \times \cdot (\hat{i}_{\hat{j}(\zeta,k)-1,k} - 7), 2^{-\hat{j}(\zeta,k)+1} \times \cdot (\hat{i}_{\hat{j}(\zeta,k)-1,k} + 6)] \text{ for all } k \in \{1, 2, \cdots, s\} \Big\}.$$

Then from Theorem 3.2 we know that $P(A_1) \geq 1 - \alpha/4$. Easy calculation shows that A_1 can also be written as

(6.129)
$$A_1 = \{ Z(f_k) \in [2^{-\hat{j}(\zeta,k)-3} \cdot 16(\hat{i}_{\hat{j}(\zeta,k)-1,k} - 7), 2^{-\hat{j}(\zeta,k)-3} \cdot 16(\hat{i}_{\hat{j}(\zeta,k)-1,k} + 6)] \}$$

Let the event $D_{2,k}$ be

$$D_{2,k} = \{ \hat{j}(\alpha/4s, k) \leq j(\alpha/4s, k) - 2 \},$$

where $j(\zeta, k)$ is defined in (6.113). By definition of $j(\zeta, k)$ we know that for $j \leq j(\zeta, k) - 1$

$$\Xi_{j,k} \cdot 2^{-\frac{j}{2}} \frac{1}{\sqrt{6}\varepsilon} > 2^{\frac{3}{2}(j(\zeta,k)-1-j)} (z_{\alpha/4s} + 1). \tag{6.130}$$

Therefore, we have

$$\begin{aligned} &P(D_{2,k} \cap \{ |\hat{i}_{\hat{j}(\zeta,k),k} - i^*_{j(\zeta,k),k}| \leq 4 \}) \\ &\leq \sum_{j=1}^{j(\alpha/4s,k)-1} P\left(\hat{j}(\zeta, k) = j, |\hat{i}_{j,k} - i^*_{j,k}| \leq 4 \right) \\ &\leq \Phi(-z_{\alpha/4s} - 1) \sum_{j=1}^{j(\alpha/4s,k)-1} P(|\hat{i}_{j,k} - i^*_{j,k}| \leq 4). \end{aligned} \tag{6.131}$$

Additionally, recall $\tilde{j}_k$ defined in (6.106), we have

(6.132)
$$\begin{aligned} &P\left(\{ |\hat{i}_{\hat{j}(\zeta,k),k} - i^*_{j,k}| \geq 5, \hat{j}(\zeta, k) \leq j(\alpha/4s, k) - 1 \} \right) \\ &\leq P\left(\tilde{j}_k \leq j(\alpha/4s, k) - 2 \right) \leq 6 \sum_{j=1}^{j(\alpha/4s,k)-2} \Phi(-2^{3\cdot(j(\alpha/4s,k)-1-j)/2}(z_{\alpha/4s} + 1) + z_{\alpha/4s}) \\ &\leq 6 \times \Phi(-z_{\alpha/4s} - 2\sqrt{2}) \times 1.000001. \end{aligned}$$

Therefore, for $\alpha \le 0.3$,

$$\begin{aligned} P(D_{2,k}) &\le \Phi(-z_{\alpha/4s} - 1) + 6.000006 \times \Phi(-z_{\alpha/4s} - 2\sqrt{2}) \\ &\le (\alpha/4s) \times \left(\frac{4}{3} \cdot \exp(-1.5) + 6.000006 \times \frac{4}{3} \exp(-4)\right) \le \alpha/8s. \end{aligned} \tag{6.133}$$

Note that for each k

$$\begin{aligned} &2^{\hat{j}(\zeta,k)+3} \times \bar{X}_{\hat{j}(\zeta,k)+3,i,k} - \int_{t_{\hat{j}(\zeta,k)+3,i-1,k}}^{t_{\hat{j}(\zeta,k)+3,i,k}} f_k(t) \cdot 2^{\hat{j}(\zeta,k)+3} \mathrm{d}t \\ &+ Y(1,1,\cdots,1) - Y(0,0\cdots,0) - f_0 - \sqrt{2}\varepsilon \int_0^1 B_k^1(x)\mathrm{d}x \Big| \hat{j}(\zeta,k) \end{aligned} \tag{6.134}$$

for $i = 1, 2, \cdots, s$ are i.i.d $N(0, 2^{\hat{j}(\zeta,k)+3} \times 3\varepsilon^2)$. And

$$\begin{aligned} &Y(1,1,\cdots,1) - Y(0,0\cdots,0) - f_0 - \\ &\sum_{k=1}^{s} \left(Y(1,1,\cdots,1) - Y(0,0\cdots,0) - f_0 - \sqrt{2}\varepsilon \int_0^1 B_k^1(x)\mathrm{d}x \right) \sim N(0, \varepsilon^2((s-1)^2 + 2s)). \end{aligned} \tag{6.135}$$

Hence we have that

$$P\left(\mathbf{f}_{hi} \le M(\mathbf{f}) \Big| A_1\right) \le \frac{\alpha}{4}. \tag{6.136}$$

Also note that on the event $A_1 \cap D_{2,k}^c$, there is a random variable such that

$$\begin{aligned} &v_k | \hat{j}(\zeta,k) \sim N(0, 3(1 - 2^{-\hat{j}(\zeta,k)-3}) 2^{\hat{j}(\zeta,k)+3} \varepsilon^2), \\ &2^{\hat{j}(\zeta,k)+3} \min_{16\cdot(\hat{i}_{\hat{j}(\zeta,k)-1,k} - 7) < i \le 16\cdot(\hat{i}_{\hat{j}(\zeta,k)-1,k} + 6)} \bar{X}_{\hat{j}(\zeta,k)+3,i,k} \\ &\quad \le M(f_k) + \rho_m(z_\zeta \varepsilon; f_k) + v_k \\ &\quad \le M(f_k) + \sqrt{3}\varepsilon z_\zeta \frac{1}{\sqrt{\rho_z(z_\zeta \varepsilon; f_k)}} + v_k, \end{aligned} \tag{6.137}$$

and $v_1, v_2, \cdots, v_k$ are independent.

Recall Lemma 6.1 and the definition of $D_{2,k}^c$, we have on the event $A_1 \cap D_{2,k}^c$

$$2^{\hat{j}(\zeta,k)+3} \min_{16\cdot(\hat{i}_{\hat{j}(\zeta,k)-1,k} - 7) < i \le 16\cdot(\hat{i}_{\hat{j}(\zeta,k)-1,k} + 6)} \bar{X}_{\hat{j}(\zeta,k)+3,i,k} \le M(f_k) + \sqrt{3}\varepsilon z_\zeta \frac{1}{\sqrt{\rho_z(z_\zeta \varepsilon; f_k)}} + v_k. \tag{6.138}$$

So we have that

$$P\left(\mathbf{f}_{lo} \geq M(\mathbf{f}) \middle| A_1 \cap \left(\cap_{k=1}^s D_{2,k}^c\right)\right) \leq \frac{\alpha}{4}. \tag{6.139}$$

Adding the components, we have

(6.140)

$$P(M(\mathbf{f}) \notin [\mathbf{f}_{lo}, \mathbf{f}_{hi}]) \leq$$
$$P(A_1^c) + \sum_{k=1}^s P(D_{2,k}) + P(\mathbf{f}_{lo} \geq M(\mathbf{f}) \Big| A_1 \cap \left(\cap_{k=1}^s D_{2,k}^c\right)) + P(\mathbf{f}_{hi} \leq M(\mathbf{f}) \Big| A_1) \leq \alpha.$$

6.11.2. *Proof of Proposition 6.10.* As $\hat{j}(\zeta,1), \hat{j}(\zeta,2), \cdots, \hat{j}(\zeta,s)$ based on independent random variables, they are independent. Hence we have

(6.141)

$$\mathbb{E}(|\mathbf{f}_{hi}-\mathbf{f}_{lo}|^2) \leq \left(2\sqrt{6}\varepsilon\left(S_{208,\alpha/8s} + z_{\alpha/4} + 2z_{\alpha/4s} + 2z_{\alpha/8}\right)\sum_{k=1}^s \mathbb{E}(2^{\frac{\hat{j}(z_{\alpha/4s},k)}{2}})\right)^2.$$

Now we will prove the following lemma.

Lemma 6.2. *For $k = 1, 2, \cdots, s$, for $\zeta \leq 0.3$,*

$$\mathbb{E}(2^{\frac{\hat{j}(\zeta,k)}{2}}) \leq 12.7 \times 2^{\frac{j(\zeta,k)}{2}}, \tag{6.142}$$

where $j(\zeta,k)$ is defined in (6.113).

Proof.

(6.143)

$$\begin{aligned}\mathbb{E}_{\mathbf{f}}(2^{\frac{\hat{j}(\zeta,k)}{2}}) &\leq \sum_{j=1}^{j(\zeta,k)+3} \mathbb{E}_{\mathbf{f}}(\hat{j}(\zeta,k) = j) \times 2^{\frac{j}{2}} + \sum_{j=j(\zeta,k)+4}^{\infty} \mathbb{E}_{\mathbf{f}}(\hat{j}(\zeta,k) = j) \times 2^{\frac{j}{2}} \\ &\leq 2^{\frac{j(\zeta,k)+5}{2}} + \sum_{j=j(\zeta,k)+4}^{\infty} 2^{\frac{j}{2}}\Phi(-z_\zeta + \frac{z_\zeta+1}{64})^{j-j(\zeta,k)-4} \\ &\leq 2^{\frac{j(\zeta,k)+5}{2}} + 2^{\frac{j(\zeta,k)+4}{2}} \times 1.74803 \leq 12.7 \times 2^{\frac{j(\zeta,k)}{2}}\end{aligned}$$

□

To bound $2^{\frac{j(\zeta,k)}{2}}$, we continue with another lemma

Lemma 6.3. *For $\zeta \leq 0.3$, and $k = 1, 2, \cdots, s$ we have*

$$2^{-j(\zeta,k)} \geq \frac{1}{4}\rho_z(z_\zeta\varepsilon; f_k). \tag{6.144}$$

Proof. Without loss of generality, assume $f_k(Z(f_k)+\rho_z(z_\zeta\varepsilon;f_k))\leq\rho_m(z_\zeta\varepsilon;f_k)$. Suppose $2^{-j}\leq\frac{1}{4}\rho_z(z_\zeta\varepsilon;f_k)$, then we have that

$$\begin{aligned}&(\bar{f}_{j,i^*_{j,k}+2,k}-\bar{f}_{j,i^*_{j,k}+1,k})\cdot 2^{-\frac{j}{2}}\frac{1}{\sqrt{6}\varepsilon}\\&\leq\rho_m(z_\zeta\varepsilon;f_k)\cdot\frac{1}{2}\sqrt{\rho_z(z_\zeta\varepsilon;f_k)}\frac{1}{\sqrt{6}\varepsilon}\leq\frac{1}{2\sqrt{2}}z_\zeta\leq z_\zeta+1.\end{aligned}\tag{6.145}$$

Therefore, $j\geq j(\zeta,k)$, thus $2^{-j(\zeta,k)}\geq\frac{1}{4}\rho_z(z_\zeta\varepsilon;f_k)$.

□

Combing Lemma 6.2 with Lemma 6.3 and getting back to Inequality (6.141), we have

$$\begin{aligned}&\mathbb{E}(|\mathbf{f}_{hi}-\mathbf{f}_{lo}|^2)\\&\leq\left(2\sqrt{6}\varepsilon\left(S_{208,\alpha/8s}+z_{\alpha/4}+2z_{\alpha/4s}+2z_{\alpha/8}\right)\sum_{k=1}^{s}12.7\times 2\frac{1}{\sqrt{\rho_z(z_{\alpha/4s}\varepsilon;f_k)}}\right)^2\\&\leq\left(8\sqrt{3}\times 12.7\times\left(S_{208,\alpha/8s}+z_{\alpha/4}+2z_{\alpha/4s}+2z_{\alpha/8}\right)\frac{1}{z_{\alpha/4s}}\sum_{k=1}^{s}\rho_m(z_{\alpha/4s}\varepsilon;f_k)\right)^2.\end{aligned}\tag{6.146}$$

Note that

$$\rho_m(z_{\alpha/4s}\varepsilon;f_k)\leq z_{\alpha/4s}\rho_m(\varepsilon;f_k),\tag{6.147}$$

and

$$\mathbb{E}(|\mathbf{f}_{hi}-\mathbf{f}_{lo}|)\leq\sqrt{\mathbb{E}(|\mathbf{f}_{hi}-\mathbf{f}_{lo}|^2)}.\tag{6.148}$$

Therefore, we have the statement.

6.12. *Analysis of Local Minimax Rates for Nonparametric Regression.* In this section, we give lower bounds for the benchmarks defined in (4.2) and (4.3).

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 or minimums. We call this error caused by discretization *discretization error*:

$$\mathfrak{D}_z(\mathbf{f};n)=\sup_{\mathbf{g}\in\mathcal{F}_s}\{\|Z(\mathbf{f})-Z(\mathbf{g})\|^2:\mathbf{f}(\tfrac{\mathbf{i}}{n})=\mathbf{g}(\tfrac{\mathbf{i}}{n})\text{ for all }\mathbf{i}\in\{0,1,\dots,n\}^s\},\tag{6.149}$$

$$\mathfrak{D}_m(\mathbf{f};n)=\sup_{\mathbf{g}\in\mathcal{F}_s}\{|M(\mathbf{f})-M(\mathbf{g})|:\mathbf{f}(\tfrac{\mathbf{i}}{n})=\mathbf{g}(\tfrac{\mathbf{i}}{n})\text{ for all }\mathbf{i}\in\{0,1,\dots,n\}^s\}.\tag{6.150}$$

(6.151)

Note that while the discretization errors are defined for $\mathbf{f} \in \mathcal{F}_s$, they are also well defined for univariate convex functions by setting $s = 1$. With a bit abuse of notation, we use them directly for univariate convex functions as well by plugging in univariate convex function f in the place of the multivariate convex function $\mathbf{f}$.

It's apparent that

(6.152)
$$\tilde{\mathtt{R}}_{z,n}(\sigma;\mathbf{f}) \geq \frac{1}{4}\mathfrak{D}_z(\mathbf{f};n), \tilde{\mathtt{R}}_{m,n}(\sigma;\mathbf{f}) \geq \frac{1}{4}\mathfrak{D}_m(\mathbf{f};n)^2, \tilde{\mathtt{L}}_{m,\alpha,n}(\sigma;\mathbf{f}) \geq (1-2\alpha))\mathfrak{D}_m(\mathbf{f};n).$$

For simplicity of notation, for $\varepsilon > 0$, we define

$$\varphi_z(\varepsilon;f) = \rho_z(\varepsilon;f)\left(1 \wedge \sqrt{n\rho_z(\varepsilon;f)}\right), \text{ for } f \in \mathcal{F}, \tag{6.153}$$

$$\varphi_m(\varepsilon;f) = \rho_m(\varepsilon;f)\left(1 \wedge \sqrt{n\rho_z(\varepsilon;f)}\right), \text{ for } f \in \mathcal{F}. \tag{6.154}$$

Now we state the lower bounds for the benchmarks, whose proof will be given later.

$$\tilde{\mathtt{R}}_{z,n}(\sigma;\mathbf{f}) \geq \left(0.1 \times \tfrac{1}{12s}\textstyle\sum_{k=1}^{s}\varphi_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2\right) \vee \tfrac{\mathfrak{D}_z(\mathbf{f};n)}{4}, \tag{6.155}$$

$$\tilde{\mathtt{L}}_{z,\alpha,n}(\sigma;\mathbf{f}) \geq \tfrac{1-\alpha-\Phi(-z_\alpha+1)}{(12s)^{s/2}}\Pi_{k=1}^{s}\left(\sqrt{\mathfrak{D}_z(f_k;n)} \vee \varphi_z(\tfrac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\right), \tag{6.156}$$

(6.157)
$$\tilde{\mathtt{R}}_{m,n}(\sigma;\mathbf{f}) \geq \left(\frac{1}{180}\sum_{k=1}^{s}\varphi_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2\frac{1}{1+\frac{s}{n}+\sum_{k=1}^{s}2\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}\right) \vee \frac{1}{2}\mathfrak{D}_m(\mathbf{f};n)^2,$$

(6.158)
$$\tilde{\mathtt{L}}_{m,\alpha,n}(\sigma;\mathbf{f}) \geq (1-\alpha-\Phi(-z_\alpha+1))\cdot \left(\frac{1}{3\sqrt{2}}\sqrt{\sum_{k=1}^{s}\varphi_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2}\sqrt{\frac{1}{1+\frac{s}{n}+\sum_{k=1}^{s}2\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}} \vee \mathfrak{D}_m(\mathbf{f};n)\right).$$

Before continue with the proofs of the lower bounds (6.155), (6.156), (6.157), and (6.158) separately, we introduce some quantities and lemmas that will be frequently used.

We introduce a function $l_n(\cdot,\cdot)$. For $f, g \in \mathcal{F}$

$$l_n(f,g) = \sqrt{\frac{\sum_{j=1}^{n}(f(\frac{i}{n}) - g(\frac{i}{n}))^2}{n+1}}. \tag{6.159}$$

l_n can be considered as a discrete L_2 norm of the difference of function f and g.

We also have the following lemma.

LEMMA 6.4. *For $f \in \mathcal{F}$, $\varepsilon > 0$, and $\delta > 0$, there exist $g \in \mathcal{F}$ such that*

$$l_n(f,g) \le \sqrt{6}\varepsilon, \tag{6.160}$$

and that

$$\begin{aligned} &|Z(f) - Z(g)| \ge \rho_z(\varepsilon; f)(1 \wedge \sqrt{2n\rho_z(\varepsilon; f)}) - \delta, \\ &M(g) - M(f) \ge \rho_m(\varepsilon; f)(1 \wedge \sqrt{2n\rho_z(\varepsilon; f)}) - \delta, \\ &g(t) \ge f(t) \text{ for } 0 \le t \le 1, \\ &\frac{1}{n+1}\sum_{i=0}^{n}(g(\frac{i}{n}) - f(\frac{i}{n})) \le l_n(f,g)\sqrt{\frac{1}{n} + 2\rho_z(\varepsilon; f)}. \end{aligned} \tag{6.161}$$

PROOF. Suppose $\eta > 0$ is a small number. For $\mu > 0$, we next define convex function $g_{\eta,\mu}$. Suppose $t_{l,\mu}, t_{r,\mu}$ are left and right end points of $\{t : f(t) \le \mu + M(f)\}$. When $t_{l,\mu} + t_{r,\mu} \ge 2Z(f)$.

$$g_{\eta,\mu}(t) = \max\{f(t), \mu + M(f) + \frac{-\eta}{t_{r,\mu} - t_{l,\mu}}(t - t_{l,\mu})\}. \tag{6.162}$$

When $t_{l,\mu} + t_{r,\mu} \le 2Z(f)$.

$$g_{\eta,\mu}(t) = \max\{f(t), \mu + M(f) + \frac{\eta}{t_{r,\mu} - t_{l,\mu}}(t - t_{r,\mu})\}. \tag{6.163}$$

For $\rho_z(\varepsilon; f) \ge \frac{1}{2n}$, we have

$$l_n(f, g_{\eta,\rho_m(\varepsilon;f)}) \le \sqrt{6}\|f - g\| \le \sqrt{6}\varepsilon, \tag{6.164}$$

for any $\eta > 0$. And we also have that

$$\lim_{\eta \to 0^+} |Z(g_{\eta,\varepsilon}) - Z(f)| \ge \rho_z(\varepsilon; f). \tag{6.165}$$

For $\rho_z(\varepsilon; f) \le \frac{1}{2n}$, we have that

$$l_n(f, g_{\eta,\rho_m(\varepsilon;f)\sqrt{2n\rho_z(\varepsilon;f)}}) \le \sqrt{6}\|f - g\| \le \sqrt{6}\varepsilon, \tag{6.166}$$

for any $\eta > 0$.

$$\lim_{\eta \to 0^+} |Z(g_{\eta,\varepsilon\sqrt{2n\rho_z(\varepsilon;f)}}) - Z(f)| \ge \rho_z(\varepsilon; f)\sqrt{2n\rho_z(\varepsilon; f)}). \tag{6.167}$$

Let $\mu = \rho_m(\varepsilon; f)(1 \wedge \sqrt{2n\rho_z(\varepsilon; f)})$.
Then we have that

(6.168)
$$
\begin{aligned}
& l_n(f, g_{\eta,\mu}) \leq 6\varepsilon^2, \\
& \lim_{\eta \to 0^+} M(g_{\eta,\mu}) - M(f) \geq \rho_m(\varepsilon; f)(1 \wedge \sqrt{2n\rho_z(\varepsilon; f)}), \\
& \lim_{\eta \to 0^+} |Z(g_{\eta,\mu}) - Z(f)| \geq \rho_z(\varepsilon; f)(1 \wedge \sqrt{2n\rho_z(\varepsilon; f)}), \\
& g_{\eta,\mu}(t) \geq f(t) \text{ for all } 0 \leq t \leq 1, \\
& \left(\frac{1}{n+1} \sum_{i=1}^{n} (g_{\eta,\mu}(\frac{i}{n}) - f(\frac{i}{n}))\right) \\
& \qquad \leq l_n(f, g_{\eta,\mu})^2 \frac{|\{i : g_{\eta,\mu}(\frac{i}{n}) > f(\frac{i}{n})\}|}{n+1} \leq l_n(f, g_{\eta,\mu})^2 \frac{2n\rho_z(\varepsilon; f) + 1}{n+1}.
\end{aligned}
$$

Take η small enough gives the statement.

□

Now we continue with analyzing the probability structure of the nonparametric regression setting.

For $\mathbf{f}, \mathbf{g} \in \mathcal{F}_s$, denote the probability distribution under $\mathbf{f}$ as $P_{\mathbf{f}}$ and that under $\mathbf{g}$ as $P_{\mathbf{g}}$. Then for observation $\{y_{\mathbf{i}}\}$, we have

$$
\log\left(\frac{P_{\mathbf{f}}}{P_{\mathbf{g}}}(\{y_{\mathbf{i}}\})\right) = \sum_{\mathbf{i} \in \{0,1,\cdots,n\}^s} \left(\frac{y_{\mathbf{i}}(\mathbf{f}(\mathbf{i}) - \mathbf{g}(\mathbf{i}))}{\sigma^2} + \frac{-\mathbf{f}(\mathbf{i})^2 + \mathbf{g}(\mathbf{i})^2}{2\sigma^2}\right). \tag{6.169}
$$

If we set $\mathbf{f}_\theta = \mathbf{f}\mathbb{1}\{\theta = 1\} + \mathbf{g}\mathbb{1}\{\theta = -1\}$, then we know that
(6.170)
$$
W = \sum_{\mathbf{i} \in \{0,1,\cdots,n\}^s} \frac{y_{\mathbf{i}}(\mathbf{f}(\mathbf{i})) - \mathbf{g}(\mathbf{i}))}{\sigma\sqrt{\sum_{\mathbf{i} \in \{0,1,\cdots,n\}^s} (\mathbf{f}(\mathbf{i})) - \mathbf{g}(\mathbf{i}))^2}} + \frac{-\mathbf{f}(\mathbf{i})^2 + \mathbf{g}(\mathbf{i})^2}{2\sigma\sqrt{\sum_{\mathbf{i} \in \{0,1,\cdots,n\}^s} (\mathbf{f}(\mathbf{i})) - \mathbf{g}(\mathbf{i}))^2}}
$$

is a sufficient statistic for θ, and

$$
W \sim N(\theta \frac{1}{2} \frac{\sqrt{\sum_{\mathbf{i} \in \{0,1,\cdots,n\}^s} (\mathbf{f}(\mathbf{i})) - \mathbf{g}(\mathbf{i}))^2/(n+1)^s}}{\sigma/(n+1)^{\frac{s}{2}}}, 1). \tag{6.171}
$$

6.12.1. *Proof of Inequality* (6.155). Recall Lemma 6.4, take $\varepsilon^2 = \frac{\sigma^2}{6(n+1)^s}\frac{1}{s}$. Take

$$
\delta < 0.001 \min_{1 \leq k \leq s} \rho_z(\varepsilon; f_k) \left(1 \wedge \sqrt{n\rho_z(\varepsilon; f_k)}\right).
$$

Take $g_{k,\delta}$ to be the function satisfying (6.161) in Lemma 6.4 for $f = f_k$. Let

$$h_{k,\delta}(t) = g_{k,\delta}(t) - \frac{1}{n+1}\sum_{i=0}^{n}(g_{k,\delta}(\frac{i}{n}) - f_k(\frac{i}{n})). \tag{6.172}$$

Let

$$\mathbf{h}_\delta(\mathbf{t}) = f_0 + \sum_{k=1}^{s} h_{k,\delta}(t_k). \tag{6.173}$$

It's easy to check $\mathbf{h}_\delta \in \mathcal{F}_s$.

Then Lemma 6.4 together with elementary calculation show that

$$\frac{\sqrt{\sum_{\mathbf{i}\in\{0,1,\cdots,n\}^s}(\mathbf{f}(\mathbf{i})) - \mathbf{g}(\mathbf{i}))^2/(n+1)^s}}{\sigma/(n+1)^{\frac{s}{2}}} \le 1, \tag{6.174}$$

and that

$$\|Z(\mathbf{h}_\delta) - Z(\mathbf{f})\|^2 \ge \sum_{k=1}^{s}\left(\rho_z(\varepsilon; f_k)\left(1 \wedge \sqrt{2n\rho_z(\varepsilon; f_k)}\right) - \delta\right)^2. \tag{6.175}$$

Recall that W defined in (6.171) is sufficient statistic for θ, we have

(6.176)

$$\begin{aligned}\tilde{\mathtt{R}}_{z,n}(\sigma;\mathbf{f}) &\ge \inf_{\hat{Z}}\max\{\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z} - Z(\mathbf{f})\|^2\right), \mathbb{E}_{\mathbf{h}_\delta}\left(\|\hat{Z} - Z(\mathbf{h}_\delta)\|^2\right)\} \ge r_2\|Z(\mathbf{f}) - Z(\mathbf{h}_\delta)\|^2, \\ &\ge r_2\sum_{k=1}^{s}\left(\rho_z(\varepsilon; f_k)\left(1 \wedge \sqrt{2n\rho_z(\varepsilon; f_k)}\right) - \delta\right)^2,\end{aligned}$$

where

$$r_2 = \inf_{\hat{\theta}}\max_{\theta=\pm1}\mathbb{E}_\theta\frac{|\hat{\theta} - \theta|^2}{4},$$

for $W \sim N(\frac{\theta}{2}, 1)$. Elementary calculation shows that $r_2 > 0.1$.

Now we take $\delta \to 0^+$, we have that

$$\begin{aligned}\tilde{\mathtt{R}}_{z,n}(\sigma;\mathbf{f}) &\ge 0.1\sum_{k=1}^{s}\rho_z(\varepsilon; f_k)^2\left(1 \wedge 2n\rho_z(\varepsilon; f_k)\right) \\ &\ge 0.1 \times \frac{1}{12s}\sum_{k=1}^{s}\varphi_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)^2,\end{aligned} \tag{6.177}$$

where the last inequality comes from Proposition 2.1.

Note that $\tilde{\mathtt{R}}_{z,n}(\sigma;\mathbf{f}) \ge \frac{\mathfrak{D}_z(\mathbf{f};n)}{4}$ apparently. We concludes the proof.

6.12.2. *Proof of Inequality* (6.156) . Take $h_{k,\delta}$ constructed in (6.172). Let $\tilde{\delta} < 0.01$ be a small positive number.

Take $f_{k,alt,\tilde{\delta}} \in \mathcal{F}$ satisfying

$$\begin{aligned} & f_{k,alt,\tilde{\delta}}(\frac{i}{n}) = f_k(\frac{i}{n}) \text{ for } 0 \le i \le n, \\ & |Z(f_{k,alt,\tilde{\delta}}) - Z(f_k)| \ge \frac{1}{2}\sqrt{(1-\tilde{\delta})\mathfrak{D}_z(f_k;n)}. \end{aligned} \tag{6.178}$$

Take

$$\begin{aligned} \mathbf{h}_{\delta,\tilde{\delta}}(\mathbf{t}) = f_0 + \sum_k^s \Bigg(& h_{k,\delta}(t_k)\mathbb{1}\{|Z(h_{k,\delta}) - Z(f_k)| \ge |Z(f_{k,alt,\tilde{\delta}}) - Z(f_k)|\} \\ & + f_{k,alt,\tilde{\delta}}(t_k)\mathbb{1}\{|Z(h_{k,\delta}) - Z(f_k)| < |Z(f_{k,alt,\tilde{\delta}}) - Z(f_k)|\}\Bigg). \end{aligned} \tag{6.179}$$

It's easy to check that $\mathbf{h}_{\delta,\tilde{\delta}}\mathcal{F}_s$.

Then we have that

$$\frac{\sqrt{\sum_{\mathbf{i}\in\{0,1,\cdots,n\}^s}(\mathbf{f}(\mathbf{i})) - \mathbf{g}(\mathbf{i}))^2/(n+1)^s}}{\sigma/(n+1)^{\frac{s}{2}}} \le 1, \tag{6.180}$$

and that

$$\|Z(\mathbf{h}_{\delta,\tilde{\delta}})_k - Z(\mathbf{f})_k\| \ge \left(\frac{1}{2}\sqrt{(1-\tilde{\delta})\mathfrak{D}_z(f_k;n)} \vee \sum_{k=1}^s \left(\rho_z(\varepsilon;f_k)\left(1 \wedge \sqrt{2n\rho_z(\varepsilon;f_k)}\right) - \delta\right)\right), \tag{6.181}$$

for $k \in \{1, 2, \cdots, s\}$.

Therefore, we have for $CI_{m,\alpha} \in \mathcal{I}_{m,\alpha,n}(\mathcal{F}_s)$,

$$\begin{aligned} \mathbb{E}_{\mathbf{f}}\left(V(CI_{m,\alpha})\right) \ge & (1-\alpha-\Phi(-z_\alpha+1)) \times \\ & \Pi_{k=1}^s \left(\frac{1}{2}\sqrt{(1-\tilde{\delta})\mathfrak{D}_z(f_k;n)} \vee \sum_{k=1}^s \left(\rho_z(\varepsilon;f_k)\left(1 \wedge \sqrt{2n\rho_z(\varepsilon;f_k)}\right) - \delta\right)\right). \end{aligned} \tag{6.182}$$

Note that $\alpha \le 0.3$ gives $1 - \alpha - \Phi(-z_\alpha + 1) > 0$.

Take $\delta, \tilde{\delta} \to 0^+$, we have

(6.183)

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(V(CI_{m,\alpha})\right)\\
&\geq (1-\alpha-\Phi(-z_\alpha+1))\Pi_{k=1}^s\left(\frac{1}{2}\sqrt{\mathfrak{D}_z(f_k;n)} \vee \left(\rho_z(\varepsilon;f_k)\left(1\wedge\sqrt{2n\rho_z(\varepsilon;f_k)}\right)\right)\right)\\
&\geq (1-\alpha-\Phi(-z_\alpha+1))\Pi_{k=1}^s\left(\frac{1}{2}\sqrt{\mathfrak{D}_z(f_k;n)} \vee \frac{1}{\sqrt{12s}}\varphi_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\right)\\
&\leq (1-\alpha-\Phi(-z_\alpha+1))(12s)^{-\frac{s}{2}}\Pi_{k=1}^s\left(\sqrt{\mathfrak{D}_z(f_k;n)} \vee \varphi_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\right).
\end{aligned}
$$

6.12.3. *Proof of Inequality* (6.157) *and Inequality* (6.158) . Let

(6.184)

$$
\varepsilon_k = \frac{\varphi_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}{\sqrt{\sum_{i=1}^s \varphi_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_i)^2}}\frac{1}{\sqrt{6}}\frac{\sigma}{(n+1)^{\frac{s}{2}}}\frac{1}{1+\frac{s}{n}+\sum_{i=1}^s 2\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_i)}.
$$

Recall Lemma 6.4. Let $\delta = \frac{0.1}{s}\cdot\min_{1\leq k\leq s}\varphi_m(\varepsilon_k;f_k)$. For each $k\in\{1,2,\cdots,s\}$, take $\varepsilon=\varepsilon_k$, and take let $g_{k,\delta}$ be the function g in Lemma 6.4.

Let $\tilde{\delta} < 0.01$ be a small positive number.

Take $f_{k,alt,\tilde{\delta}} \in \mathcal{F}$ satisfying

$$
\begin{aligned}
&f_{k,alt,\tilde{\delta}}(\frac{i}{n}) = f_k(\frac{i}{n}) \text{ for } 0\leq i\leq n,\\
&|M(f_{k,alt,\tilde{\delta}}) - M(f_k)| \geq \frac{1}{2}(1-\tilde{\delta})\mathfrak{D}_m(f_k;n).
\end{aligned}
\tag{6.185}
$$

Let

$$
\mathbf{g}_\delta(\mathbf{t}) = f_0 + \sum_{k=1}^s g_{k,\delta}(t_k). \tag{6.186}
$$

Clearly $\mathbf{g}_\delta \in \mathcal{F}_s$.

With a bit abuse of notation, in this proof let

$$
\Delta_k = \frac{1}{n+1}\sum_{i=0}^n g_{k,\delta}(\frac{i}{n}) - f_k(\frac{i}{n}) \tag{6.187}
$$

Then we have that

$$
\begin{aligned}
&\frac{\sqrt{\sum_{\mathbf{i}\in\{0,1,\cdots,n\}^s}(\mathbf{f}(\mathbf{i}))-\mathbf{g}_\delta(\mathbf{i}))^2/(n+1)^s}}{\sigma/(n+1)^{\frac{s}{2}}}\\
&=\frac{\sqrt{(\sum_{k=1}^s\Delta_k)^2+\sum_{k=1}^s l_n(f_k,g_{k,\delta})-\Delta_k)^2}}{\sigma/(n+1)^{\frac{s}{2}}}\\
&\le\frac{\sqrt{\sum_{k=1}^s l_n(f_k,g_{k,\delta})-\Delta_k)^2}\sqrt{1+\frac{s}{n}+\sum_{k=1}^s 2\rho_z(\varepsilon_i;f_k)}}{\sigma/(n+1)^{\frac{s}{2}}}\\
&\le\frac{\sqrt{\sum_{k=1}^s 6\varepsilon_k^2}\sqrt{1+\frac{s}{n}+\sum_{k=1}^s 2\rho_z(\varepsilon;f_k)}}{\sigma/(n+1)^{\frac{s}{2}}}.\\
&\le 1
\end{aligned}
\tag{6.188}
$$

Also, by Lemma 6.4, we have that

(6.189)

$$
\begin{aligned}
M(\mathbf{g}_\delta)-M(\mathbf{f})&=\sum_{k=1}^s M(g_{k,\delta})-M(f_k)\ge\sum_{k=1}^s\rho_m(\varepsilon_k;f)(1\wedge\sqrt{2n\rho_z(\varepsilon_k;f)})-\delta\\
&\ge\sum_{k=1}^s\sqrt{\frac{1}{3}}\frac{\varepsilon_k}{\sigma/(n+1)^{\frac{s}{2}}}\varphi_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)-\delta\\
&\ge\frac{1}{3\sqrt{2}}\sqrt{\sum_{k=1}^s\varphi_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2}\sqrt{\frac{1}{1+\frac{s}{n}+\sum_{k=1}^s 2\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}}-s\delta.
\end{aligned}
$$

Recall the sufficient statistic W given in (6.171).

So we have

$$
\begin{aligned}
\tilde{\mathtt{R}}_{m,n}(\sigma;\mathbf{f})&\ge\inf_{\hat{M}}\max\{\mathbb{E}_{\mathbf{f}}(|\hat{M}-M(\mathbf{f})|^2),\mathbb{E}_{\mathbf{g}_\delta}(|\hat{M}-M(\mathbf{g}_\delta)|^2)\}\\
&\ge r_2|M(\mathbf{f})-M(\mathbf{g}_\delta)|^2,
\end{aligned}
\tag{6.190}
$$

where

$$
r_2=\inf_{\hat{\theta}}\max_{\theta=\pm1}\mathbb{E}_\theta\frac{|\hat{\theta}-\theta|^2}{4},
$$

for $W\sim N(\frac{\theta}{2},1)$. Elementary calculation shows that $r_2>0.1$.

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

(6.191)

$$
\tilde{\mathtt{R}}_{m,n}(\sigma;\mathbf{f})\ge\frac{1}{180}\sum_{k=1}^s\varphi_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2\frac{1}{1+\frac{s}{n}+\sum_{k=1}^s 2\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}.
$$

It's apparent that $\tilde{\mathtt{R}}_{m,n}(\sigma;\mathbf{f}) \geq \frac{1}{4}\mathfrak{D}_m(\mathbf{f};n)^2$. This concludes the proof of Inequality (6.157). We now turn to the proof of Inequality (6.158) .

Let $\tilde{\delta} < 0.01$ be a small positive number. Then there exist $\tilde{\mathbf{f}}_1, \tilde{\mathbf{f}}_2 \in \mathcal{F}_s$ such that

(6.192)
$$\tilde{\mathbf{f}}_1(\frac{\mathbf{i}}{n}) = \mathbf{f}(\frac{\mathbf{i}}{n}) = \tilde{\mathbf{f}}_2(\frac{\mathbf{i}}{n}) \text{ for } \mathbf{i} \in \{0,1,\cdots,n\}^s, \qquad |M(\tilde{\mathbf{f}}_1) - M(\tilde{\mathbf{f}}_2)| \geq (1-\tilde{\delta})\mathfrak{D}_m(\mathbf{f};n),$$

Suppose $CI_{m,\alpha} \in \mathcal{I}_{m,\alpha,n}(\mathcal{F}_s)$.

It's clear that $CI_{m,\alpha} \in \mathcal{I}_{m,\alpha,n}(\{\mathbf{f}, \mathbf{g}_\delta\})$, $CI_{m,\alpha} \in \mathcal{I}_{m,\alpha,n}(\{\tilde{\mathbf{f}}_2, \tilde{\mathbf{f}}_1\})$. Therefore, we have that

$$\tilde{\mathtt{L}}_{m,\alpha,n}(\sigma;\mathbf{f}) \geq (1-2\alpha)\cdot(1-\tilde{\delta})\mathfrak{D}_m(\mathbf{f};n), \tag{6.193}$$

and that

(6.194)
$$\begin{aligned}
&\tilde{\mathtt{L}}_{m,\alpha,n}(\sigma;\mathbf{f}) \\
&\geq (1-\alpha-\Phi(-z_\alpha+1))\cdot|M(\mathbf{f}) - M(\mathbf{g}_\delta)| \\
&\geq (1-\alpha-\Phi(-z_\alpha+1))\cdot\frac{1}{3\sqrt{2}}\sqrt{\sum_{k=1}^{s}\varphi_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2}\sqrt{\frac{1}{1+\frac{s}{n}+\sum_{k=1}^{s}2\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}} \\
&\qquad - s\delta.
\end{aligned}$$

Letting $\delta, \tilde{\delta} \to 0^+$ gives Inequality (6.158).

6.13. *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.

6.14. *Proof of Theorem 4.1.* We have

$$\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z} - Z(\mathbf{f})\|^2\right) \leq \sum_{k=1}^{s}\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z}_k - Z(f_k)\|^2\right). \tag{6.195}$$

Note that Proposition 2.1 gives

$$\rho_z((z_\zeta+1)\frac{\sqrt{6}\sigma}{\sqrt{n}(n+1)^{\frac{s-1}{2}}};f_k) \leq \left(3\times 4\sqrt{3}\right)^{\frac{2}{3}}\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k) \tag{6.196}$$

for $\zeta \le \Phi(-2)$. Also note that $\mathfrak{D}_z(\mathbf{f};n) = \sum_{k=1}^{s} \mathfrak{D}_z(f_k;n)$.

Recall the lower bound for $\tilde{\mathtt{R}}_{z,n}(\sigma;\mathbf{f})$ given in Inequality (6.155).

So it is sufficient to prove that for $\zeta \le 0.15$ the following holds

$$\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(\|\hat{Z}_k - Z(f_k)\|^2\right) \le \\
&\check{C}_2 \rho_z((z_\zeta+1)\frac{\sqrt{6}\sigma}{\sqrt{n}(n+1)^{\frac{s-1}{2}}}; f_k)^2 \sqrt{n\rho_z((z_\zeta+1)\frac{\sqrt{6}\sigma}{\sqrt{n}(n+1)^{\frac{s-1}{2}}}; f_k) \vee 1} + 2\mathfrak{D}_z(f_k;n),
\end{aligned} \tag{6.197}$$

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

Now we proceed with proving it.

First we introduce a quantity for a general $\zeta > 0$:

$$\begin{aligned}
\xi_k(\zeta) = \sup\Bigg\{\xi : \min\Bigg\{&\sqrt{\xi}\left[f_k(Z(f_k)+\xi) - M(f_k)\right], \\
&\sqrt{\xi}\left[f_k\left(Z(f_k)-\xi\right) - M(f_k)\right]\Bigg\} \times \frac{\sqrt{n}}{\sqrt{6}\sigma/(n+1)^{\frac{s-1}{2}}} \le z_\zeta + 1\Bigg\}.
\end{aligned} \tag{6.198}$$

Then let

$$\mathtt{j}_k(\zeta) = \max\{j : \frac{2^{J-j}}{n} > \xi_k(\zeta)\}. \tag{6.199}$$

We further introduce the following quantities.

$$\begin{aligned}
\mathtt{i}^*_{k,j} &= \max\{i : Z(f_k) \in \left[\frac{2^{J-j}\cdot(i-1)}{n} - \frac{1}{2n}, \frac{2^{J-j}\cdot i}{n} - \frac{1}{2n}\right]\} \\
\tilde{\mathtt{j}}_k &= \min\left(\{j : |\hat{\mathtt{i}}_{k,j} - \mathtt{i}^*_{k,j}| \ge 2\} \cup \infty\right), \\
\acute{\mathtt{j}}_k &= \min\left(\{j : |\hat{\mathtt{i}}_{k,j} - \mathtt{i}^*_{k,j}| \ge 5\} \cup \infty\right), \\
\grave{\mathtt{j}}_k &= \min\left(\{j : |\hat{\mathtt{i}}_{k,j} - \mathtt{i}^*_{k,j}| \ge 7\} \cup \infty\right).
\end{aligned} \tag{6.200}$$

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

LEMMA 6.5. *For $j \le \min\{J, \mathtt{j}_k(\zeta)\}$, we have*

$$\frac{1}{\tilde{\sigma}_{k,j}} \sum_{h=(\mathtt{i}^*_{k,j}+1)2^{J-j}}^{(\mathtt{i}^*_{k,j}+2)2^{J-j}-1} \left(f_k(\frac{h}{n}) - f_k(\frac{h-2^{J-j}}{n})\right) \ge 2^{\frac{3}{2}(\mathtt{j}_k(\zeta)-j)}\left(z_\zeta+1\right), \tag{6.201}$$

and

$$\frac{1}{\tilde{\sigma}_{k,j}} \sum_{h=(\mathtt{i}^*_{k,j}-2)2^{J-j}}^{(\mathtt{i}^*_{k,j}-1)2^{J-j}-1} \left(f_k(\frac{h-2^{J-j}}{n}) - f_k(\frac{h}{n}) \right) \geq 2^{\frac{3}{2}(\mathtt{j}_k(\zeta)-j)} \left(z_\zeta + 1 \right). \tag{6.202}$$

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

$$\begin{aligned} &\mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}+2} \leq \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}+1}, \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}+3} \leq \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}+1}, \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}+4} \leq \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}+1}, \\ &\mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}-2} \leq \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}-1}, \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}-3} \leq \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}-1}, \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}-4} \leq \mathtt{Y}^l_{k,j,\mathtt{i}^*_{k,j}-1}. \end{aligned} \tag{6.203}$$

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

LEMMA 6.6. *Suppose* $\zeta \leq 0.5$*.*

$$\mathbb{E}_{\mathbf{f}} \left(2^{-2\tilde{\mathtt{j}}_k} \mathbb{1}\{\tilde{\mathtt{j}}_k \leq J\} \right) \leq \check{C}_0 2^{-2\mathtt{j}_k(\zeta)} \left(1 \wedge 2^{J-\mathtt{j}_k(\zeta)} \right), \tag{6.204}$$

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

REMARK 6.1. Note that the left hand side of Inequality (6.204) does not depend on ζ, but we state this more general lemma.

LEMMA 6.7. *Suppose* $\zeta \leq 0.5$*.*

$$\mathbb{E}_{\mathbf{f}} \left(2^{-2\check{\mathtt{j}}_k(\zeta)} \mathbb{1}\{\check{\mathtt{j}}_k(\zeta) < \infty\} \mathbb{1}\{\tilde{\mathtt{j}}_k > \check{\mathtt{j}}_k(\zeta)\} \right) \leq \check{C}_0 2^{-2\mathtt{j}_k(\zeta)} \left(1 \wedge 2^{J-\mathtt{j}_k(\zeta)} \right), \tag{6.205}$$

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

LEMMA 6.8. *Suppose* $\zeta \leq 0.5$*.*

$$\begin{aligned} &\mathbb{E}_{\mathbf{f}} \left(|\hat{Z}_k - Z(f_k)|^2 \mathbb{1}\{\check{\mathtt{j}}_k(\zeta) = \infty, \tilde{\mathtt{j}}_k > J\} \right) \\ &\leq 64 \cdot 2^{-2\mathtt{j}_k(\zeta)} \left(1 \wedge 2^{J-\mathtt{j}_k(\zeta)} \right) + 2\mathfrak{D}_z(f_k; n). \end{aligned} \tag{6.206}$$

With these lemmas, we have that

$$\mathbb{E}_{\mathbf{f}} \left(|\hat{Z}_k - Z(f_k)|^2 \right) \leq \check{C}_1 \cdot 2^{-2\mathtt{j}_k(\zeta)} \left(1 \wedge 2^{J-\mathtt{j}_k(\zeta)} \right) + 2\mathfrak{D}_z(f_k; n), \tag{6.207}$$

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

Now we introduce the following lemma about $\xi_k(\zeta)$ and $\mathtt{j}_k(\zeta)$, which immediately concludes the proof of Theorem 4.1.

LEMMA 6.9. *For $\zeta > 0$, we have*

$$2\rho_z((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k) \geq \xi_k(\zeta) \geq \frac{1}{2}\rho_z((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k). \tag{6.208}$$

$$\frac{n+2}{2} \leq 2^J \leq n+1. \tag{6.209}$$

$$2^{-\mathtt{j}_k(\zeta)} \leq \frac{2n}{2^J}\xi_k(\zeta) \leq 8\rho_z((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k). \tag{6.210}$$

6.14.1. *Proof of Lemma 6.6.* A basic property of normal tail bound is that $\frac{\Phi(-2\sqrt{2}x)}{\Phi(-x)}$ decreases with $x > 0$ increasing.

$$\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(2^{-2\tilde{\mathtt{j}}_k}\mathbb{1}\{\tilde{\mathtt{j}}_k \leq J\}\right)\\
&\leq \sum_{j=1}^{J} 2^{-2\mathtt{j}_k(\zeta)} \cdot 2^{-2j+2\mathtt{j}_k(\zeta)}\left(\Phi(-2^{\frac{3}{2}(\mathtt{j}_k(\zeta)-j)}(z_\zeta+1))\mathbb{1}\{j \leq \mathtt{j}_k(\zeta)\} + \mathbb{1}\{j > \mathtt{j}_k(\zeta)\}\right)\\
&\leq \mathbb{1}\{J \leq \mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)} \cdot 2^{-2J+2\mathtt{j}_k(\zeta)}\Phi(-2^{\frac{3}{2}(\mathtt{j}_k(\zeta)-J)}(z_\zeta+1))\frac{1}{1-4\frac{\Phi(-2\sqrt{2})}{\Phi(-1)}}\\
&\quad + \mathbb{1}\{J > \mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)}\left(\frac{1}{1-4\frac{\Phi(-2\sqrt{2})}{\Phi(-1)}} + \frac{1}{3}\right)\\
&\leq \mathbb{1}\{J \leq \mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)} \cdot 2^{J-\mathtt{j}_k(\zeta)}\sup_{x\geq 1} 2x^2\Phi(-x) + 2\cdot\mathbb{1}\{J > \mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)}
\end{aligned} \tag{6.211}$$

Let $\check{C}_0 = \max\{\sup_{x\geq 1} 2x^2\Phi(-x), 2\}$, then we have the lemma.

6.14.2. *Proof of Lemma 6.7.* By our stopping rule, apparently $\check{\mathtt{j}}_k(\zeta) \geq 1$.

(6.212)

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(2^{-2\check{\mathtt{j}}_k(\zeta)}\mathbb{1}\{\check{\mathtt{j}}_k(\zeta)<\infty\}\mathbb{1}\{\tilde{\mathtt{j}}_k>\check{\mathtt{j}}_k(\zeta)\}\right)\\
&=\sum_{j=1}^{J}2^{-2j}\mathbb{E}_{\mathbf{f}}\left(\mathbb{E}_{\mathbf{f}}\left(\mathbb{1}\{\tilde{\mathtt{j}}_k>\check{\mathtt{j}}_k(\zeta)=j\}\middle|\nu^l_{k,i}\right)\right)\\
&\le\sum_{j=1}^{J}2^{-2\mathtt{j}_k(\zeta)}\cdot 2^{-2j+2\mathtt{j}_k(\zeta)}\left(\Phi(-2^{\frac{3}{2}(\mathtt{j}_k(\zeta)-j)}(z_\zeta+1))\mathbb{1}\{j\le\mathtt{j}_k(\zeta)\}+\mathbb{1}\{j>\mathtt{j}_k(\zeta)\}\right)\\
&\le\mathbb{1}\{J\le\mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)}\cdot 2^{-2J+2\mathtt{j}_k(\zeta)}\Phi(-2^{\frac{3}{2}(\mathtt{j}_k(\zeta)-J)}(z_\zeta+1))\frac{1}{1-4\frac{\Phi(-2\sqrt{2})}{\Phi(-1)}}\\
&\quad+\mathbb{1}\{J>\mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)}\left(\frac{1}{1-4\frac{\Phi(-2\sqrt{2})}{\Phi(-1)}}+\frac{1}{3}\right)\\
&\le\mathbb{1}\{J\le\mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)}\cdot 2^{J-\mathtt{j}_k(\zeta)}\sup_{x\ge1}2x^2\Phi(-x)+2\mathbb{1}\{J>\mathtt{j}_k(\zeta)\}2^{-2\mathtt{j}_k(\zeta)}
\end{aligned}
$$

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

6.14.3. *Proof of Lemma 6.8.* Note that $\check{\mathtt{j}}_k(\zeta)=\infty,\tilde{\mathtt{j}}_k>J$ means that

$$\{i:f_k(\frac{i}{n})=\min_{l\in\{0,1,\cdots,n\}}\}\subset\{\hat{\mathtt{i}}_{k,J}-3,\hat{\mathtt{i}}_{k,J}-2,\hat{\mathtt{i}}_{k,J}-1,\hat{\mathtt{i}}_{k,J},\hat{\mathtt{i}}_{k,J}+1\},\tag{6.213}$$

and that

$$Z(f_k)\in[\frac{\hat{\mathtt{i}}_{k,J}-3}{n},\frac{\hat{\mathtt{i}}_{k,J}+1}{n}].\tag{6.214}$$

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

(6.215)

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

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 6.5 we know that

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(|\hat{Z}_k - Z(f_k)|^2 \mathbb{1}\{\check{\mathrm{j}}_k(\zeta) = \infty, \tilde{\mathrm{j}}_k > J\}\right) \\
&\leq 2\mathbb{E}_{\mathbf{f}}\left(|\hat{Z}_k - \frac{i_m}{n}|^2\right) + 2\mathfrak{D}_z(f_k; n) \\
&\leq 2 \times \frac{16}{n^2} \times 4\Phi(-2^{\frac{3}{2}(\mathrm{j}_k(\zeta)-J)}(z_\zeta + 1)) + 2\mathfrak{D}_z(f_k; n) \\
&\leq 128\left(\frac{n+1}{n}\right)^2 2^{-2J}\Phi(-2^{\frac{3}{2}(\mathrm{j}_k(\zeta)-J)}) + 2\mathfrak{D}_z(f_k; n) \\
&\leq 128\left(\frac{n+1}{n}\right)^2 2^{-2\mathrm{j}_k(\zeta)} \cdot 2^{J-\mathrm{j}_k(\zeta)} \cdot 2^3\Phi(-\sqrt{8}) + 2\mathfrak{D}_z(f_k; n) \\
&< 10 \cdot 2^{-2\mathrm{j}_k(\zeta)} \cdot 2^{J-\mathrm{j}_k(\zeta)} + 2\mathfrak{D}_z(f_k; n)
\end{aligned}
\tag{6.216}
$$

Hence we concludes the proof.

6.14.4. *Proof of Lemma 6.9.* 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

$$
\begin{aligned}
&\Delta_{1,k}\Delta_{2,k}^2 \\
&\leq \|f_k - \max\{f_k, M(f_k) + \rho_m((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\}\|^2 \\
&= \left((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}\right)^2.
\end{aligned}
\tag{6.217}
$$

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

$$\Delta_{4,k} = \min\{f_k(Z(f_k) + \Delta_{3,k}), f_k(Z(f_k) - \Delta_{3,k})\} - M(f_k).$$

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

$$\begin{aligned} &\Delta_{3,k}\Delta_{4,k}^2 \\ &\geq \|f_k - \max\{f_k, M(f_k) + \rho_m((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\}\|^2 \\ &= \left((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}\right)^2. \end{aligned} \tag{6.218}$$

6.15. *Proof of Theorem 4.2 .* Note that the coordinates of the hyper cube $CI_{z,\alpha}$ are independence from each other, so the following two propositions are sufficient to give the statement of the theorem.

PROPOSITION 6.11. *For $CI_{k,\alpha}$ defined in (4.14)*

$$\mathbb{E}_{\mathbf{f}}\left(\mathbb{1}\{Z(f_k) \notin CI_{k,\alpha}\}\right) \leq \alpha/s, \tag{6.219}$$

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

PROPOSITION 6.12. *For $CI_{k,\alpha}$ defined in (4.14)*

$$\mathbb{E}_{\mathbf{f}}\left(|t_{k,hi} - t_{k,lo}|^2\right) \leq C_5\rho_z(z_{\alpha/s}\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)^2\left(1 \wedge n\rho_z(z_{\alpha/s}\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\right) + 9\mathfrak{D}_z(f_k; n), \tag{6.220}$$

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

The reason Proposition 6.12 implies the statement of expected volume in Theorem 4.2 is as follows. Proposition (6.12) implies that

$$\mathbb{E}_{\mathbf{f}}\left(|t_{k,hi} - t_{k,lo}|\right) \leq \sqrt{C_5}\cdot(2z_{\alpha/s})\cdot\varphi_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k) + 3\sqrt{\mathfrak{D}_z(f_k; n)}, \tag{6.221}$$

where $\varphi_z(\cdot,\cdot)$ is defined in Equation (6.153). This further gives that

$$\mathbb{E}_{\mathbf{f}}\left(V(CI_{z,\alpha})\right) \leq \left(3 + \sqrt{C_5}\cdot(2z_{\alpha/s})\right)^s \Pi_{k=1}^s \left(\varphi_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k) \vee \sqrt{\mathfrak{D}_z(f_k; n)}\right). \tag{6.222}$$

This combined with the lower bound for $\tilde{\mathtt{L}}_{z,\alpha,n}(\sigma; \mathbf{f})$ given in (6.156) gives the statement about expected volume.

Before we continue with the proofs of the propositions, recall the quantities we defined in Equation (6.200) and (6.199).

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

(6.223)
$$i_{m,l} = \min\{i : f(\frac{i}{n}) = \min_{h\in\{0,1,\cdots,n\}} f(\frac{h}{n})\}, i_{m,r} = \max\{i : f(\frac{i}{n}) = \min_{h\in\{0,1,\cdots,n\}} f(\frac{h}{n})\}.$$

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

$$B_1 = \{i_l \geq i_{m,l} + 1\} \cup \{i_r \leq i_{m,2} - 2\}. \tag{6.224}$$

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 $i_l \leq i_r + 1$.

We recollect the quantities defined in Equations (6.200), (6.199).

6.15.1. *Proof of Proposition 6.11.* The event that $\{Z(f_k) \notin CI_{k,\alpha}\}$ can be partitioned into the followings

(6.225)
$$\begin{aligned}\{Z(f_k) \notin CI_{k,\alpha}\} \subset &\{\grave{\mathtt{j}}_k \leq \hat{\mathtt{j}}_k(\alpha/2s) - 1\} \\ &\cup \left(\{\grave{\mathtt{j}}_k \geq \hat{\mathtt{j}}_k(\alpha/2s), \check{\mathtt{j}}_k(\alpha/2s) = \infty\} \cap B_1\right) \\ &\cup \left(\left(\{\grave{\mathtt{j}}_k \geq \hat{\mathtt{j}}_k(\alpha/2s), \check{\mathtt{j}}_k(\alpha/2s) = \infty\} \cap B_1^c\right) \cap \{Z(f_k) \notin CI_{k,\alpha}\}\right).\end{aligned}$$

We will bound them separately.

(6.226)
$$\mathbb{E}_{\mathbf{f}}\left(\mathbb{1}\{\grave{\mathtt{j}}_k \leq \hat{\mathtt{j}}_k(\alpha/2s) - 1\}\right) \leq \mathbb{E}_{\mathbf{f}}\left(\left(\mathbb{1}\{\mathtt{T}_{k,\grave{\mathtt{j}}_k} \geq \tilde{\sigma}_{k,\grave{\mathtt{j}}_k}(z_{\alpha/2s})\}\middle|\nu^l_{k,\cdot}\right)\right) \leq \alpha/2s.$$

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

(6.227)
$$\begin{aligned}&\mathbb{E}_{\mathbf{f}}\left(\{\grave{\mathtt{j}}_k \geq \hat{\mathtt{j}}_k(\alpha/2s), \check{\mathtt{j}}_k(\alpha/2s) = \infty\} \cap B_1\right) \\ &\leq P(\nu^e_{k,i_{m,l}} - \nu^e_{k,i_{m,l}+1} + \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}}\left(z^3_{k,i_{m,l}} - z^3_{k,i_{m,l}+1}\right) > 2\sqrt{3}\frac{\sigma}{(n+1)^{\frac{s-1}{2}}} z_{\alpha_1}) \\ &\quad + P(\nu^e_{k,i_{m,r}-1} - \nu^e_{k,i_{m,r}} + \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}}\left(z^3_{k,i_{m,r}-1} - z^3_{k,i_{m,r}}\right) < -\frac{2\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}} z_{\alpha_1}) \\ &\leq 2\alpha_1 \leq \alpha/4s.\end{aligned}$$

On the event $\{\grave{\mathrm{j}}_k \geq \hat{\mathrm{j}}_k(\alpha/2s), \check{\mathrm{j}}_k(\alpha/2s) = \infty\} \cap B_1^c$, 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

$$\begin{aligned} &P_{\mathbf{f}}\left(\left(\{\grave{\mathrm{j}}_k \geq \hat{\mathrm{j}}_k(\alpha/2s), \check{\mathrm{j}}_k(\alpha/2s) = \infty\} \cap B_1^c\right) \cap \{Z(f_k) \notin CI_{k,\alpha}\}\right) \\ &\leq \mathbb{E}_{\mathbf{f}}\left(\mathbb{1}\{i_m = i_l = i_r + 1 \leq n-1, t_{k,hi} < Z(f_k)\}\right) \\ &\quad + \mathbb{E}_{\mathbf{f}}\left(\mathbb{1}\{i_m = i_l = i_r + 1 \geq 1, t_{k,lo} > Z(f_k)\}\right). \end{aligned} \tag{6.228}$$

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

$$l_1 : y = f(\frac{i_m}{n}), l_2 : y(t) = f(\frac{i_m+1}{n}) + \frac{f(\frac{i_m+2}{n}) - f(\frac{i_m+1}{n})}{1/n}(t - \frac{i_m+1}{n}). \tag{6.229}$$

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

Basic calculation shows that

$$t_{k,r} = \frac{f_k(\frac{i_m}{n}) - f_k(\frac{i_m+1}{n})}{n(f_k(\frac{i_m+2}{n}) - f_k(\frac{i_m+1}{n}))} + \frac{i_m+1}{n}. \tag{6.230}$$

It is easy to check that the distribution of

$$\begin{aligned} &\Bigg(\nu^e_{k,i_m} - \nu^e_{k,i_m+1} - \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}}\left(z^3_{k,i_m} - z^3_{k,i_m+1} - 2\sqrt{2}z_{\alpha_2}\right), \\ &\nu^e_{k,i_m+2} - \nu^e_{k,i_m+1} - \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}}\left(z^3_{k,i_m+2} - z^3_{k,i_m+1} - 2\sqrt{2}z_{\alpha_2}\right)\Bigg) \end{aligned} \tag{6.231}$$

is the same with the following

$$\begin{aligned} &\Bigg(f_k(\frac{i_m}{n}) + \frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}} \cdot \eta_0 - f_k(\frac{i_m+1}{n}) - \frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}} \cdot \eta_1 + \frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}} \cdot 2z_{\alpha_2}, \\ &f_k(\frac{i_m+2}{n}) + \frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}} \cdot \eta_2 - f_k(\frac{i_m+1}{n}) - \frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}} \cdot \eta_1 + \frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}} \cdot 2z_{\alpha_2}\Bigg), \end{aligned} \tag{6.232}$$

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

$$\begin{aligned}&\mathbb{E}_{\mathbf{f}}\left(\mathbb{1}\{i_m = i_l = i_r + 1 \leq n-1, t_{k,hi} < Z(f_k)\}\right)\\ \leq & P(\eta_0 < -z_{\alpha_2}) + P(\eta_1 > z_{\alpha_2}) + P(\eta_2 < -z_{\alpha_2}) \leq 3\alpha_2 = \frac{\alpha}{8s}.\end{aligned} \tag{6.233}$$

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

$$P_{\mathbf{f}}(Z(f_k) \notin CI_k) \leq \alpha/2s + 2\alpha_1 + 6\alpha_2 = \alpha/s. \tag{6.234}$$

6.15.2. *Proof of Proposition 6.12.*

$$\begin{aligned}&\mathbb{E}_{\mathbf{f}}\left(|CI_k|^2\right)\\ &\leq 26^2\mathbb{E}_{\mathbf{f}}\left(\frac{2^{2J-2\hat{\mathrm{j}}_k(\alpha/2s)}}{n^2}\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s) < \infty, \check{\mathrm{j}}_k(\alpha/2s) < \tilde{\mathrm{j}}_k\}\right)\\ &+ 28^2\mathbb{E}_{\mathbf{f}}\left(\frac{2^{2J-2\tilde{\mathrm{j}}_k}}{n^2}\mathbb{1}\{\tilde{\mathrm{j}}_k \leq \hat{\mathrm{j}}_k(\alpha/2s)\}\right)\\ &+ \mathbb{E}_{\mathbf{f}}\left(|CI_k|^2\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s) = \infty, \tilde{\mathrm{j}}_k > J\}\right)\end{aligned} \tag{6.235}$$

Recall Lemma 6.6, 6.7 and 6.9, we have first two terms being bounded by multiple times $\rho_z((z_{\alpha/2s}+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\left(1 \wedge \sqrt{n\rho_z((z_{\alpha/2s}+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)}\right)$, specifically,

$$\begin{aligned}&E_{\mathbf{f}}\left(|CI_k|^2\right)\\ &\leq \check{C}_3\rho_z((z_{\alpha/2s}+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)^2\left(1 \wedge n\rho_z((z_{\alpha/2s}+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\right)\\ &\quad + \mathbb{E}_{\mathbf{f}}\left(|CI_k|^2\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s) = \infty, \tilde{\mathrm{j}}_k > J\}\right),\end{aligned} \tag{6.236}$$

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

Note that $\frac{z_{\alpha/2s}+1}{z_{\alpha/s}} < 4$, and invoke Proposition 2.1 by Cai et al. (2023a), it suffices to bound the remaining term.

We proceed to bound the remaining term. Note that

(6.237)

$$\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k) \le \left(2\frac{z_{\alpha/8s}}{z_{\alpha/s}}\cdot 4\sqrt{3}\sqrt{\frac{n+1}{n}}\right)^{\frac{2}{3}}\rho_z(z_{\alpha/s}\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k),$$

$$\frac{n+1}{n}\le 2, \qquad \frac{z_{\alpha/8s}}{z_{\alpha/s}} < 4 \text{ for } \alpha \le 0.3.$$

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

Lemma 6.10.

(6.238)

$$\mathbb{E}_{\mathbf{f}}\left(|CI_k|^2\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s)=\infty,\tilde{\mathrm{j}}_k>J\}\right)\le$$

$$\check{C}_4\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)^2\left(1\wedge n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right)+9\mathfrak{D}_z(f_k;n)$$

where $\check{C}_4 > 28^2$ is an absolute constant.

Proof. When

$$\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\ge\frac{1}{n}, \tag{6.239}$$

lemma 6.10 holds.

Now we consider the case that

$$\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)<\frac{1}{n}. \tag{6.240}$$

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

$$\begin{aligned} f_k(\frac{i+1}{n})-f_k(\frac{i}{n}) &\ge \frac{1}{n}\frac{\rho_m(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)}{\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)} \\ &\ge \frac{1}{\sqrt{2}}z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}}\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right)^{-\frac{3}{2}}. \end{aligned} \tag{6.241}$$

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

(6.242)

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

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

$$B_2 = \{i_l \le i_{m,l} - 1\} \cup \{i_r \ge i_{m,r}\}. \tag{6.243}$$

Then we know that

$$\begin{aligned} &P_{\mathbf{f}}(B_2 \cap \{\check{\mathrm{j}}_k(\alpha/2s) = \infty, \tilde{\mathrm{j}}_k > J\}) \\ &\le 28\Phi\left(-\sqrt{2} z_{\alpha/8s}\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\right)^{-\frac{3}{2}} + z_{\alpha_1}\right). \end{aligned} \tag{6.244}$$

On the other hand, for the bad event B_1 defined in (6.224), we have

$$\begin{aligned} &P_{\mathbf{f}}(B_1 \cap \{\check{\mathrm{j}}_k(\alpha/2s) = \infty, \tilde{\mathrm{j}}_k > J\}) \\ &\le \Phi\left(-\sqrt{2} z_{\alpha/8s}\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\right)^{-\frac{3}{2}} - z_{\alpha_1}\right). \end{aligned} \tag{6.245}$$

Note that we have $z_{\alpha/8s} > 1$ for $0 < \alpha \le 1$. Hence we have

(6.246)

$$\begin{aligned} &\mathbb{E}_{\mathbf{f}}\left(|CI_k|^2 \mathbb{1}\{B_1 \cup B_2\} \mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s) = \infty, \tilde{\mathrm{j}}_k > J\}\right) \\ &\le \frac{28^2}{n^2} \times 40\Phi\left(-(\sqrt{2}-1) z_{\alpha/8s}\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\right)^{-\frac{3}{2}}\right) \\ &\le \check{C}_5 \rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)^2 \left(1 \wedge n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\right), \end{aligned}$$

where $\check{C}_5 = 28^2 \times 40 \times \sup_{x>1} x^2 \Phi(-(\sqrt{2}-1)x)$.

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 $i_{m,l} \ne 1$ and $i_{m,l} \ne n-1$.

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

$$\begin{aligned} &\mathbb{E}_{\mathbf{f}}\left(|CI_k|^2 \mathbb{1}\{(B_1 \cup B_2)^c\} \mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s) = \infty, \tilde{\mathrm{j}}_k > J\}\right) \\ &\le \frac{9}{n^2} \le 9\mathfrak{D}_z(f_k; n). \end{aligned} \tag{6.247}$$

Combining with Inequality (6.246), we have lemma 6.10.

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

(6.248)
$$\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(|CI_k|^2 \mathbb{1}\{(B_1\cup B_2)^c\}\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s)=\infty, \tilde{\mathrm{j}}_k > J\}\right)\\
&\le \mathbb{E}_{\mathbf{f}}\left(2(t_{k,hi}-i_m)^2 \mathbb{1}\{(B_1\cup B_2)^c\}\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s)=\infty, \tilde{\mathrm{j}}_k > J, i_m \le n-2\}\right)\\
&\quad + \mathbb{E}_{\mathbf{f}}\left(2(t_{k,lo}-i_m)^2 \mathbb{1}\{(B_1\cup B_2)^c\}\mathbb{1}\{\check{\mathrm{j}}_k(\alpha/2s)=\infty, \tilde{\mathrm{j}}_k > J, i_m \ge 2\}\right).
\end{aligned}$$

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 (6.230), for simplicity of notation, denote

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

we have

$$\begin{aligned}
&\mathbb{E}_{\mathbf{f}}\left(2(t_{k,hi}-i_m)^2\mathbb{1}\{D\}\right)\\
&\le \mathbb{E}_{\mathbf{f}}\left(\left(4(t_{k,hi}-t_{k,r})_+^2 + 4(t_{k,r}-\frac{i_m}{n})^2\right)\mathbb{1}\{D\}\right) \qquad (6.249)\\
&\le 4\mathfrak{D}_z(f_k;n) + 4\mathbb{E}_{\mathbf{f}}\left((t_{k,hi}-t_{k,r})_+^2\mathbb{1}\{D\}\right).
\end{aligned}$$

To bound the second term, we will split event D into $D\cap A$ and $D\cap A^c$, where A is an event define later. We will consider the expectation on these two events.

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 (6.232), 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 upper bounded:

(6.250)
$$\begin{aligned}
&t_{k,hi}-t_{k,r} \le\\
&\frac{\varepsilon\eta_0\left(f_k(\frac{i_m+2}{n})-f_k(\frac{i_m+1}{n})\right)+\varepsilon\eta_1\left(f_k(\frac{i_m}{n})-f_k(\frac{i_m+2}{n})\right)+\varepsilon\eta_2\left(f_k(\frac{i_m+1}{n})-f_k(\frac{i_m}{n})\right)}{n\left(f_k(\frac{i_m+2}{n})-f_k(\frac{i_m+1}{n})+\varepsilon\eta_2-\varepsilon\eta_1+2\varepsilon z_{\alpha_2}\right)\left(f_k(\frac{i_m+2}{n})-f_k(\frac{i_m+1}{n})\right)}\\
&+\frac{2z_{\alpha_2}\varepsilon\left(f_k(\frac{i_m+2}{n})-f_k(\frac{i_m}{n})\right)}{n\left(f_k(\frac{i_m+2}{n})-f_k(\frac{i_m+1}{n})+\varepsilon\eta_2-\varepsilon\eta_1+2\varepsilon z_{\alpha_2}\right)\left(f_k(\frac{i_m+2}{n})-f_k(\frac{i_m+1}{n})\right)}.
\end{aligned}$$

The reason it is not an equation is due to the possibility of upper truncation if $t_{k,hi}$ by $\frac{i_m+1}{n}$

Recall that we define η_0, η_1, η_2 in Equation (6.232).

Now we consider a "good" event

$$A = \{\eta_1 \le \frac{f_k(\frac{i_m+2}{n}) - f_k(\frac{i_m+1}{n})}{6\varepsilon} + \frac{1}{2}\varepsilon z_{\alpha_2}, \eta_2 \ge -\frac{f_k(\frac{i_m+2}{n}) - f_k(\frac{i_m+1}{n})}{6\varepsilon} - \frac{1}{2}\varepsilon z_{\alpha_2}\}. \tag{6.251}$$

Under this good event A, we have

$$f_k(\frac{i_m+2}{n}) - f_k(\frac{i_m+1}{n}) + \varepsilon\eta_2 - \varepsilon\eta_1 + 2\varepsilon z_{\alpha_2} \ge \frac{2}{3}\left(f_k(\frac{i_m+2}{n}) - f_k(\frac{i_m+1}{n})\right) + \varepsilon z_{\alpha_2}. \tag{6.252}$$

Then we have that

$$\begin{aligned} &\mathbb{E}_{\mathbf{f}}\left((t_{k,hi} - t_{k,r})_+^2 \mathbb{1}\{D \cap A\}\right) \\ &\le 4\frac{1}{n^2}\left(\frac{\varepsilon}{\frac{2}{3}\left(f_k(\frac{i_m+2}{n}) - f_k(\frac{i_m+1}{n})\right) + \varepsilon z_{\alpha_2}}\right)^2 \left(1 + 4 + 1 + 16 z_{\alpha_2}^2\right) \\ &\le 4\frac{1}{n^2}\left(\frac{1}{\frac{2}{3}\cdot 2 z_{\alpha/8s}\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\right)^{-\frac{3}{2}} + z_{\alpha/24s}}\right)^2 \left(6 + 16 z_{\alpha/24s}^2\right) \\ &\le \rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)^2 \cdot \left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\right)\left(13.5 + 36\left(\frac{z_{\alpha/24s}}{z_{\alpha/8s}}\right)^2\right). \end{aligned} \tag{6.253}$$

The second inequality is due to Inequality (6.241).

Also note that $\frac{z_{\alpha/24s}}{z_{\alpha/8s}} < 2$ for $\alpha < 1$. Hence we have that

$$\begin{aligned} &\mathbb{E}_{\mathbf{f}}\left((t_{k,hi} - t_{k,r})_+^2 \mathbb{1}\{D \cap A\}\right) \\ &< 86\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)^2 \cdot \left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\right). \end{aligned} \tag{6.254}$$

For event $A^c \cap D$, we have

$$P(A^c \cap D) \le 2\Phi\left(-\frac{z_{\alpha/8s}}{3}\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\right)^{-\frac{3}{2}}\right). \tag{6.255}$$

Therefore we have

(6.256)
$$\mathbb{E}_{\mathbf{f}}\left((t_{k,hi}-t_{k,r})_+^2\mathbb{1}\{D\cap A^c\}\right) \leq 18\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)^2\cdot\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right).$$

Adding up the expection on event $D\cap A^c$ and $D\cap A$ and going back to Inequality (6.249), we have the first term in (6.248) bounded. Using similar arguments, the second term can be bounded by the same bound. So we have

(6.257)
$$\mathbb{E}_{\mathbf{f}}\left(|CI_k|^2\mathbb{1}\{(B_1\cup B_2)^c\}\mathbb{1}\{\check{\mathtt{j}}_k(\alpha/2s)=\infty,\tilde{\mathtt{j}}_k>J\}\right) \leq 8D(f_k;n)+832\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)^2\cdot\left(n\rho_z(z_{\alpha/8s}\frac{2\sqrt{12}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\right).$$

This concludes case 2, thus the proof of the lemma.

□

6.16. *Proof of Theorem 4.3.* Note that $\mathfrak{D}_m(\mathbf{f};n)\geq\sum_{k=1}^s\left(\min\{f_k(\frac{i}{n}):0\leq i\leq n\}-M(f_k)\right)$. Recall the lower bound of $\tilde{\mathtt{L}}_{m,\alpha,n}(\sigma;\mathbf{f})$ given in Equation (6.157). Note that $\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\leq 1$ for all $k\in\{1,2,\cdots,s\}$. Using Cauchy-Schwartz inequality, we know that it suffices to prove that

(6.258)
$$\mathbb{E}\left(\hat{M}-M(\mathbf{f})\right)^2\leq\left(C_m\sum_{k=1}^s\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\left(1\wedge\sqrt{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}\right)+\sum_{k=1}^s\left(\min\{f_k(\frac{i}{n}):0\leq i\leq n\}-M(f_k)\right)\right)^2,$$

for some positive absolute constant C_m.

Now we will prove this statement.

Recall that $\zeta=\Phi(-2)<0.1$.

For simplicity of notation, denote

$$\hat{\mathtt{f}}_{k,i}=\frac{1}{2^{\hat{\mathtt{j}}_k(\zeta)}}\sum_{w=2^{\hat{\mathtt{j}}_k(\zeta)}\cdot(i-1)}^{2^{\hat{\mathtt{j}}_k(\zeta)}\cdot i-1}f_k(\frac{w}{n}).$$

Note that $\{\nu_{k,h}^u : 1 \leq k \leq s, 0 \leq h \leq n, u = l, r, e\}$ are independent. So we have that $2^{\hat{\mathtt{j}}_k(\zeta)-J}\mathtt{Y}^e_{k,\hat{\mathtt{j}}_k(\zeta),\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\zeta)}+2\Delta_k} - \hat{\mathtt{f}}_{k,\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\zeta)}+2\Delta_k}\Big|(\nu^l_{\cdot,\cdot}, \nu^r_{\cdot,\cdot}) \sim N(0, (1-2^{\hat{\mathtt{j}}_k(\zeta)-J})2^{\hat{\mathtt{j}}_k(\zeta)-J} \cdot 3\frac{\sigma^2}{(n+1)^{s-1}})$.

Also recall the independence between $\mathbf{er}(\{y_\mathbf{i}\})$ and $\{\nu_{k,h}^u : 1 \leq k \leq s, 0 \leq h \leq n, u = l, r, e\}$. So we have that

(6.259)

$$
\begin{aligned}
&\mathbb{E}\left(\hat{M} - M(\mathbf{f})\right)^2 \leq \\
&\leq \left(\sqrt{\mathbb{E}\left(\frac{1}{(n+1)^s}\sum_{\mathbf{i}\in\{0,1,2,\cdots,n\}^s}\mathbf{er}(\{y_\mathbf{i}\}) - f_0\right)^2} + \sum_{k=1}^s \sqrt{\mathbb{E}\left(\hat{M}_k - M(f_k)\right)^2}\right)^2 \\
&\leq \left(\sqrt{\mathbb{E}\left(\left(\frac{1}{(n+1)^s}\sum_{\mathbf{i}\in\{0,1,2,\cdots,n\}^s}\mathbf{er}(\{y_\mathbf{i}\})\right) - f_0\right)^2}\right. \\
&\quad + \sum_{k=1}^s \left(\sqrt{\mathbb{E}\left(\left(\hat{M}_k - \hat{\mathtt{f}}_{k,\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\zeta)}+2\Delta_k}\right)^2 \mathbb{1}\{\check{\mathtt{j}}_k(\zeta) < \infty\}\right)}\right. \\
&\qquad + \sqrt{\mathbb{E}\left(\left(\hat{\mathtt{f}}_{k,\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\zeta)}+2\Delta_k} - M(f_k)\right)^2 \mathbb{1}\{\check{\mathtt{j}}_k(\zeta) < \infty\}\right)} \\
&\qquad \left.\left. + \sqrt{\mathbb{E}\left(\mathbb{1}\{\check{\mathtt{j}}_k(\zeta) = \infty\}(\hat{M}_k - M(f_k))\right)^2}\right)\right)^2 \\
&\leq \left(\frac{\sigma}{(n+1)^{\frac{s}{2}}} + \sum_{k=1}^s \left(\sqrt{\frac{3\sigma^2}{(n+1)^{s-1}}}\sqrt{\mathbb{E}(2^{\hat{\mathtt{j}}_k(\zeta)-J}\mathbb{1}\{\check{\mathtt{j}}_k(\zeta) < \infty\})} + \right.\right. \\
&\qquad \sqrt{\mathbb{E}\left(\left(\hat{\mathtt{f}}_{k,\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\zeta)}+2\Delta_k} - M(f_k)\right)^2 \mathbb{1}\{\check{\mathtt{j}}_k(\zeta) < \infty\}\right)} + \\
&\qquad \left.\left.\sqrt{\mathbb{E}\left(\left(\hat{M}_k - M(f_k)\right)^2 \mathbb{1}\{\check{\mathtt{j}}_k(\zeta) = \infty\}\right)}\right)\right)^2.
\end{aligned}
$$

Now we will continue with bounding the terms in Inequality (6.259) separately.

We introduce the following lemma, which we will prove later, to bound the first term in the summation.

Lemma 6.11. *For $\zeta \leq 0.1$, we have*

(6.260)
$$\mathbb{E}(2^{\hat{\mathtt{j}}_k(\zeta)}\mathbb{1}\{\check{\mathtt{j}}_k(\zeta) < \infty\}) \leq 37 \cdot 2^{\mathtt{j}_k(\zeta)}$$

for $k = 1, 2, \cdots, s$, *where* $\mathtt{j}_k(\zeta)$ *is defined in Equation 6.199.*

By definition of $\mathtt{j}_k(\zeta)$, we know that

$$\frac{2^{J-\mathtt{j}_k(\zeta)}}{n} > \xi_k(\zeta). \tag{6.261}$$

By Lemma 6.9, we have that

$$\frac{2^{J-\mathtt{j}_k(\zeta)}}{n} > \xi_k(\zeta) \geq \frac{1}{2}\rho_z((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k). \tag{6.262}$$

Recall that we have $\zeta < 0.1$ (because $\zeta = \Phi(-2)$ here).
This combined with Lemma 6.11 we have that

$$\mathbb{E}(2^{\hat{\mathtt{j}}_k(\zeta)-J}\mathbb{1}\{\check{\mathtt{j}}_k(\zeta) < \infty\}) \leq 37 \cdot 2^{\mathtt{j}_k(\zeta)-J} \leq \frac{148}{n}\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)^2 \cdot \left(\frac{\sigma}{(n+1)^{\frac{s}{2}}}\right)^{-2}. \tag{6.263}$$

The second inequality is due to that

$$\begin{aligned}
&\frac{1}{2}\left((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}\right)^2 \\
&\leq \rho_z((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)\rho_m((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k)^2 \\
&\leq \rho_z((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k) \cdot \left((z_\zeta + 1)\sqrt{6}\sqrt{\frac{n+1}{n}}\right)^2 \rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)^2.
\end{aligned} \tag{6.264}$$

Therefore, we have the first term in the summation in Inequality (6.259) upper bounded, which we summarize into the following lemma.

Lemma 6.12.

$$\begin{aligned}
&\sqrt{\frac{3\sigma^2}{(n+1)^{s-1}}}\sqrt{\mathbb{E}(2^{\hat{\mathtt{j}}_k(\zeta)-J}\mathbb{1}\{\check{\mathtt{j}}_k(\zeta) < \infty\})} \\
&\leq \min\left\{\sqrt{\frac{3\sigma^2}{(n+1)^{s-1}}}, \sqrt{\frac{3 \cdot 148(n+1)}{n}}\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\right\} \\
&\leq \min\left\{\sqrt{\frac{6(n+1)}{n}}\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\sqrt{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)}, \sqrt{\frac{444(n+1)}{n}}\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\right\}.
\end{aligned} \tag{6.265}$$

Now we continue with bounding the second term in the summation in Inequality (6.259).

Note that our localization step and stopping rule for each coordinate parallel that in Cai et al. (2023a), but with noise level $\frac{\sigma}{(n+1)^{\frac{s}{2}}}$. So according to Lemma C.42 and Lemma C.45 in Cai et al. (2023b), we have that

$$\begin{aligned}
&\mathbb{E}\left(\left(\hat{\mathtt{f}}_{k,\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\zeta)}+2\Delta_k} - M(f_k)\right)^2 \mathbb{1}\{\check{\mathtt{j}}_k(\zeta) < \infty\}\right)\\
&\le \min\left\{c_{m2}\rho_m(\frac{\sigma}{(n+1)^{\frac{s-1}{2}}}\frac{1}{\sqrt{n}}; f_k)^2, \check{c}_{m2}\frac{\sigma^2}{(n+1)^{s-1}}\right\} \qquad (6.266)\\
&\le c_m\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)^2\left(1 \wedge n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\right),
\end{aligned}$$

where c_{m2} and $\check{c}_{m2}$ are from Lemma C.42 and C.45 in Cai et al. (2023b), and c_m is an absolute positive constant.

Now we turn to the third term in the summation in Inequality (6.259).

Recall that $\{\nu^e_{k,h}\}$ is independent from $\{\nu^l_{k,h}\}\cup\{\nu^r_{k,h}\}$. Let $\tilde{f}_k = \min_{\hat{\mathtt{i}}_{k,J}-2\le i\le \hat{\mathtt{i}}_{k,J}+2} f_k(\frac{i-1}{n})$. Elementary calculation show that

(6.267)
$$\begin{aligned}
&\mathbb{E}\left(\left(\hat{M}_k - M(f_k)\right)^2 \mathbb{1}\{\check{\mathtt{j}}_k(\zeta) = \infty\}\right)\\
&\le 2\cdot 5\cdot \frac{\sigma^2}{(n+1)^{s-1}} P(\check{\mathtt{j}}_k(\zeta) = \infty) + 2\mathbb{E}\left(\left(\tilde{f}_k - M(f_k)\right)^2 \mathbb{1}\{\check{\mathtt{j}}_k(\zeta) = \infty\}\right).
\end{aligned}$$

Again, note that the localization procedure and stopping rule for each coordinate parallels that in Cai et al. (2023a), by Lemma C.46 and Lemma C.43 in Cai et al. (2023b), we have that

$$\begin{aligned}
&\mathbb{E}\left(\left(\tilde{f}_k - M(f_k)\right)^2 \mathbb{1}\{\check{\mathtt{j}}_k(\zeta) = \infty\}\right)\\
&\le \min\{\check{c}_{m3}^2\frac{\sigma^2}{(n+1)^{s-1}}, c_{m6}\cdot 2\rho_m(\sqrt{\frac{\sigma^2}{(n+1)^s}}; f_k)^2\} \qquad (6.268)\\
&\quad + \left(\min\{f_k(\frac{i}{n}) : 0 \le i \le n\} - M(f_k)\right)^2.
\end{aligned}$$

And by Lemma C.44 in Cai et al. (2023b), we have that

$$\frac{\sigma^2}{(n+1)^{s-1}} P(\check{\mathtt{j}}_k(\zeta) = \infty) \le 64\rho_m(\sqrt{\frac{\sigma^2}{(n+1)^s}}; f_k)^2. \qquad (6.269)$$

Also note that $\frac{\sigma^2}{(n+1)^{s-1}} \le 4\rho_m(\sqrt{\frac{\sigma^2}{(n+1)^s}}; f_k)^2 \cdot n\rho_z(\sqrt{\frac{\sigma^2}{(n+1)^s}}; f_k)$ and that

$$\frac{\sigma}{(n+1)^{\frac{s}{2}}} \le \sqrt{2}\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\sqrt{\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)}.$$

Adding the three parts together, and going back to Inequality (6.259), we have that

(6.270)
$$\mathbb{E}\left(\hat{M} - M(\mathbf{f})\right)^2 \le \left(C_m \sum_{k=1}^{s} \rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\left(1 \wedge \sqrt{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)}\right) + \sum_{k=1}^{s}\left(\min\{f_k(\frac{i}{n}) : 0 \le i \le n\} - M(f_k)\right)\right)^2,$$

where C_m is a positive absolute constant. This concludes the proof of the theorem.

Now we give the proof of Lemma 6.11.

6.16.1. *Proof of Lemma 6.11.* By the definition of $\mathtt{j}_k(\zeta)$, we immediately have the following facts that we summarize into a lemma

LEMMA 6.13. *For $J \ge j \ge \mathtt{j}_k(\zeta) + 5$, we have that*

(6.271)
$$\frac{1}{\tilde{\sigma}_{k,j}} \sum_{h=(\mathtt{i}^*_{k,j}+13)2^{J-j}}^{(\mathtt{i}^*_{k,j}+14)2^{J-j}-1} \left(f_k(\frac{h}{n}) - f_k(\frac{h-2^{J-j}}{n})\right) \le 2^{-2} \times 2^{\frac{3}{2}(5+\mathtt{j}_k(\zeta)-j)}\left(z_\zeta + 1\right),$$

or

(6.272)
$$\frac{1}{\tilde{\sigma}_{k,j}} \sum_{h=(\mathtt{i}^*_{k,j}-14)2^{J-j}}^{(\mathtt{i}^*_{k,j}-13)2^{J-j}-1} \left(f_k(\frac{h-2^{J-j}}{n}) - f_k(\frac{h}{n})\right) \le 2^{-2} \times 2^{\frac{3}{2}(5+\mathtt{j}_k(\zeta)-j)}\left(z_\zeta + 1\right).$$

Therefore, we have that

(6.273)
$$\begin{aligned}&\mathbb{E}(2^{\hat{\mathtt{j}}_k(\zeta)}\mathbb{1}\{\check{\mathtt{j}}_k(\zeta) < \infty\})\\ &\le 2^{\mathtt{j}_k(\zeta)} \le 16 \cdot 2^{\mathtt{j}_k(\zeta)} + \sum_{j=\mathtt{j}_k(\zeta)+5}^{J} 2^j \Phi(-z_\zeta + \frac{z_\zeta+1}{4} \cdot 2^{\frac{3}{2}(5+\mathtt{j}_k(\zeta)-j)}) \le 37 \cdot 2^{\mathtt{j}_k(\zeta)}.\end{aligned}$$

The last inequality is based on elementary calculation.

6.17. *Proof of Theorem 4.4.* Recall the lower bound of $\tilde{\mathtt{L}}_{m,\alpha,n}(\sigma;\mathbf{f})$ given in Inequality (6.158). Using Cauchy-Schwartz inequality, it suffices to prove the following two propositions.

Proposition 6.13 (Coverage). *For $0 < \alpha \le 0.3$, $CI_{m,\alpha}$ defined in (4.25) is a $1-\alpha$ level confidence interval for $M(\mathbf{f})$.*

Proposition 6.14 (Expected Length). *Suppose $\alpha \le 0.3$. For $CI_{m,\alpha}$ defined in (4.25), we have*

(6.274)
$$\mathbb{E}(|CI_{m,\alpha}|) \le \mathfrak{D}_m(\mathbf{f};n)+\bar{C}_{m,\alpha,s}\sum_{k=1}^{s}\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\left(1\wedge\sqrt{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}\right),$$

where

(6.275)
$$\bar{C}_{m,\alpha,s} = \left(2\sqrt{3}S_{210,\alpha/8s}+3(z_{\alpha/4s}+1)\right)\sqrt{8\cdot 148\cdot 2}+\left(\sqrt{3}S_{210,\alpha/8s}+2\right)\cdot 32+ (6+S_{212,\alpha/24s}+z_{\alpha/48s}/\sqrt{2})\cdot 210\cdot\sqrt{3}\cdot 32+z_{\alpha/8}4\sqrt{6},$$

and $\mathfrak{D}_m(\mathbf{f};n)$ is defined in (6.150).

6.17.1. *Proof of Proposition 6.13.* Denote

$$\mathtt{j}_k^* = j_{F,k}\wedge J. \tag{6.276}$$

Note that $\zeta = \alpha/4s$ and recall Theorem 4.15, we have that for the event A_1 defined by

(6.277)
$$A_1 = \{Z(\mathbf{f})_k \in [\frac{2^{J-\hat{\mathtt{j}}_k(\alpha/4s)+1}}{n}\times(\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\alpha/4s)-1}-7)-\frac{1}{2n},\frac{2^{J-\hat{\mathtt{j}}_k(\alpha/4s)+1}}{n}\times(\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\alpha/4s)-1}+6)-\frac{1}{2n}] \text{ for } k=1,2\cdots,s\},$$

its probability satisfies

$$P(A_1)\ge 1-\alpha/4. \tag{6.278}$$

Note that

(6.279)

$$\left\{2^{\mathtt{j}_k^*-J}\left(\mathtt{Y}^e_{k,\mathtt{j}_k^*,i}-\sum_{w=2^{J-\mathtt{j}_k^*}(i-1)}^{i\cdot 2^{J-\mathtt{j}_k^*}-1} f_k(\frac{w}{n})\right)+\sqrt{2}\frac{\sigma}{(n+1)^{\frac{s-1}{2}}}\frac{\sum_{l=0}^n z_{k,l}^1}{n+1}+\frac{1}{(n+1)^s}\sum_{\mathbf{i}}\mathbf{er}(\{y_{\mathbf{i}}\})-f_0 : 0\le i\le n\right\}\Bigg|\left(\hat{\mathtt{j}}_k(\zeta),\hat{\mathtt{i}}_{k,\hat{\mathtt{j}}_k(\zeta)}\right)\overset{i.i.d}{\sim} N(0,2^{\mathtt{j}_k^*-J}\cdot 3\frac{\sigma^2}{(n+1)^{s-1}}),$$

for $i=0,1,2,\cdots,n$. This fact together with the fact that on event A_1,

(6.280)

$$\min_{I_{k,lo}\le i\le I_{k,hi}} 2^{\mathtt{j}_k^*-J}\sum_{w=2^{J-\mathtt{j}_k^*}(i-1)}^{i\cdot 2^{J-\mathtt{j}_k^*}-1} f_k(\frac{w}{n})=\min_{0\le i\le n} 2^{\mathtt{j}_k^*-J}\sum_{w=2^{J-\mathtt{j}_k^*}(i-1)}^{i\cdot 2^{J-\mathtt{j}_k^*}-1} f_k(\frac{w}{n}),$$

gives

(6.281)

$$P\Bigg(\tilde{M}_{k,md}+\frac{1}{(n+1)^s}\sum_{\mathbf{i}}\mathbf{er}(\{y_{\mathbf{i}}\})-f_0-M(f_k)+\sqrt{2}\frac{\sigma}{(n+1)^{\frac{s-1}{2}}}\frac{\sum_{l=0}^n z_{k,l}^1}{n+1}$$
$$\le -S_{210,\alpha/8s}\times\frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}}\times 2^{\frac{\mathtt{j}_k^*-J}{2}}\Bigg|A_1\Bigg)\le \alpha/8s.$$

Also note that $\frac{1}{(n+1)^s}\sum_{\mathbf{i}}\mathbf{er}(\{y_{\mathbf{i}}\})-f_0\sim N(0,\frac{\sigma^2}{(n+1)^s})$, elementary calculation on the remainder terms of $\tilde{M}_{hi}$ gives

$$P\left(\tilde{M}_{hi}\le M(\mathbf{f})|A_1\right)\le\frac{\alpha}{8}+\frac{\alpha}{8}. \tag{6.282}$$

Recollect quantities introduced in (6.200) and (6.199).
Lemma 6.9 and the definition of $\mathtt{j}_k(\zeta)$ gives

$$\frac{2^{J-\mathtt{j}_k(\zeta)}}{n}\le 4\rho_z((z_\zeta+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k).$$

Therefore

(6.283)

$$2\sqrt{3}(z_\zeta+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}\sqrt{\frac{n}{2^{J-\mathtt{j}_k(\zeta)}}}\ge\rho_m((z_\zeta+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k).$$

This means for $j \geq \mathtt{j}_k(\zeta) + 3$,

$$\frac{3\sigma(z_\zeta + 1)}{(n+1)^{\frac{s-1}{2}}}\sqrt{\frac{1}{2^{J-j}}} \geq \rho_m((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k), \tag{6.284}$$

and if further $j \leq J$,

$$\min_{w\in\{-2,-1,0\}} \sum_{h=(\mathtt{i}^*_{k,j}+w)2^{J-j}}^{(\mathtt{i}^*_{k,j}+w+1)2^{J-j}-1} f_k(\frac{h}{n}) \leq M(f_k) + \rho_m((z_\zeta + 1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}}; f_k). \tag{6.285}$$

Now we define an event

$$D_{2,k} = \{\check{\mathtt{j}}_k(\zeta) \leq \mathtt{j}_k(\zeta) - 1\}. \tag{6.286}$$

Lemma 6.5 gives that for $\zeta \leq 0.1$

$$\begin{aligned} &P(D_{2,k}) \leq P(\tilde{\mathtt{j}}_k \leq \mathtt{j}_k(\zeta) - 1) + P(\check{\mathtt{j}}_k(\zeta) \leq \mathtt{j}_k(\zeta) - 1, \tilde{\mathtt{j}}_k \geq \mathtt{j}_k(\zeta)) \\ &\leq 6\Phi(-z_\zeta - 2) \times \frac{1}{1-0.001} + \Phi(-z_\zeta - 2)\frac{1}{1-0.001} \leq \zeta \cdot \frac{7}{1-0.001} \cdot \exp(-4) \cdot \frac{4}{3} \leq 0.5\zeta. \end{aligned} \tag{6.287}$$

Note that $\zeta = \alpha/4s$, hence $P(D_{2,k}) \leq \alpha/8s$ and $P(\cup_{k=1}^s D_{2,k}) \leq \alpha/8$.

Equations (6.279), (6.280), (6.284), (6.285) together with the apparent fact that

$$\min\{v_1, \cdots, v_w\} \leq \max\{v_1, \cdots, v_w\}$$

, we have that

$$\begin{aligned} &P\Bigg(\tilde{M}_{k,lo} + \frac{1}{(n+1)^s}\sum_{\mathbf{i}} \mathbf{er}(\{y_{\mathbf{i}}\}) - f_0 + \sqrt{2}\frac{\sigma}{(n+1)^{\frac{s-1}{2}}}\frac{\sum_{l=0}^n z^1_{k,l}}{n+1} - M(f_k) \geq 0 \\ &\qquad\qquad \Bigg| A_1 \cap D^c_{2,k} \cap \{j_{F,k} \leq J\}\Bigg) \leq \alpha/8s. \end{aligned} \tag{6.288}$$

Now we introduce a lemma.

Lemma 6.14.

$$P\left(\tilde{M}_{k,lo} \geq M(f_k) \middle| A_1 \cap D^c_{2,k} \cap \{j_{F,k} \geq J+1\}\right) \leq \alpha/8s, \tag{6.289}$$

for $k = 1, 2, \cdots, s$.

PROOF. We prove the inequality for any fixed $k \in \{1, 2, \cdots, s\}$. Denote $\delta_i = \nu^e_{k,i} - f_k(\frac{i}{n})$

Note that $\{\nu^e_{\cdot,\cdot}\}$ is independent with $\{\nu^l_{\cdot,\cdot}, \nu^r_{\cdot,\cdot}\}$, elementary calculation show that

(6.290)
$$P(\max\{|\delta_i| : (k_l-1)\vee 0 \le i \le (k_r+2)\wedge n\} \le H \Big| \nu^l_{\cdot,\cdot}, \nu^r_{\cdot,\cdot}) \ge 1-2\cdot\alpha/24s-2\cdot\alpha/48s = 1-\alpha/8s.$$

Denote event

$$B = \max\{|\delta_i| : k_l \vee 0 \le i \le k_r + 2 \wedge n\} \le H$$

.

On event A_1, we know that $\frac{k_l}{n} \le Z(f_k) \le \frac{k_r+1}{n}$.

Recall a geometric fact: for $t \in [i/n, (i+1)/n]$, where $1 \le i \le n-2$, we have that

(6.291)
$$f_k(t) \ge \max\{\frac{f_k(\frac{i}{n}) - f_k(\frac{i-1}{n})}{1/n}(t-\frac{i}{n})+f_k(\frac{i}{n}), \frac{f_k(\frac{i+2}{n}) - f_k\frac{i+1}{n})}{1/n}(t-\frac{i+1}{n})+f_k(\frac{i+1}{n})\}$$

and the right hand side are also attainable for some f_k when $\{f_k(\frac{i}{n}) : i = 0, 1, \cdots, n\}$ are given.

For $0 < t \le 1/n$, we have that

$$f_k(t) \ge \frac{f(2/n) - f(1/n)}{1/n}(t - 1/n) + f(1/n) \tag{6.292}$$

and the right hand side is attainable for some f_k when $\{f_k(\frac{i}{n}) : i = 0, 1, \cdots, n\}$ are given.

For $1 > t \le n - 1/n$, we have that

$$f_k(t) \ge \frac{f((n-2)/n) - f((n-1)/n)}{1/n}(t-(n-1)/n)+f((n-1)/n). \tag{6.293}$$

On event B, we have that

$$h(i) \le \min_{t\in[\frac{i}{n},\frac{i+1}{n}]} f_k(t), \tag{6.294}$$

for $i = t_l, \cdots, t_r$.

Therefore, on event $A_1 \cap B$, we have that

$$\tilde{M}_{k,lo} \le f_k(t). \tag{6.295}$$

Also we have

(6.296)

$$P(A_1 \cap B \big| A_1 \cap D_{2,k}^c \cap \{j_{F,k} \geq J+1\})$$
$$= \mathbb{E}\left(\mathbb{E}(\mathbb{1}\{B\}|\{\nu_{k,\cdot}^l, \nu_{k,\cdot}^r\})\mathbb{1}\{A_1 \cap D_{2,k}^c \cap \{j_{F,k} \geq J+1\}\}\right)/P(A_1 \cap D_{2,k}^c \cap \{j_{F,k} \geq J+1\})$$
$$\geq 1 - \alpha/8s,$$

which gives the statement of the lemma.

□

Write $\tilde{M}_{lo}$ in the form

(6.297)

$$\tilde{M}_{lo} =$$
$$f_0 + \Bigg((|\{k : j_{F,k} \leq J\}| - 1) \cdot \left(f_0 - \frac{1}{(n+1)^s} \sum_{\mathbf{i} \in \{0,1,2,\cdots,n\}^s} \mathbf{er}(\{y_{\mathbf{i}}\}) \right) -$$
$$\sum_{k=1}^{s} \mathbb{1}\{j_{F,k} \leq J\} \sqrt{2} \frac{\sigma}{(n+1)^{\frac{s-1}{2}}} \frac{\sum_{l=0}^{n} z_{k,l}^1}{n+1} - z_{\alpha/8} \cdot 2\sqrt{3} \frac{\sigma}{(n+1)^{\frac{s}{2}}} s \Bigg)$$
$$+ \sum_{k=1}^{s} \Bigg(\tilde{M}_{k,lo} +$$
$$\mathbb{1}\{j_{F,k} \leq J\} \left(\frac{1}{(n+1)^s} \sum_{\mathbf{i} \in \{0,1,2,\cdots,n\}^s} \mathbf{er}(\{y_{\mathbf{i}}\}) - f_0 + \sqrt{2} \frac{\sigma}{(n+1)^{\frac{s-1}{2}}} \frac{\sum_{l=0}^{n} z_{k,l}^1}{n+1} \right) \Bigg),$$

we have

(6.298)

$$
\begin{aligned}
&P\left(\tilde{M}_{lo} > M(\mathbf{f})\Big|A_1 \cap \left(\cap_{k=1}^{s} D_{2,k}^c\right)\right) \\
&\leq P\Bigg(\left(\left|\{k : j_{F,k} \leq J\}\right| - 1\right) \left(f_0 - \frac{1}{(n+1)^s} \sum_{\mathbf{i}\in\{0,1,2,\cdots,n\}^s} \mathbf{er}(\{y_{\mathbf{i}}\})\right) \\
&\qquad - \sum_{k=1}^{s} \mathbb{1}\{j_{F,k} \leq J\}\sqrt{2}\frac{\sigma}{(n+1)^{\frac{s-1}{2}}} \frac{\sum_{l=0}^{n} z_{k,l}^1}{n+1} - z_{\alpha/8}\cdot 2\sqrt{3}\frac{\sigma}{(n+1)^{\frac{s}{2}}} s > 0 \\
&\qquad\qquad \Bigg| A_1 \cap \left(\cap_{k=1}^{s} D_{2,k}^c\right)\Bigg) \\
&\quad + \sum_{k=1}^{s} \Bigg(P\left(\tilde{M}_{k,lo} + \frac{1}{(n+1)^s}\sum_{\mathbf{i}} \mathbf{er}(\{y_{\mathbf{i}}\}) - f_0 - M(f_k) \geq 0 \Big| A_1 \cap D_{2,k}^c \cap \{j_{F,k} \leq J\}\right) \\
&\qquad\qquad \times P\left(A_1 \cap D_{2,k}^c \cap \{j_{F,k} \leq J\} \big| A_1 \cap \left(\cap_{k=1}^{s} D_{2,k}^c\right)\right) \\
&\quad + P\left(\tilde{M}_{k,lo} \geq M(f_k) \Big| A_1 \cap D_{2,k}^c \cap \{j_{F,k} \geq J+1\}\right) \\
&\qquad\qquad \times P\left(A_1 \cap D_{2,k}^c \cap \{j_{F,k} \geq J+1\} \big| A_1 \cap \left(\cap_{k=1}^{s} D_{2,k}^c\right)\right)\Bigg).
\end{aligned}
$$

Inequality (6.288) and Lemma 6.14 gives that the sum of the terms in the summation is upper bounded by $\alpha/8s$ for each k.

For the first term, split it into summation of conditional probability on $A_1 \cap \left(\cap_{k=1}^{s} D_{2,k}^c\right) \cap \{j_{F,k} = j_k : k = 1, 2, \cdots, s\}$ times $P(A_1 \cap \left(\cap_{k=1}^{s} D_{2,k}^c\right) \cap \{j_{F,k} = j_k : k = 1, 2, \cdots, s\} \Big| A_1 \cap \left(\cap_{k=1}^{s} D_{2,k}^c\right))$ for legitimate j. Elementary calculation show that the conditional probability on $A_1 \cap \left(\cap_{k=1}^{s} D_{2,k}^c\right) \cap \{j_{F,k} = j_k : k = 1, 2, \cdots, s\}$ is upper bounded by $\alpha/8$.

Therefore

$$
P\left(\tilde{M}_{lo} > M(\mathbf{f})\Big|A_1 \cap \left(\cap_{k=1}^{s} D_{2,k}^c\right)\right) \leq \alpha/8 + \alpha/8 = \alpha/4.
$$

Therefore,

(6.299)

$$
\begin{aligned}
P(M(\mathbf{f}) \notin [\tilde{M}_{lo}, \tilde{M}_{hi}]) \leq & P(A_1^c) + \sum_{k=1}^{s} P(D_{2,k}) + P\left(\tilde{M}_{lo} > M(\mathbf{f})\Big|A_1 \cap \left(\cap_{k=1}^{s} D_{2,k}^c\right)\right) \\
& + P\left(\tilde{M}_{hi} < M(\mathbf{f})\Big|A_1 \cap \left(\cap_{k=1}^{s} D_{2,k}^c\right)\right) \leq \alpha.
\end{aligned}
$$

6.17.2. *Proof of Proposition 6.14.*

(6.300)

$$\mathbb{E}(\tilde{M}_{hi} - \tilde{M}_{lo}) = z_{\alpha/8}\frac{4\sqrt{3}\sigma}{(n+1)^{\frac{s}{2}}}s + \sum_{k=1}^{s}\mathbb{E}(\tilde{M}_{k,hi} - \tilde{M}_{k,lo})$$
$$\leq z_{\alpha/8}4\sqrt{6}\sum_{k=1}^{s}\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\sqrt{\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)} + \sum_{k=1}^{s}\mathbb{E}(\tilde{M}_{k,hi} - \tilde{M}_{k,lo}).$$

Recall that $\mathfrak{D}_m(\mathbf{f}; n)$ defined in (6.150) also applies to univariate case by setting $s = 1$, more specifically,

(6.301)

$$\mathfrak{D}_m(f_k; n) = \min\{f_k(\frac{i}{n}) : 0 \leq i \leq n\} - \min\{M(h) : h(\frac{i}{n}) = f_k(\frac{i}{n}) \text{ for } 0 \leq i \leq n, h \in \mathcal{F}\}.$$

Then it is easy to see that

$$\mathfrak{D}_m(\mathbf{f}; n) = \sum_{k=1}^{s}\mathfrak{D}_m(f_k; n). \tag{6.302}$$

So it is sufficient to prove that the following holds for any $k \in \{1, 2, \cdots, s\}$

(6.303)

$$\mathbb{E}(\tilde{M}_{k,hi} - \tilde{M}_{k,lo}) \leq \mathfrak{D}_m(f_k; n) + \tilde{C}_{m,\alpha,s}\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\left(1 \wedge \sqrt{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)}\right),$$

where

(6.304)

$$\tilde{C}_{m,\alpha,s} = \left(2\sqrt{3}S_{210,\alpha/8s} + 3(z_{\alpha/4s} + 1)\right)\sqrt{8 \cdot 148 \cdot 2} + \left(\sqrt{3}S_{210,\alpha/8s} + 2\right) \cdot 32 +$$
$$(6 + S_{212,\alpha/24s} + z_{\alpha/48s}/\sqrt{2}) \cdot 210 \cdot \sqrt{3} \cdot 32.$$

This gives the statement of the proposition by taking $\bar{C}_{m,\alpha,s} = z_{\alpha/8}4\sqrt{6} + \tilde{C}_{m,\alpha,s}$.

Next we will prove Inequality (6.303).

We have

$$\begin{aligned}&\mathbb{E}(\tilde{M}_{k,hi} - \tilde{M}_{k,lo}) \leq \\ &\mathbb{E}((\tilde{M}_{k,hi} - \tilde{M}_{k,lo})\mathbb{1}\{j_{F,k} \leq J\}) + \mathbb{E}((\tilde{M}_{k,hi} - \tilde{M}_{k,lo})\mathbb{1}\{j_{F,k} > J\}).\end{aligned} \tag{6.305}$$

For the first term we have

(6.306)
$$\begin{aligned}
&\mathbb{E}((\tilde{M}_{k,hi} - \tilde{M}_{k,lo})\mathbb{1}\{j_{F,k} \le J\}) \\
&= \left(2\sqrt{3}S_{210,\alpha/8s} + 3(z_{\alpha/4s}+1)\right)\frac{\sigma}{(n+1)^{s-1}}\mathbb{E}(2^{\frac{j_{F,k}-J}{2}}\mathbb{1}\{j_{F,k} \le J\}) \\
&\le \left(2\sqrt{3}S_{210,\alpha/8s} + 3(z_{\alpha/4s}+1)\right)\frac{\sigma}{(n+1)^{s-1}}\left(\mathbb{E}(2^{\frac{\hat{\mathrm{j}}_k(\zeta)+3-J}{2}}) \wedge 1\right) \\
&\le \left(2\sqrt{3}S_{210,\alpha/8s} + 3(z_{\alpha/4s}+1)\right)\frac{\sigma}{(n+1)^{s-1}} \\
&\qquad \left(\sqrt{8\cdot\frac{148}{n}\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)^2\cdot\left(\frac{\sigma}{(n+1)^{\frac{s}{2}}}\right)^{-2}} \wedge 1\right) \\
&\le \left(2\sqrt{3}S_{210,\alpha/8s} + 3(z_{\alpha/4s}+1)\right)\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k) \\
&\qquad \left(\sqrt{8\cdot\frac{148(n+1)}{n}} \wedge \sqrt{\frac{2(n+1)}{n}}\sqrt{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}\right).
\end{aligned}$$

The second to last inequality is due to Inequality (6.263).

Let $\tilde{C}_{m,s,\alpha,0} = \left(2\sqrt{3}S_{210,\alpha/8s} + 3(z_{\alpha/4s}+1)\right)\sqrt{8\cdot 148\cdot 2}$, we have

(6.307)
$$\mathbb{E}((\tilde{M}_{k,hi} - \tilde{M}_{k,lo})\mathbb{1}\{j_{F,k} \le J\}) \le \tilde{C}_{m,s,\alpha,0}\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\left(1 \wedge \sqrt{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}\right).$$

Now we turn to the second term in Equation (6.305). We introduce two quantities first.

(6.308)
$$\tilde{f}_k = \min_{(I_{k,lo}-1)\wedge 0 \le i \le (I_{k,hi}-1)\vee n} f_k(\frac{i}{n}), \qquad \tilde{i}_{k,m} = \operatorname*{arg\,min}_{(I_{k,lo}-1)\wedge 0 \le i \le (I_{k,hi}-1)\vee n} f_k(\frac{i}{n}).$$

Note that these two quantities depend on $\{\nu^l_{k,\cdot}, \nu^r_{k,\cdot}\}$.

(6.309)
$$\begin{aligned}
&\mathbb{E}(\left(\tilde{M}_{k,hi} - \tilde{M}_{k,lo}\right)\mathbb{1}\{j_{F,k} > J\}) \\
&\le \mathbb{E}\left(\left(\tilde{M}_{k,hi} - \tilde{f}_k\right)_+ \mathbb{1}\{j_{F,k} > J\}\right) + \mathbb{E}\left(\left(\tilde{f}_k - \tilde{M}_{k,lo}\right)_+ \mathbb{1}\{j_{F,k} > J\}\right).
\end{aligned}$$

Note that

$$\tilde{M}_{k,hi} \le \nu^e_{k,\tilde{i}_{k,m}} + S_{210,\alpha/8s} \times \sqrt{3}\frac{\sigma}{(n+1)^{\frac{s-1}{2}}}, \tag{6.310}$$

hence we have that

(6.311)

$$\mathbb{E}\left(\left(\tilde{M}_{k,hi}-\tilde{f}_k\right)_+\mathbb{1}\{j_{F,k}>J\}\right)\le P(j_{F,k}>J)\left(\frac{\sqrt{3}\sigma}{(n+1)^{\frac{s}{2}}}+S_{210,\alpha/8s}\frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}}\right).$$

Lemma 6.15.

(6.312)

$$\frac{\sigma}{(n+1)^{\frac{s-1}{2}}}P(j_{F,k}>J)\le 32\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\left(1\wedge\sqrt{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}\right).$$

Proof. Recall that $\zeta=\alpha/4s\le 0.25$. According to Lemma 6.13, we know that when $J\ge \mathtt{j}_k(\zeta)+8$,

(6.313)

$$P(j_{F,k}>J)\le \Pi_{j=\mathtt{j}_k(\zeta)+5}^{J-3}\Phi(-z_\zeta+2^{\frac{3}{2}(\mathtt{j}_k(\zeta)+5-j)}\frac{z_{\zeta+1}}{4})<0.4^{J-\mathtt{j}_k(\zeta)-7}.$$

By Lemma 6.9 and the definition of $\mathtt{j}_k(\zeta)$, we have that

(6.314)

$$0.4^{J-\mathtt{j}_k(\zeta)-7}<2^7\cdot 2^{\mathtt{j}_k(\zeta)-J}<2^7\cdot\frac{1}{n\xi_k(\zeta)}\le 2^8\frac{1}{n\rho_z((z_\zeta+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)}$$

When $n\rho_z((z_\zeta+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)\ge 2^8$, we have that

$$2^{\mathtt{j}_k(\zeta)-J+8}<\frac{1}{n\xi_k(\zeta)}\cdot 2^8\le 2^9\cdot\frac{1}{n\rho_z((z_\zeta+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)}\le 2. \tag{6.315}$$

Note that $2^{\mathtt{j}_k(\zeta)-J+8}$ only takes integer value, hence we have $\mathtt{j}_k(\zeta)-J+8\le 0$. Hence

(6.316)

$$\begin{aligned}\frac{\sigma}{(n+1)^{\frac{s-1}{2}}}P(j_{F,k}>J)&\le\sqrt{2}\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\cdot 2^8\frac{1}{\sqrt{n\rho_z((z_\zeta+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k)}}\cdot\sqrt{2}\\&\le 32\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k).\end{aligned}$$

Also, we always have

(6.317)
$$\frac{\sigma}{(n+1)^{\frac{s-1}{2}}}P(j_{F,k}>J) \le \frac{\sigma}{(n+1)^{\frac{s-1}{2}}}$$
$$\le \sqrt{2}\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\sqrt{\frac{n+1}{n}}\sqrt{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}$$
$$\le 32\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\sqrt{\frac{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}{2^8}}$$

Note that when $\sqrt{\frac{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}{2^8}} \ge 1$, we have $n\rho_z((z_\zeta+1)\frac{\sqrt{6}\sigma}{(n+1)^{\frac{s-1}{2}}\sqrt{n}};f_k) \ge 2^8$, in which case we have Inequality (6.316) holds.

So we have

(6.318)
$$\frac{\sigma}{(n+1)^{\frac{s-1}{2}}}P(j_{F,k}>J) \le 32\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\left(1\wedge\sqrt{\frac{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}{2^8}}\right)$$
$$\le 32\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\left(1\wedge\sqrt{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}\right).$$

□

With Lemma 6.15, going back to inequality (6.311), we have

(6.319)
$$\mathbb{E}\left(\left(\tilde{M}_{k,hi}-\tilde{f}_k\right)_+\mathbb{1}\{j_{F,k}>J\}\right)\le$$
$$\left(\sqrt{3}S_{210,\alpha/8s}+2\right)\cdot 32\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)\left(1\wedge\sqrt{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}};f_k)}\right).$$

Now we turn to the second term in Inequality (6.309).

We have the following lemma

LEMMA 6.16. *Let $\mathfrak{D}_m(f_k;n)$ be defined in (6.301). Then we have*

(6.320)
$$\mathbb{E}\left(\left(\tilde{f}_k - \tilde{M}_{k,lo}\right)_+ \mathbb{1}\{j_{F,k} > J\}\right) \leq \mathfrak{D}_m(f_k;n) + (6 + S_{212,\alpha/24s} + z_{\alpha/48s}/\sqrt{2}) \cdot 210 \cdot \sqrt{3} \times 32\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\left(1 \wedge \sqrt{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)}\right)$$

PROOF. We first recall a basic geometry property of univariate convex functions. Suppose f is a convex function. For any $0 \leq i \leq j \leq n$, we have that

$$\min_{i \leq l \leq j}\{f(\frac{l}{n})\} - \min_{\frac{i}{n} \leq t \leq \frac{j}{n}} f(t) \leq \min_{0 \leq l \leq n}\{f(\frac{l}{n})\} - \min_{0 \leq t \leq 1} f(t). \tag{6.321}$$

For $0 \leq i \leq n-1$, we define a reference number $\tilde{h}(i)$, which is the smallest number a function h could achieve on $[i/n, (i+1)/n]$ when it has the same values with f_k on the grid points (i.e $0, 1/n, 2/n, \cdots, 1$).

(6.322)
$$\tilde{h}(i) = \min_{i/n \leq t \leq (i+1)/n} \max\Big\{f_k(\frac{i+1}{n}) + \frac{f_k(\frac{i+2}{n}) - f_k(\frac{i+1}{n})}{1/n}(t - \frac{i+1}{n}), f_k(\frac{i}{n}) + \frac{f_k(\frac{i-1}{n}) - f_k(\frac{i}{n})}{1/n}(t - \frac{i}{n})\Big\},$$

where $f(-1/n) = \infty = f(\frac{n+1}{n})$ and $\infty \times 0$ is set to 0.

Therefore, we have that

(6.323)
$$\begin{aligned}
&\mathbb{E}\left(\left(\tilde{f}_k - \tilde{M}_{k,lo}\right)_+ \mathbb{1}\{j_{F,k} > J\}\right) \\
&\leq \mathbb{E}\left(\mathbb{E}\left((\tilde{f}_k - \min_{t_l \leq i \leq t_r} \tilde{h}(i)) + \sum_{i=k_l}^{k_r}(\tilde{h}(i) - h(i))_+ \Big| \{\nu^r_{\cdot,\cdot}, \nu^l_{\cdot,\cdot}\}\right) \mathbb{1}\{j_{F,k} > J\}\right) \\
&\leq \mathfrak{D}_m(f_k;n) P(j_{F,k} > J) + \mathbb{E}\left(\sum_{i=k_l}^{k_r} \mathbb{E}\left((\tilde{h}(i) - h(i))_+ \Big| \{\nu^r_{\cdot,\cdot}, \nu^l_{\cdot,\cdot}\}\right) \mathbb{1}\{j_{F,k} > J\}\right).
\end{aligned}$$

Now we are left with bounding the second term.

Recollect the notation $\delta_i = \nu^e_{k,i} - f_k(\frac{i}{n})$ for $0 \leq i \leq n$, and $\delta_i = 0$ for $i \notin \{0, 1, \cdots, n\}$.

Elementary calculation shows that

$$(\tilde{h}(i) - h(i))_+ \leq 2|\delta_i| + 2|\delta_{i+1}| + |\delta_{i-1}| + |\delta_{i+2}| + 3H. \tag{6.324}$$

And note that for fixed i , $\delta_{i-1}, \delta_i, \delta_{i+1}, \delta_{i+2}$ are independent from $\{\nu^l_{\cdot,\cdot}, \nu^r_{\cdot,\cdot}\}$.
Also $\delta_i \sim N(0, \frac{n}{n+1}\frac{3\sigma^2}{(n+1)^{s-1}})$.
Therefore, we have that

$$\begin{aligned}
&\mathbb{E}\left(\sum_{i=k_l}^{k_r} \mathbb{E}\left((\tilde{h}(i) - h(i))_+ \middle| \{\nu^r_{\cdot,\cdot}, \nu^l_{\cdot,\cdot}\}\right) \mathbb{1}\{j_{F,k} > J\}\right) \\
&\leq \frac{\sqrt{3}\sigma}{(n+1)^{\frac{s-1}{2}}}(6 + S_{212,\alpha/24s} + z_{\alpha/48s}/\sqrt{2}) \cdot 210 P(j_{F,k} > J) \\
&\leq (6 + S_{212,\alpha/24s} + z_{\alpha/48s}/\sqrt{2}) \cdot 210 \cdot \sqrt{3} \times \\
&\qquad 32\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\left(1 \wedge \sqrt{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)}\right).
\end{aligned} \tag{6.325}$$

The last inequality comes from Lemma 6.15.

This concludes the proof of Lemma 6.16.

□

Now, combining Lemma 6.16, Inequality (6.309), Inequality (6.319) and Inequality (6.307), we have that

$$\mathbb{E}(\tilde{M}_{k,hi} - \tilde{M}_{k,lo}) \leq \mathfrak{D}_m(f_k; n) + \tilde{C}_{m,\alpha,s}\rho_m(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)\left(1 \wedge \sqrt{n\rho_z(\frac{\sigma}{(n+1)^{\frac{s}{2}}}; f_k)}\right), \tag{6.326}$$

where

$$\begin{aligned}
\tilde{C}_{m,\alpha,s} = &\left(2\sqrt{3}S_{210,\alpha/8s} + 3(z_{\alpha/4s} + 1)\right)\sqrt{8 \cdot 148 \cdot 2} + \left(\sqrt{3}S_{210,\alpha/8s} + 2\right) \cdot 32 + \\
&(6 + S_{212,\alpha/24s} + z_{\alpha/48s}/\sqrt{2}) \cdot 210 \cdot \sqrt{3} \cdot 32.
\end{aligned} \tag{6.327}$$

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. (2023a). Estimation and inference for minimizer and minimum of convex functions: Adaptivity, optimality, and uncertainty principles. *The Annals of Statistics, under review.*

Cai, T. T., Chen, R., and Zhu, Y. (2023b). Supplement to "Estimation and Inference for Minimizer and Minimum of Convex Functions: Optimality, Adaptivity, and Uncertainty Principles".

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