SUPPLEMENT TO "ESTIMATION AND INFERENCE FOR MINIMIZER AND MINIMUM OF CONVEX FUNCTIONS: OPTIMALITY, ADAPTIVITY, AND UNCERTAINTY PRINCIPLES"

By T. Tony Cai, Ran Chen, and Yuancheng Zhu

University of Pennsylvania

This is the supplement to the paper *"Estimation and Inference for Minimizer and Minimum of Convex Functions: Optimality, Adaptivity, and Uncertainty Principles"*. It is organized into four sections. Section A presents the simulation results. Section B offers a comparison between our procedures and the methods based on convexity-constrained least squares for the minimizer, along with a discussion of the connections with the classical minimax framework. Section C provides the proofs of the main results, and Section D contains the proofs of supporting technical lemmas.

Contents

APPENDIX A: SIMULATION RESULTS

Our proposed algorithms for non-parametric regression are easy to implement and computationally fast. We implement the algorithms in R and the code is available at `https://github.com/chenrancece/MMCF`. This section presents the numerical results for our algorithms. The data splitting procedure in our proposed algorithm was introduced in the main paper to create independence, which is purely for technical reasons. In simulation, we also include a variant of our method without the data splitting step. That is, the original data set is used in the localization, stopping, and estimation/inference steps. Simulation studies are carried out to examine the numerical performance of the proposed algorithms, including the non-split variant. Comparisons are made with $CLSCI_\alpha$ in (B.2) proposed by Deng et al. (2020) and the CLS estimator for the minimizer.

The simulation studies use 7 test functions with different levels of smoothness around the minimizer, 6 sample sizes ranging from 100 to 50,000, 5 confidence levels for the confidence intervals, and 100 replications. We compared the proposed methods, their non-split variant, and the CLS methods in terms of computational time, average absolute error (for the estimators), and coverage probability and length (for the confidence intervals). We also investigated the relationship with the benchmarks when the benchmarks can be calculated explicitly. The results can be summarized as follows.

- **Computational cost:** Our methods are significantly faster than CLS methods.[1] For small sample sizes, all methods are relatively fast. For $n \geq 5000$, our procedures are at least 10 times faster than the CLS methods for all functions. In many cases, they are more than 100 times faster. This gap is further increased as the sample size grows.

- **Confidence interval for the minimizer:** Our methods achieve the nominal coverage consistently and the empirical lengths are proportional to the benchmark. In comparison, the coverage probability of $CLSCI_\alpha$ can be far below the nominal level for a variety of functions, including functions that are not differentiable at the minimizer or have vanishing second order derivative around the minimizer. For a piecewise linear function such as $100 \cdot |2x - 1|$, $CLSCI_\alpha$ is long and its length remains roughly constant as the sample size increases, while the benchmark goes to zero. [2]
- **Estimation of the minimizer:** The numerical performances of our methods and the CLS estimator are comparable. Interestingly, in the cases where the benchmarks can be calculated explicitly, the performance of the CLS estimator relative to the benchmarks (and our methods) deteriorates with increasing smoothness of the function around the minimizer, while the performance of our estimator remains steady relative to the benchmarks.
- **Estimation and CI for the minimum:** For estimation and inference for the minimum, we are unaware of CLS based procedures that have theoretical guarantees, so we only examined the performance of our methods. The empirical absolute error for estimator and the lengths of the confidence intervals for the minimum exhibit linear relationship with the corresponding benchmarks (when calculable). The nominal coverages of the confidence intervals are achieved in all the settings.

A.1. Experiment Design. To generate the data, we set noise level $\sigma = 1$. We use test functions with different smoothness, minimizer location, and symmetry. We tested on sample sizes 100, 500, 1000, 5000, 10000, and 50000. For inference, we take 5 confidence levels, namely 0.8, 0.9, 0.95, 0.98, and 0.99, which correspond to $\alpha = 0.2, 0.1, 0.05, 0.02, 0.01$. For each test function and each sample size, we performed 100 replicates and calculated averages accordingly

In experiments evaluating our methods' behavior compared with theoretical

[1]This is also supported by complexity analysis. Time complexity of our algorithms is $O(n)$. Time complexity for CLS itself scales as $\mathcal{O}(n^3)$ for generic quadratic programming solvers or $\mathcal{O}(n^2)$ per iteration for first-order methods, according to Simonetto (2021).

[2]The behavior of the CLS based confidence interval is not surprising due to its asymptotic nature of coverage and high dependency on the second-order derivative.

results, we include functions with calculable benchmarks, along with sample sizes facilitating the examination of the relationship, which we will discuss in detail in Section A.3. Now we focus on the general functions and comparison.

We implement and compare three methods, as summarized in Table 1.

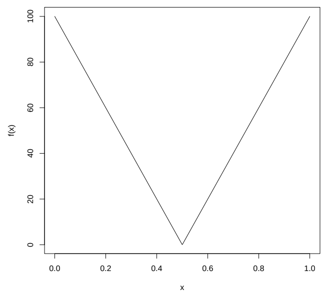
Method	Estimation		Inference	
	Minimizer	Minimum	Minimizer	Minimum
Proposed (split)	✓	✓	✓	✓
Variant (non-split & stop)	✓	✓	✓	✓
CLS based	✓		✓	

TABLE 1
List of the methods to be compared and their applicable scenario.

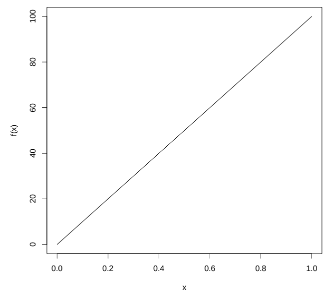
We investigate the following metrics.

- Running time of the methods.
- Empirical risks for estimating the minimizer and minimum.
- Coverage and length of confidence intervals for the minimizer and the minimum. In particular, we construct confidence interval with 5 different confidence levels with α ranging from 0.2 to 0.01.

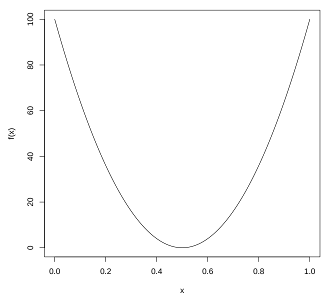
We have 7 test functions, as shown in Equation (A.1). We scale the functions by 100 so that reasonable sample sizes can cover from sample-scarce region to sample-rich region. Figure 1 shows the plots of those functions (in the order $1, 2, 3, 4, 5, 6, 7$ from left top to right bottom), grouped based on smoothness. Note that we include functions of different smoothness around the minimizer (e.g., of the types x, x^2, x^4, $\exp(-1/x)$, represented by f_1, f_3, f_5, f_6), with both symmetric (i.e., f_1, f_3, f_5, f_6) and asymmetric configurations (i.e., f_2, f_4, f_7). We also include functions with the minimizer at boundary (i.e., f_2, f_4). Using similar arguments as in the proof of Proposition B.1, we can convolve the true function with a smooth kernel concentrated enough to the center to have a function that is both smooth (i.e., differentiable to any order) and arbitrarily close to the original true function, regardless of the smoothness of the true function. Therefore, the phenomenon shown here also carries to the non-asymptotic region (i.e., small to medium sample sizes) of functions differentiable to any order.



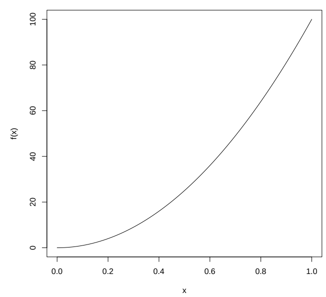
(a) f_1



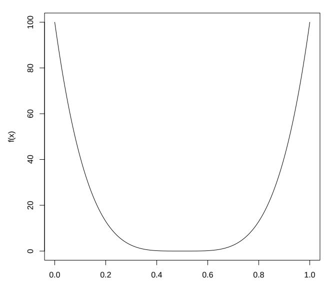
(b) f_2



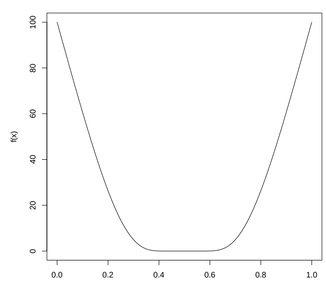
(c) f_3



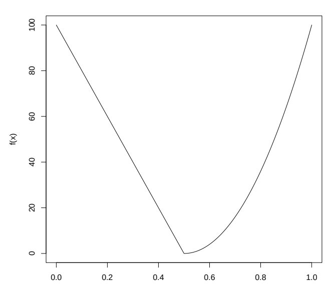
(d) f_4



(e) f_5



(f) f_6



(g) f_7

Fig 1: Plot of True Functions

(A.1)

$$f_1(x) = 100|2x-1|$$

non-differentiable, symmetric, linear

$$f_2(x) = 100x$$

asymmetric, linear, with minimizer at the boundary

$$f_3(x) = 100(|2x-1|)^2$$

twice differentiable with positive second order derivative

$$f_4(x) = 100x^2$$

twice differentiable, asymmetric, with minimizer at the boundary

$$f_5(x) = 100(|2x-1|)^4$$

fourth-order differentiable with vanishing second order derivative

$$f_6(x) = 100\exp\left(2 - \frac{1}{|x-0.5|}\right)$$

arbitrarily differentiable with vanishing derivatives of any order

$$f_7(x) = 100|2x-1|\mathbb{1}\{x < 0.5\} + 100|2x-1|^2\mathbb{1}\{x \geq 0.5\}$$

non-differentiable, non-symmetric.

A.2. Numerical Results and Comparison with CLS Methods. Now we present the simulation results using the 7 test functions. In particular, we compare our methods with the CLS methods for estimation and confidence intervals for the minimizer.

A.2.1. *Results Presentation and Results for Four Tasks.*

Plots and Tables. Before we discuss the results, we explain how we present the results for each function. For each true function, we provide the following plots: the true function, the time vs log sample size plot (for all three methods), the log empirical risk vs log sample size plot for estimation of the minimizer, the log empirical expected length vs log sample size plot for inference of the minimizer, the log empirical risk vs log sample size plot for estimation of the minimum, and the log empirical expected length vs log sample size plot for inference of the minimum. For empirical expected lengths, we plot for $\alpha = 0.01$, other confidence levels are similar. We also provide tables for CLS empirical coverage for the minimizer, log risk for the minimizer, log length for the minimizer for $\alpha = 0.01$, and our non-split version CI's empirical coverage for the minimum. The plots and tables are shown in Figures 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13. In Section A.2.2, we give detailed explanations for each function separately in addition to the task-wise discussion in Section A.2.1.

Estimation of Minimizer. In general, our methods tie with the CLS method for the estimation of the minimizer.

Our methods behave better than CLS for functions with higher smoothness (e.g., the third row in Figure 1: $x^4, \exp(-1/x)$ type). For less smooth functions (e.g., linear, half-side-quadratic), CLS behaves better. For quadratic function with minimizer away from boundary, our methods tie with CLS. This sensitivity to smoothness is due to CLS rather than our methods. We show in Section A.3 that our methods are stable compared to the benchmarks and hence are insensitive to the smoothness of the true functions.

Inference for Minimizer. For the inference of minimizer, both our methods achieve the nominal coverage (empirical coverages are at 0.99 or 1). CLS confidence interval does not achieve nominal coverage consistently. For all the functions except the linear functions and the quadratic function with minimizer at the middle, the CLS confidence interval misses the nominal coverage by far. For linear functions, the expected length of $CLSCI$ converges extremely slowly with the increase of the sample size (if converges at all).

In Section A.3, we provide a more detailed discussion of the comparison with the theoretical results for our methods.

Estimation for Minimum. The plots show nice decreasing patterns. For the polynomial type functions, we can see a nice linear relationship between log empirical risk and log sample size, which is a good indicator of a linear relationship between the empirical risk and the benchmark, as the benchmark of a polynomial function is a power function (with negative exponent) of sample size. A detailed comparison with theoretical results is in Section A.3.

Inference for Minimum. Both our methods achieve nominal coverages in all settings (shown in Table (d) in the corresponding figure). The plots on empirical expected length show a nice decreasing pattern. Comparison with theoretical results is discussed in Section A.3.

Computing Time. Our methods are significantly faster than CLS based methods. For our methods, we measure the total time used for producing all four results, while for CLS based methods, we only measure the time taken to fit a CLS. The time for each function is the sum of times used for 100 replicates. Although this measurement of the computing time favors CLS based methods, our methods still take much less time.

A.2.2. *Figures, Tables, and Detailed Discussion.*

Discussion for results of $f_1(x) = 100|2x-1|$ *and* $f_2(x) = 100x$. Both functions are piecewise linear functions, and their smoothness is the lowest among all test functions.

In terms of coverage of confidence interval, our methods achieve nominal coverage consistently, $CLSCI_\alpha$ achieves the nominal coverage in most cases, but not consistently — it fails in some cases for f_2. Therefore, we turn to the expected lengths of the confidence intervals.

Piecewise linear functions are prototypes for supporting examples for sub-optimality of $CLSCI_\alpha$, in both rigorous proof and intuitive reasoning that we present in Section B.1. The simulation results on the length, as shown below, go along with the theoretical analysis.

The fourth plot in Figure 2 shows the log empirical expected length for the minimizer vs log sample size for f_1, which clearly shows that the empirical expected length of $CLSCI_\alpha$ shrinks much slower than our methods, supporting our intuitive reasoning in Section B.1. Further, extended experiments on even larger sample sizes show that the log empirical expected length eventually fluctuated around -2.3 rather than converging to $-\infty$.

For f_2, the fourth plot in Figure 4 show the expected length of $CLSCI_\alpha$ hardly converges, while those of our methods clearly converge.

For estimating the minimizer, the piecewise linear function f_1 is in favor of the CLS estimator for the minimizer, as discussed in section B.1. The results indeed shows that the empirical risk for CLS is around 0.6 times that of our method, although all the methods show the same rate. Similar phenomena also holds for f_2.

For tasks involving the minimum, we primarily focus on the relationship with theoretical benchmarks and the empirical coverage for the minimum. The nominal coverages are consistently achieved. A detailed comparison with theoretical results is deferred to Section A.3.

Discussion for results of $f_3(x) = 100(|2x-1|)^2$ *and* $f_4(x) = 100x^2$. The quadratic function f_3 belongs to the prototype function class that $CLSCI_\alpha$ is designed for. It has higher smoothness than f_1 but lower smoothness than f_5 and f_6.

From Table (a) in Figure 7, we can see that $CLSCI_\alpha$ does not consistently achieve nominal coverages for f_3. However, its coverage behavior for f_3 is much better than that for other test functions except piece-wise linear functions. Nevertheless, it is worth mentioning that for another quadratic function $f : x \mapsto (x-1/2)^2$, most of the empirical coverages of $CLSCI_\alpha$ are far below the nominal coverages. An explanation is that the difference in the scale between f and f_3 leads to different signal-to-noise balances — f is too weak a signal so that reasonable sample sizes do not reach the asymptotic region of $CLSCI_\alpha$ for f. This instability in coverage is an issue for $CLSCI_\alpha$, as the true underlying function is always unknown. In contrast, our methods achieve the nominal coverage consistently.

The fourth plot in Figure 6 shows that the convergence rates of length for f_3 are almost the same for all methods, but the empirical expected length for $CLSCI_\alpha$ is shorter than that of our methods by a constant multiplier. This is not surprising as our goal in this paper is to propose methods that can achieve the benchmarks up to a constant multiplier. The details of the building blocks in our procedures have flexibilities for further improvement of the constant, which we leave to future investigation.

For estimating the minimizer of f_3, the performance of the CLS estimator is between our two versions.

For the half-quadratic function f_4, the performances of estimation of minimizer are similar to that of linear function, while the performance of inference of the minimizer is different from both quadratic and linear functions. Our methods achieve the nominal coverage consistently, but $CLSCI_\alpha$ misses the nominal coverage by far.

Discussion for results of $f_5(x) = 100(|2x-1|)^4$. f_5 has relatively higher smoothness. For the inference of the minimizer, Table (a) in Figure 11 shows that $CLSCI_\alpha$ has empirical coverages that fall significantly below the nominal coverages. In contrast, our procedures attain nominal coverages.

For the estimation of the minimizer, all methods have similar rates yet both our methods have smaller empirical risks.

Discussion for results of $f_6(x) = 100\exp\left(2 - \frac{1}{|x-0.5|}\right)$. f_6 has the highest smoothness. Its arbitrary-order derivative at the minimizer $x = 0.5$ is 0.

For the inference of the minimizer, Table (a) in Figure 13 shows that the empirical coverages of $CLSCI_\alpha$ fall significantly below the nominal coverages. In contrast, our procedures achieve nominal coverages.

For estimating the minimizer, the CLS estimator has larger empirical risks than ours and does not show a clear trend of convergence. Ours have already shown a clear pattern of converging to 0 (a.k.a to $-\infty$ on the log scale).

Discussion for results of $f_7 = 100|2x-1|\mathbb{1}\{x < 0.5\} + 100|2x-1|^2\mathbb{1}\{x \geq 0.5\}$. f_7 is asymmetric, non-differentiable, and differing in smoothness on two sides. For estimation of the minimizer, all methods have similar behavior with CLS being slightly better. For inference of the minimizer, $CLSCI_\alpha$ has empirical coverages that fall significantly below the nominal coverages. All of our confidence intervals achieve the nominal coverage and has empirical expected lengths showing nice decreasing patterns.

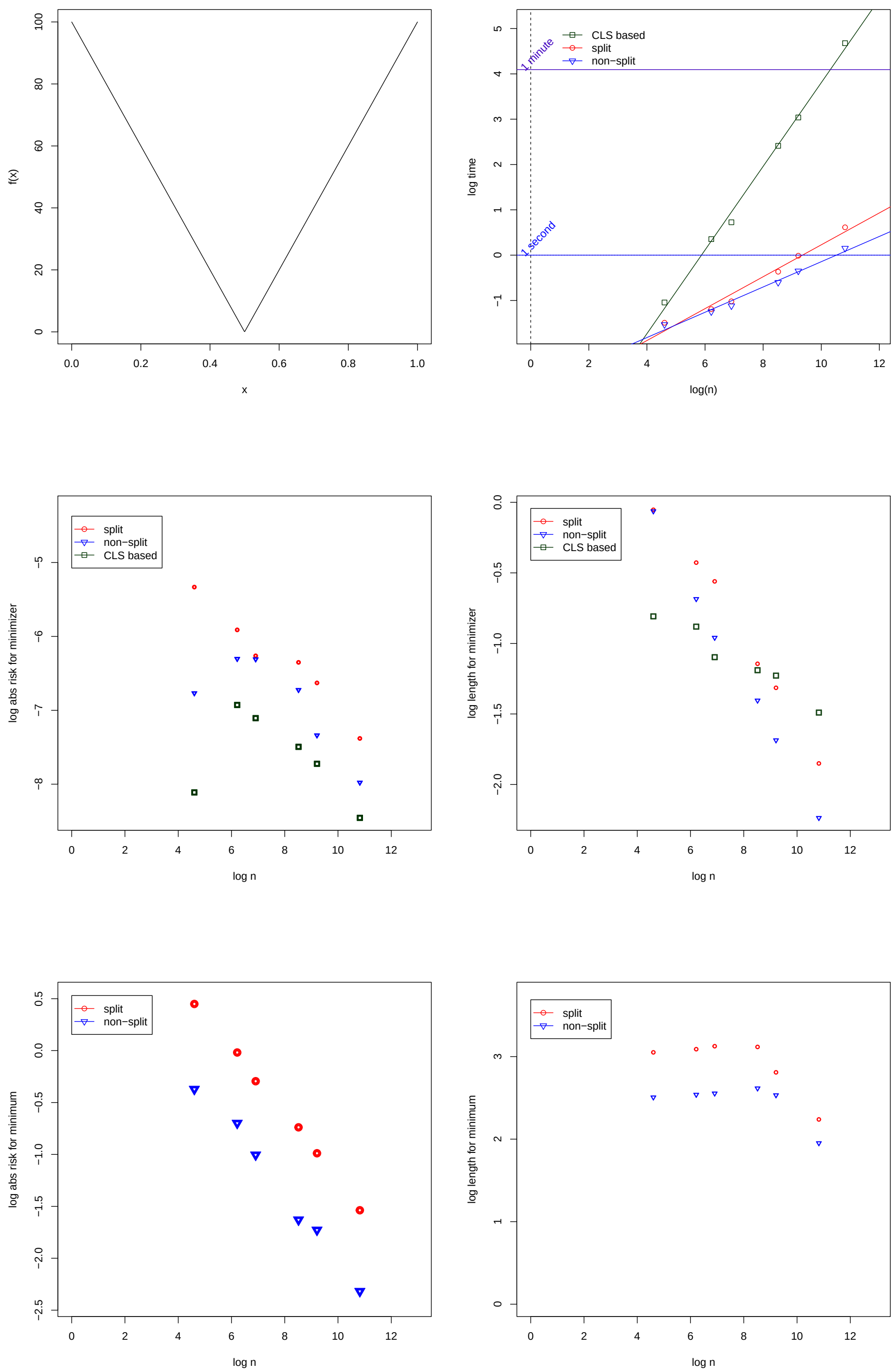


Fig 2: Plots for $f_1(x) = 100|2x - 1|$

	100	500	1000	5000	10000	50000
0.8	1	1	0.99	0.95	0.94	0.97
0.9	1	1	0.99	0.97	0.97	0.98
0.95	1	1	0.99	0.98	1	0.99
0.98	1	1	0.99	1	1	0.99
0.99	1	1	1	1	1	1

(a) Empirical coverage of CLS confidence interval for minimizer

	100	500	1000	5000	10000	50000
split	−0.053	−0.427	−0.56	−1.144	−1.315	−1.851
non−split	−0.063	−0.685	−0.959	−1.404	−1.686	−2.236
CLS based	−0.808	−0.881	−1.097	−1.19	−1.228	−1.49

(b) Log empirical length of confidence interval for minimizer

	100	500	1000	5000	10000	50000
split	−5.332	−5.912	−6.263	−6.351	−6.63	−7.381
non−split	−6.768	−6.303	−6.309	−6.724	−7.339	−7.978
CLS based	−8.112	−6.928	−7.106	−7.495	−7.724	−8.456

(c) Log empirical risk for minimizer

	100	500	1000	5000	10000	50000
0.8	0.98	1	1	0.99	1	1
0.9	0.98	1	1	1	1	1
0.95	1	1	1	1	1	1
0.98	1	1	1	1	1	1
0.99	1	1	1	1	1	1

(d) Empirical coverage of our confidence interval for the minimum

Fig 3: Tables for $f_1(x) = 100|2x - 1|$

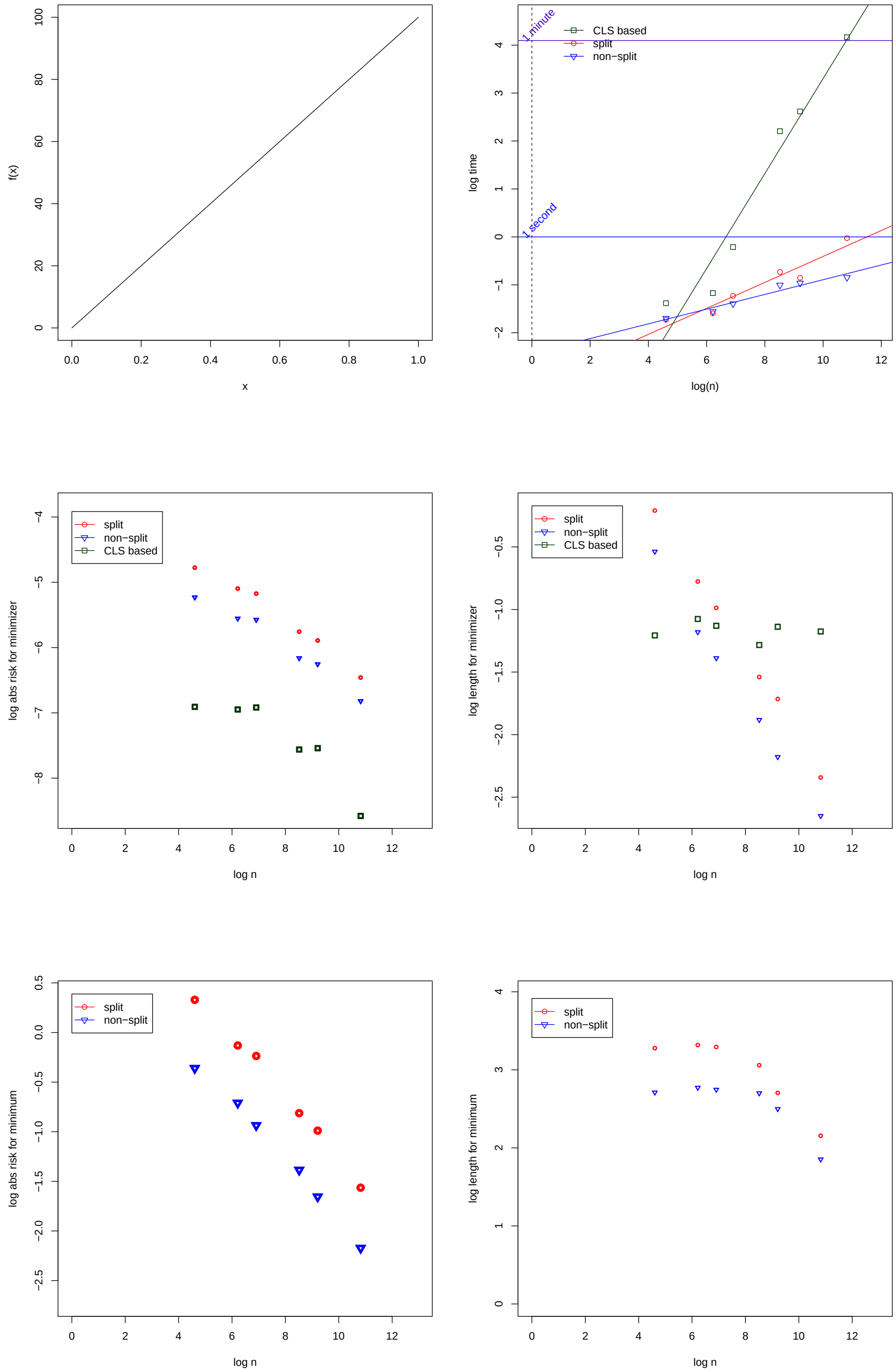


Fig 4: Plots for $f_2(x) = 100x$

	100	500	1000	5000	10000	50000
0.8	1	0.97	0.95	0.94	0.96	0.96
0.9	1	0.97	0.98	0.96	0.96	0.98
0.95	1	0.99	0.99	0.97	0.96	0.99
0.98	1	1	1	0.99	**0.96**	0.99
0.99	1	1	1	1	**0.99**	1

(a) Empirical coverage of CLS confidence interval for minimizer

	100	500	1000	5000	10000	50000
split	−0.209	−0.776	−0.987	−1.54	−1.715	−2.343
non−split	−0.537	−1.181	−1.389	−1.882	−2.179	−2.65
CLS based	−1.207	−1.076	−1.131	−1.284	−1.138	−1.176

(b) Log empirical length of confidence interval for minimizer

	100	500	1000	5000	10000	50000
split	−4.776	−5.096	−5.174	−5.757	−5.892	−6.459
non−split	−5.231	−5.555	−5.574	−6.161	−6.254	−6.82
CLS based	−6.908	−6.949	−6.918	−7.562	−7.541	−8.579

(c) Log empirical risk for minimizer

	100	500	1000	5000	10000	50000
0.8	1	1	1	1	1	1
0.9	1	1	1	1	1	1
0.95	1	1	1	1	1	1
0.98	1	1	1	1	1	1
0.99	1	1	1	1	1	1

(d) Empirical coverage of our confidence interval for the minimum

Fig 5: Tables for $f_2(x) = 100x$

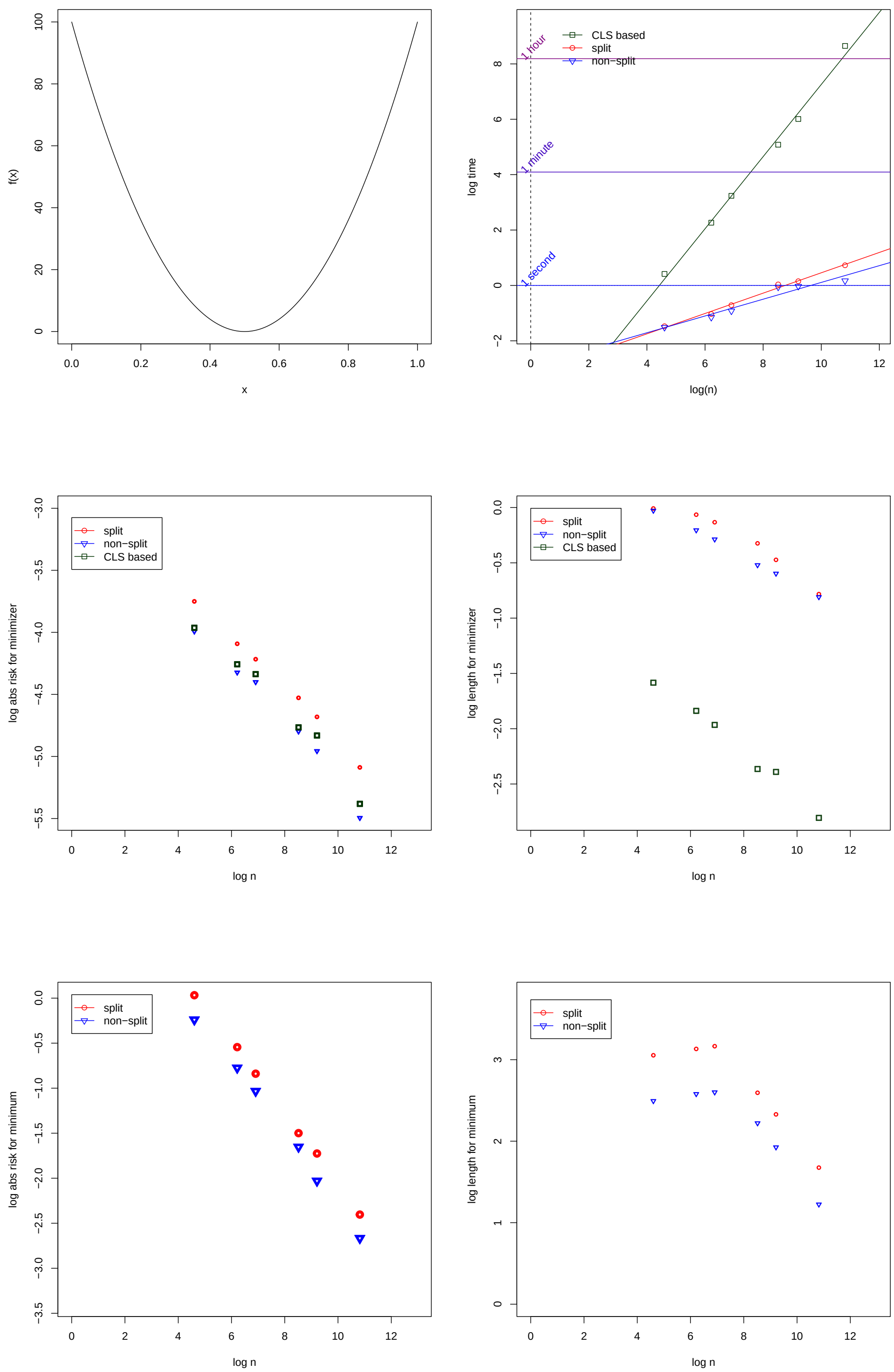


Fig 6: Plots for $f_3(x) = 100(|2x - 1|)^2$

	100	500	1000	5000	10000	50000
0.8	0.81	**0.79**	0.82	0.84	**0.76**	0.82
0.9	**0.88**	0.92	0.91	0.91	**0.89**	0.93
0.95	0.96	0.96	0.97	**0.95**	0.97	0.97
0.98	0.99	0.99	1	0.99	0.99	0.99
0.99	1	1	1	1	**0.99**	1

(a) Empirical coverage of CLS confidence interval for minimizer

	100	500	1000	5000	10000	50000
split	−0.009	−0.065	−0.133	−0.324	−0.473	−0.783
non−split	−0.028	−0.205	−0.287	−0.52	−0.597	−0.809
CLS based	−1.584	−1.838	−1.965	−2.364	−2.391	−2.806

(b) Log empirical length of confidence interval for minimizer

	100	500	1000	5000	10000	50000
split	−3.75	−4.092	−4.216	−4.528	−4.681	−5.088
non−split	−3.993	−4.323	−4.401	−4.799	−4.956	−5.495
CLS based	−3.963	−4.257	−4.337	−4.766	−4.831	−5.382

(c) Log empirical risk for minimizer

	100	500	1000	5000	10000	50000
0.8	1	0.98	1	0.99	0.98	0.98
0.9	1	0.99	1	0.99	0.99	1
0.95	1	1	1	1	0.99	1
0.98	1	1	1	1	0.99	1
0.99	1	1	1	1	1	1

(d) Empirical coverage of our confidence interval for the minimum

Fig 7: Tables for $f_3(x) = 100(|2x - 1|)^2$

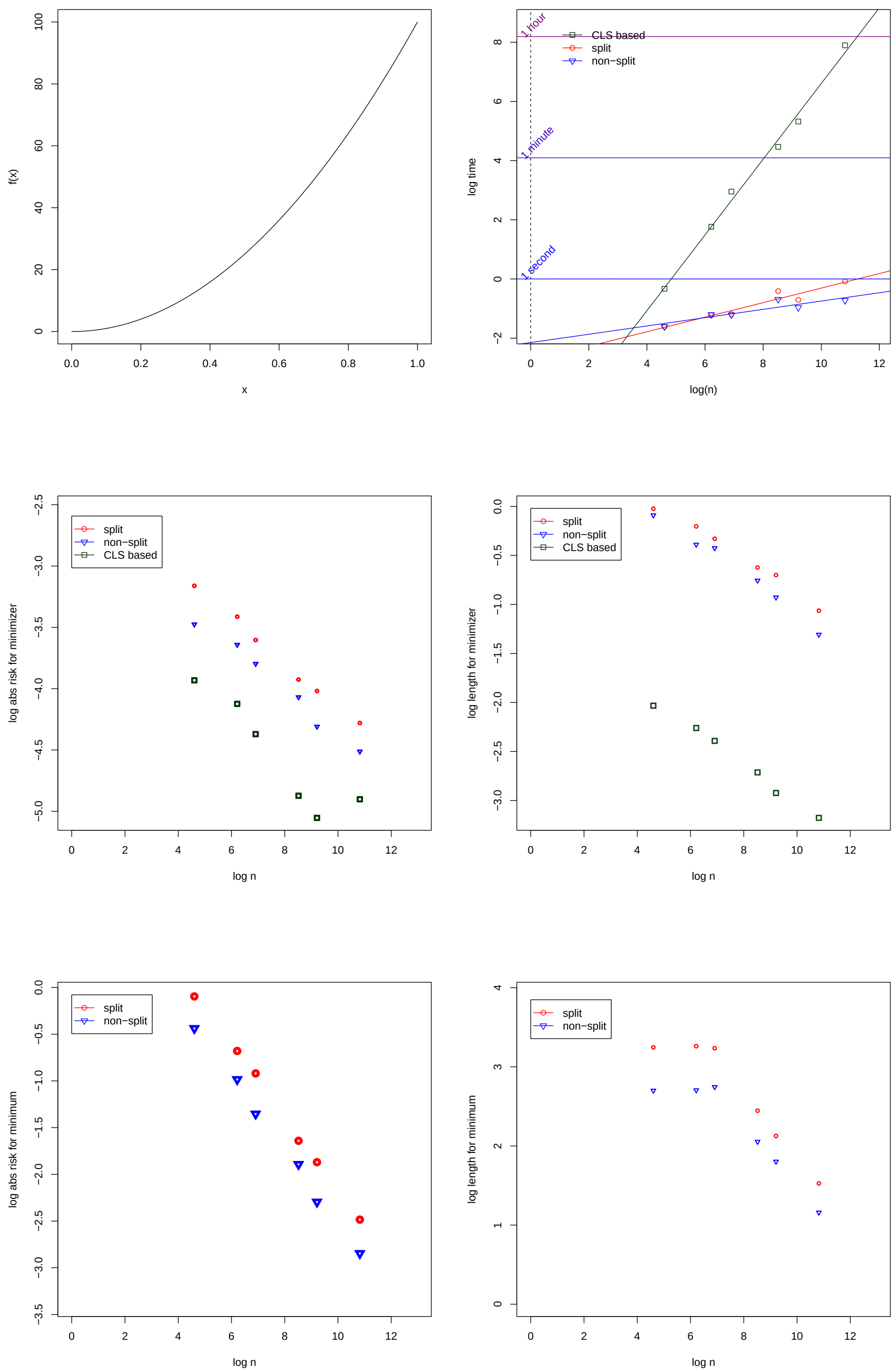


Fig 8: Plots for $f_4(x) = 100x^2$

	100	500	1000	5000	10000	50000
0.8	**0.73**	**0.72**	**0.77**	0.87	0.84	**0.68**
0.9	**0.8**	**0.81**	**0.82**	**0.88**	**0.89**	**0.77**
0.95	**0.89**	**0.91**	**0.89**	0.96	**0.94**	**0.88**
0.98	**0.96**	**0.95**	**0.94**	0.98	**0.96**	**0.92**
0.99	**0.97**	0.99	0.99	0.99	**0.98**	0.99

(a) Empirical coverage of CLS confidence interval for minimizer

	100	500	1000	5000	10000	50000
split	−0.024	−0.204	−0.33	−0.624	−0.7	−1.064
non−split	−0.09	−0.389	−0.425	−0.756	−0.928	−1.308
CLS based	−2.033	−2.261	−2.392	−2.714	−2.924	−3.177

(b) Log empirical length of confidence interval for minimizer

	100	500	1000	5000	10000	50000
split	−3.161	−3.413	−3.603	−3.926	−4.019	−4.28
non−split	−3.475	−3.642	−3.797	−4.07	−4.309	−4.512
CLS based	−3.932	−4.124	−4.371	−4.874	−5.054	−4.902

(c) Log empirical risk for minimizer

	100	500	1000	5000	10000	50000
0.8	1	0.98	0.98	0.97	0.95	0.94
0.9	1	0.99	1	0.97	0.97	0.95
0.95	1	0.99	1	0.99	0.99	0.99
0.98	1	0.99	1	1	0.99	0.99
0.99	1	1	1	1	1	0.99

(d) Empirical coverage of our confidence interval for the minimum

Fig 9: Tables for $f_4(x) = 100x^2$

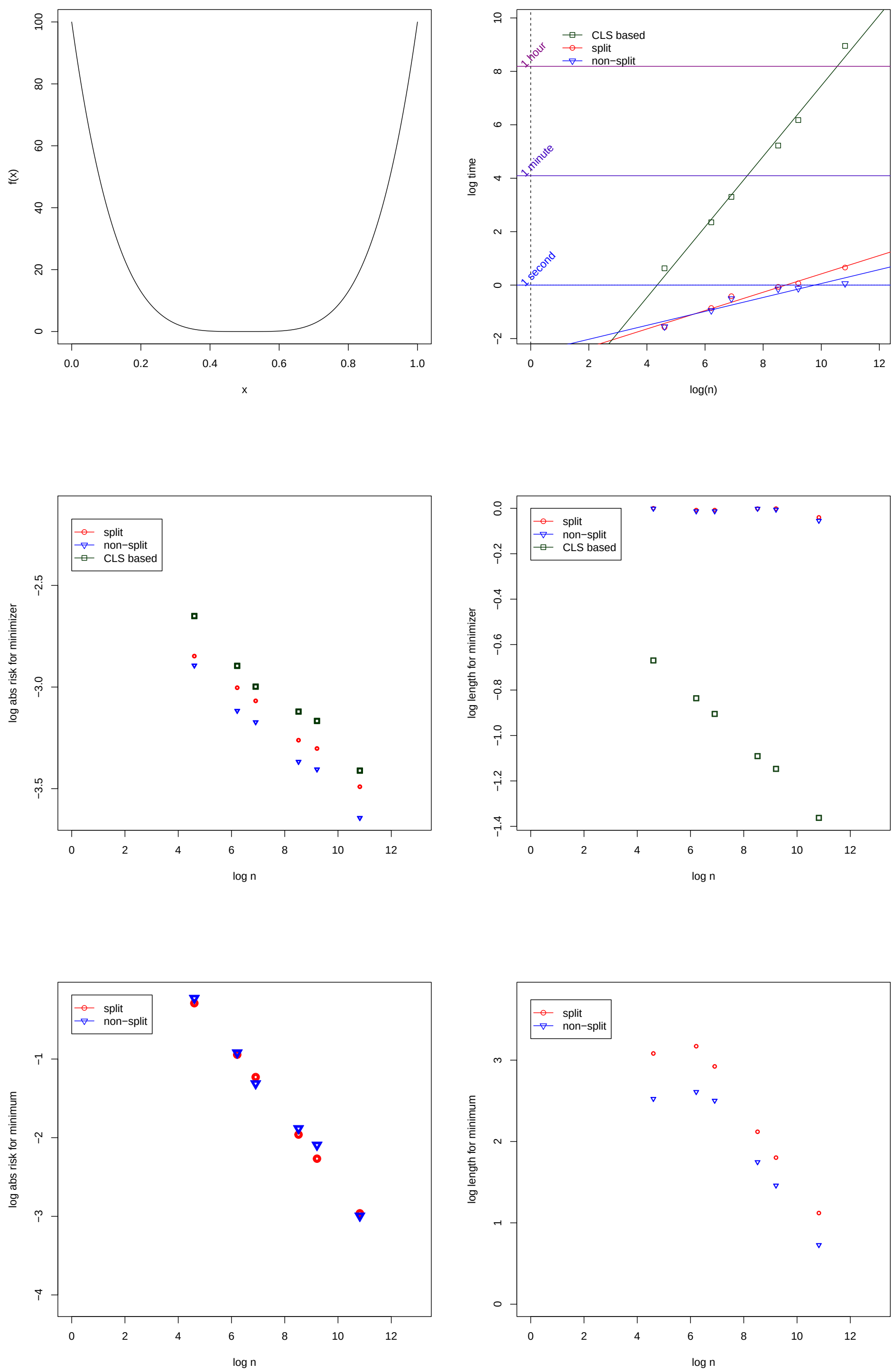


Fig 10: Plots for $f_5(x) = 100(|2x-1|)^4$

	100	500	1000	5000	10000	50000
0.8	**0.57**	**0.61**	**0.67**	**0.56**	**0.56**	**0.62**
0.9	**0.82**	**0.83**	**0.82**	**0.83**	**0.77**	**0.82**
0.95	**0.89**	**0.91**	**0.91**	**0.91**	**0.9**	**0.94**
0.98	**0.95**	**0.95**	**0.96**	**0.96**	**0.95**	**0.96**
0.99	**0.96**	**0.99**	**0.98**	**0.99**	**0.97**	**0.97**

(a) Empirical coverage of CLS confidence interval for minimizer

	100	500	1000	5000	10000	50000
split	0	−0.008	−0.008	−0.002	−0.002	−0.04
non−split	−0.001	−0.013	−0.013	−0.002	−0.006	−0.055
CLS based	−0.67	−0.836	−0.905	−1.091	−1.147	−1.363

(b) Log empirical length of confidence interval for minimizer

	100	500	1000	5000	10000	50000
split	−2.848	−3.003	−3.068	−3.262	−3.302	−3.49
non−split	−2.894	−3.117	−3.173	−3.368	−3.405	−3.644
CLS based	−2.651	−2.896	−2.998	−3.121	−3.167	−3.411

(c) Log empirical risk for minimizer

	100	500	1000	5000	10000	50000
0.8	0.97	0.95	0.99	0.97	0.96	0.97
0.9	0.97	0.97	0.99	0.98	0.98	0.98
0.95	1	0.99	1	1	1	1
0.98	1	0.99	1	1	1	1
0.99	1	**0.99**	1	1	1	1

(d) Empirical coverage of our confidence interval for the minimum

Fig 11: Tables for $f_5(x) = 100(|2x - 1|)^4$

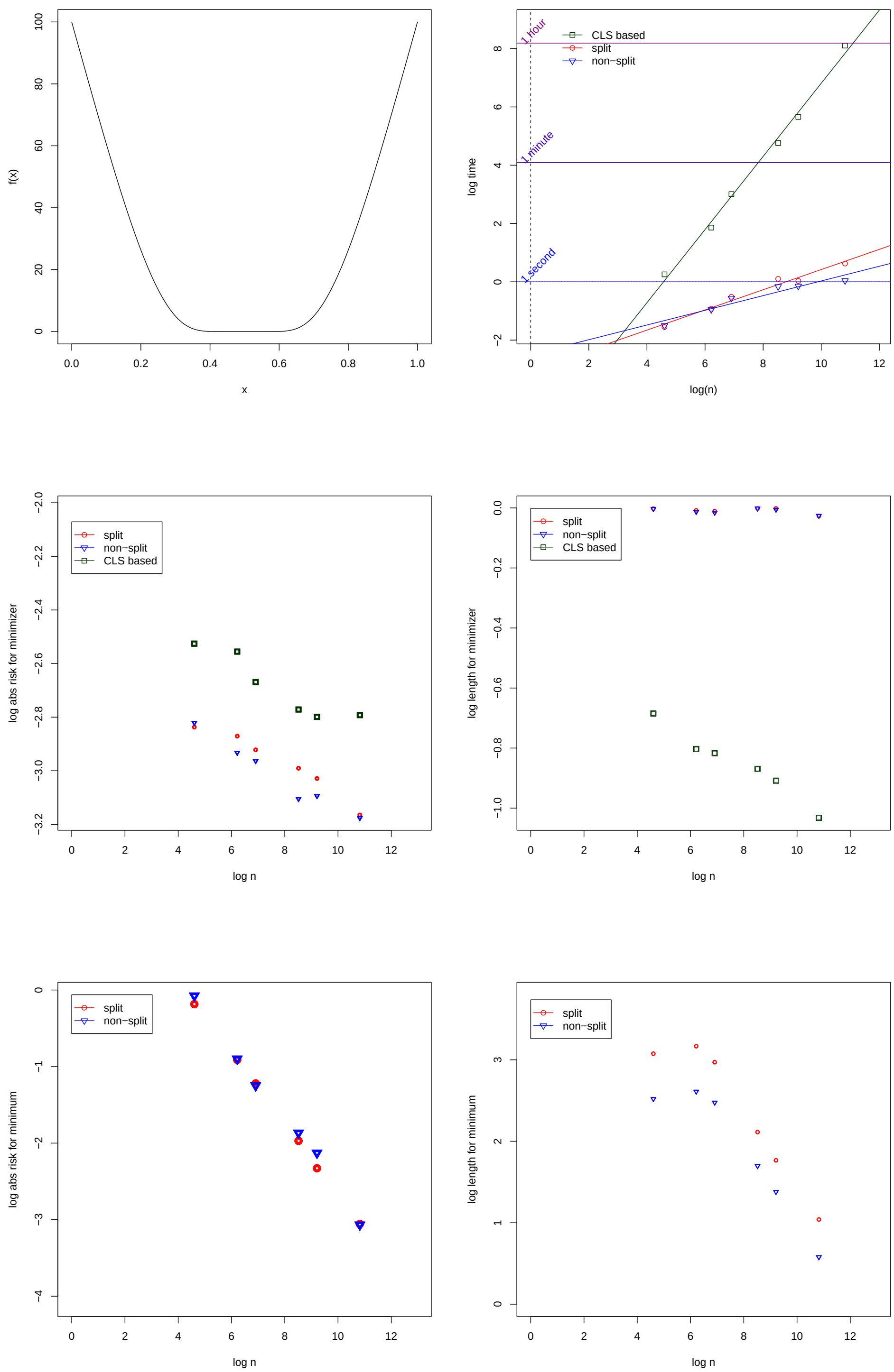


Fig 12: Plots for $f_6(x) = 100 \exp\left(2 - \frac{1}{|x-0.5|}\right)$

	100	500	1000	5000	10000	50000
0.8	**0.44**	**0.39**	**0.47**	**0.46**	**0.46**	**0.33**
0.9	**0.73**	**0.66**	**0.71**	**0.76**	**0.74**	**0.69**
0.95	**0.89**	**0.84**	**0.84**	**0.93**	**0.91**	**0.86**
0.98	**0.93**	**0.9**	**0.92**	**0.97**	**0.95**	**0.93**
0.99	**0.95**	**0.93**	**0.96**	**0.97**	**0.96**	**0.95**

(a) Empirical coverage of CLS confidence interval for minimizer

	100	500	1000	5000	10000	50000
split	−0.003	−0.008	−0.011	−0.002	−0.002	−0.027
non−split	−0.003	−0.013	−0.016	−0.002	−0.006	−0.026
CLS based	−0.685	−0.803	−0.817	−0.869	−0.909	−1.033

(b) Log empirical length of confidence interval for minimizer

	100	500	1000	5000	10000	50000
split	−2.837	−2.871	−2.922	−2.991	−3.029	−3.165
non−split	−2.821	−2.933	−2.963	−3.105	−3.094	−3.176
CLS based	−2.526	−2.555	−2.669	−2.772	−2.799	−2.792

(c) Log empirical risk for minimizer

	100	500	1000	5000	10000	50000
0.8	0.98	0.93	0.98	0.94	0.97	0.96
0.9	0.98	0.96	0.99	0.97	0.98	0.96
0.95	0.99	0.99	0.99	1	1	1
0.98	1	0.99	1	1	1	1
0.99	1	**0.99**	1	1	1	1

(d) Empirical coverage of our confidence interval for the minimum

Fig 13: Tables for $f_6(x) = 100 \exp\left(2 - \frac{1}{|x-0.5|}\right)$

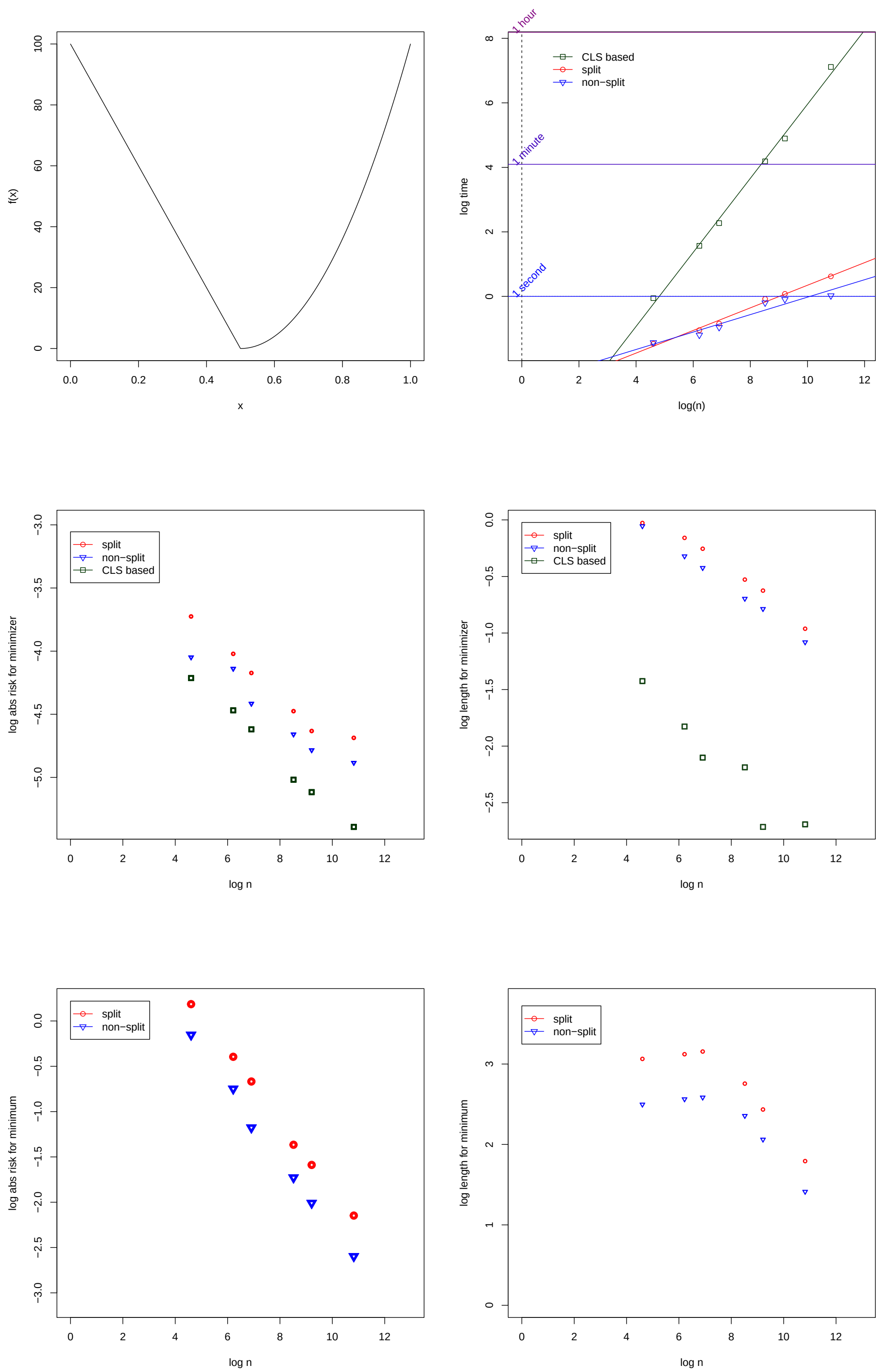


Fig 14: Plots for $f_7(x) = 100|2x-1|\mathbb{1}\{x<0.5\} + 100|2x-1|^2\mathbb{1}\{x\geq 0.5\}$

	100	500	1000	5000	10000	50000
0.8	**0.71**	**0.71**	**0.67**	**0.66**	**0.68**	**0.69**
0.9	**0.78**	**0.8**	**0.75**	**0.85**	**0.82**	**0.78**
0.95	**0.89**	**0.88**	**0.83**	**0.88**	**0.87**	**0.83**
0.98	**0.95**	**0.94**	**0.91**	**0.97**	**0.97**	**0.93**
0.99	**0.99**	**0.98**	**0.97**	**0.99**	1	**0.99**

(a) Empirical coverage of CLS confidence interval for minimizer

	100	500	1000	5000	10000	50000
split	−0.027	−0.159	−0.255	−0.528	−0.625	−0.962
non−split	−0.055	−0.321	−0.424	−0.696	−0.787	−1.081
CLS based	−1.425	−1.827	−2.102	−2.187	−2.714	−2.691

(b) Log empirical length of confidence interval for minimizer

	100	500	1000	5000	10000	50000
split	−3.725	−4.021	−4.173	−4.475	−4.633	−4.687
non−split	−4.048	−4.139	−4.416	−4.659	−4.784	−4.884
CLS based	−4.213	−4.469	−4.619	−5.018	−5.117	−5.392

(c) Log empirical risk for minimizer

	100	500	1000	5000	10000	50000
0.8	1	1	0.99	0.96	1	0.99
0.9	1	1	0.99	0.97	1	0.99
0.95	1	1	0.99	0.99	1	1
0.98	1	1	1	0.99	1	1
0.99	1	1	1	1	1	1

(d) Empirical coverage of our confidence interval for the minimum

Fig 15: Tables for $f_7(x) = 100|2x-1|\mathbb{1}\{x < 0.5\} + 100|2x-1|^2\mathbb{1}\{x \geq 0.5\}$

A.3. Comparison with Benchmarks. In this subsection, we consider the functions for which benchmarks can be explicitly calculated. The primary task is to investigate the relationship between empirical risks/lengths and the benchmarks.

We consider a different set of functions whose benchmarks can be easily calculated:

$$\begin{aligned} h_1(t) &= 100\big|t-0.5\big|, \\ h_2(t) &= 200\big|2(t-0.5)\big|^{\frac{3}{2}}, \\ h_3(t) &= 200\big|2(t-0.5)\big|^{2}, \\ h_4(t) &= 200\big|2(t-0.5)\big|^{3}, \\ h_5(t) &= 200\big|2(t-0.5)\big|^{4}. \end{aligned} \tag{A.2}$$

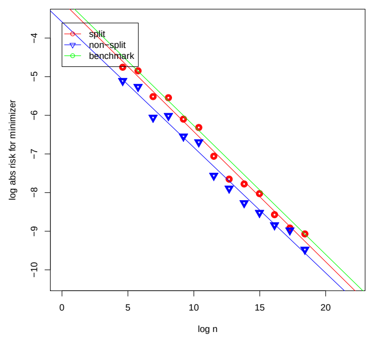
All other settings remain the same as before, except that we take roughly exponentially equally spaced sample sizes.

We calculated the corresponding benchmarks (the discretization errors in these examples are negligible): $\rho_z(\sqrt{1/n}; f)$ and $\rho_m(\sqrt{1/n}; f)$.

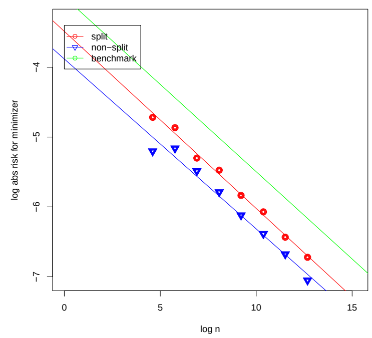
The log risk/length vs. log sample size plots for the minimizer and minimum with the reference line of the benchmark are shown in Figures 16, 17, 18, and 19. For the estimation of the minimizer, in addition to the almost identical slope with the reference line (i.e., linear relationship between empirical risk and benchmark), the intercept difference of the reference line and the log risk of non-split version ranges between 0.6472699 and 1.036388, meaning that $\frac{\rho_z(\sqrt{1/n};f)}{\text{empirical risk for minimizer}}$ for non-split version ranges in $[1.910318, 2.819016]$, implying that the performance of non-split version is quite robust when smoothness varies.

For the other three tasks, excluding the outlier points that are clearly influenced by the truncation for confidence interval, the slopes of the methods and the reference line are almost identical.

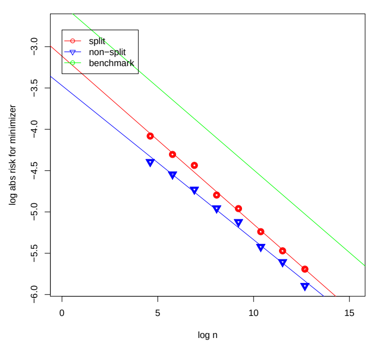
The empirical performances, therefore, agree with the theoretical results.



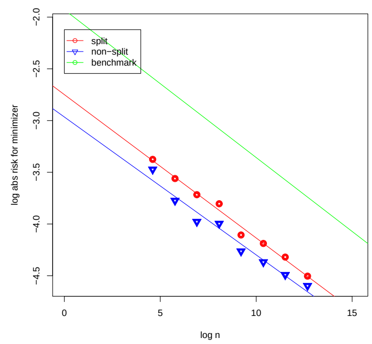
(a) h_1



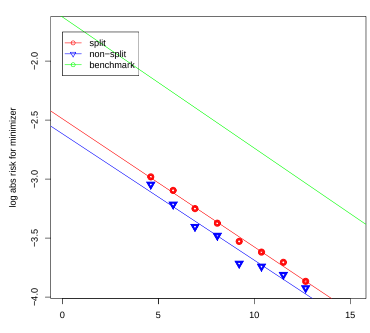
(b) h_2



(c) h_3

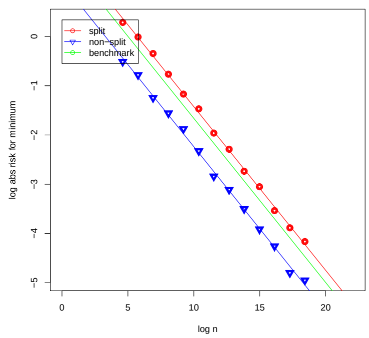


(d) h_4

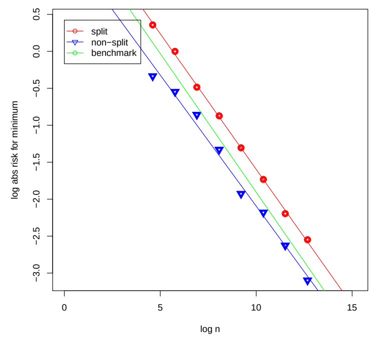


(e) h_5

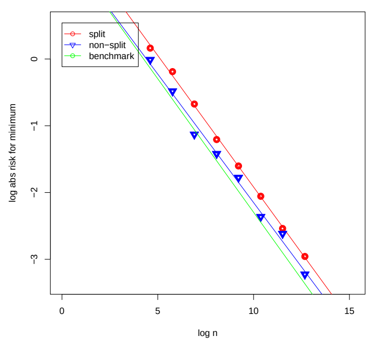
Fig 16: Empirical risks for minimizer



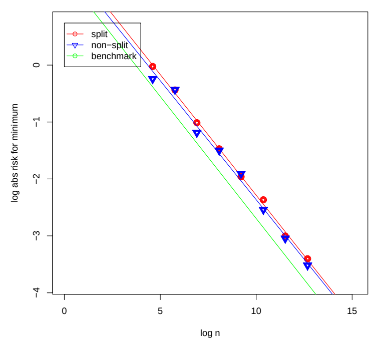
(a) h_1



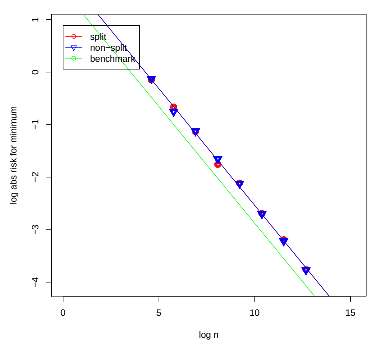
(b) h_2



(c) h_3

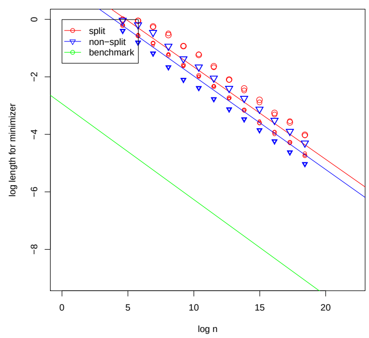


(d) h_4

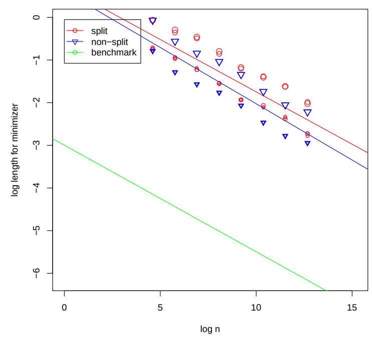


(e) h_5

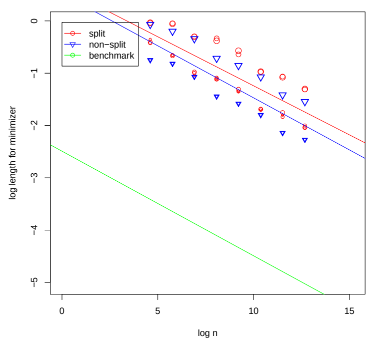
Fig 17: Empirical risks for minimum



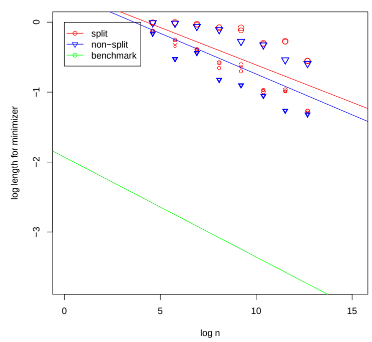
(a) h_1



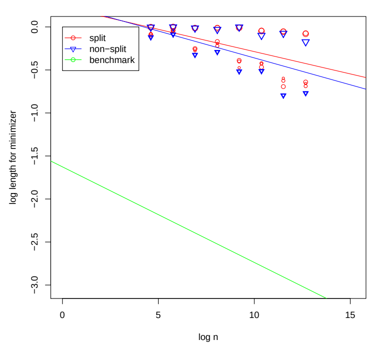
(b) h_2



(c) h_3

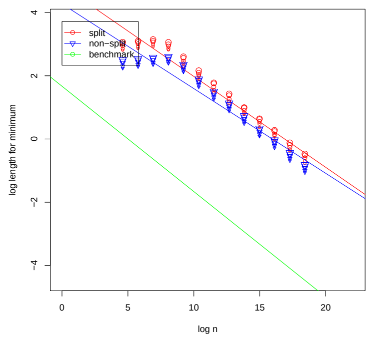


(d) h_4

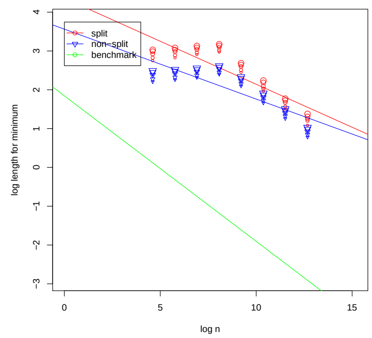


(e) h_5

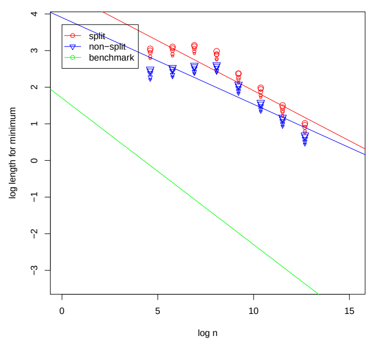
Fig 18: Empirical lengths for minimizer



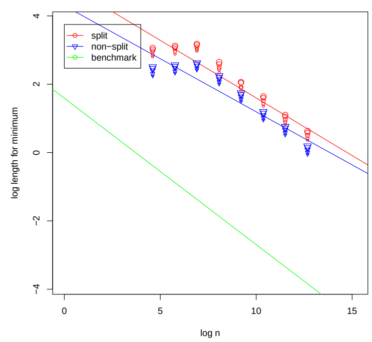
(a) h_1



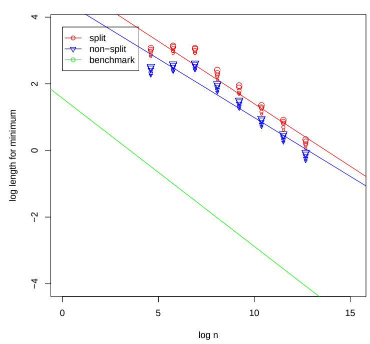
(b) h_2



(c) h_3



(d) h_4



(e) h_5

Fig 19: Empirical lengths for minimum

APPENDIX B: COMPARISON WITH CLS METHODS AND CONNECTIONS WITH THE CLASSICAL MINIMAX FRAMEWORK

In this section, we compare our procedures with the convexity-constrained least squares (CLS) methods for the minimizer (see Section B.1 for inference, Section B.4 for estimation, and the corresponding numerical results are in Section A), discuss the connections between local minimax framework and the classical minimax framework for problems considered in this paper, and elaborate on the generality of the Uncertainty Principle.

In particular, we prove that the CLS confidence interval for the minimizer proposed in Deng et al. (2020) is sub-optimal under the local minimax framework. We also provide a larger class of functions that potentially also lead to sub-optimality and provide the intuitive reasoning behind, which is validated through numerical results. Through investigating the connection with the classical minimax framework, we established that optimal procedures under the local minimax framework (e.g., our algorithms) are also optimal under the classical minimax framework (see Section B.2 for details). Implications of these results include that our algorithms are optimal under the setting that CLS is theoretically investigated. In addition, we provide more settings where the Uncertainty Principle holds in Section B.3.

B.1. Comparison with CLS Confidence Interval (CLSCI): Sub-optimality of CLSCI Under Non-asymptotic Local Minimax Framework and Optimality of Our Algorithms Under Several Frameworks. The convexity-constrained least squares (CLS) estimator is widely used for estimating a convex regression function globally. While CLS estimation and inference methods for the minimizer have been proposed and studied in the literature (e.g., Shoung and Zhang (2001); Ghosal and Sen (2017); Deng et al. (2020)), the theoretical analyses usually assume the existence of second or higher order derivatives with an even order derivative being positive and all lower order derivatives being zero at the minimizer. However, it is unclear how the CLS estimator behaves under our non-asymptotic framework or even asymptotically in general when the underlying convex function is nonsmooth at the minimizer. As for the minimum, to the best of our knowledge, no CLS-based method for estimation or inference with theoretical guarantees exists.

It is interesting to compare with the CLS confidence interval for the minimizer proposed in Deng et al. (2020). Let

$$\hat{f}_n = \min_{f\in\mathcal{F}} \sum_{i=1}^{n} (y_i - f(x_i))^2 \tag{B.1}$$

be the CLS estimator. Let $\hat{m}_n$ be the anti-mode of $\hat{f}_n$, $\hat{v}_m$ (resp. $\hat{u}_m$) be the first kink of $\hat{f}_n$ to the right (resp. left) of $\hat{m}_n$. Under the assumption that the second order derivative exists and is positive around the minimizer, Deng et al. (2020) introduce the following $(1-\alpha)$-level confidence interval,

$$CLSCI_\alpha = [\hat{m}_n \pm c_\alpha^m(\hat{v}_m - \hat{u}_m)] \cap [0,1], \tag{B.2}$$

where c_α^m is a constant depending on α only.

For positive integer k and positive number A, denote k-smooth A-bounded convex function class by

(B.3)

$$\mathcal{F}_{k,A} = \begin{cases} \left\{f\in\mathcal{F} : f \text{ is } k\text{-differentiable, } |f^{(k)}(Z(f))| < A\right\}, & k \text{ is odd} \\ \left\{f\in\mathcal{F} : f \text{ is } k\text{-differentiable, } 0 < |f^{(k)}(Z(f))| < A\right\}, & k \text{ is even} \end{cases}.$$

The parameter space described, with the exception of convexity, was also considered in the estimation of the mode for unimodal smooth functions (not necessarily convex) in Shoung and Zhang (2001).

Clearly the collection of convex functions with continuous positive second order derivative around the minimizer, denoted by $\mathcal{F}_2$, can be expressed as $\mathcal{F}_2 = \cup_{A>0}\mathcal{F}_{2,A}$. Deng et al. (2020) shows that the confidence interval $CLSCI_\alpha$ has desired coverage probability asymptotically over $\mathcal{F}_2$. The following result shows that $CLSCI_\alpha$ defined in (B.2) is sub-optimal over $\mathcal{F}_{k,A}$ for any k and A under the local minimax framework.

PROPOSITION B.1. *For positive integer k and positive number A, for any sample size $n \geq 5$,*

$$\sup_{f\in\mathcal{F}_{k,A}} \frac{\mathbb{E}_f L(CLSCI_\alpha)}{\tilde{L}_{z,\alpha,n}(\sigma; f)} = \infty. \tag{B.4}$$

where $\tilde{L}_{z,\alpha,n}(\sigma; f)$ is the benchmark defined in Equation (4.2).

Proposition B.1 shows that for any $n \geq 5$, there exists $f \in \mathcal{F}_{k,A}$ such that the length of $CLSCI_\alpha$ at f is much larger than the local minimax benchmark. In contrast, our proposed confidence interval $\mathtt{CI}_{z,\alpha}$ achieves the benchmark up to an absolute constant for all $f \in \mathcal{F}$. This phenomenon can be attributed to the nonasymptotic nature of our framework compared to the asymptotic nature of $CLSCI_\alpha$. In summary, the CLS construction, which only takes into account the kinks, fails to make full use of the convexity property.

Now we continue with proving Proposition B.1, providing additional scenarios that CLSCI potentially remains sub-optimal with intuitive reasoning whose associated numerical validation is in Section A, and showing that our algorithms are optimal under the setting CLSCI is theoretically investigated.

PROPOSITION B.2. *For positive integer k and positive number A, for any function $r(n) \geq 1$, for any integer $n \geq 5$, $\exists f_n \in \mathcal{F}_{k,A}$ such that*

$$\frac{\mathbb{E}_{f_n} L(CLSCI_\alpha)}{\tilde{L}_{z,\alpha,n}(\sigma; f)} \geq r(n), \tag{B.5}$$

where $\tilde{L}_{z,\alpha,n}(\sigma; f)$ is defined in Equation (4.2), $\mathcal{F}_{k,A}$ is defined in Equation (B.3).

PROOF. Suppose k and A are fixed. Recall that in the proof of Theorem 4.2, we have

$$\mathbb{E}_f L(\mathtt{CI}_{z,\alpha}) \leq C_{2,\alpha}\left(\sup_{h \in \mathcal{G}_n(f)} \rho_z(\tfrac{\sigma}{\sqrt{n}}; h)\left(1 \wedge \sqrt{n\rho_z(\tfrac{\sigma}{\sqrt{n}}; h)}\right) + \tfrac{(1-2\alpha)}{2}\mathfrak{D}_z(n,f)\right),$$

where the definition of $\mathcal{G}_n(f)$ is given in Equation (C.105).

Combining this inequality with the lower bound of the local minimax length of the confidence interval that we established in Proposition C.4, namely

$$\tilde{L}_{z,\alpha,n}(\sigma; f) \geq \tilde{C}_{z,\alpha}\left(\sup_{g \in \mathcal{G}_n(f)} \rho_z(\tfrac{\sigma}{\sqrt{n}}; g)\left(1 \wedge \sqrt{n\rho_z(\tfrac{\sigma}{\sqrt{n}}; g)}\right) + \tfrac{(1-2\alpha)}{2}\mathfrak{D}_z(n,f)\right),$$

we see that it suffices to show that for any $r(n) > 0$, there exists $f \in \mathcal{F}_{k,A}$ such that

$$\frac{\mathbb{E}_f L(CLSCI_\alpha)}{\left(\sup_{g \in \mathcal{G}_n(f)} \rho_z(\tfrac{\sigma}{\sqrt{n}}; g)\left(1 \wedge \sqrt{n\rho_z(\tfrac{\sigma}{\sqrt{n}}; g)}\right) + \tfrac{(1-2\alpha)}{2}\mathfrak{D}_z(n,f)\right)} \geq r(n).$$

Note that $L(CLSCI_\alpha) \geq \frac{1}{n}$, we only need to find $f \in \mathcal{F}_{k,A}$ such that

$$\mathfrak{D}_z(n,f) \leq \frac{1}{2nr(n)} \text{ and } \sup_{g \in \mathcal{G}_n(f)} \rho_z(\frac{\sigma}{\sqrt{n}}; g) \leq \frac{1}{2n(r(n)+1)}. \tag{B.6}$$

Consider function $f_0 : x \mapsto 4n(r(n)+1)^{\frac{3}{2}}(\sigma+1)|x - \frac{\lfloor n/2 \rfloor}{n}|$, for which we have

$$\mathfrak{D}_z(n, f_0) = 0, \qquad \sup_{g \in \mathcal{G}_n(f_0)} \rho_z(\frac{\sigma}{\sqrt{n}}; g) \le \frac{(\frac{3}{4})^{\frac{1}{3}}}{2} \frac{1}{n(r(n)+1)}.$$

The conditions mentioned in Inequality (B.6) are met, but f_0 is not in $\mathcal{F}_{k,A}$. Now we will proceed to construct $f_1 \in \mathcal{F}_{k,A}$ such that the conditions in Inequality (B.6) are still met for $f = f_1$.

For function f defined on $[0,1]$, define the following transformation.

$$\tilde{f}(x) = \begin{cases} f(x), & x \in [0,1] \\ f(1) + \sup\limits_{t \to 1^-} \dfrac{f(1) - f(t)}{1-t}(x-1), & x > 1 \\ f(0) + \sup\limits_{t \to 0^+} \dfrac{f(t) - f(0)}{t} x, & x < 0 \end{cases}. \tag{B.7}$$

Then consider the following class of transformations of function f:

$$T(f;\delta)(x) = \int \tilde{f}(t) \frac{1}{\sqrt{2\pi}\delta} \exp\left(-\frac{(x-t)^2}{2\delta^2}\right) dt. \tag{B.8}$$

It is easy to check that this transformation preserves convexity: if f is a convex function on $[0,1]$, then $T(f;\delta)$ is a convex function on $\mathbb{R}$. In addition, this transformation is a smoothing transformation: if f is continuous on $[0,1]$, then $T(f;\delta)$ is infinitely differentiable on $\mathbb{R}$. Further, for any given f, the sequence of transformed functions $\{T(f;\delta)\}_{\delta>0}$ converges uniformly to $\tilde{f}$ as $\delta \to 0^+$. When we focus on the interval $[0,1]$, we have that for fixed f, $\lim_{\delta \to 0^+} \sup_{x \in [0,1]} \left| T(f;\delta)(x) - f(x) \right| = 0$.

Now we list some basic properties of the transformed function $T(f_0;\delta)$.

Clearly, $T(f_0;\delta)$ is convex and infinitely differentiable. Clearly, we have uniform convergence: $\lim_{\delta \to 0^+} \sup_{x \in [0,1]} \left| T(f_0;\delta)(x) - f_0(x) \right| = 0$.

Now we consider $T(f_0;\delta)^{(k)}(\frac{\lfloor \frac{n}{2} \rfloor}{n})$. Let $a_i^{(m)}$ denote coefficient corresponding to the term $x^i \exp(-x^2/2)$ in the m order derivative of function $u(x) = \exp(-x^2/2)$. For $i < 0$, $a_i^{(m)} = 0$.

Calculations show that

$$T(f_0;\delta)^{(k)}(\frac{\lfloor \frac{n}{2} \rfloor}{n}) = \begin{cases} 0 & k \text{ is odd} \\ v(n)\frac{a_0^{(k-2)}}{\delta^{k-1}} = v(n)\frac{(-1)^{(k-2)/2}(k-3)!!}{\delta^{k-1}} & k \text{ is even and } k \ge 4 \\ v(n)\delta & k = 0 \\ v(n)/\delta & k = 2 \end{cases}, \tag{B.9}$$

where $v(n) = 2 \cdot 4n(r(n)+1)^{\frac{3}{2}}(\sigma+1)/\sqrt{2\pi}$.

The calculations are primarily based on the following facts.

$$\int_{-\infty}^{\infty} |t| t^k \exp(-t^2/2) dt = \begin{cases} 0 & \text{for odd } k \\ 2 \times k!! & \text{for even } k \end{cases},$$

$$a_i^{(m+1)} = (i+1)a_{i+1}^{(m)} - a_{i-1}^{(m)} = (i+1)(i+2)a_{i+2}^{(m-1)} - (2i+1)a_i^{(m-1)} + a_{i-1}^{(m-1)}.$$

Taylor expansion of $u(x)$ gives $a_0^{(m)}$.

Now we proceed with construction of the target f_1.

Let a class of transformations of f be

$$T_2(f;\eta)(x) = \max\{0, f(x) + \eta(|2x-1| - 0.5)\} \tag{B.10}$$

for $\eta > 0$. This transformation clearly preserves convexity. Consider $T(T_2(f_0;\eta);\delta)$. We start with showing that it converges uniformly to at f_0 on $[0,1]$ as $\delta, \eta \to 0^+$. Let $g(x) = \big||2x-1| - 0.5\big|$. Clearly,

$$\begin{aligned} &\sup_{x\in[0,1]} |T(T_2(f_0;\eta);\delta)(x) - f_0(x)| \\ &\quad \le \sup_{x\in[0,1]} |T(f_0;\delta) - f_0(x)| + \eta \sup_{x\in[0,1]} \Big|T(g;\delta)(x)\Big|. \end{aligned}$$

Therefore, for any $\nu > 0$, there exist $\eta(\nu, f_0), \delta(\nu, f_0) > 0$ such that for any positive $\delta < \delta(\nu, f_0), \eta < \eta(\nu, f_0)$,

$$\sup_{x\in[0,1]} |T(T_2(f_0;\eta);\delta)(x) - f_0(x)| < \nu. \tag{B.11}$$

This uniform convergence gives that

(B.12a)

$$\lim_{\delta,\eta\to 0^+} \sup_{g\in\mathcal{G}_n(T(T_2(f_0;\eta);\delta))} \rho_z(\frac{\sigma}{\sqrt{n}}; g) = \sup_{g\in\mathcal{G}_n(f_0)} \rho_z(\frac{\sigma}{\sqrt{n}}; g) < \frac{1}{2n(r(n)+1)},$$

(B.12b)

$$\lim_{\delta,\eta\to 0^+} \mathfrak{D}_z(n, T(T_2(f_0;\eta);\delta)) = \mathfrak{D}_z(n, f_0) = 0,$$

(B.12c)

$$\lim_{\delta,\eta\to 0^+} Z(T(T_2(f_0;\eta);\delta)) = Z(f_0) \in (0,1)$$

For k-th order derivative of $T(T_2(f_0;\eta);\delta)$, elementary calculation shows that

$$\lim_{\delta\to 0^+} \left.\frac{\partial^k T(T_2(f_0;\eta);\delta)}{\partial x^k}\right|_{x=Z(T(T_2(f_0;\eta);\delta))} = 0, \text{ for all } \eta > 0.$$

and that

$$\lim_{\eta\to 0^+} \frac{\partial^k T(T_2(f_0;\eta);\delta)}{\partial x^k}\bigg|_{x=Z(T(T_2(f_0;\eta);\delta))} = T(f_0;\delta)^{(k)}(\frac{\lfloor \frac{n}{2}\rfloor}{n}),\ \text{for all } \delta > 0.$$

Recall the expression of $T(f_0;\delta)^{(k)}(\frac{\lfloor \frac{n}{2}\rfloor}{n})$ in Equation (B.9). Therefore, there exist $\eta > 0$ and $\delta > 0$ depending on f_0 , such that the function $f_1 = T(T_2(f_0;\eta);\delta)\Big|_{[0,1]}$ both satisfies conditions in Inequality (B.6) and is in $\mathcal{F}_{k,A}$.

Details of choosing such η and δ are as follows. For an even k, choose a small enough $\delta = \delta_0$ such that the limits in Equations (B.12) can be approximately achieved by all $\delta \le \delta_0$ and $\eta \le \eta_0$, such that Inequality (B.6) hold. Then choose a small enough $\eta \le \eta_0$ such that $\frac{\partial^k T(T_2(f_0;\eta);\delta)}{\partial x^k}\big|_{x=Z(T(T_2(f_0;\eta);\delta))} > 0$ and Inequality (B.6) holds. Then fix this η and select $\delta < \delta_0$ that is small enough such that $A > \frac{\partial^k T(T_2(f_0;\eta);\delta)}{\partial x^k}\big|_{x=Z(T(T_2(f_0;\eta);\delta))} > 0$. For an odd k, choose a small enough η (for Inequality (B.6) to hold for some small δ) and then a small enough δ such that Inequality (B.6) holds and $\frac{\partial^k T(T_2(f_0;\eta);\delta)}{\partial x^k}\big|_{x=Z(T(T_2(f_0;\eta);\delta))} < A$.

□

In the proof, we can observe the strength of non-asymptotic and non-localized results. A significant distinction between $\rho_z(\frac{\sigma}{\sqrt{n}};f)$, as featured in our theorem, and the second-order derivative, heavily relied upon by $CLSCI_\alpha$ in both method and theoretical guarantees, is that $\rho_z(\frac{\sigma}{\sqrt{n}};f)$ does not require any form of limit, whereas the second-order derivative does.

In this regard, unlike $\rho_z(\frac{\sigma}{\sqrt{n}};f)$, the second-order derivative exclusively characterizes the local behavior of a function within an infinitely small interval around a point. It is a localized quantity and demands twice differentiability. Consequently, an asymptotic procedure based on a localized quantity encounters the issue that, for some functions, regardless of how large n becomes, it remains outside the scope of locality.

To demonstrate the sub-optimality of $CLSCI_\alpha$, let us turn our attention to convex piecewise linear functions. Simulation results provide compelling evidence supporting the sub-optimality of CLSCI for this class of functions. In our simulations, we included two representative functions: $f_1(x) = 100|2x-1|$, and $f_2(x) = 100x$. These functions serve as prototypes for all piecewise linear functions. f_1 is a 1-kink piecewise linear function, and Figure 2 clearly illustrates that the length of $CLSCI$ converges much more slowly than our confidence interval. This slow convergence indicates sub-optimality, as our method is theoretically and empirically shown to align with the minimax rate, as detailed in Section 4.3 of the main paper and Section A.3 of the supplement. On the other hand, $f_2(x) = 100x$ is a linear function, and Figure 4 demonstrates that the lengths of CLSCI hardly converge at all. A rigorous analysis of CLSCI's behavior for convex piecewise linear functions, or non-smooth functions in general, in a non-asymptotic context is a formidable challenge and necessitates the development of new analytical tools. This topic is of independent interest and falls beyond the scope of this paper. However, we provide intuitive reasoning to explain the slow convergence of CLSCI, if it converges at all.

If a convex piecewise linear function f has all its kinks at rational points, there are infinitely many sample sizes n for which f belongs to the function class that convex least squares can precisely estimate. That is, the expectation version of the CLS estimator for those n,

$$\hat{f}_{oracle,n} = \underset{f \text{ is piecewise linear convex function}}{\arg\min} \mathbb{E}_f\Big(\sum_i (f(x_i) - y_i)^2\Big), \tag{B.13}$$

gives $\hat{f}_{oracle,n} = f$. Recall the construction of $CLSCI_\alpha$ in Deng et al. (2020), which we summarized in Equation (B.2). Consider a 1-kink convex piecewise linear function $f = f_1$ and assume n is even. In this case, $\hat{f}_{oracle,n}$ as defined in Equation (B.13) equals f. This implies that the left and right nearest kinks to the minimizer are located at 0 and 1, respectively. Consequently, if $CLSCI_\alpha$ were based on $\hat{f}_{oracle,n}$, it would have a constant length for all even n. Although the actual CLS estimator (i.e., $\hat{f}_n$ defined in Equation (B.1)) produces kinks that are slightly closer to the minimizer compared to the oracle version of CLS (i.e., $\hat{f}_{oracle,n}$), resulting in a shorter expected length of the confidence interval $CLSCI_\alpha$, this improvement is unlikely to completely resolve the issue. As a result, we anticipate that CLSCI remains sub-optimal. The numerical results mentioned earlier provide support for this assertion. Similar arguments apply to linear functions as well.

For a general piecewise linear convex function with the minimizer taking a rational value $\frac{p}{q}$, these arguments hold when n is a multiple of q and sufficiently large. For piecewise linear convex functions f with irrational minimizers, consider linear interpolation on rational grids for f. The same arguments apply in these cases. It is worth noting that these arguments also highlight a conflict between the estimation and inference of the minimizer in CLS-based methods for convex piecewise linear functions. Better estimation from CLS implies a longer length of $CLSCI_\alpha$, so the construction of $CLSCI_\alpha$ after CLS estimation also contributes to sub-optimality.

These examples, in addition to the one we provide in the proof of sub-optimality, demonstrate that the convex least squares component of confidence interval construction is not the sole reason for sub-optimality. However, it does make it challenging to fully exploit the convexity of the true function when constructing a CLS-based confidence interval. In contrast, Algorithm 1 offers a means to fully leverage the convexity of the true function.

On the other hand, when considering our methods under asymptotic conditions or within the classical minimax framework for the class of smooth convex functions (defined in Section B.2), both of which have coarser criteria than the local minimax framework, we achieve the optimal rates of $n^{-\frac{1}{2k+1}}$ for the minimizer. The connection between the local minimax framework and the classical minimax framework is discussed in Section B.2. Further discussion involving CLS (for the estimation of the minimizer) and our estimator for the minimizer is provided in the latter part of Section B.2 and Section B.4.

B.2. Connections With the Classical Minimax Framework: Lower Bounds, Optimality, and Characteristics. In this part, we relate local minimax rates to classical minimax rates, which captures the worst case for a certain function class.

Before going into details, we elaborate on a general comparison. The lower bound provided by our non-asymptotic local minimax framework over a certain function class is no larger than the classical minimax lower bound over the same function class. Because in the classical minimax framework, the Le Cam two-point reduction, in a way, can be considered as a two-point case of Assouad's or Fano's Lemma, which are typical tools for establishing lower bounds for the classical minimax framework. This makes the local minimax rate a stricter criterion, which preserves more information before taking supremum over the function class (i.e., individual functions are treated individually). This strictness/information-preserving property increases the difficulty for constructing adaptive optimal procedures (i.e., attaining the potentially smaller lower bound) but enables characterizing the difficulty of estimating of individual functions and makes establishing the non-superefficiency type of results conceptually possible.

As an illustration, we consider the convex function class with additional smoothness conditions, as in literature the classical minimax rates for both smooth functions and smooth convex functions are extensively investigated. We walk through the procedures translating local minimax rates to classical minimax lower bounds and highlights the following additional implications.

- For the same class of functions, all optimal procedures under non-asymptotic local minimax benchmarks are optimal in the classical sense.

- The local minimax rates established for one class of functions (e.g., convex functions) can be useful for establishing classical minimax lower bounds for another function class (e.g., smooth functions).

Further, we demonstrate that the classical minimax rates for the convex function class are meaningless, which shows the advantage of the non-asymptotic local minimax framework.

The smoothness condition we consider is local smoothness defined around the minimizer. For $k > 1$ and $B \geq B_1 > 0$, the locally smooth convex function class $\Gamma_1(k; B_1, B)$ is defined as

(B.14)
$$\Gamma_1(k; B_1, B) = \left\{ f \in \mathcal{F} : B_1 \leq \varliminf_{t \to Z(f)} \frac{|f(t) - f(Z(f))|}{|t - Z(f)|^k} \leq \varlimsup_{t \to Z(f)} \frac{|f(t) - f(Z(f))|}{|t - Z(f)|^k} \leq B \right\}.$$

A similar smoothness class has been studied by Shoung and Zhang (2001), with the difference being that their smoothness requires the limit to exist and be exactly B (i.e., $B_1 = B$). Later in this section, We will also briefly discuss a global version of smoothness.

The moduli of continuity for the locally smooth convex function class are given by,

$$\hat{\omega}_z(\varepsilon; f) = \sup\{|Z(f) - Z(g)| : \|f - g\|_2 \leq \varepsilon, g \in \Gamma_1(k; B_1, B)\},$$
$$\hat{\omega}_m(\varepsilon; f) = \sup\{|M(f) - M(g)| : \|f - g\|_2 \leq \varepsilon, g \in \Gamma_1(k; B_1, B)\},$$

for any locally smooth convex function $f \in \Gamma_1(k; B_1, B)$.

Further, similar to the proof of Proposition 2.2, we can show that

$$\hat{\omega}_z(\varepsilon; f) \geq \rho_z(\varepsilon; f), \hat{\omega}_m(\varepsilon; f) \geq \rho_m(\varepsilon; f). \tag{B.15}$$

We defer the proof of this inequality to the last part of this section.

Consider function $f_1 : t \mapsto B\left|t - \frac{1}{2}\right|^k$, which is in $\Gamma_1(k; B_1, B)$. Then we can lower bound the classical minimax rate of estimating the minimum for the function class $\Gamma_1(k; B_1, B)$ by:

$$\begin{aligned}
&\inf_{\hat{M}} \sup_{f \in \Gamma_1(k; B_1, B)} E_f|\hat{M} - M(f)| \\
&\geq \sup_{f \in \Gamma_1(k; B_1, B)} \sup_{g \in \Gamma_1(k; B_1, B)} \inf_{\hat{M}} \max_{h \in \{f, g\}} \mathbb{E}_h|\hat{M} - M(h)| \\
&\geq \sup_{g \in \Gamma_1(k; B_1, B)} \inf_{\hat{M}} \max_{h \in \{f_1, g\}} \mathbb{E}_h|\hat{M} - M(h)| \\
&\geq a_1 \rho_m(\varepsilon; f_1) \\
&= a_1 c_{B,k} \varepsilon^{\frac{2k}{2k+1}},
\end{aligned} \tag{B.16}$$

where $c_{B,k} = (\frac{(2k+1)(k+1)}{4k^2})^{\frac{k}{2k+1}} B^{\frac{1}{2k+1}}$.

Similarly, for estimating the minimizer, take $f_1 : t \mapsto B_1 \left|t - \frac{1}{2}\right|^k$, we can lower bound the classical minimax rate by

(B.17)
$$\inf_{\hat{Z}} \sup_{f \in \Gamma_1(k;B_1,B)} E_f|\hat{Z} - Z(f)| \geq a_1 B_1^{-\frac{2}{2k+1}} \left(\frac{(2k+1)(k+1)}{4k^2}\right)^{\frac{1}{2k+1}} \varepsilon^{\frac{2}{2k+1}}.$$

Note that the class of locally smooth convex functions $\Gamma_1(k; B_1, B)$ is a subset of the class of locally smooth functions. Therefore, the lower bounds for $\Gamma_1(k; B_1, B)$ also hold for the class of locally smooth functions. This implies that our local minimax rates, while are primarily based on the properties of convex functions, can also be used to establish lower bounds for the class of locally smooth functions.

Moreover, this technique of establishing lower bounds for one functions class under the classical minimax framework using the local minimax lower bound for another function class has wider applicability. To illustrate this point, we use this approach to establish lower bounds for estimating the minimum for globally smooth function, which is also extensively studied in the literature.

The globally smooth convex function class $\Gamma_2(B, k)$ is defined as

$$\Gamma_2(B, k) = \{f \in \mathcal{F} : |f(t) - f(Z(f))| \leq B|t - Z(f)|^k, \forall t \in [0, 1]\}. \tag{B.18}$$

Note that the global smoothness imposes conditions on the behavior of a function not just around its minimizer, which makes the globally smooth convex function class smaller than the locally smooth convex function class (if we can let $B_1 = 0$ to allow the same form).

The continuity modulus for the globally smooth convex function class can be similarly defined as

$$\tilde{\omega}_m(\varepsilon; f) = \sup\left\{|M(f) - M(g)| : \|f - g\|_2 \leq \varepsilon, g \in \Gamma_2(B, k)\right\}, \tag{B.19}$$

for $f \in \Gamma_2(B, k)$.

Similarly, we can show that

$$\tilde{\omega}_m(\varepsilon; f) \geq \rho_m(\varepsilon; f), \tag{B.20}$$

the proof of which is deferred to the last part.

Inequality (B.20) and similar arguments as in Inequality (B.16) give that the minimax rate for estimation of the minimum for function class $\Gamma_2(B, k)$ is lower bounded by $a_1 c_{B,k} \varepsilon^{\frac{2k}{2k+1}}$ (where $c_{B,k} = (\frac{(2k+1)(k+1)}{4k^2})^{\frac{k}{2k+1}} B^{\frac{1}{2k+1}}$), which automatically serves as a lower bound for the globally smooth function class.

Our discussion on establishing lower bounds under the classical minimax framework and transferring rates from local minimax framework to the classical minimax framework for the white noise model can be extended to the non-parametric regression. Despite the large volume of literature on non-parametric regression, the lower bounds for various smooth classes are not well known. For instance, the lower bound for isotropic Hölder class is not known until lately (Belitser et al., 2021). To get the analogous discussion for the non-parametric regression, we only need to replace ε in the white noise model with $\frac{\sigma^2}{n}$, as the discretization error is always dominated by the noise-induced error for commonly seen smooth function classes in the classical minimax framework.

Now we proceed to see the advantage of local minimax benchmarks compared with classical minimax rates. Consider a collection of functions $f_\delta : t \mapsto \delta|t - \frac{1}{2}|$, for $\delta > 0$. This collection of functions is convex. We have lower bounds (up to some absolute constants) for classical minimax rates for convex functions, given by

$$\lim_{\delta \to 0^+} \rho_z(\varepsilon; f_\delta) = \frac{1}{2},$$
$$\lim_{\delta \to +\infty} \rho_m(\varepsilon; f_\delta) = \infty.$$

Any procedure will be optimal under the classical minimax framework, which makes the classical minimax framework meaningless in this setting.

Finally, we are ready to show that our methods are adaptively optimal for function classes for which the CLS estimator and CLSCI are investigated. The function class for which the CLS estimator and CLSCI are investigated in (Ghosal and Sen, 2017; Deng et al., 2020) can be written as $\cup_{B>0}\Gamma_1(k; B, B)$ for even integer $k \geq 2$. Previous discussion established that our procedures not only achieve the optimal minimax rate in the classical sense (in terms of n) for $\Gamma_1(k; B, B)$ but also have a risk/length smaller than a universal constant multiple of the lower bound for each and every B and k. Our procedures do not depend on B or k, meaning that our procedures are adaptively optimal under the classical setting.

Proof of Inequality (B.15) and Inequality (B.20). The proofs are similar to the proof of Proposition 2.2. Using the same notation as in Proposition 2.2, t_l and t_r are the left and right endpoints of the interval $\{t : f(t) \leq M(f) + \rho_m(\varepsilon; f)\}$.

To prove Inequality (B.20) (i.e., $\tilde{\omega}_m(\varepsilon; f) \geq \rho_m(\varepsilon; f)$), we only need to replace $g_\delta(t)$ in the proof of Proposition 2.2 by $\bar{g}_\delta$ defined as

$$\bar{g}_\delta(x) = \begin{cases} f(x) & x \notin [t_l, t_r] \\ \mu_\varepsilon + \delta(\left|\frac{x-Z(f)}{t_l-Z(f)}\right|^k - 1) & t_l \le x \le Z(f) \\ \mu_\varepsilon + \delta(\left|\frac{x-Z(f)}{t_r-Z(f)}\right|^k - 1) & Z(f) \le x \le t_r \end{cases}, \tag{B.21}$$

for $k \ge 1$ and $0 < \delta \le \min\{B|t_l - Z(f)|^k, B|t_r - Z(f)|^k, \frac{\rho_m(\varepsilon;f)}{k}\}$. It is easy to see that this new $\bar{g}_\delta \in \Gamma_2(B,k)$, $\|\bar{g}_\delta - f\| \le \varepsilon$ and $\lim_{\delta\to 0} |M(\bar{g}_\delta) - M(f)| = \rho_m(\varepsilon; f)$. When $k < 1$, we just replace the k in the newly constructed $\bar{g}_\delta$ with 1.

To prove Inequality (B.15) (i.e., $\hat{\omega}_z(\varepsilon; f) \ge \rho_z(\varepsilon; f)$ and $\hat{\omega}_m(\varepsilon; f) \ge \rho_m(\varepsilon; f)$), without loss of generality, we assume $t_r - Z(f) = \rho_z(\varepsilon; f)$. Note that $k > 1$. We only need to replace $g_\delta(t)$ in the proof of Proposition 2.2 to be $\tilde{g}_\delta(t)$, which is defined in the following way: let $h_s(t) = B|t - t_r + \delta|^k + s$, as when δ is small enough, $\forall t > t_r - \delta$, $\frac{f(t)-f(t_r-\delta)}{t-t_r+\delta}$ is lower bounded by $\frac{\lim_{t\to t_r^-} f(t_r)-f(t)}{2}$, so $\exists s$ such that $h_s(t)$ and $g_\delta(t)$ has an intersection $t_1 \in (t_l, t_r - \delta)$ and an intersection $t_2 \in (t_r - \delta, t_r)$, which satisfy $h_s(t) > g_\delta(t), \forall t \in (t_1, t_2)$ and $h_s(t) < g_\delta(t)$ for a small neighborhood outside (t_1, t_2) on both sides.

Define $\tilde{g}_\delta$ by

$$\tilde{g}_\delta(t) = \begin{cases} g_\delta(t) & t \in [0,1]\backslash(t_1, t_2) \\ h_s(t) & t \in (t_1, t_2) \end{cases}. \tag{B.22}$$

Then $\tilde{g}_\delta \in \Gamma_1(k; B_1, B)$, $\|\tilde{g} - f\| \le \varepsilon$, $\lim_{\delta\to 0} |Z(\tilde{g}_\delta) - Z(f)| = \rho_z(\varepsilon; f)$, and $\lim_{\delta\to 0} |M(\tilde{g}_\delta) - M(f)| \ge \lim_{\delta\to 0} |M(g_\delta) - M(f)| = \rho_m(\varepsilon; f)$. □

B.3. More on the Uncertainty Principle. In this subsection, we elaborate on the generality of the Uncertainty Principle. We start with the convex smoothness class we discussed in Section B.2. Uncertainty principle still holds for the function class $\Gamma_1(k; B_1, B)$ defined in (B.14), which contains all the functions $f \in \mathcal{F}$ satisfying

$$B_1 \le \varliminf_{t\to Z(f)} \frac{|f(t) - f(Z(f))|}{|t - Z(f)|^k} \le \varlimsup_{t\to Z(f)} \frac{|f(t) - f(Z(f))|}{|t - Z(f)|^k} \le B.$$

It follows from Inequality (B.15) that the moduli of continuity for the minimizer and minimum over the function class $\Gamma_1(k; B_1, B)$ have the following relationship.

$$\hat{\omega}_z(\varepsilon; f)\hat{\omega}_m(\varepsilon; f)^2 \ge \rho_z(\varepsilon; f)\rho_m(\varepsilon; f)^2 \ge \frac{\varepsilon^2}{2}. \tag{B.23}$$

So the Uncertainty Principle also holds for $\Gamma_1(k; B_1, B)$.

Further, applying the smoothing technique in Equation (B.8) in the proof of Proposition B.2 to $\tilde{g}_\delta$ defined by Equation (B.22), we know that the Uncertainty Principle also holds for the k-th order differentiable convex function class for any k.

Therefore, there are many choices of subclasses of $\mathcal{F}$ where the Uncertainty Principle holds. Interested readers can further explore other possible choices. Further, given the prevalent occurrence of tension between different quantities (e.g., minimizer and minimum in our case), we believe that similar Uncertainty Principles can be developed in diverse scenarios.

B.4. Comparison with the CLS Estimator for Minimizer. We now turn to a comparison of the CLS estimator and our proposed estimator for the minimizer.

Analyzing the behavior of the CLS estimator within our framework poses significant challenges. On the other hand, the theoretical analysis for our method easily gives a corollary that our estimator achieves the same optimal rate as the CLS estimator under a coarser criterion — in asymptotic sense with functions that have positive second order derivatives — the same context in which the CLS estimator is typically studied. Numerical results demonstrate that the behavior of the CLS estimator aligns with our methods, albeit with sensitivity to the smoothness of the functions. Now we proceed with details.

Existing theoretical results for the CLS estimator of the minimizer are both asymptotic and for a fixed function with strong regularity assumptions such as twice differentiability. The tools used in establishing the performance of the CLS estimator in the literature are insufficient for studying its behavior under our non-asymptotic local minimax framework for general convex functions without smoothness conditions. Therefore, the properties of the CLS estimator under our framework for convex function class are unclear and difficult to analyze.

More precisely, existing analyses of the CLS estimator are based on the limiting distribution, which is usually obtained by performing a second-order Taylor expansion of the function around the minimizer and analyzing the resulting empirical process. However, the limiting distribution only holds with the sample size going to infinity for a fixed function, so similar arguments can not lead to results that hold uniformly for all functions within a function class, regardless of whether the sample size is fixed or growing. Additionally, the Taylor expansion approach is not applicable when the second-order derivative does not exist at the minimizer. As a result, analyzing the behavior of the

CLS estimator under our local minimax framework requires new tools and is of separate interest.

On the other hand, theoretical results for our estimator can be easily transferred into one that uses the same criterion used in theoretical results for the CLS estimator — asymptotic results for convex functions with positive second order derivatives at the minimizer. Under this criterion, our estimator and the CLS estimator have the same optimal rate.

For functions that are twice differentiable around the minimizer with a positive second-order derivative at the minimizer, the boundedness of the CLS estimator for the minimizer $\hat{Z}_{\text{cvx}}$ (i.e., $\hat{Z}_{\text{cvx}} \in [0, 1]$) and its limiting distribution (Theorem 2.9 in Deng et al. (2020)) give that

$$\limsup_{n\to\infty} \mathbb{E}(|\hat{Z}_{\text{cvx}} - Z(f)|)(n/\sigma^2)^{1/5} \leq \left(\frac{1}{f''(Z(f))}\right)^{2/5} \text{const}_1,$$

where const_1 is an absolute (positive) constant, and that

$$\liminf_{n\to\infty} \mathbb{E}(|\hat{Z}_{\text{cvx}} - Z(f)|)(n/\sigma^2)^{1/5} \geq \left(\frac{1}{f''(Z(f))}\right)^{2/5} \text{const}_2,$$

where const_2 is another absolute (positive) constant. Note that for functions twice differentiable at the minimizer with a positive second-order derivative, the key part of the benchmark for the minimizer in our framework $\rho_z(\frac{\sigma}{\sqrt{n}}; f)$ is of the order $(\sigma^2/n)^{1/5}\left(\frac{1}{f''(Z(f))}\right)^{2/5}$ when n goes to infinity. Although the benchmark has a discretization part as shown in Section C.11, it can be easily verified that the order of the benchmark remains $(\sigma^2/n)^{1/5}\left(\frac{1}{f''(Z(f))}\right)^{2/5}$ when f is fixed and n goes to infinity. In this asymptotic sense, the CLS estimator matches our rate, which is also the optimal rate, for functions twice differentiable at the minimizer with a positive second-order derivative (the lower bound provided in Section B.2). However, this match in rate is under a coarser criterion and does not imply optimality for $\hat{Z}_{\text{cvx}}$ under our non-asymptotic framework.

Now we look at the numerical experiments we have shown in Section A. Figure 6 shows that the CLS estimator and our methods have almost the same behavior for $f(x) = 100(|2x-1|)^2$, a function with positive second order derivative. However, the performance of the CLS estimator, when compared with that of our estimator, deteriorates as the smoothness of the underlying function grows, as shown in Figure 10 and Figure 12, and improves as the smoothness of functions decreases, as shown in Figure 2 and Figure 4. In contrast, our method is stable in terms of the smoothness of the functions, as shown in the comparison with the theoretical benchmarks in Section A.3.

APPENDIX C: PROOFS OF THE RESULTS IN THE MAIN PAPER

This section provides the proofs of all the main results presented in the paper "Estimation and Inference for Minimizer and Minimum of Convex Functions: Optimality, Adaptivity, and Uncertainty Principles".

C.1. Notation, Lemmas and Basic Properties.

We begin by introducing and recollecting notation that will be frequently used in the proofs.

Note that Y_l, Y_s, and Y_e are defined on the same probability space. We use $\mathbb{E}_s$ to denote the expectation with respect to the distribution of Y_s and so on. We denote by i_j^* the index for the subinterval at level j that contains the minimizer $Z(f)$; we denote by $\tilde{j}$ the index for the level where the chosen interval is at least two blocks away from the subinterval containing the minimizer, i.e.,

$$i_j^* = \max\{i : Z(f) \in [t_{j,i-1}, t_{j,i}]\}, \qquad \tilde{j} = \min\{j : |\hat{i}_j - i_j^*| \geq 2\}. \tag{C.1}$$

It is easy to see that $\tilde{j} \geq 2$, and $\tilde{j}$ only depends on Y_l. In addition, we let

$$j^* = \min\{j : m_j \leq \frac{\rho_z(\varepsilon; f)}{4}\}. \tag{C.2}$$

Then by this definition, $\frac{\rho_z(\varepsilon;f)}{8} < m_{j^*} \leq \frac{\rho_z(\varepsilon;f)}{4}$. Furthermore, $\mu_{j,i}$ denotes the average of f on interval $[t_{j,i-1}, t_{j,i}]$, i.e.,

$$\mu_{j,i} = \frac{1}{m_j} \int_{t_{j,i-1}}^{t_{j,i}} f(t) dt. \tag{C.3}$$

We now list the notation that is used throughout the proofs of theorems in Section 3, in case readers get lost in the middle of reading a proof.

$$\begin{aligned}
i_j^* =& \max\{i : Z(f) \in [t_{j,i-1}, t_{j,i}]\}, \quad \tilde{j} = \min\{j : |\hat{i}_j - i_j^*| \geq 2\}, \\
\mu_{j,i} =& \frac{1}{m_j} \int_{t_{j,i-1}}^{t_{j,i}} f(t) dt, \quad j^* = \min\{j : m_j \leq \frac{\rho_z(\varepsilon; f)}{4}\}, \\
\mathcal{E}_{j,i} =& \frac{1}{\sqrt{m_j}} \left(W_2(t_{j,i}) - 2W_2(t_{j,i-1}) + W_2(t_{j,i-2}) \right), \\
j^w =& \min\{j : |\hat{i}_j - i_j^*| \geq 5\}, \quad \hat{f} = \frac{1}{m_{\hat{j}}} \int_{t_{\hat{i}_{\hat{j}}+\Delta-1}}^{t_{\hat{i}_{\hat{j}}+\Delta}} f(t) dt, \\
\Delta =& 2 \left(\mathbb{1}\{\tilde{X}_{\hat{j},\hat{i}_{\hat{j}}+6} - \tilde{X}_{j,\hat{i}_{\hat{j}}+5} \leq 2\sigma_j\} - \mathbb{1}\{\tilde{X}_{\hat{j},\hat{i}_{\hat{j}}-6} - \tilde{X}_{j,\hat{i}_{\hat{j}}-5} \leq 2\sigma_j\} \right).
\end{aligned} \tag{C.4}$$

In the data splitting step of the white noise model, we obtain three independent copies of the observations: Y_l, Y_s, and Y_e. While we let them have equal variance ($3\varepsilon^2$), it is not necessary. We denote the variances of Y_l, Y_s, and Y_e as $c_l^2\varepsilon^2$, $c_s^2\varepsilon^2$, and $c_e^2\varepsilon^2$, respectively, in the supplement. This helps demonstrate how the results depend on variance and allows for easy derivation of analogous results for modified splitting procedures.

Similar for regression model, the splitting procedure for the regression model can be modified to allow different variances of the three copies $\{y_{l,\cdot}\}$, $\{y_{s,\cdot}\}$, and $\{y_{e,\cdot}\}$. We denote the scaling factors for the three copies as γ_l, γ_s, and γ_e, respectively, i.e., for all i, $\mathrm{Var}(y_{l,i}) = \gamma_l^2\sigma^2$, $\mathrm{Var}(y_{s,i}) = \gamma_s^2\sigma^2$, and $\mathrm{Var}(y_{e,i}) = \gamma_e^2\sigma^2$.

For the regression model, we use similar notion for the length of subinterval, the index of the interval in which the minimizer lies, etc. The following notation is used in the proofs of the results for regression model.

$$
\begin{aligned}
& m_j = \frac{2^{J-j}}{n}, \qquad \mathtt{t}_{j,i} = i \cdot m_j - \frac{1}{n}, \\
& \mathtt{i}_j^* = \max\{i : Z(f) \in [\mathtt{t}_{j,i-1} + \frac{1}{2n}, \mathtt{t}_{j,i} + \frac{1}{2n}]\}, \\
& \tilde{\mathtt{j}} = \min\{\min\{j : |\hat{\mathtt{i}}_j - \mathtt{i}_j^*| \geq 2\}, \infty\}, \\
& \mathtt{j}^* = \min\{j : m_j \leq \frac{\rho_z(\frac{\sigma}{\sqrt{n}}; f)}{4}\}, \qquad \mathtt{j}^{\mathtt{w}} = \min\{j : \left|\hat{\mathtt{i}}_j - \mathtt{i}_j^*\right| \geq 5\}, \\
& \mathtt{Y}_x = \{y_{x,0}, y_{x,1}, \cdots, y_{x,n}\}, \text{ for } x = l, s, e, \\
& ave_f(j,i) = \frac{1}{2^{J-j}} \sum_{k=2^{J-j}(i-1)}^{2^{J-j}\cdot i-1} f(x_k), \\
& \mathfrak{E}_{j,i,x} = Y_{j,i,x} - ave_f(j,i) \cdot 2^{J-j}, \qquad \hat{\mathtt{f}} = ave_f(\hat{\mathtt{j}}, \tilde{\mathtt{i}}_{\tilde{\mathtt{j}}}).
\end{aligned} \tag{C.5}
$$

To keep the logic flow neat, additional notation for non-parametric regression are introduced in Section C.11.

We finish this part by recalling some of the basic properties that are frequently used in the proofs. The proofs for these properties are deferred to later sections. Firstly, we revisit a basic property for convex functions.

LEMMA C.1. *For a convex function f, and any $0 \leq x_1 < x_2 < x_3 \leq 1$, we have*

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(x_3) - f(x_1)}{x_3 - x_1} \leq \frac{f(x_3) - f(x_2)}{x_3 - x_2}.$$

Next we introduce the following lemma that helps with detailed calculation.

Lemma C.2. *For $x > 6^{1/3}$, we have*

$$\frac{2x\Phi(2-(2x)^{\frac{3}{2}}\sqrt{2/3})}{x\Phi(2-\sqrt{2/3}x^{3/2})} < 0.008,$$

where Φ is the Cumulative Density Function (CDF) of a standard normal distribution.

We further introduce two quantities that will be often used in the proofs of the theorems in Section 3 of the main paper. Let

$$Q = \sup_{x\geq 0} x^2\Phi(-x) \quad \text{and} \quad V = \sup_{x\geq 0} x^2\Phi(2-x), \tag{C.6}$$

for which we have the following results.

Lemma C.3.

$$Q = \sup_{x\geq 0} x^2\Phi(-x) \leq 0.169, \quad V = \sup_{x\geq 0} x^2\Phi(2-x)V < 2.0555. \tag{C.7}$$

C.2. Proof of Proposition 2.1. We begin by proving the statement with respect to the minimum. That is, for $\varepsilon > 0$, $f \in \mathcal{F}$, and $c \in (0,1)$,

$$c \leq \frac{\rho_m(c\varepsilon; f)}{\rho_m(\varepsilon; f)} \leq c^{\frac{2}{3}}. \tag{C.8}$$

Proof. Without loss of generality, we assume $M(f) = 0$. We first prove the left hand side. Define the β-indexed function on $[0,1]$, g_β, as

$$g_\beta := t \mapsto \max\{f(t), \rho_m(\beta\varepsilon; f)\},$$

and it is not difficult to see that

$$\|g_1 - f\|^2 = \varepsilon^2, \qquad \|g_c - f\|^2 = c^2\varepsilon^2. \tag{C.9}$$

Define function $\tilde{g}$ on $[0,1]$ as $\tilde{g} := t \mapsto \max\{f(t), c\rho_m(\varepsilon; f)\}$. Let $t_{l,m}$ and $t_{r,m}$ be the left and right end point of the interval $\{t \in [0,1] : f(t) \leq \tilde{g}(t)\} = \{t \in [0,1] : f(t) \leq c\rho_m(\varepsilon; f)\}$. Clearly, $[t_{l,m}, t_{r,m}] \subset \{t : f(t) \leq \rho_m(\varepsilon; f)\}$, which gives

$$\|\tilde{g} - f\|^2 = \int_{t_{l,m}}^{t_{r,m}} (c\rho_m(\varepsilon; f) - f(t))^2 dt \tag{C.10}$$

$$\leq \int_{t_{l,m}}^{t_{r,m}} c^2(\rho_m(\varepsilon; f) - f(t))^2 dt \tag{C.11}$$

$$\leq c^2\|g_1 - f\|^2 = c^2\varepsilon^2 = \|g_c - f\|^2. \tag{C.12}$$

Therefore, $\tilde{g} \leq g_c$ at all the points. This gives $c\rho_m(\varepsilon; f) \leq \rho_m(c\varepsilon; f)$.

Now we turn to the right hand side, which can be reduced to finding the value of

$$\inf_{f\in\mathcal{F}} \frac{\rho_m(\varepsilon; f)}{\rho_m(c\varepsilon; f)}.$$

Let the left side and right side of the "water area" with "water level" $\rho_m(c\varepsilon; f)$ be

$$x_{l,m} = \min\{x : g_c(x) \geq f(x)\}, x_{r,m} = \max\{x : g_c(x) \geq f(x)\}. \tag{C.13}$$

The rest of the proofs can be divided into four steps.

1) The first step shows that taking the infimum of $\frac{\rho_m(\varepsilon;f)}{\rho_m(c\varepsilon;f)}$ over $\mathcal{F}$ is the same as over the function class

(C.14)

$$\mathcal{F}_l = \big\{f \in \mathcal{F} : f\big|_{[0,x_{l,m}]}, f\big|_{[x_{r,m},1]} \text{ are linear functions with slopes } f'(x_{l,m}+), \text{ and} f'(x_{r,l}-)\big\}.$$

2) The second step shows that it is further no smaller than taking the infimum over the function class

$$\mathcal{F}_{ll} = \big\{f \in \mathcal{F} : f\big|_{[0,Z(f)]}, f\big|_{[Z(f),1]} \text{ are piece-wise linear functions with at most two pieces}, f\big|_{[0,x_{l,m}]}, f\big|_{[x_{r,m},1]} \text{ are linear functions } \big\}.$$

3) In the third step, we define two extended function spaces

$$\tilde{\mathcal{F}}_c = \big\{f \text{ is convex function with unique minimizer on } (-\infty,\infty) : f\big|_{(-\infty,0]}, f\big|_{[1,\infty)]} \text{ are linear functions}, f\big|_{[0,1]} \in \mathcal{F}\big\},$$

$$\tilde{\mathcal{F}}_{ll} = \big\{f \in \tilde{\mathcal{F}}_c : f\big|_{(-\infty,Z(f)]} \text{ and } f\big|_{[Z(f),\infty)} \text{are piece-wise linear functions with at most three pieces } \big\}.$$

as well as two extended geometric quantities $\tilde{\rho}_z(\varepsilon; f)$, $\tilde{\rho}_m(\varepsilon; f)$ for $f \in \tilde{\mathcal{F}}_c$:

$$\tilde{\rho}_z(\varepsilon; f) = \max\{|t - Z(f)| : f(t) \leq \mu(\varepsilon; f)\},\ \tilde{\rho}_m(\varepsilon; f) = \mu(\varepsilon; f) - M(f),$$

where $\mu(\varepsilon; f)$ satisfies that for function f_μ defined on $\mathbb{R}$ as $f_\mu : t \mapsto \max\{\mu(\varepsilon; f), f(t)\}$, the following holds: $\|f_\mu - f\|^2 = \varepsilon^2$. In this step, we show that

$$\inf_{f\in\mathcal{F}_{ll}} \frac{\rho_m(\varepsilon; f)}{\rho_m(c\varepsilon; f)} \geq \inf_{f\in\tilde{\mathcal{F}}_{ll}} \frac{\tilde{\rho}_m(\varepsilon; f)}{\tilde{\rho}_m(c\varepsilon; f)}.$$

4) Finally, in the fourth step, we show that

$$\inf_{f\in\tilde{\mathcal{F}}_{ll}} \frac{\tilde{\rho}_m(\varepsilon; f)}{\tilde{\rho}_m(c\varepsilon; f)} \geq \inf_{f\in\tilde{\mathcal{F}}_{L}} \frac{\tilde{\rho}_m(\varepsilon; f)}{\tilde{\rho}_m(c\varepsilon; f)} = c^{-\frac{2}{3}},$$

where $\tilde{\mathcal{F}}_L = \{f \in \tilde{\mathcal{F}}_{ll} : f\big|_{(-\infty,Z(f)]}$ and $f\big|_{[Z(f),\infty)}$ are linear functions$\}$.

Step 1. We start with defining a mapping $L_1 : \mathcal{F} \longrightarrow \mathcal{F}_l$, i.e., L_1 maps a function in $\mathcal{F}$ to a function in $\mathcal{F}_l$. Then we will show that

$$\frac{\rho_m(\varepsilon; f)}{\rho_m(c\varepsilon; f)} \geq \frac{\rho_m(\varepsilon; L_1(f))}{\rho_m(c\varepsilon; L_1(f))}, \tag{C.15}$$

by showing

$$\rho_m(c\varepsilon; f) = \rho_m(c\varepsilon; L_1(f)), \text{ and} \tag{C.16a}$$

$$\rho_m(\varepsilon; f) \geq \rho_m(\varepsilon; L_1(f)). \tag{C.16b}$$

Then Inequality (C.15) gives

$$\inf_{f\in\mathcal{F}} \frac{\rho_m(\varepsilon; f)}{\rho_m(c\varepsilon; f)} \geq \inf_{f\in\mathcal{F}} \frac{\rho_m(\varepsilon; L_1(f))}{\rho_m(c\varepsilon; L_1(f))} \geq \inf_{f\in\mathcal{F}_l} \frac{\rho_m(\varepsilon; f)}{\rho_m(c\varepsilon; f)}. \tag{C.17}$$

Granting Inequality (C.17) holds, to prove the statement of the first step, we only need to show that

$$\inf_{f\in\mathcal{F}} \frac{\rho_m(\varepsilon; f)}{\rho_m(c\varepsilon; f)} \leq \inf_{f\in\mathcal{F}_l} \frac{\rho_m(\varepsilon; f)}{\rho_m(c\varepsilon; f)}. \tag{C.18}$$

Inequality (C.18) indeed holds as $\mathcal{F}_l \subset \mathcal{F}$.

Now we give the precise definition of L_1 by giving the value of $L_1(f)(t)$ for $[0,1]$, and show that Inequalities (C.16) hold. Convexity of f ensures the existence of one-sided derivatives. Let

$$(L_1(f))(t) = \begin{cases} f(x_{l,m}) + f'(x_{l,m}+) \cdot (t - x_{l,m}) & 0 \leq t < x_{l,m} \\ f(t) & t \in [x_{l,m}, x_{r,m}] \\ f(x_{r,m}) + f'(x_{r,m}-) \cdot (t - x_{r,m}) & 1 \geq t > x_{r,m} \end{cases}.$$

Clearly, Inequality (C.16a) holds. Now we will show that Inequality (C.16b) holds. Without loss of generality, we can assume $M(f) = 0$. It is clear that $M(L_1(f)) = 0, \quad L_1(f)(t) \leq f(t)\, \forall t \in [0,1]$. Define a function $\tilde{L_1}(f)$ on $[0,1]$ as $\tilde{L_1}(f) : t \mapsto \max\{L_1(f)(t), \rho_m(\varepsilon; f)(t)\}$. Then we have

$$\begin{aligned} \|\tilde{L_1}(f) - L_1(f)\|^2 &= \int_0^1 \left((\rho_m(\varepsilon; f) - L_1(f)(t))_+\right)^2 dt \\ &\geq \int_0^1 \left((\rho_m(\varepsilon; f) - f(t))_+\right)^2 dt = \varepsilon^2. \end{aligned} \tag{C.19}$$

This inequality gives Inequality (C.16b).

Step 2. Similarly to Step 1, we will define a mapping $L_2 : \mathcal{F}_l \longrightarrow \mathcal{F}_{ll}$ that maps a function in $\mathcal{F}_l$ to a function in $\mathcal{F}_{ll}$, such that the following inequalities hold.

$$\rho_m(c\varepsilon; f) = \rho_m(c\varepsilon; L_2(f)), \text{ and} \tag{C.20a}$$
$$\rho_m(\varepsilon; f) \geq \rho_m(\varepsilon; L_2(f)). \tag{C.20b}$$

Then, similar arguments as in Step 1 will give the statement of Step 2. To define $L_2(f)$, we define a sequence of functions $\{h(\delta; f)\}_{\delta>0}$ and then pick one in this sequence. We first introduce two quantities:

$$l(\delta; f) := \min\{t \in [0,1] : f(t) \leq \delta + M(f)\},$$
$$r(\delta; f) := \max\{t \in [0,1] : f(t) \leq \delta + M(f)\}.$$

When there is no ambiguity, we will omit f, resulting in $l(\delta)$, and $r(\delta)$. Now we define four functions l_1, $l_{2,\delta}$, $l_{3,\delta}$, and l_4 on $\mathbb{R}$. Recall the definition of $x_{l,m}$ and $x_{r,m}$ in (C.13).

$$l_1(t) = \begin{cases} \frac{f(x_{l,m})-f(0)}{x_{l,m}} t + f(0), & \text{if } x_{l,m} > 0, \\ (t - x_{l,m}) \lim_{s\to 0^+} \frac{f(x_{l,m}+s)-f(x_{l,m})}{s} + f(x_{l,m}), & \text{if } Z(f) > x_{l,m} = 0, \\ M(f), & \text{if } Z(f) = x_{l,m} = 0, \end{cases}$$

$$l_{2,\delta}(t) = \begin{cases} \frac{\delta}{l(\delta)-Z(f)}(t - Z(f)) + M(f), & \text{if } Z(f) > 0, \\ M(f), & \text{if } Z(f) = 0, \end{cases}$$

$$l_{3,\delta}(t) = \begin{cases} \frac{\delta}{r(\delta)-Z(f)}(t - Z(f)) + M(f), & \text{if } Z(f) < 1, \\ M(f), & \text{if } Z(f) = 1, \end{cases}$$

$$l_4(t) = \begin{cases} \frac{f(1)-f(x_{r,m})}{1-x_{r,m}}(t - x_{r,m}) + f(x_{r,m}), & \text{if } x_{r,m} < 1, \\ (x_{r,m} - t) \lim_{s\to 0^+} \frac{f(x_{r,m}-s)-f(x_{r,m})}{s} + f(x_{r,m}), & \text{if } Z(f) < x_{r,m} = 1, \\ M(f), & \text{if } Z(f) = x_{r,m} = 1. \end{cases}$$

Based on these four functions, we define a new function $h(\delta; f)$ on $[0,1]$:

$$h(\delta; f) : t \mapsto \max\{l_1(t), l_{2,\delta}(t), l_{3,\delta}(t), l_4(t)\}.$$

When there is no ambiguity on f, we will denote $h(\delta; f)$ as $h(\delta)$. Clearly,

$$h(\delta_1)(t) \geq h(\delta_2)(t), \text{ for all } t \in [0,1], \text{ when } \delta_1 \geq \delta_2.$$

This, along with the continuity of f implies that $\rho_m(c\varepsilon; h(\delta))$ increases continuously as δ increases. Further, for $\delta = \rho_m(c\varepsilon; f)$ and $\delta \to 0^+$, we have

$$\rho_m(c\varepsilon; h(\rho_m(c\varepsilon; f))) \geq \rho_m(c\varepsilon; f), \qquad \lim_{\delta\to 0^+} \rho_m(c\varepsilon; h(\delta)) \leq \rho_m(c\varepsilon; f).$$

Further, inequality $\lim_{\delta\to 0^+}\rho_m(c\varepsilon; h(\delta)) \le \rho_m(c\varepsilon; f)$ takes equality only when both $f|_{[0,Z(f)]}$ and $f|_{[Z(f),1]}$ are linear functions. Therefore, $\exists\delta_0 \in (0, \rho_m(c\varepsilon; f)]$ such that $\rho_m(c\varepsilon; h(\delta_0)) = \rho_m(c\varepsilon; f)$. We define $L_2(f)$ to be $h(\delta_0)$. Consequently, Inequality (C.20a) holds. It is also easy to check $L_2(f) \in \mathcal{F}_{ll}$. We use the following shorthand $h := h(\delta_0)$ when there is no need to emphasize δ_0. Now we will prove $\rho_m(\varepsilon; h(\delta_0)) \le \rho_m(\varepsilon; f)$ (Inequality (C.20b)) by proving $\|h - g_1\| \ge \|f - g_1\| = \varepsilon$. By $\rho_m(c\varepsilon; h(\delta_0)) = \rho_m(c\varepsilon; f)$, $\delta_0 \le \rho_m(c\varepsilon; f)$, and the construction of $h(\delta)$, we have

$$\begin{aligned}
&\{t : h(\delta_0)(t) \le \rho_m(c\varepsilon; f)\} = [x_{l,m}, x_{r,m}],\\
&[0,1]/[x_{l,m}, x_{r,m}] \subset \{t : h(\delta_0)(t) = f(t)\}.
\end{aligned}$$

Further, by the construction of $h(\delta)$, we have

$$f(t) \le (h(\delta))(t), \text{ for } t \in [l(\delta), r(\delta)], \quad f(t) \ge (h(\delta))(t), \text{ for } t \notin [l(\delta), r(\delta)],$$

Therefore, we have

$$\begin{aligned}
0 =& \|f - g_c\|^2 - \|h(\delta_0) - g_c\|^2\\
=& \int_{x_{l,m}}^{x_{r,m}} \left((f(t) - g_c(t))^2 - ((h(\delta_0)(t) - g_c(t))^2\right) dt\\
=& \int_{x_{l,m}}^{x_{r,m}} (h - f)(2g_c - f - h)dt\\
\ge& \int_{(x_{l,m}, l(\delta_0))\cup(r(\delta_0), x_{r,m})} 2(h - f)(\rho_m(c\varepsilon; f) - \delta_0)dt\\
&+ \int_{[l(\delta_0), r(\delta_0)]} 2(h - f)(\rho_m(c\varepsilon; f) - \delta_0)dt\\
\ge& 2(\rho_m(c\varepsilon; f) - \delta_0)\int_0^1 (h - f)dt.
\end{aligned} \tag{C.21}$$

It then follows that

$$\begin{aligned}
&\|h - g_1\|^2 - \|f - g_1\|^2\\
=& \int_{x_{l,m}}^{x_{r,m}} \left((h - g_1)^2 - (f - g_1)^2\right) dt\\
=& \int_{x_{l,m}}^{x_{r,m}} (h - f)(f + h - 2g_1)dt\\
=& \int_{x_{l,m}}^{x_{r,m}} (h - f)(f + h - 2g_c)dt + \int_{x_{l,m}}^{x_{r,m}} 2(h - f)(g_c - g_1)dt\\
=& \|h - g_c\|^2 - \|f - g_c\|^2 + 2(\rho_m(c\varepsilon; f) - \rho_m(\varepsilon; f))\int_0^1 (h - f)dt \ge 0.
\end{aligned} \tag{C.22}$$

As a result, $\rho_m(\varepsilon; h) \le \rho_m(\varepsilon; f)$, which is $\rho_m(\varepsilon; L_2(f)) \le \rho_m(\varepsilon; f)$. This is Inequality (C.20b).

Step 3. First, we show that $\tilde{\rho}_z(\varepsilon; f)$ and $\tilde{\rho}_m(\varepsilon; f)$ are well defined for functions in $\tilde{\mathcal{F}}_{ll}$. As $\tilde{\mathcal{F}}_{ll} \subset \tilde{\mathcal{F}}_c$, it is sufficient to show that $\tilde{\rho}_z(\varepsilon; f)$ and $\tilde{\rho}_m(\varepsilon; f)$ are well defined for functions in $\tilde{\mathcal{F}}_c$. This holds true as for any function $f \in \tilde{\mathcal{F}}_c$, f has a unique minimizer.

Now for each $f \in \mathcal{F}_{ll}$, we will define a class of functions $L_3(f) = \{\tilde{f}_{\delta_1,\delta_2} \in \tilde{\mathcal{F}}_{ll} : \delta_1 > 0, \delta_2 > 0\}$ such that

$$\tilde{\rho}_m(\varepsilon; \tilde{f}_{\delta_1,\delta_2}) \le \rho_m(\varepsilon; f), \quad \liminf_{\max\{\delta_1,\delta_2\}\to 0^+} \tilde{\rho}_m(c\varepsilon; \tilde{f}_{\delta_1,\delta_2}) \ge \rho_m(c\varepsilon; f). \tag{C.23}$$

We define function $\tilde{f}_{\delta_1,\delta_2}$ by defining its values on three intervals, $(-\infty, 0)$, $[0,1]$, and $(1,\infty)$. Specifically, for $t \in [0,1]$,

$$\tilde{f}_{\delta_1,\delta_2}(t) = f(t),$$

for $t \in (-\infty, 0)$,

$$\tilde{f}_{\delta_1,\delta_2}(t) = \begin{cases} f(0) + \frac{f(x_{l,m}) - f(0)}{x_{l,m}} t, & x_{l,m} > 0 \\ f(0) + \min\{-\delta_1^{-1}, \lim_{s\to 0^+} \frac{f(s)-f(0)}{s}\} t, & x_{l,m} = 0 \end{cases},$$

and for $t \in (1,\infty)$,

$$\tilde{f}_{\delta_1,\delta_2}(t) = \begin{cases} f(1) + \frac{f(x_{r,m}) - f(1)}{x_{r,m}-1}(t-1), & x_{r,m} < 1 \\ f(1) + \max\{\delta_r^{-1}, \lim_{s\to 0^+} \frac{f(1)-f(1-s)}{s}\}(t-1), & x_{l,m} = 1 \end{cases}.$$

Clearly, $\tilde{f}_{\delta_1,\delta_2} \in \tilde{\mathcal{F}}_{ll}$. For ease of presentation, we extend the meaning of $\max\{\cdot,\cdot\}$ to allow function-value arguments in the remaining of this proof. Suppose g is a function defined on $\mathcal{X}$ and C is a constant, then $\max\{g, C\}$ or $\max\{C, g\}$ gives a function on $\mathcal{X}$: $x \mapsto \max\{g(x), C\}$.

We proceed with showing Inequality (C.23) using the definition of $L_3(f)$.

Note that for any $\xi \in (0, \rho_m(c\varepsilon; f))$,

$$\begin{aligned} &\lim_{\max\{\delta_1,\delta_2\}\to 0^+} \|\max\{\tilde{f}_{\delta_1,\delta_2}, M(f) + \rho_m(c\varepsilon; f) - \xi\} - \tilde{f}_{\delta_1,\delta_2}\| \\ &= \|\max\{f, M(f) + \rho_m(c\varepsilon; f) - \xi\} - f\| < c\varepsilon. \end{aligned}$$

Therefore,

$$\liminf_{\max\{\delta_1,\delta_2\}\to 0^+} \tilde{\rho}_m(c\varepsilon; \tilde{f}_{\delta_1,\delta_2}) \ge \rho_m(c\varepsilon; f) - \xi.$$

Since it holds for any $\xi \in (0, \rho_m(c\varepsilon; f))$, let $\xi \to 0^+$, we have

$$\liminf_{\max\{\delta_1,\delta_2\}\to 0^+} \tilde{\rho}_m(c\varepsilon; \tilde{f}_{\delta_1,\delta_2}) \geq \rho_m(c\varepsilon; f). \tag{C.24}$$

For any $\delta_1, \delta_2 > 0$,

$$\|\max\{\tilde{f}_{\delta_1,\delta_2}, M(f)+\rho_m(\varepsilon; f)\}-\tilde{f}_{\delta_1,\delta_2}\| \geq \|\max\{f, M(f)+\rho_m(\varepsilon; f)\}-f\| \geq \varepsilon,$$

which yields that

$$\tilde{\rho}_m(\varepsilon; \tilde{f}_{\delta_1,\delta_2}) \leq \rho_m(\varepsilon; f).$$

Now we have Inequality (C.23). Since $L_3(f) \subset \tilde{\mathcal{F}}_{ll}$, we get

$$\inf_{f\in\mathcal{F}_{ll}} \frac{\rho_m(\varepsilon; f)}{\rho_m(c\varepsilon; f)} \geq \inf_{f\in\tilde{\mathcal{F}}_{ll}} \frac{\tilde{\rho}_m(\varepsilon; f)}{\tilde{\rho}_m(c\varepsilon; f)}.$$

Step 4. We begin by defining several sets of functions such that $\tilde{\mathcal{F}}_{ll}$ is the disjoint union of them. Let

$$\begin{aligned}\tilde{G}(k_1, k_2) = \{f \in \tilde{\mathcal{F}}_{ll} :& f\big|_{(-\infty,Z(f))} \text{ is } k_1\text{-piece linear function,}\\ & f\big|_{(Z(f),\infty)} \text{ is } k_2\text{-piece linear function}\}.\end{aligned} \tag{C.25}$$

Then

$$\tilde{\mathcal{F}}_{ll} = \bigcup_{1\leq k_1,k_2\leq 3} \tilde{G}(k_1, k_2).$$

Clearly, $\tilde{\mathcal{F}}_L = \tilde{G}(1, 1)$, and

$$\frac{\tilde{\rho}_m(\varepsilon; f)}{\tilde{\rho}_m(c\varepsilon; f)} = c^{-\frac{2}{3}}, \forall f \in \tilde{\mathcal{F}}_L.$$

It remains to prove that

$$\inf_{f\in\tilde{\mathcal{F}}_{ll}} \frac{\tilde{\rho}_m(\varepsilon; f)}{\tilde{\rho}_m(c\varepsilon; f)} \geq \inf_{f\in\tilde{\mathcal{F}}_L} \frac{\tilde{\rho}_m(\varepsilon; f)}{\tilde{\rho}_m(c\varepsilon; f)}. \tag{C.26}$$

Let

$$G(k) = \bigcup_{k_1+k_2=k} \tilde{G}(k_1, k_2), \text{ for } k = 2, 3, 4, 5, 6.$$

It suffices to prove that for $k \geq 3$

$$\inf_{f\in G(k)} \frac{\tilde{\rho}_m(\varepsilon; f)}{\tilde{\rho}_m(c\varepsilon; f)} \geq \inf_{f\in G(k-1)} \frac{\tilde{\rho}_m(\varepsilon; f)}{\tilde{\rho}_m(c\varepsilon; f)},$$

which gives Inequality (C.26) and complete the proof of step 4.

Similar to the arguments in previous steps, we prove it by constructing mappings $L_4 : G(k) \longrightarrow G(k-1)$ and $L_5 : G(k) \longrightarrow G(k-1)$ such that for any $f \in G(k)$, at least one of the following holds: $\frac{\tilde{\rho}_m(\varepsilon;f)}{\tilde{\rho}_m(c\varepsilon;f)} \geq \frac{\tilde{\rho}_m(\varepsilon;L_4(f))}{\tilde{\rho}_m(c\varepsilon;L_4(f))}$, $\frac{\tilde{\rho}_m(\varepsilon;f)}{\tilde{\rho}_m(c\varepsilon;f)} \geq \frac{\tilde{\rho}_m(\varepsilon;L_5(f))}{\tilde{\rho}_m(c\varepsilon;L_5(f))}$.

In this step, we keep using the extended definition of $\max\{\cdot,\cdot\}$ for function-value arguments defined in Step 3.

Let S_t be the set of the knots of $f \in \tilde{\mathcal{F}}_{ll}$, then $\left|S_t/\{Z(f)\}\right| = k-2 \geq 1$. Let

$$x^* = \max\{x \in S_t : f(x) = \max\{f(t) : t \in S_t\}\}, \quad t_l = \min S_t, \quad t_r = \max S_t.$$

Clearly, $x^* \neq Z(f)$. Without loss of generality, assume $x^* > Z(f)$. Then by definition of x^*, $f\big|_{[x^*,\infty)}$ is a linear function. We define a function $L_4(f) \in G(k-1)$. Convexity of f ensures the existence of the left derivative $f'(x^*-)$. Further, by definition of x^*, $f'(x^*-) > 0$. For $t \in \mathbb{R}$, $L_4(f)(t)$ is defined by

$$(L_4(f))(t) = \begin{cases} f(t), & t < x^* \\ f(x^*) + f'(x^*-)(t-x^*), & t \geq x^* \end{cases}. \tag{C.27}$$

If $f(x^*) \geq M(f) + \tilde{\rho}_m(c\varepsilon; f)$, we have

$$\tilde{\rho}_m(c\varepsilon; L_4(f)) = \tilde{\rho}_m(c\varepsilon; f), \qquad \tilde{\rho}_m(\varepsilon; L_4(f)) \leq \tilde{\rho}_m(\varepsilon; f),$$

which implies that $\frac{\tilde{\rho}_m(\varepsilon;f)}{\tilde{\rho}_m(c\varepsilon;f)} \geq \frac{\tilde{\rho}_m(\varepsilon;L_4(f))}{\tilde{\rho}_m(c\varepsilon;L_4(f))}$.

If $f(x^*) < M(f) + \tilde{\rho}_m(c\varepsilon; f)$, we have $f(t_l) \leq f(x^*) < M(f) + \tilde{\rho}_m(c\varepsilon; f)$. Denote p_l, p_r to be the left and right root of $f(t) = M(f)+\tilde{\rho}_m(\varepsilon; f)$. Then $p_l < x_{l,m} < t_l \leq Z(f) < x^* < x_{r,m} < p_r$. Now we will first prove Inequality (C.34), and then construct a new function $L_5(f) \in G(k-1)$ such that $\frac{\tilde{\rho}_m(\varepsilon;L_5(f))}{\tilde{\rho}_m(c\varepsilon;L_5(f))} \leq \frac{\tilde{\rho}_m(\varepsilon;f)}{\tilde{\rho}_m(c\varepsilon;f)}$. We start with splitting $\|\max\{L_4(f), M(f)+\tilde{\rho}_m(\varepsilon; f)\}-L_4(f)\|^2$ into three parts of integration. We introduce the shorthand $\tau = \frac{M(f)+\tilde{\rho}_m(\varepsilon;f)-f(x^*)}{f'(x^*-)}$ and note that

$$\Big[\max\{L_4(f), M(f) + \tilde{\rho}_m(\varepsilon; f)\} - L_4(f)\Big](t) =$$
$$\begin{cases} 0 & t \notin [p_l, x^* + \tau] \\ M(f) + \tilde{\rho}_m(\varepsilon; f) - f(t) & t \in [p_l, t_l] \cup [t_l, x^*] \\ M(f) + \tilde{\rho}_m(\varepsilon; f) - \Big[f(x^*) + f'(x^*-)(t-x^*)\Big] & t \in [x^*, x^* + \tau] \end{cases}.$$

We have

$$\| \max\{L_4(f), M(f) + \tilde{\rho}_m(\varepsilon; f)\} - L_4(f)\|^2 = \tag{C.28}$$
$$\underbrace{\int_{p_l}^{t_l} (M(f) + \tilde{\rho}_m(\varepsilon; f) - f(t))^2 dt}_{\Gamma_1} + \underbrace{\int_{t_l}^{x^*} (M(f) + \tilde{\rho}_m(\varepsilon; f) - f(t))^2 dt}_{\Gamma_2}$$
$$+ \underbrace{\frac{1}{f'(x^*-)} \frac{(\tilde{\rho}_m(\varepsilon; f) + M(f) - f(x^*))^3}{3}}_{\Gamma_3}$$

Similarly, $\| \max\{L_4(f), M(f) + \tilde{\rho}_m(c\varepsilon; f)\} - L_4(f)\|^2$ can be split into 3 parts as well.

$$\| \max\{L_4(f), M(f) + \tilde{\rho}_m(c\varepsilon; f)\} - L_4(f)\|^2 \tag{C.29}$$
$$= \underbrace{\int_{x_{l,m}}^{t_l} (M(f) + \tilde{\rho}_m(c\varepsilon; f) - f)^2 dt}_{\gamma_1} + \underbrace{\int_{t_l}^{x^*} (M(f) + \tilde{\rho}_m(c\varepsilon; f) - f)^2 dt}_{\gamma_2} +$$
$$\underbrace{\frac{1}{f'(x^*-)} \frac{(\tilde{\rho}_m(c\varepsilon; f) + M(f) - f(x^*))^3}{3}}_{\gamma_3}.$$

We will compare $\frac{\| \max\{L_4(f), M(f)+\tilde{\rho}_m(\varepsilon; f)\} - L_4(f)\|^2}{\| \max\{L_4(f), M(f)+\tilde{\rho}_m(c\varepsilon; f)\} - L_4(f)\|^2}$ with $\frac{\| \max\{f, M(f)+\tilde{\rho}_m(\varepsilon; f)\} - f\|^2}{\| \max\{f, M(f)+\tilde{\rho}_m(c\varepsilon; f)\} - f\|^2} = \frac{1}{c^2}$. Now we split $\| \max\{f, M(f) + \tilde{\rho}_m(c\varepsilon; f)\} - f\|^2$ and $\| \max\{f, M(f) + \tilde{\rho}_m(\varepsilon; f)\} - f\|^2$ into 3 parts for each. Further, some of the parts equal to the aforementioned parts.

$$\| \max\{f, M(f) + \tilde{\rho}_m(c\varepsilon; f)\} - f\|^2 \tag{C.30}$$
$$= \underbrace{\int_{x_{l,m}}^{t_l} (M(f) + \tilde{\rho}_m(c\varepsilon; f) - f)^2 dt}_{\gamma_1} + \underbrace{\int_{t_l}^{x^*} (M(f) + \tilde{\rho}_m(c\varepsilon; f) - f)^2 dt}_{\gamma_2}$$
$$+ \underbrace{\frac{1}{f'(x^*+)} \frac{(\tilde{\rho}_m(c\varepsilon; f) + M(f) - f(x^*))^3}{3}}_{\gamma_4}.$$

$$\begin{aligned}(\text{C.31})\quad &\|\max\{f, M(f)+\tilde{\rho}_m(\varepsilon;f)\}-f\|^2 \\ &= \underbrace{\int_{p_l}^{t_l}(M(f)+\tilde{\rho}_m(\varepsilon;f)-f)^2dt}_{\Gamma_1} + \underbrace{\int_{t_l}^{x^*}(M(f)+\tilde{\rho}_m(\varepsilon;f)-f)^2dt}_{\Gamma_2} + \\ &\qquad \underbrace{\frac{1}{f'(x^*+)}\frac{(\tilde{\rho}_m(\varepsilon;f)+M(f)-f(x^*))^3}{3}}_{\Gamma_4}.\end{aligned}$$

Elementary calculation gives

(C.32a)

$$\frac{\Gamma_1}{\gamma_1} \le \left(\frac{M(f)+\tilde{\rho}_m(\varepsilon;f)-f(t_l)}{M(f)+\tilde{\rho}_m(c\varepsilon;f)-f(t_l)}\right)^3 \le \left(\frac{\tilde{\rho}_m(\varepsilon;f)+M(f)-f(x^*)}{\tilde{\rho}_m(c\varepsilon;f)+M(f)-f(x^*)}\right)^3,$$

(C.32b)

$$\frac{\Gamma_2}{\gamma_2} \le \left(\frac{\tilde{\rho}_m(\varepsilon;f)+M(f)-f(x^*)}{\tilde{\rho}_m(c\varepsilon;f)+M(f)-f(x^*)}\right)^2 < \left(\frac{\tilde{\rho}_m(\varepsilon;f)+M(f)-f(x^*)}{\tilde{\rho}_m(c\varepsilon;f)+M(f)-f(x^*)}\right)^3,$$

(C.32c)

$$\frac{\Gamma_4}{\gamma_4} = \left(\frac{\tilde{\rho}_m(\varepsilon;f)+M(f)-f(x^*)}{\tilde{\rho}_m(c\varepsilon;f)+M(f)-f(x^*)}\right)^3 = \frac{\Gamma_3}{\gamma_3}.$$

Further, $f'(x^*+) \ge f'(x^*-) > 0$ implies that $\Gamma_3 \ge \Gamma_4 > 0$. Consequently, we have

$$\frac{\Gamma_1+\Gamma_2+\Gamma_3}{\gamma_1+\gamma_2+\gamma_3} \ge \frac{\Gamma_1+\Gamma_2+\Gamma_4}{\gamma_1+\gamma_2+\gamma_4}, \tag{C.33}$$

which follows from the fact that $\frac{a+c}{b+d} \ge \frac{a}{b}$ if $a,b,c,d>0$ and $\frac{c}{d} \ge \frac{a}{b}$. Note that the terms in Inequality (C.33) are exactly the split parts of the quantities in Equation (C.28), Equation (C.29), Equation (C.30), and Equation (C.31). Consequently, we have

$$\frac{\|\max\{L_4(f), M(f)+\tilde{\rho}_m(\varepsilon;f)\}-L_4(f)\|^2}{\|\max\{L_4(f), M(f)+\tilde{\rho}_m(c\varepsilon;f)\}-L_4(f)\|^2} \ge \frac{1}{c^2}. \tag{C.34}$$

Define function $L_5(f) \in G(k-1)$ by scaling $L_4(f)$ horizontally with scaling factor $\lambda = \frac{c\varepsilon}{\|\max\{L_4(f), M(f)+\tilde{\rho}_m(c\varepsilon;f)\}-L_4(f)\|}$.

$$L_5(f) : t \mapsto M(f) + \left[(L_4(f))(t) - M(f)\right]\lambda. \tag{C.35}$$

Clearly,

$$\tilde{\rho}_m(c\varepsilon; L_5(f)) = \lambda\tilde{\rho}_m(c\varepsilon;f), \quad \tilde{\rho}_m(\varepsilon; L_5(f)) \le \lambda\tilde{\rho}_m(\varepsilon;f).$$

Thus the statement is proved. □

Now let us turn to the proof of the geometric property of the minimizer, namely, for $\varepsilon > 0$, $c \in (0,1)$, and $f \in \mathcal{F}$,

$$\max\{(c/2)^{\frac{2}{3}}, c\} \leq \frac{\rho_z(c\varepsilon; f)}{\rho_z(\varepsilon; f)} \leq 1. \tag{C.36}$$

Proof. The right hand side of the inequality is straightforward. For the left hand side, we prove a stronger version,

$$c^{-2} \geq \frac{3}{4}\left(\frac{\rho_z(\varepsilon; f)}{\rho_z(c\varepsilon; f)}\right)^2 + \frac{1}{4}\left(\frac{\rho_z(\varepsilon; f)}{\rho_z(c\varepsilon; f)}\right)^3. \tag{C.37}$$

Similar to Step 3 in the previous proof for the minimum, for any $f \in \mathcal{F}$, we have a class of functions $\{\tilde{f}_{\delta_1,\delta_2} : \delta_1, \delta_2\}$, but with a bit of abuse of notation, we define $\tilde{f}_{\delta_1,\delta_2}$ here as

$$\tilde{f}_{\delta_1,\delta_2}(t) = \begin{cases} f(t), & t \in [0,1] \\ f(0) + \min\{-\delta_1^{-1}, \lim_{s\to 0^+} \frac{f(s)-f(0)}{s}\}t, & t \in (-\infty, 0) \\ f(1) + \max\{\delta_2^{-1}, \lim_{s\to 0^+} \frac{f(1)-f(1-s)}{s}\}(t-1), & t \in (1,\infty) \end{cases}.$$

Similarly , we have

$$\lim_{\max\{\delta_1,\delta_2\}\to 0^+} \tilde{\rho}_z(\varepsilon; \tilde{f}_{\delta_1,\delta_2}) = \rho_z(\varepsilon; f), \qquad \lim_{\max\{\delta_1,\delta_2\}\to 0^+} \tilde{\rho}_z(c\varepsilon; \tilde{f}_{\delta_1,\delta_2}) = \rho_z(c\varepsilon; f).$$

Hence

$$\sup_{f\in\mathcal{F}} \frac{\rho_z(\varepsilon; f)}{\rho_z(c\varepsilon; f)} \leq \sup_{f\in\tilde{\mathcal{F}}_c} \frac{\tilde{\rho}_z(\varepsilon; f)}{\tilde{\rho}_z(c\varepsilon; f)}. \tag{C.38}$$

Similar to the proof of the minimum, for $f \in \tilde{\mathcal{F}}_c$, denote p_l, p_r as the two roots of $f(t) = M(f) + \tilde{\rho}_m(\varepsilon; f)$, and q_l, q_r as the two roots of $f(t) = M(f) + \tilde{\rho}_m(c\varepsilon; f)$. Without loss of generality, we assume $p_r = Z(f) + \tilde{\rho}_z(\varepsilon; f)$. We define four quantities:

$$\begin{aligned} \Delta_1 &= \int_{p_l}^{Z(f)} (\tilde{\rho}_m(\varepsilon; f) + M(f) - f)^2 dt, \\ \Delta_2 &= \int_{q_l}^{Z(f)} (\tilde{\rho}_m(c\varepsilon; f) + M(f) - f)^2 dt, \\ \Delta_3 &= \int_{Z(f)}^{q_r} (\tilde{\rho}_m(c\varepsilon; f) + M(f) - f)^2 dt, \\ \Delta_4 &= \int_{Z(f)}^{p_r} (\tilde{\rho}_m(\varepsilon; f) + M(f) - f)^2 dt. \end{aligned} \tag{C.39}$$

Then we know that

$$\varepsilon^2 = \| \max\{f, M(f) + \tilde{\rho}_m(\varepsilon; f)\} - f\|^2 = \Delta_1 + \Delta_4, \tag{C.40}$$

and that

$$c^2\varepsilon^2 = \| \max\{f, M(f) + \tilde{\rho}_m(c\varepsilon; f)\} - f\|^2 = \Delta_2 + \Delta_3. \tag{C.41}$$

We also have

$$\frac{\Delta_1}{\Delta_2} \geq \left(\frac{\tilde{\rho}_m(\varepsilon; f)}{\tilde{\rho}_m(c\varepsilon; f)}\right)^2 \geq \left(\frac{p_r - Z(f)}{q_r - Z(f)}\right)^2 \geq \left(\frac{\tilde{\rho}_z(\varepsilon; f)}{\tilde{\rho}_z(c\varepsilon; f)}\right)^2. \tag{C.42}$$

Next we will show that

$$\frac{\Delta_4}{\Delta_3} \geq \left(\frac{p_r - Z(f)}{q_r - Z(f)}\right)^3 \geq \left(\frac{\tilde{\rho}_z(\varepsilon; f)}{\tilde{\rho}_z(c\varepsilon; f)}\right)^3.$$

For the ease of presentation, let us define four quantities $w_1 = p_r - Z(f) = \tilde{\rho}_z(\varepsilon; f), w_2 = q_r - Z(f) \leq \tilde{\rho}_z(c\varepsilon; f), v_1 = \tilde{\rho}_m(\varepsilon; f), v_2 = \tilde{\rho}_m(c\varepsilon; f)$. Using this notation, we can rewrite the expression for Δ_4/Δ_3 as follows:

$$\begin{aligned} \frac{\Delta_4}{\Delta_3} &= \frac{\int_0^{w_1}(v_1 + M(f) - f(p_r - t))^2 dt}{\int_0^{w_2}(v_2 + M(f) - f(q_r - t))^2 dt} \\ &= \frac{w_1 \int_0^1 (v_1 + M(f) - f(p_r - w_1 \cdot t))^2 dt}{w_2 \int_0^1 (v_2 + M(f) - f(q_r - w_2 \cdot t))^2 dt}. \end{aligned} \tag{C.43}$$

We also have the following inequality:

$$\begin{aligned} & M(f) + v_1 - f(p_r - w_1 \cdot t) = f(p_r) - f(p_r - w_1 \cdot t) \\ =& \frac{f(p_r) - f(p_r - w_1 \cdot t)}{w_1 \cdot t} w_1 \cdot t \overset{(iii)}{\geq} \frac{f(q_r) - f(q_r - w_2 \cdot t)}{w_2 \cdot t} w_1 \cdot t \\ & = \frac{w_1}{w_2}(f(q_r) - f(q_r - w_2 \cdot t)), \end{aligned} \tag{C.44}$$

where step (iii) follows from the convexity of f and the facts that $p_r > q_r$, and $p_r - w_1 \cdot t \geq q_r - w_2 \cdot t$. Continuing with Inequality (C.43), we have

$$\frac{\Delta_4}{\Delta_3} \geq \frac{w_1 \int_0^1 \left(\frac{w_1}{w_2}(f(q_r) - f(q_r - w_2 \cdot t))\right)^2 dt}{w_2 \int_0^1 (f(q_r) - f(q_r - w_2 \cdot t))^2 dt} = \left(\frac{w_1}{w_2}\right)^3. \tag{C.45}$$

In addition, we have

$$\frac{\Delta_3}{\Delta_2} \geq \frac{1}{3} \frac{w_2}{\tilde{\rho}_z(c\varepsilon; f)} \tag{C.46}$$

Therefore,

$$
\begin{aligned}
c^{-2} &= \frac{\Delta_1+\Delta_4}{\Delta_2+\Delta_3} \geq \frac{\left(\frac{w_1}{w_2}\right)^2\Delta_2+\Delta_3\left(\frac{w_1}{w_2}\right)^3}{\Delta_2+\Delta_3} \\
&\geq \frac{1+\frac{1}{3}\frac{w_2}{\tilde\rho_z(c\varepsilon;f)}\frac{w_1}{w_2}}{1+\frac{1}{3}\frac{w_2}{\tilde\rho_z(c\varepsilon;f)}}\left(\frac{w_1}{w_2}\right)^2 \geq \frac{1+\frac{1}{3}\frac{\tilde\rho_z(\varepsilon;f)}{\tilde\rho_z(c\varepsilon;f)}}{\frac{4}{3}}\left(\frac{\tilde\rho_z(\varepsilon;f)}{\tilde\rho_z(c\varepsilon;f)}\right)^2 \\
&= \frac{3}{4}\left(\frac{\tilde\rho_z(\varepsilon;f)}{\tilde\rho_z(c\varepsilon;f)}\right)^2+\frac{1}{4}\left(\frac{\tilde\rho_z(\varepsilon;f)}{\tilde\rho_z(c\varepsilon;f)}\right)^3.
\end{aligned} \tag{C.47}
$$

Since this inequality holds for all $f \in \tilde{\mathcal{F}}_c$ and together with Inequality (C.38), we obtain

$$
c^{-2} \geq \frac{3}{4}\left(\sup_{f\in\mathcal{F}}\frac{\tilde\rho_z(\varepsilon;f)}{\tilde\rho_z(c\varepsilon;f)}\right)^2+\frac{1}{4}\left(\sup_{f\in\mathcal{F}}\frac{\tilde\rho_z(\varepsilon;f)}{\tilde\rho_z(c\varepsilon;f)}\right)^3.
$$

□

C.3. Proof of Proposition 2.2. We begin by establishing the lower bound on the local modulus of continuity $\omega_z(\varepsilon;f)$, namely, $\rho_z(\varepsilon;f)$. We define $u_\varepsilon = \sup\{u : \|f-f_u\|_2 \leq \varepsilon\}$ for given f and ε. Let t_ℓ and t_r ($t_\ell < Z(f) < t_r$) be the two end points of the interval $\{t : f(t) \leq u_\varepsilon\}$. Without loss of generality we assume that $|t_r - Z(f)| \geq |t_\ell - Z(f)|$, which implies that $\rho_z(\varepsilon;f) = t_r - Z(f)$. For any $\delta \in (0, t_r - t_\ell)$, consider the function g_δ defined as

$$
g_\delta : t \mapsto \max\left\{f(t), u_\varepsilon - \frac{u_\varepsilon - f(t_r-\delta)}{t_r-t_\ell-\delta}(t-t_\ell)\right\}. \tag{C.48}
$$

It is easy to verify that g_δ is convex with minimum point at $t_r - \delta$, and that $\|f-g_\delta\| \leq \|f-f_{u_\varepsilon}\| \leq \varepsilon$. See a graphical illustration in Figure 20. Therefore, taking $\delta \to 0+$ gives

$$
\omega_z(\varepsilon;f) \geq \lim_{\delta\to0+}(t_r-\delta) = \rho_z(\varepsilon;f).
$$

Now we switch to the upper bound. Suppose g is a function such that $\|f-g\| \leq \varepsilon$, with minimum point at $Z(g) > Z(f)$. We will use proof by contradiction.

If $Z(g) > Z(f) + 3\rho_z(\varepsilon;f)$, then $1 \geq Z(f) + 3\rho_z(\varepsilon;f)$. Recycling our notation, we define $t_\ell(u_\varepsilon) = \inf\{t : f(t) \leq u_\varepsilon\}$ and $t_r(u_\varepsilon) = \sup\{t : f(t) \leq$

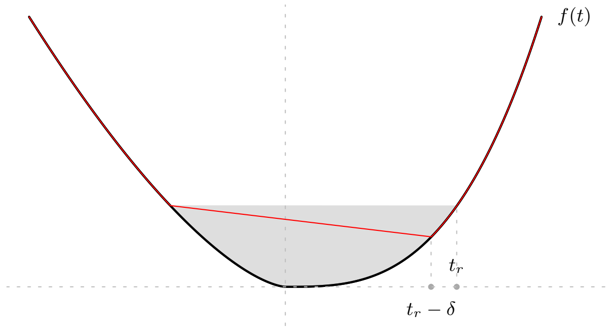


Fig 20: Illustration of construction of g_δ, colored red in the plot

$u_\varepsilon\}$. Convexity of f implies that f is continuous, hence $f(t_r(u_\varepsilon)) = u_\varepsilon$. We have two cases: 1, $g(t_r(u_\varepsilon)) > u_\varepsilon$, 2, $g(t_r(u_\varepsilon)) \leq u_\varepsilon$.

For case 1, we know $g(t) > u_\varepsilon$ for $t_l(u_\varepsilon) \leq t \leq t_r(u_\varepsilon)$, so

$$\|f - g\|^2 > \int_{t_l(u_\varepsilon)}^{t_r(u_\varepsilon)} (u_\varepsilon - f(t))^2 dt = \varepsilon^2.$$

For case 2, we know $g(t) \leq u_\varepsilon$ for $t_r(u_\varepsilon) \leq t \leq t_r(u_\varepsilon) + 2\rho_z(\varepsilon; f)$, so

$$\begin{aligned}\|f - g\|^2 &\geq \int_{t_r(u_\varepsilon)}^{t_r(u_\varepsilon)+2\rho_z(\varepsilon;f)} \left(\frac{u_\varepsilon}{t_r(u_\varepsilon) - Z(f)}(t - t_r(u_\varepsilon))\right)^2 dt \\ &\geq \frac{u^2}{(t_r(u_\varepsilon) - Z(f))^2}\frac{8\rho_z(\varepsilon; f)^3}{3} = \frac{8}{3}\rho_z(\varepsilon; f)u^2 \geq \frac{4\varepsilon^2}{3}\end{aligned}$$

Either case, the there is a contradiction. Therefore, $Z(g) \leq Z(f) + 3\rho_z(\varepsilon; f)$.

Let us now turn to $\omega_m(\varepsilon; f)$ and firstly show that $\omega_m(\varepsilon; f) \geq \rho_m(\varepsilon; f)$. In fact, if we take the convex function g_δ as defined in Equation (C.48), we have that $\|f - g_\delta\| \leq \varepsilon$ and that

$$\lim_{\delta \to 0^+} \min_t g_\delta(t) - Z(f) = \rho_m(\varepsilon; f),$$

which completes the proof.

Next, we will show that $\omega_m(\varepsilon; f)$ can be upper bounded by $\rho_m(\varepsilon; f)$ up to a constant factor of 3.

For any $g \in \mathcal{F}$ such that $\|f - g\| \leq \varepsilon$, we can immediately obtain

$$M(g) - M(f) \leq \rho_m(\varepsilon; f).$$

Otherwise, if $M(g) - M(f) > \rho_m(\varepsilon; f)$, then $g(t) > \rho_m(\varepsilon; f) + M(f)$ for all t, and hence $\varepsilon^2 \geq \|f - g\|^2 \geq (M(g) - M(f) - \rho_m(\varepsilon; f))^2 (t_r(u_\varepsilon) - t_l(u_\varepsilon)) + \|f_{u_\varepsilon} - f\|^2 > \varepsilon^2$.

On the other hand, we need to show the minimum value of g cannot be too small compared to $M(f)$. For the ease of presentation, we assume that $M(f) = 0$ only for this part. As in the previous parts, we write $t_\ell = \inf\{t : f(t) \leq u_\varepsilon\}$, $t_r = \sup\{t : f(t) \leq u_\varepsilon\}$, and $v_\varepsilon = t_r - t_\ell$. Graphically, v_ε is the width of the water-filling surface. Suppose that $M(g) = -\alpha u_\varepsilon$ for some $\alpha > 0$. Consider the width of the set $\{t : g(t) \leq 0\}$, which we denote as γv_ε for some $\gamma > 0$. From Figure 21, we see that the integral $\|f - g\|_2^2$ has to contain the ℓ_2 area of the three shaded triangles (the two triangles on the side might not exist). Given that $M(g) = -a u_\varepsilon$ and $|\{t : g(t) \leq 0\}| = \gamma v_\varepsilon$, some calculation shows that

$$\begin{aligned} \|f - g\|^2 &\geq u_\varepsilon^2 v_\varepsilon \cdot \frac{1}{3}\alpha^2\gamma \left(1 + \left(\frac{1}{\gamma} - \frac{\alpha+1}{\alpha}\right)^3 \vee 0\right) \\ &\geq \varepsilon^2 \cdot \frac{1}{3}\alpha^2\gamma \left(1 + \left(\frac{1}{\gamma} - \frac{\alpha+1}{\alpha}\right)^3 \vee 0\right) \end{aligned}$$

where the second inequality follows from $u_\varepsilon^2 v_\varepsilon \geq \varepsilon^2$. Fixing α and minimizing over γ, we have that if $\alpha > 3$, $\|f - g\|^2 > \varepsilon^2$, which is contradictory. Therefore, we have

$$M(f) - M(g) \leq 3\rho_m(\varepsilon; f).$$

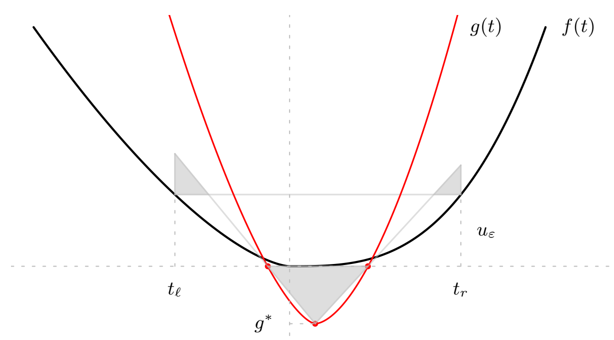


Fig 21: Illustration of upper bound proof

C.4. Proof of Theorem 2.1. We begin with the lower bounds by first proving that $R_z(\varepsilon; f) \geq \Phi(-0.5)\omega_z(\varepsilon; f)$. The proof for $R_m(\varepsilon; f) \geq$

$\Phi(-0.5)\omega_m(\varepsilon; f)$ is analogous and will hence be omitted.

Let $f \in \mathcal{F}$. Let $g \in \mathcal{F}$, which we will specify later. Take $\theta \in \{1, -1\}$ as a parameter to be estimated and let $f_1 = f$ and $f_{-1} = g$.

Any estimator $\hat{Z}$ of the minimizer $Z(f_\theta)$ gives an estimator of θ by

$$\hat{\theta} = \frac{\hat{Z} - \frac{Z(f_1)+Z(f_{-1})}{2}}{\frac{Z(f_1)-Z(f_{-1})}{2}},$$

and therefore $\mathbb{E}_\theta|\hat{Z} - Z(f_\theta)| = |Z(f_1) - Z(f_{-1})|\mathbb{E}_\theta \frac{|\hat{\theta}-\theta|}{2}$. On the other hand, a sufficient statistic for θ is given by

$$W = \frac{\int_0^1 (f_1(t) - f_{-1}(t))dY(t) - \frac{1}{2}\int_0^1 (f_1(t)^2 - f_{-1}(t)^2)dt}{\varepsilon\|f_1 - f_{-1}\|}. \tag{C.49}$$

Let $\mathbb{P}_\theta$ be the probability measure associated with the white noise model corresponding to f_θ. Then $W \sim N\left(\frac{\theta}{2} \cdot \frac{\|f_1 - f_{-1}\|}{\varepsilon}, 1\right)$ under $\mathbb{P}_\theta$.

Note that for any $\omega_z(\varepsilon; f) > \delta > 0$ there exists $h_\delta \in \mathcal{F}$ such that $\|f - h_\delta\|_2 = \varepsilon$ and $|Z(f) - Z(h_\delta)| \geq \omega_z(\varepsilon; f) - \delta$. Let $g = h_\delta$. Then we have $R_z(\varepsilon; f) \geq (\omega_z(\varepsilon; f) - \delta) \cdot r_1$, where r_1 is the minimax risk of the two-point problem based on an observation $X \sim N(\frac{\theta}{2}, 1)$, i.e., $r_1 = \inf_{\hat{\theta}} \max_{\theta=\pm 1} \mathbb{E}_\theta \frac{|\hat{\theta}-\theta|}{2}$. It is easy to see that $r_1 = \Phi(-0.5)$. Taking $\delta \to 0^+$, we have $R_z(\varepsilon; f) \geq \Phi(-0.5)\omega_z(\varepsilon; f)$, so $a_1 \geq \Phi(-0.5) \approx 0.309$.

Next, we show for $0 < \alpha < 0.3$ that $L_{z,\alpha}(\varepsilon; f) \geq b_\alpha \omega_z(\varepsilon/3; f)$ where $b_\alpha = 0.6 - 2\alpha$. A lower bound for $L_{m,\alpha}(\varepsilon; f)$ can be derived following a similar argument. We begin by recalling a lemma from Cai and Guo (2017).

LEMMA C.4 (Cai and Guo, 2017). *For any $CI \in \mathcal{I}_{z,\alpha}(\{f, g\})$,*

$$\mathbb{E}_f L(CI) \geq |Z(f) - Z(g)|(1 - 2\alpha - \mathrm{TV}(P_f, P_g)),$$

where TV *denotes the total variation distance between the two distributions of the white noise models corresponding to f and g. Similarly, for any $CI \in \mathcal{I}_{m,\alpha}(\{f, g\})$,*

$$\mathbb{E}_f L(CI) \geq |M(f) - M(g)|(1 - 2\alpha - \mathrm{TV}(P_f, P_g)).$$

Again let $g \in \mathcal{F}$. Then for $CI \in \mathcal{I}_{z,\alpha}(\{f, g\})$, by Lemma C.4,

$$\mathbb{E}_f L(CI) \geq |Z(f) - Z(g)|(1 - 2\alpha - \mathrm{TV}(P_f, P_g)).$$

Note that $\mathrm{TV}(P_f, P_g) \le \sqrt{\chi^2(P_f, P_g)}$, where $\chi^2(P_f, P_g) = \int \left(\frac{dP_f}{dP_g}\right)^2 dP_g - 1$ is the χ^2 distance between P_f and P_g. Girsanov's theorem yields that $\frac{dP_f}{dP_g} = \exp\left(\int \frac{f(t)-g(t)}{\varepsilon^2} dY(t) - \frac{1}{2}\int \frac{f(t)^2-g(t)^2}{\varepsilon^2} dt\right)$, and hence

$$\begin{aligned}\chi^2(P_f, P_g) &= \int \exp(2\int \frac{f(t)-g(t)}{\varepsilon^2} dY(t) - \int \frac{f(t)^2-g(t)^2}{\varepsilon^2} dt) dP_g - 1 \\ &= \exp(\frac{\|f-g\|^2}{\varepsilon^2}) - 1.\end{aligned}$$

Using it to bound the total variation distance, we get

$$\mathbb{E}_f L(CI) \ge |Z(f) - Z(g)| \left(1 - 2\alpha - \sqrt{\exp\left(\frac{\|f-g\|^2}{\varepsilon^2}\right) - 1}\right).$$

We continue by specifying g. For any $\omega_z(\varepsilon/3; f) > \delta > 0$, picking $g = g_\delta \in \mathcal{F}$ such that $\|f - g_\delta\| = \varepsilon/3$ and $|Z(f) - Z(g_\delta)| \ge \omega_z(\varepsilon/3; f) - \delta$, we have $\mathbb{E}_f L(CI) \ge (0.6 - 2\alpha)\left(\omega_z(\varepsilon/3; f) - \delta\right)$. By taking $\delta \to 0^+$, we have

$$L_{z,\alpha}(\varepsilon; f) \ge (0.6 - 2\alpha)\, \omega_z(\varepsilon/3; f).$$

Now we turn to the upper bounds and introduce two lemmas, one for the minimum and another for the minimizer, that will be proved later.

LEMMA C.5. *For $0 < \alpha \le 0.3$ and any $f \in \mathcal{F}$,*

$$R_m(\varepsilon; f) \le A_m \rho_m(\varepsilon; f) \le A_m \omega_m(\varepsilon; f), \tag{C.50}$$

$$L_{m,\alpha}(\varepsilon; f) \le B_{m,\alpha} \rho_m(\varepsilon; f) \le B_{m,\alpha} \omega_m(\varepsilon; f), \tag{C.51}$$

where $A_m = 1.03$ and $0 < B_{m,\alpha} \le 3(1 - 2\alpha) z_\alpha$.

LEMMA C.6. *For $0 < \alpha \le 0.3$ and any $f \in \mathcal{F}$,*

$$R_z(\varepsilon; f) \le A_z \rho_z(\varepsilon; f) \le A_z \omega_z(\varepsilon; f), \tag{C.52}$$

$$L_{z,\alpha}(\varepsilon; f) \le B_{z,\alpha} \rho_z(\varepsilon; f) \le B_{z,\alpha} \omega_z(\varepsilon; f), \tag{C.53}$$

where $A_z = 1.5$ and $0 < B_{z,\alpha} \le 3(1 - 2\alpha) \min\{z_\alpha, (2z_\alpha)^{2/3}\}$.

The theorem follows as $B_\alpha \ge \max\{B_{z,\alpha}, B_{m,\alpha}\}$ and $A_1 \ge \max\{A_m, A_z\}$. □

PROOF OF LEMMA C.5. For any function $g \in \mathcal{F}$, define f_θ with $\theta \in \{-1, 1\}$ and $f_{-1} = f$ and $f_1 = g$. Recall that for W defined in (C.49), $W \sim N(\theta \cdot \frac{\|f_1 - f_{-1}\|}{2\varepsilon}, 1)$. Let $\hat{M} = \text{sign}(W) \cdot \frac{M(g)-M(f)}{2} + \frac{M(g)+M(f)}{2}$. Then $\mathbb{E}_f(|\hat{M} - M(f)|) = |M(f) - M(g)|\Phi(-\frac{\|g-f\|}{2\varepsilon}) = \mathbb{E}_g(|\hat{M} - M(g)|)$. Therefore,

$$\begin{aligned}
R_m(\varepsilon; f) &\leq \sup_{g \in \mathcal{F}} |M(f) - M(g)|\Phi(-\frac{\|g - f\|}{2\varepsilon}) \overset{\text{(i)}}{\leq} \sup_{c>0} \omega_m(c\varepsilon; f)\Phi(-\frac{c}{2}) \\
&\overset{\text{(ii)}}{\leq} \max\{3\rho_m(\varepsilon; f) \sup_{0<c\leq 1} c^{\frac{2}{3}}\Phi(-\frac{c}{2}), \sup_{c\geq 1} \omega_m(c\varepsilon; f)\Phi(-\frac{c}{2})\} \\
&\overset{\text{(iii)}}{\leq} \max\{3\rho_m(\varepsilon; f)\Phi(-\frac{1}{2}), \sup_{c\geq 1} \omega_m(c\varepsilon; f)\Phi(-\frac{c}{2})\},
\end{aligned}$$

where (i) is due to the definition of $\omega_m(c\varepsilon; f)$ in Equation (2.2), (ii) follows from Proposition 2.1, (iii) is due to the fact that $c^{\frac{2}{3}}\Phi(-\frac{c}{2})$ increases in $c \in [0, 1]$. Furthermore we have,

$$\begin{aligned}
\sup_{c\geq 1} \omega_m(c\varepsilon; f)\Phi(-\frac{c}{2})\} &\overset{\text{(iv)}}{\leq} \sup_{c\geq 1} 3\rho_m(c\varepsilon; f)\Phi(-\frac{c}{2}) \overset{\text{(v)}}{\leq} 3\rho_m(\varepsilon; f) \cdot \sup_{c\geq 1} c\Phi(-\frac{c}{2}) \\
&\overset{\text{(vi)}}{\leq} 3\rho_m(\varepsilon; f) \times 0.3423 \overset{\text{(vii)}}{\leq} 1.03\omega_m(\varepsilon; f),
\end{aligned}$$

where (iv) is due to Proposition 2.2, (v) and (vii) are due to Proposition 2.1, and (vi) is due to a bound for $\sup_{c\geq 1} c\Phi(-\frac{c}{2})$, which follows from the elementary inequalities: $\Phi(-c/2) \leq \frac{1}{c}\sqrt{\frac{2}{\pi}} \exp(-\frac{c^2}{8})$ for $c > 0$; $\frac{\partial(c\Phi(-c/2))}{\partial c} = \Phi(-c/2) - \frac{c}{2}\sqrt{\frac{1}{2\pi}} \exp(-\frac{c^2}{8}) < 0$ for $c > 2$; and $\sup_{c\in[k/100,(k+1)/100]} c\Phi(-c/2) \leq 0.01(k+1)\Phi(-0.01 \times k/2)$ for $k = \{100, 101, \cdots, 200\}$. Therefore, we can take $A_m = \max\{3\Phi(-1/2), 1.03\} = 1.03$.

For inference of the minimum, consider the following confidence interval:

$$CI_{m,\alpha} = \begin{cases} \{M(f)\} & W < -z_\alpha + \frac{\|f-g\|}{2\varepsilon} \\ \{M(g)\} & W \geq (z_\alpha - \frac{\|f-g\|}{2\varepsilon}) \vee (-z_\alpha + \frac{\|f-g\|}{2\varepsilon}) \\ [M(f) \wedge M(g), M(f) \vee M(g)] & \text{otherwise} \end{cases}.$$

Note that $P_h(M(h) \notin CI_{m,\alpha}) \leq \alpha$ for $h \in \{f, g\}$ and for $\theta \in \{0, 1\}$,

$$\begin{aligned}
\mathbb{E}_{f_\theta} L(CI_{m,\alpha}) &\leq |M(f) - M(g)|P_{f_\theta}(-z_\alpha + 0.5\frac{\|f - g\|}{\varepsilon} \leq W < z_\alpha - 0.5\frac{\|f - g\|}{\varepsilon}) \\
&\leq |M(f) - M(g)|(\Phi(z_\alpha - \frac{\|f - g\|}{\varepsilon}) - \alpha)_+.
\end{aligned}$$

Therefore, it follows from Proposition 2.1 that

$$\begin{aligned}
&L_{m,\alpha}(\varepsilon; f)\\
&\le \sup_{g\in\mathcal{F}} |M(f)-M(g)|(\Phi(z_\alpha - \frac{\|f-g\|}{\varepsilon}) - \alpha)_+ \le \sup_{c>0} \omega_m(c\varepsilon; f)(\Phi(z_\alpha - c) - \alpha)_+\\
&\le \max\{\omega_m(\varepsilon; f)(\Phi(z_\alpha) - \alpha)_+, \sup_{c>1} \omega_m(c\varepsilon; f)(\Phi(z_\alpha - c) - \alpha)_+\}\\
&= \max\{\omega_m(\varepsilon; f)(1-2\alpha), \sup_{c>1} \omega_m(c\varepsilon; f)(\Phi(z_\alpha - c) - \alpha)_+\}.
\end{aligned}$$

Further, recalling $\alpha < 0.3$, we have $2z_\alpha > 1$, thus

$$\begin{aligned}
&\sup_{c>1} \omega_m(c\varepsilon; f)(\Phi(z_\alpha - c) - \alpha)_+ \le \sup_{c>1} 3\rho_m(c\varepsilon; f)(\Phi(z_\alpha - c) - \alpha)_+\\
&\le 3\rho_m(\varepsilon; f) \sup_{c>1} c(\Phi(z_\alpha - c) - \alpha)_+ = 3\rho_m(\varepsilon; f) \sup_{2z_\alpha > c > 1} c(\Phi(z_\alpha - c) - \alpha)\\
&\overset{\text{(viii)}}{\le} 3\rho_m(\varepsilon; f)\,[(1-2\alpha)z_\alpha \mathbb{1}\{z_\alpha \ge 1\} + (0.5-\alpha)\cdot 2z_\alpha \mathbb{1}\{z_\alpha < 1\}]\\
&\le 3\omega_m(\varepsilon; f)(1-2\alpha)z_\alpha,
\end{aligned}$$

where (viii) follows from $\sup_{c\in[A,B]} c(\Phi(z_\alpha - c) - \alpha) \le B(\Phi(z_\alpha - A) - \alpha)$ for any $1 \le A \le B \le 2z_\alpha$. In conclusion, $L_{m,\alpha}(\varepsilon; f) \le 3(1-2\alpha)z_\alpha\rho_m(\varepsilon; f) \le 3(1-2\alpha)z_\alpha\omega_m(\varepsilon; f)$. □

PROOF OF LEMMA C.6. For any $g \in \mathcal{F}$, consider f_θ with $\theta \in \{-1, 1\}$, $f_{-1} = f$ and $f_1 = g$. Recall that for W defined in (C.49), $W \sim N(\theta \cdot \frac{\|f_1 - f_{-1}\|}{2\varepsilon}, 1)$. Let $\hat{Z} = \text{sign}(W)\cdot\frac{Z(g)-Z(f)}{2} + \frac{Z(g)+Z(f)}{2}$. Then $\mathbb{E}_f(|\hat{Z} - Z(f)|) = |Z(f) - Z(g)|\Phi(-\frac{\|g-f\|}{2\varepsilon}) = \mathbb{E}_g(|\hat{Z} - Z(g)|)$. Therefore,

$$\begin{aligned}
R_z(\varepsilon; f) &\le \sup_{g\in\mathcal{F}} |Z(f) - Z(g)|\Phi(-\frac{\|g-f\|}{2\varepsilon}) \le \sup_{c>0} \omega_z(c\varepsilon; f)\Phi(-\frac{c}{2})\\
&\le \max\{0.5\omega_z(\varepsilon; f), \sup_{c\ge 1} \omega_z(c\varepsilon; f)\Phi(-\frac{c}{2})\}.
\end{aligned} \tag{C.54}$$

In addition,

$$\begin{aligned}
&\sup_{c\ge1} \omega_z(c\varepsilon; f)\Phi(-\frac{c}{2})\} \le \sup_{c\ge1} 3\rho_z(c\varepsilon; f)\Phi(-\frac{c}{2})\\
&\le 3\sup_{c\ge1} \min\{c, (2c)^{\frac{2}{3}}\}\rho_z(\varepsilon; f)\Phi(-\frac{c}{2}) \le 1.03\rho_z(\varepsilon; f).
\end{aligned} \tag{C.55}$$

Inequalities (C.55) and (C.54) together with Proposition 2.1 show that we can take $A_z = 1.5$.

For inference of the minimizer, let

$$CI_{z,\alpha} = \begin{cases} \{Z(f)\} & W < -z_\alpha + 0.5\frac{\|f-g\|}{\varepsilon} \\ \{Z(g)\} & W \geq (z_\alpha - \frac{\|f-g\|}{2\varepsilon}) \vee (-z_\alpha + \frac{\|f-g\|}{2\varepsilon}) \\ [Z(f) \wedge Z(g), Z(f) \vee Z(g)] & \text{otherwise} \end{cases} .$$

Clearly, we have $P_f(Z(f) \notin CI_{z,\alpha}) \leq \alpha, P_g(Z(g) \notin CI_{z,\alpha}) \leq \alpha$.

For the expected length, similar to the proof for Lemma C.5, we have for $\theta \in \{-1, 1\}$,

$$\mathbb{E}_{f_\theta} L(CI_{z,\alpha}) \leq |Z(f) - Z(g)|(\Phi(z_\alpha - \frac{\|f-g\|}{\varepsilon}) - \alpha)_+. \tag{C.56}$$

Therefore

$$\begin{aligned} & L_{z,\alpha}(\varepsilon; f) \\ & \leq \sup_{g\in\mathcal{F}} |Z(f) - Z(g)|(\Phi(z_\alpha - \frac{\|f-g\|}{\varepsilon}) - \alpha)_+ \leq \sup_{c>0} \omega_z(c\varepsilon; f)(\Phi(z_\alpha - c) - \alpha)_+ \\ & \leq \max\{\omega_z(\varepsilon; f)(\Phi(z_\alpha) - \alpha)_+, \sup_{c>1} \omega_z(c\varepsilon; f)(\Phi(z_\alpha - c) - \alpha)_+\} \\ & \leq \max\{\omega_z(\varepsilon; f)(1 - 2\alpha), \sup_{c>1} \omega_z(c\varepsilon; f)(\Phi(z_\alpha - c) - \alpha)_+\}. \end{aligned}$$

Note that $0 < \alpha < 0.3$ implies $2z_\alpha > 1$. Hence

$$\begin{aligned} & \sup_{c>1} \omega_z(c\varepsilon; f)(\Phi(z_\alpha - c) - \alpha)_+ \leq \sup_{c>1} 3\rho_z(c\varepsilon; f)(\Phi(z_\alpha - c) - \alpha)_+ \\ & \leq 3\rho_z(\varepsilon; f) \sup_{c>1} \min\{c, (2c)^{2/3}\}(\Phi(z_\alpha - c) - \alpha)_+ \\ & \leq 3\rho_z(\varepsilon; f) \max\left\{(1 - 2\alpha) \min\{z_\alpha, (2z_\alpha)^{2/3}\}\mathbb{1}\{z_\alpha \geq 1\}, (0.5 - \alpha) \min\{2z_\alpha, (4z_\alpha)^{2/3}\}\right\} \\ & \leq 3\omega_z(\varepsilon; f)(1 - 2\alpha) \min\{z_\alpha, (2z_\alpha)^{2/3}\}. \end{aligned}$$

In conclusion, $L_{z,\alpha}(\varepsilon; f) \leq 3(1 - 2\alpha) \min\{z_\alpha, (2z_\alpha)^{2/3}\}\omega_z(\varepsilon; f)$. □

C.5. Proof of Theorem 2.2. It follows from Theorem 2.1 and Proposition 2.2 that $A_1^3\omega_z(\varepsilon; f) \cdot \omega_m(\varepsilon; f)^2 \geq R_z(\varepsilon; f) \cdot R_m(\varepsilon; f)^2 \geq a_1^3\omega_z(\varepsilon; f) \cdot \omega_m(\varepsilon; f)^2$ and $\rho_z(\varepsilon; f)\cdot\rho_m(\varepsilon; f)^2 \leq \omega_z(\varepsilon; f)\cdot\omega_m(\varepsilon; f)^2 \leq 27\rho_z(\varepsilon; f)\cdot\rho_m(\varepsilon; f)^2$. Furthermore,

$$\frac{\varepsilon^2}{2} \leq \rho_z(\varepsilon; f) \cdot \rho_m(\varepsilon; f)^2 \leq 3\varepsilon^2. \tag{C.57}$$

This can be shown as follows. Let $u = \rho_m(\varepsilon; f) + M(f)$ and define $f_u(t) = \max\{f(t), u\}$ as in Section 2.1. Note that $\|f - f_u\|_\infty \leq \rho_m(\varepsilon; f)$ and it follows from the definition of $\rho_m(\varepsilon; f)$ that $\|f - f_u\|_2 = \varepsilon$. As illustrated in Figure 1 in Section 2.1 (with special attention to the rectangle ABCD and the triangle EDF),

$$2\rho_z(\varepsilon; f) \cdot \rho_m(\varepsilon; f)^2 \geq \int_0^1 (f(t) - f_u(t))^2 dt = \varepsilon^2$$
$$\geq \max\left\{\int_0^{Z(f)} (f(t) - f_u(t))^2 dt, \int_{Z(f)}^1 (f(t) - f_u(t))^2 dt\right\} \geq \frac{1}{3}\rho_z(\varepsilon; f) \cdot \rho_m(\varepsilon; f)^2.$$

To conclude, we have for any $f \in \mathcal{F}$

$$274\varepsilon^2 > 81A_1^3\varepsilon^2 \geq R_z(\varepsilon; f) \cdot R_m(\varepsilon; f)^2 \geq \frac{a_1^3}{2}\varepsilon^2 \geq \frac{\Phi(-0.5)^3}{2}\varepsilon^2.$$

Similarly, we have

$$L_{z,\alpha}(\varepsilon; f) \cdot L_{m,\alpha}(\varepsilon; f)^2 \geq (0.6 - 2\alpha)^3 \cdot \omega_z(\frac{\varepsilon}{3}; f) \cdot \omega_m(\frac{\varepsilon}{3}; f)^2 \geq \frac{(0.6 - 2\alpha)^3}{18}\varepsilon^2,$$

and $L_{z,\alpha}(\varepsilon; f) \cdot L_{m,\alpha}(\varepsilon; f)^2 \leq B_\alpha^3 \omega_z(\varepsilon; f)\omega_m(\varepsilon; f)^2 \leq 3^7 \cdot (1 - 2\alpha)^3\varepsilon^2$. □

C.6. Proof of Theorem 2.3 . We will first introduce two propositions, the proofs of which are deferred to the next section. Based on these two propositions, we will complete the proof of the theorem.

PROPOSITION C.1 (Penalty for super-efficiency in estimation of the minimizer). *For any estimator $\hat{Z}$, if $\exists f \in \mathcal{F}$ such that $\mathbb{E}_f|\hat{Z} - Z(f)| \leq cR_z(\varepsilon; f)$, then $\exists f_1 \in \mathcal{F}$, such that*

$$\mathbb{E}_{f_1}(|\hat{Z} - Z(f_1)|) \geq h_z(c)R_z(\varepsilon; f_1),$$

for $0 < c < \frac{2}{15}$, where $h_z(c)$ is a constant only depending on c satisfying that $h_z(c) \geq \mathbb{1}\{0.0007 \leq c < \frac{2}{15}\}0.111\left(1 - \Phi(1 + \Phi^{-1}(3c))\right) + \mathbb{1}\{0 < c < 0.0007\} \max\{\frac{1}{24}\Phi^{-1}(1 - 3c)^{\frac{2}{3}}, 0.111\left(1 - \Phi(1 + \Phi^{-1}(3c))\right)\}$.

PROPOSITION C.2 (Penalty for super-efficiency in estimation of the minimum). *For any estimator $\hat{M}$, if $\exists f \in \mathcal{F}$ such that $\mathbb{E}|\hat{M} - M(f)| \leq cR_m(\varepsilon; f)$, then $\exists f_1 \in \mathcal{F}$, such that*

$$\mathbb{E}_{f_1}|\hat{M} - M(f_1)| \geq h_m(c)R_m(\varepsilon; f_1),$$

for $0 < c < 0.1$, where $h_m(c)$ is a constant only depending on c satisfying $h_m(c) \geq \mathbb{1}\{0.1 > c \geq \frac{\Phi(-1)}{2.06}\}0.208118 + \mathbb{1}\{0 < c < \frac{\Phi(-1)}{2.06}\}z_{2.06c}^{\frac{2}{3}}/4.12$.

In the propositions, we will use c instead of γ as used in the main paper. We will keep this change in the remaining proof of the theorem to avoid confusion with the usage of γ in the proofs of supporting lemmas that we deferred to the next section.

By Proposition C.1, we have

$$h_z(c) \geq \frac{1}{24} z_{3c}^{\frac{2}{3}}, \text{ for } c < 0.0007.$$

Suppose $\bar{h}_z(c) = 0.111(1-\Phi(1-z_{3c}))$. Clearly, $\bar{h}_z(c)$ decreases as c increases. Moreover, we know that $\left(\log\left(\frac{1}{c}\right)\right)^{\frac{1}{3}}$ also decreases as c increases, when $c \in (0, 0.1)$. Thus,

$$\inf_{c\in[0.0007,0.1]} \frac{\bar{h}_z(c)}{\left(\log\left(\frac{1}{c}\right)\right)^{\frac{1}{3}}} \geq \min_{7\leq k\leq 999} \inf_{c\in[\frac{k}{10000},\frac{k+1}{10000}]} \frac{\bar{h}_z(c)}{\left(\log\left(\frac{1}{c}\right)\right)^{\frac{1}{3}}}$$
$$\geq \min_{7\leq k\leq 999} \frac{\bar{h}_z(\frac{k+1}{10000})}{\left(\log\left(\frac{10000}{k}\right)\right)^{\frac{1}{3}}} \geq 0.0266 > \frac{1}{38}.$$

By Proposition C.2, we have

$$\inf_{c\in[\frac{\Phi(-1)}{2.06},0.1)} \frac{h_m(c)}{\left(\log\left(\frac{1}{c}\right)\right)^{\frac{1}{3}}} \geq \frac{0.208118}{\left(\log\left(\frac{2.06}{\Phi(-1)}\right)\right)^{\frac{1}{3}}} \geq 0.1520614 > \frac{1}{7}.$$

Therefore, it remains to understand relationships between $z_{2.06c}^{\frac{2}{3}}$, $z_{3c}^{\frac{2}{3}}$ and $\left(\log\left(\frac{1}{c}\right)\right)^{\frac{1}{3}}$. We have the following lemma that we will prove in Section D on page 108.

Lemma C.7. *For $\alpha < 0.08$, $z_{2.06\alpha} \geq 0.61\sqrt{\log 1/\alpha}$. For $\alpha < 0.005$, $z_{3\alpha} \geq 0.599\sqrt{\log 1/\alpha}$.*

Since $0.08 > \frac{\Phi(-1)}{2.06}$, we have for $c < 0.1$,

$$h_m(c) \geq \min\{\frac{1}{7}, 0.61^{\frac{2}{3}}/4.12\} \left(\log\frac{1}{c}\right)^{\frac{1}{3}} = \frac{1}{7}\left(\log\frac{1}{c}\right)^{\frac{1}{3}}.$$

For $h_z(c)$, we have, for $c < 0.1$,

$$h_z(c) \geq \min\{\frac{1}{38}, 0.599^{\frac{2}{3}}\frac{1}{24}\} \left(\log\frac{1}{c}\right)^{\frac{1}{3}} = \frac{1}{38}\left(\log\frac{1}{c}\right)^{\frac{1}{3}}.$$

Now we prove the propositions.

Proof of Proposition C.1. We have the following two lemmas, which we will prove in Section D on page 109 and 115.

Lemma C.8. *For any estimator $\hat{Z}$, if $\exists f \in \mathcal{F}$ such that $\mathbb{E}_f|\hat{Z} - Z(f)| \leq c\rho_z(\varepsilon; f)$, then $\exists f_1 \in \mathcal{F}$, such that*

$$\mathbb{E}_{f_1}(|\hat{Z} - Z(f_1)|) \geq \tilde{h}_z(c)\rho_z(\varepsilon; f_1)$$

for $c < 1$. For $0 < c < 0.0011$, $\tilde{h}_z(c) \geq \frac{1}{16}\Phi^{-1}(1-2c)^{\frac{2}{3}}$.

Lemma C.9. *For any estimator $\hat{Z}$, if $\exists f \in \mathcal{F}$ such that $\mathbb{E}_f|\hat{Z} - Z(f)| \leq c\rho_z(\varepsilon; f)$, then $\exists f_1 \in \mathcal{F}$, such that*

$$\mathbb{E}_{f_1}(|\hat{Z} - Z(f_1)|) \geq \tilde{h}_z(c)\rho_z(\varepsilon; f_1)$$

for $c < 1$. For $0 < c < 0.2$, $\tilde{h}_z(c) \geq 0.1666\left(1 - \Phi(1 + \Phi^{-1}(2c))\right)$.

Recall that, by Lemma C.6, $0.308\rho_z(\varepsilon; f) \leq R_z(\varepsilon; f) \leq \frac{3}{2}\rho_z(\varepsilon; f)$. Therefore, for any estimator $\hat{Z}$, if $\exists f \in \mathcal{F}$ such that $\mathbb{E}_f|\hat{Z} - Z(f)| \leq cR_z(\varepsilon; f)$, then $\exists f_1 \in \mathcal{F}$, such that

$$\mathbb{E}_{f_1}(|\hat{Z} - Z(f_1)|) \geq h_z(c)R_z(\varepsilon; f_1),$$

for $c < \frac{2}{15}$. $h_z(c) \geq \mathbb{1}\{0.0007 \leq c < \frac{2}{15}\}0.111\left(1 - \Phi(1 + \Phi^{-1}(3c))\right) + \mathbb{1}\{c < 0.0007\}\max\{\frac{1}{24}\Phi^{-1}(1-3c)^{\frac{2}{3}}, 0.111\left(1 - \Phi(1 + \Phi^{-1}(3c))\right)\}$.

□

Proof of Proposition C.2. Again we introduce a lemma and prove it in Section D on page 115.

Lemma C.10. *For any estimator $\hat{M}$, if $\exists f \in \mathcal{F}$ such that $\mathbb{E}_f|\hat{M} - M(f)| \leq c\rho_m(\varepsilon; f)$, then $\exists f_1 \in \mathcal{F}$, such that*

$$\mathbb{E}_{f_1}(|\hat{M} - M(f_1)|) \geq \tilde{h}_m(c)\rho_m(\varepsilon; f_1)$$

for $c < 1$. For $c \leq 0.103$, $\tilde{h}_m(c) \geq \mathbb{1}\{0.103 \geq c \geq \frac{\Phi(-1)}{2}\}0.214362 + \mathbb{1}\{c < \frac{\Phi(-1)}{2}\}z_{2c}^{\frac{2}{3}}/4$.

According to Lemma C.5, we have $R_m(\varepsilon; f) \leq 1.03\rho_m(\varepsilon; f)$. Therefore, we have, for any estimator $\hat{M}$, if $\exists f \in \mathcal{F}$ such that $\mathbb{E}_f|\hat{M} - M(f)| \leq cR_m(\varepsilon; f)$, then $\exists f_1 \in \mathcal{F}$, such that

$$\mathbb{E}_{f_1}|\hat{M} - M(f_1)| \geq h_m(c)R_m(\varepsilon; f_1)$$

for $c < 1$. For $c < 0.1$, $h_m(c) \geq \mathbb{1}\{0.1 > c \geq \frac{\Phi(-1)}{2.06}\}0.208118 + \mathbb{1}\{c < \frac{\Phi(-1)}{2.06}\}z_{2.06c}^{\frac{2}{3}}/4.12$.

□

C.7. Proof of Theorem 3.1 . Recall that $\tilde{j}$ is defined in Equation (C.1) and only depends on Y_l. Then, We have

$$\begin{aligned} \mathbb{E}(|\hat{Z}-Z(f)|) =& \mathbb{E}(\mathbb{1}\{\hat{j}<\tilde{j}\}|\hat{Z}-Z(f)|)+\mathbb{E}(\mathbb{1}\{\hat{j}\geq\tilde{j}\}|\hat{Z}-Z(f)|) \\ \leq& \mathbb{E}(\mathbb{1}\{\hat{j}<\tilde{j}\}1.5m_{\hat{j}})+\mathbb{E}(\mathbb{1}\{\hat{j}\geq\tilde{j}\}|\hat{Z}-Z(f)|). \end{aligned} \tag{C.58}$$

We will show separately that

$$\mathbb{E}(\mathbb{1}\{\hat{j}<\tilde{j}\}1.5m_{\hat{j}}) \leq 32.1\rho_z(\varepsilon;f), \tag{C.59a}$$

$$\mathbb{E}(\mathbb{1}\{\hat{j}\geq\tilde{j}\}|\hat{Z}-Z(f)|) \leq 20.9\rho_z(\varepsilon;f). \tag{C.59b}$$

Bounds in Inequality (C.59) combined with Theorem 2.1 and Proposition 2.2 give the statement:

$$\mathbb{E}(|\hat{Z}-Z(f)|) < 53\rho_z(\varepsilon;f) \leq \frac{53}{a_1}R_z(\varepsilon;f). \tag{C.60}$$

C.7.1. *Proof of Bound* (C.59a).

$$\begin{aligned} &\mathbb{E}_{l,s}(\mathbb{1}\{\hat{j}<\tilde{j}\}m_{\hat{j}}) \\ &= \sum_{j_1=3}^{j^*-1} m_{j_1} \underbrace{\mathbb{E}_{l,s}(\mathbb{1}\{\hat{j}<\tilde{j},\hat{j}=j_1\})}_{\eta_0(j_1)} + \underbrace{\sum_{j_1=j^*}^{\infty} m_{j_1}\mathbb{E}_{l,s}(\mathbb{1}\{\hat{j}<\tilde{j},\hat{j}=j_1\})}_{\kappa} \end{aligned} \tag{C.61}$$

Next, we will analysis κ and $\eta_0(j_1)$ for $j_1 \leq j^*-1$ separately.

Analysis of κ. Clearly,

$$\kappa \leq \sum_{j_1=j^*}^{\infty} m_{j^*}\mathbb{E}_{l,s}(\mathbb{1}\{\hat{j}<\tilde{j},\hat{j}=j_1\}) \leq m_{j^*}P(j^*\leq\hat{j}<\tilde{j}). \tag{C.62}$$

Analysis of $\eta_0(j_1)$. Note that we have the following relationship between events:

$$\begin{aligned} \{j_1<\tilde{j},\hat{j}=j_1\} \subset& \{T_{j_1}\leq 2\sqrt{6}\varepsilon\sqrt{m_{j_1}}, j_1<\tilde{j}\} \\ \subset& \{\tilde{X}_{j_1,\hat{i}_{j_1}+6}-\tilde{X}_{j_1,\hat{i}_{j_1}+5}\leq 2\sqrt{6}\varepsilon\sqrt{m_{j_1}}, j_1<\tilde{j}\}\cup \\ & \{\tilde{X}_{j_1,\hat{i}_{j_1}-6}-\tilde{X}_{j_1,\hat{i}_{j_1}-5}\leq 2\sqrt{6}\varepsilon\sqrt{m_{j_1}}, j_1<\tilde{j}\}. \end{aligned} \tag{C.63}$$

Therefore, the expectation of indicator function of the first event is no larger than that of the last event (the union). Further, taking conditional expectation first with respect to Y_j gives

$$\eta_0(j_1) \leq \underbrace{\mathbb{E}_l\Big(\mathbb{E}_s\big(\mathbb{1}\{\tilde{X}_{j_1,\hat{i}_{j_1}+6} - \tilde{X}_{j_1,\hat{i}_{j_1}+5} \leq 2\sqrt{6}\varepsilon\sqrt{m_{j_1}}, j_1 < \tilde{j}\}|Y_l\big)\Big)}_{\eta_1(j_1)} + \underbrace{\mathbb{E}_l\Big(\mathbb{E}_s\big(\mathbb{1}\{\tilde{X}_{j_1,\hat{i}_{j_1}-6} - \tilde{X}_{j_1,\hat{i}_{j_1}-5} \leq 2\sqrt{6}\varepsilon\sqrt{m_{j_1}}, j_1 < \tilde{j}\}|Y_l\big)\Big)}_{\eta_2(j_1)}. \tag{C.64}$$

Bounding $\eta_1(j_1)$ and $\eta_2(j_1)$ for $j_1 \leq j^* - 1$ take similar steps, so we only walk through the steps for $\eta_1(j_1)$ for $j_1 \leq j^* - 1$. Note that only when $\hat{i}_{j_1} + 6 \leq 2^{j_1}$ the indicator function in the expectation can take 1, so in the following we have indicator function $\mathbb{1}\{\hat{i}_{j_1} + 6 \leq 2^{j_1}\}$ in the expectation without writing it out.

We introduce the following quantity for the (partly standardized) noise part of the statistic defined in stopping-rule Section 3.1.3.

$$\mathcal{E}_{j,i} = \frac{1}{\sqrt{m_j}}\left(W_2(t_{j,i}) - 2W_2(t_{j,i-1}) + W_2(t_{j,i-2})\right), \tag{C.65}$$

where W_2 is define in Equation (3.2).

Then for $2 \leq i \leq 2^j$, we have

$$\mathcal{E}_{j,i} \sim N(0, 6\varepsilon^2).$$

Hence for $j_1 \leq j^* - 1$ we have

(C.66)

$$\begin{aligned}
&\eta_1(j_1) = \mathbb{E}_l(\mathbb{E}_s(\mathbb{1}\{\tilde{X}_{j_1,\hat{i}_{j_1}+6} - \tilde{X}_{j_1,\hat{i}_{j_1}+5} \leq 2\sqrt{6}\varepsilon\sqrt{m_{j_1}}, j_1 < \tilde{j}\}|Y_l)) \\
&= \mathbb{E}_l(\mathbb{E}_s(\mathbb{1}\{(\mu_{j_1,\hat{i}_{j_1}+6} - \mu_{j_1,\hat{i}_{j_1}+5})\sqrt{m_{j_1}} - 2\sqrt{6}\varepsilon \leq -\mathcal{E}_{j_1,\hat{i}_{j_1}+6}\}|Y_l)\mathbb{1}\{j_1 < \tilde{j}\}) \\
&\leq \mathbb{E}_l(\mathbb{E}_s(\mathbb{1}\{(\mu_{j_1,i^*_{j_1}+5} - \mu_{j_1,i^*_{j_1}+4})\sqrt{m_{j_1}} - 2\sqrt{6}\varepsilon \leq -\mathcal{E}_{j_1,\hat{i}_{j_1}+6}\}|Y_l)\mathbb{1}\{j_1 < \tilde{j}\}).
\end{aligned}$$

Further, for $(\mu_{j_1,i^*_{j_1}+5} - \mu_{j_1,i^*_{j_1}+4})\sqrt{m_{j_1}}$, we have

$$\begin{aligned}
(\mu_{j_1,i^*_{j_1}+5} - \mu_{j_1,i^*_{j_1}+4})\sqrt{m_{j_1}} &\overset{(i)}{\geq} (\frac{\rho_m(\varepsilon;f)}{\rho_z(\varepsilon;f)}m_{j_1})\sqrt{m_{j_1}} \\
&= \rho_m(\varepsilon;f)\sqrt{\rho_z(\varepsilon;f)}2^{\frac{3}{2}(j^*-j_1)}\left(\frac{m_{j^*}}{\rho_z(\varepsilon;f)}\right)^{\frac{3}{2}} \\
&\overset{(ii)}{\geq} \frac{1}{\sqrt{2}}\varepsilon 2^{\frac{3}{2}(j^*-j_1)}\left(\frac{m_{j^*}}{\rho_z(\varepsilon;f)}\right)^{\frac{3}{2}} \overset{(iii)}{\geq} \frac{1}{\sqrt{2}}\varepsilon 2^{\frac{3}{2}(j^*-j_1)}2^{-\frac{9}{2}},
\end{aligned} \tag{C.67}$$

where step(i) follows from $j_1 \leq j^* - 1$, step (ii) follows from Inequality (C.57), step (iii) is by the definition of j^* in Equation (C.2). We will use both bounds after step (ii) and step (iii) in Inequality (C.67) later. Continuing with Inequality (C.66), we have

$$\begin{aligned}
&\eta_1(j_1)\\
&\leq \mathbb{E}_l\Big(\mathbb{E}_s\big(\mathbb{1}\{\frac{1}{\sqrt{2}}\varepsilon 2^{\frac{3}{2}(j^*-j_1)}2^{-\frac{9}{2}}\left(\frac{8m_{j^*}}{\rho_z(\varepsilon;f)}\right)^{\frac{3}{2}} - 2\sqrt{6}\varepsilon \leq -\mathcal{E}_{j_1,\hat{i}_{j_1}+6}\}|Y_l\big)\mathbb{1}\{j_1 < \tilde{j}\}\Big)\\
&= \underbrace{\mathbb{E}_l\left(\Phi(2 - 2^{\frac{3}{2}(j^*-j_1-3)-1}\frac{1}{\sqrt{3}}\left(\frac{8m_{j^*}}{\rho_z(\varepsilon;f)}\right)^{\frac{3}{2}})\mathbb{1}\{j_1 < \tilde{j}\}\right)}_{\eta_3(j_1)}\\
&\leq \underbrace{\mathbb{E}_l\left(\Phi(2 - 2^{\frac{3}{2}(j^*-j_1-3)-1}\frac{1}{\sqrt{3}})\mathbb{1}\{j_1 < \tilde{j}\}\right)}_{\eta_4(j_1)}.
\end{aligned}$$

Therefore, we have $\eta_1(j_1) \leq \eta_3(j_1) \leq \eta_4(j_1)$ for $j_1 \leq j^* - 1$. Following almost the same steps, we also have $\eta_2(j_1) \leq \eta_3(j_1) \leq \eta_4(j_1)$ for $j_1 \leq j^* - 1$. This gives

$$\eta_0(j_1) \leq 2\eta_3(j_1) \leq 2\eta_4(j_1) \text{ for } j_1 \leq j^* - 1. \tag{C.68}$$

Plugging in the bounds for $\eta_0(j_1)$ ($j_1 \leq j^* - 1$) and κ back into Inequality (C.61), together with the facts that $m_{j_1} = 2^{j^*-j_1}m_{j^*}$ and $2 - 2^{\frac{3}{2}(j^*-j_1-3)-1}\frac{1}{\sqrt{3}} \leq 0$ for $j_1 \leq j^* - 5$, we have that

$$\begin{aligned}
&\mathbb{E}_{l,s}(\mathbb{1}\{\hat{j} < \tilde{j}\}m_{\hat{j}})\\
&\leq m_{j^*}P(j^* \leq \hat{j} < \tilde{j}) + \sum_{j_1=3}^{j^*-5} 2^{j^*-j_1}m_{j^*}\cdot 2\eta_4(j_1) + \sum_{j_1=j^*-4}^{j^*-1} 2^{j^*-j_1}m_{j^*}\cdot 2\eta_3(j_1)\\
&\leq m_{j^*} \times 2^5 \sum_{j_1=3}^{j^*-6} 2^{j^*-j_1-4}\Phi(2 - 2^{\frac{3}{2}(j^*-j_1-4)}\cdot\sqrt{\frac{2}{3}}) + 2\Phi(2 - 2^{\frac{3}{2}}\cdot\sqrt{\frac{2}{3}})2^5 m_{j^*}\\
&\qquad + \frac{\rho_z(\varepsilon;f)}{8} \times 32 \times (\sum_{k=0}^{3}\frac{2^{-k}8m_{j^*}}{\rho_z(\varepsilon;f)}\Phi(2 - \sqrt{\frac{2}{3}}\left(\frac{2^{-k}8m_{j^*}}{\rho_z(\varepsilon;f)}\right)^{\frac{3}{2}})) + m_{j^*}\\
&\overset{(a)}{\leq} m_{j^*} \times 2^5 \times \frac{4\Phi(2 - 8\times\sqrt{2/3})}{1 - 0.008} + 24.3m_{j^*} + 2\rho_z(\varepsilon;f)(2 + 1 + 0.5 + 0.25) + m_{j^*}\\
&< 25.4m_{j^*} + 15\rho_z(\varepsilon;f) \leq 21.4\rho_z(\varepsilon;f).
\end{aligned} \tag{C.69}$$

Therefore,

$$\mathbb{E}_{l,s}(\mathbb{1}\{\hat{j} < \tilde{j}\}1.5m_{\hat{j}}) \le 32.1\rho_z(\varepsilon; f). \tag{C.70}$$

The detailed calculations of step (a) are based on Lemma C.2.

C.7.2. *Proof of Bound* (C.59b).

$$\begin{aligned}
&\mathbb{E}(\mathbb{1}\{\hat{j} \ge \tilde{j}\}|\hat{Z} - Z(f)|) \\
&\overset{(i)}{\le} \sum_{j=1}^{j^*-4} \mathbb{E}(\mathbb{1}\{\hat{j} \ge j, \tilde{j} = j\}|\hat{Z} - Z(f)|) + \sum_{j=j^*-3}^{\infty} \mathbb{E}(\mathbb{1}\{\hat{j} \ge j, \tilde{j} = j\}|\hat{Z} - Z(f)|) \\
&\overset{(ii)}{\le} \sum_{j=1}^{j^*-4} 2^{j^*-j} m_{j^*} \mathbb{E}\Big(\mathbb{1}\{\hat{j} \ge j\}\Big(\Big[5\mathbb{1}\{X_{j,i_j^*-3} \le X_{j,i_j^*-1}\} + 4\mathbb{1}\{X_{j,i_j^*-2} \le X_{j,i_j^*-1}\} \\
&+ 6\mathbb{1}\{X_{j,i_j^*-4} \le X_{j,i_j^*-1}\}\Big]\mathbb{1}\{t_{j,i_j^*-1} \ge m_j\} + \Big[5\mathbb{1}\{X_{j,i_j^*+3} \le X_{j,i_j^*+1}\} \\
&+ 6\mathbb{1}\{X_{j,i_j^*+4} \le X_{j,i_j^*+1}\} + 4\mathbb{1}\{X_{j,i_j^*+2} \le X_{j,i_j^*+1}\}\Big]\mathbb{1}\{t_{j,i_j^*+1} \le 1\}\Big)\Big) + 6 \times 8 \times m_{j^*} \\
&\overset{(iii)}{\le} 2\sum_{j=1}^{j^*-4} 2^{j^*-j} m_{j^*}\big(4\Phi(-\frac{\rho_m(\varepsilon; f)}{\rho_z(\varepsilon; f)} m_j \frac{\sqrt{m_j}}{\sqrt{6}\varepsilon}) + 5\Phi(-2\frac{\rho_m(\varepsilon; f)}{\rho_z(\varepsilon; f)} m_j \frac{\sqrt{m_j}}{\sqrt{6}\varepsilon}) + \\
&\qquad 6\Phi(-3\frac{\rho_m(\varepsilon; f)}{\rho_z(\varepsilon; f)} m_j \frac{\sqrt{m_j}}{\sqrt{6}\varepsilon})\big) + 48 m_{j^*} \\
&\le 48 m_{j^*} + 2\sum_{j=1}^{j^*-4} 2^{j^*-j} m_{j^*}\big(4\Phi(-\sqrt{2/3}(\frac{1}{8})^{\frac{3}{2}} \times 2^{\frac{3}{2}(j^*-j-1)}) + \\
&\quad 5\Phi(-\sqrt{2/3}\cdot 2(\frac{1}{8})^{\frac{3}{2}} \times 2^{\frac{3}{2}(j^*-j-1)}) + 6\Phi(-\sqrt{2/3}\cdot 3(\frac{1}{8})^{\frac{3}{2}} \times 2^{\frac{3}{2}(j^*-j-1)})\big) \\
&\overset{(a)}{\le} 48 m_{j^*} + 2 m_{j^*}\Bigg(16\Phi(-\sqrt{\frac{2}{3}}) \times 4 + 32\Phi(-4\sqrt{\frac{1}{3}})\frac{1}{1 - 2 \times \frac{\Phi(-8\sqrt{\frac{2}{3}})}{\Phi(-4/\sqrt{3})}} \times 4 + \\
&\quad 5 \times \Phi(-2\sqrt{\frac{2}{3}}) \times 16 \times \frac{1}{1 - 2 \times \frac{\Phi(-8/\sqrt{3})}{\Phi(-2\sqrt{2/3})}} + 6 \times \Phi(-\sqrt{6}) \times \frac{1}{1 - 2 \times \frac{\Phi(-4\sqrt{3})}{\Phi(-\sqrt{6})}}\Bigg) \\
&< 20.9\rho_z(\varepsilon; f).
\end{aligned}$$

Step (i)(ii)(iii) follows from splitting, simplifying, and analyzing the events in the indicator functions. Step (a) follows from the fact that $\frac{\Phi(-2\sqrt{2}x)}{\Phi(-x)}$ decreases as $x > 0$ increases.

C.8. Proof of Theorem 3.2. We will show the coverage guarantee and the upper bound of the expected length separately.

C.8.1. *Proof of Coverage Guarantee.* Recalling that we introduced the notation j^w to denote the step that the localization procedure chooses an interval relatively far away from the right one:

$$j^w = \min\{j : |\hat{i}_j - i_j^*| \geq 5\}. \tag{C.71}$$

Then we know that $|\hat{i}_{j^w-1} - i^*_{j^w-1}| \leq 4$, so we have that $|\hat{i}_{j^w+k} - i^*_{j^w+k}| \leq 6 * 2^{k+1} - 2$ for all $k \geq -1$. We now introduce a lemma that provides an upper bound on the probability of stopping at least $K+1$ steps after reaching j^w.

Lemma C.11. *For j^w defined in Equation (C.71), and for $K \geq 0$, we have*

$$P(\hat{j} \geq j^w + K + 1) \leq \Phi(-2)^K.$$

In particular, for $K_\alpha = \lceil \frac{\log \alpha}{\log \Phi(-2)} \rceil$, $P(\hat{j} \geq j^w + K_\alpha + 1) \leq \alpha$.

Note that when $\hat{j} \leq j^w + K_\alpha$, we have $|\hat{i}_{\hat{j}} - i^*_{\hat{j}}| \leq 12 \cdot 2^{K_\alpha} - 2$, implying that $Z(f) \in [L, U]$. Therefore, we have

$$P(Z(f) \in CI_{z,\alpha}) \geq P(\hat{j} \leq j^w + K_\alpha) = 1 - P(\hat{j} \geq j^w + K_\alpha + 1) \geq 1 - \alpha. \quad \square$$

It remains to prove Lemma C.11.

Proof of Lemma C.11. Now we will calculate the probability that the stopping rule does not stop K steps after j^w. When $j^w = \infty$, $\hat{j}$ can never be larger that j^w, so it suffices to consider the event $\{j^w < \infty\}$.

$$\begin{aligned}
&\mathbb{E}_{l,s}\Big(\mathbb{1}\{\hat{j} \geq j^w + K + 1\}\mathbb{1}\{j^w < \infty\}\Big) \\
&= \mathbb{E}_{l,s}\Big(\sum_{j_1=3}^{\infty} \mathbb{1}\{\hat{j} \geq j_1 + K + 1\}\mathbb{1}\{j^w = j_1\}\Big) \\
&= \mathbb{E}_{l}\Big(\sum_{j_1=3}^{\infty} \mathbb{E}_s(\mathbb{1}\{\hat{j} \geq j_1 + K + 1\}|Y_l)\mathbb{1}\{j^w = j_1\}\Big) \\
&\overset{(i)}{\leq} \mathbb{E}_{l}\Big(\sum_{j_1=3}^{\infty} \Phi(-2)^K \mathbb{1}\{j^w = j_1\}\Big) \leq \Phi(-2)^K.
\end{aligned} \tag{C.72}$$

The rationale for step (i) in Equation (C.72) is as follows. Define the set of possible localization sequences with $j^w = j_1$ truncated at step $j_1 + K + 1$:

$$\mathfrak{S}(j_1, K+1) = \Big\{(i_0, \ldots, i_{j_1+1+K}) \Big| (1 \vee 2i_j - 2) \leq i_{j+1} \leq (2i_j + 1 \wedge 2^{j+1}),$$
$$\forall 0 \leq j \leq j_1 + K, \& \, i_0 = 1, \& \, |i_{j_1} - i^*_{j_1}| \geq 5, \& \, |i_j - i^*_j| \leq 4, \forall 0 \leq j < j_1 \Big\}.$$

$\forall s \in \mathfrak{S}(j_1, K+1)$, denote $(i_l, \dots, i_h)$ in s as $s(l,h)$, and denote the sequence $(\hat{i}_l, \dots, \hat{i}_h)$ produced by the localization procedure as $\hat{s}(l,h)$. If $l = h$, we will abbreviate $s(l,l)$ into $s(l)$ and $\hat{s}(l,l)$ into $\hat{s}(l)$. Then we know that for $s \in \mathfrak{S}(j_1, K+1)$ with $s(j_1) \leq i^*_{j_1} - 5$, we have $s(j) + 6 < i^*_j$ for $j = j_1 + 1, \dots, K+1$, therefore, $\mu_{j,s(j)+6} - \mu_{j,s(j)+5} \leq 0$. On the other hand, for $s \in \mathfrak{S}(j_1, K+1)$ with $s(j_1) \geq i^*_{j_1} + 5$, we have $\mu_{j,s(j)-6} - \mu_{j,s(j)-5} \leq 0$. Now we define a sign function indicating which side $s(j)$ is on to i^*_j,

$$\mathfrak{sg}(s,j) = \text{sign}\{i^*_j - s(j)\}.$$

We introduce the shorthand $\tau_{j,i} = W_2(t_{j,i}) - W_2(t_{j,i-1})$. Now we proceed to the analysis of the first inequality in Equation (C.72), and without confusion, we write $\mathfrak{S}(j,K+1)$ as $\mathfrak{S}$ and $\mathfrak{sg}(s,j)$ as $\mathfrak{sg}$.

$$\begin{aligned}
&\mathbb{E}_s\big(\mathbb{1}\{\hat{j} \geq j_1 + K + 1\}|Y_l\big)\mathbb{1}\{j^w = j_1\} \\
&= \mathbb{E}_s\big(\sum_{s\in\mathfrak{S}} \mathbb{1}\{\hat{j} \geq j_1 + K + 1, \hat{s}(0, j_1 + 1 + K) = s\}|Y_l\big)\mathbb{1}\{j^w = j_1\} \\
&\leq \mathbb{E}_s\big(\sum_{s\in\mathfrak{S}} \mathbb{1}\big\{min\{\tilde{X}_{j,s(j)+6} - \tilde{X}_{j,s(j)+5}, \tilde{X}_{j,s(j)-6} - \tilde{X}_{j,s(j)-5}\} \geq 2\sqrt{2}c_s\varepsilon\sqrt{m_j}, \\
&\qquad \forall j = j_1 + 1, \cdots, j_1 + K\big\}\mathbb{1}\{\hat{s}(0, j_1 + 1 + K) = s\}|Y_l\big)\mathbb{1}\{j^w = j_1\} \\
&\leq \sum_{s\in\mathfrak{S}} \mathbb{E}_s\big(\mathbb{1}\{\tilde{X}_{j,s(j)+6\mathfrak{sg}} - \tilde{X}_{j,s(j)+5\mathfrak{sg}} \geq 2\sqrt{2}c_s\varepsilon\sqrt{m_j}, \\
&\forall j = j_1 + 1, \cdots, j_1 + K\}\mathbb{1}\{\hat{s}(0, j_1 + 1 + K) = s\}|Y_l\big)\mathbb{1}\{j^w = j_1\} \\
&\leq \sum_{s\in\mathfrak{S}} \mathbb{E}_s\big(\mathbb{1}\{m_j \cdot \mu_{j,s(j)+6\mathfrak{sg}} - m_j \cdot \mu_{j,s(j)+5\mathfrak{sg}} + \tau_{j,s(j)+6\mathfrak{sg}} - \tau_{j,s(j)+5\mathfrak{sg}} \geq 2\sqrt{2}c_s\varepsilon\sqrt{m_j}, \\
&\qquad \forall j = j_1 + 1, \cdots, j_1 + K\}\mathbb{1}\{\hat{s}(0, j_1 + 1 + K) = s\}|Y_l\big)\mathbb{1}\{j^w = j_1\} \\
&\leq \sum_{s\in\mathfrak{S}} \mathbb{E}_s\big(\mathbb{1}\{\tau_{j,s(j)+6\mathfrak{sg}} - \tau_{j,s(j)+5\mathfrak{sg}} \geq 2\sqrt{2}c_s\varepsilon\sqrt{m_j}, \\
&\qquad \forall j = j_1 + 1, \cdots, j_1 + K\}\mathbb{1}\{\hat{s}(0, j_1 + 1 + K) = s\}|Y_l\big)\mathbb{1}\{j^w = j_1\} \\
&= \sum_{s\in\mathfrak{S}} \Phi(-2)^K \mathbb{E}_s\big(\mathbb{1}\{\hat{s}(0, j_1 + 1 + K) = s\}|Y_l\big)\mathbb{1}\{j^w = j_1\} \\
&= \Phi(-2)^K \mathbb{1}\{j^w = j_1\}.
\end{aligned}$$

□

C.8.2. *Proof of Upper Bound of Expected Length.* We have the following lemma for the length of the confidence interval for the minimizer.

LEMMA C.12 (Length of Confidence Interval for the Minimizer). *For* $0 < \alpha < 0.3$*, the expected length of the confidence interval given in* (3.6) *satisfies*

$$\mathbb{E}(CI_{z,\alpha}(Y)) \leq (24 \times 2^{K_\alpha} - 3) \times 17.5 \times \rho_z(\varepsilon; f) \leq C_{z,\alpha}L_{z,\alpha}(\varepsilon; f).$$

PROOF OF LEMMA C.12. Recall that we denote by $\tilde{j}$ the stage where the localization procedure start choosing an interval not close to the target:

$$\tilde{j} = \min\{j : |\hat{i}_j - i_j^*| \geq 2\}.$$

To prove the lemma, we only need to upper bound $\mathbb{E}(m_{\hat{j}})$. Splitting the entire probability space into smaller events gives

$$\begin{aligned}
&\mathbb{E}(m_{\hat{j}}) = \mathbb{E}(m_{\hat{j}}\mathbb{1}\{\hat{j} \geq j^* - 3\}) + \mathbb{E}(m_{\hat{j}}\mathbb{1}\{\hat{j} \leq j^* - 4\}) \\
&\leq 8m_{j^*} + \mathbb{E}(m_{\hat{j}}\mathbb{1}\{\hat{j} \geq \tilde{j}, \hat{j} \leq j^* - 4\}) + \mathbb{E}(m_{\hat{j}}\mathbb{1}\{\hat{j} \leq \tilde{j} - 1, \hat{j} \leq j^* - 4\}) \\
&\leq 2\rho_z(\varepsilon; f) + \underbrace{\mathbb{E}(m_{\tilde{j}}\mathbb{1}\{\tilde{j} \leq \hat{j} \leq j^* - 4\})}_{\eta_1} + \underbrace{\sum_{j=1}^{j^*-4} m_j \mathbb{E}(\mathbb{1}\{\hat{j} = j, \tilde{j} \geq j+1\})}_{\eta_2}.
\end{aligned} \tag{C.73}$$

We bound η_1 and η_2 separately as follows. We start with η_1.

$$\begin{aligned}
\eta_1 &= \mathbb{E}(m_{\tilde{j}}\mathbb{1}\{\tilde{j} \leq \hat{j} \leq j^* - 4\}) \leq \mathbb{E}(m_{\tilde{j}}\mathbb{1}\{\tilde{j} \leq j^* - 4\}) = \sum_{j=1}^{j^*-4} m_j \mathbb{E}(\mathbb{1}\{\tilde{j} = j\}) \\
&\leq \sum_{j=1}^{j^*-4} m_j \mathbb{E}(\mathbb{1}\{X_{j,i_j^*+3} \leq X_{j,i_j^*+1}, t_{j,i_j^*+3} \leq 1\} + \mathbb{1}\{X_{j,i_j^*+2} \leq X_{j,i_j^*+1}, t_{j,i_j^*+1} \leq 1\} \\
&\quad + \mathbb{1}\{X_{j,i_j^*-3} \leq X_{j,i_j^*-1}, t_{j,i_j^*-3} \geq m_j\ \} + \mathbb{1}\{X_{j,i_j^*-2} \leq X_{j,i_j^*-1}, t_{j,i_j^*-2} \geq m_j\}) \\
&\quad + \mathbb{1}\{X_{j,i_j^*-4} \leq X_{j,i_j^*-1}, t_{j,i_j^*-4} \geq m_j\} + \mathbb{1}\{X_{j,i_j^*+4} \leq X_{j,i_j^*+1}, t_{j,i_j^*+4} \leq 1\}) \\
&\leq \sum_{j=1}^{j^*-4} 2^{j^*-j} m_{j^*} \times 2(\Phi(-\frac{\rho_m(\varepsilon; f)}{\rho_z(\varepsilon; f)} m_j \sqrt{m_j} \frac{1}{c_l\sqrt{2}\varepsilon}) + \Phi(-\frac{\rho_m(\varepsilon; f)}{\rho_z(\varepsilon; f)} 2m_j \sqrt{m_j} \frac{1}{c_l\sqrt{2}\varepsilon}) \\
&\quad + \Phi(-\frac{\rho_m(\varepsilon; f)}{\rho_z(\varepsilon; f)} 3m_j \sqrt{m_j} \frac{1}{c_l\sqrt{2}\varepsilon})) \\
&\leq \sum_{j=1}^{j^*-4} 2^{j^*-j} \frac{\rho_z(\varepsilon; f)}{2} \Big(\Phi(-2^{\frac{3}{2}(j^*-j)} (\frac{1}{8})^{\frac{3}{2}} \frac{1}{2c_l}) + \Phi(-2^{\frac{3}{2}(j^*-j)} \times 2 \times (\frac{1}{8})^{\frac{3}{2}} \frac{1}{2c_l}) \\
&\quad + \Phi(-2^{\frac{3}{2}(j^*-j)} \times 3 \times (\frac{1}{8})^{\frac{3}{2}} \frac{1}{2c_l})\Big) \\
&= 4 \sum_{j=1}^{j^*-4} 2^{j^*-j-3} \rho_z(\varepsilon; f) \Big(\Phi(-2^{\frac{3}{2}(j^*-j-4)} \sqrt{2/3}) + \Phi(-\frac{1}{\sqrt{2}} 2^{\frac{3}{2}(j^*-j-3)} \sqrt{2/3}) \\
&\quad + \Phi(-3 \times 2^{\frac{3}{2}(j^*-j-3)} \sqrt{2/3})\Big) \\
&\overset{(a)}{\leq} 8\rho_z(\varepsilon; f) \times \big(\Phi(-\sqrt{2/3}) + \Phi(-2\sqrt{2/3}) + \Phi(-\sqrt{6}) + \\
&\quad [2\Phi(-4/\sqrt{3}) + 2\Phi(-8/\sqrt{3}) + 2\Phi(-12/\sqrt{3})] \frac{1}{1 - 2\frac{\Phi(-8\sqrt{2/3})}{\Phi(-4/\sqrt{3})}}\big).
\end{aligned}$$

Step (a) follows from the fact that $\frac{\Phi(-2\sqrt{2}x)}{\Phi(-x)}$ decreases as $x > 0$ increases.

Now we bound η_2 in Equation (C.73).

$$\eta_2 = \sum_{j=1}^{j^*-4} m_j \mathbb{E}(\mathbb{1}\{\hat{j} = j, \tilde{j} \geq j+1\}) = \sum_{j=1}^{j^*-4} m_j \mathbb{E}_l\big(\mathbb{E}_s(\hat{j} = j | Y_l)\mathbb{1}\{\tilde{j} \geq j+1\}\big)$$
$$\leq \sum_{j=1}^{j^*-4} m_j \mathbb{E}_l\big(\mathbb{E}_s(\tilde{X}_{j,\hat{i}_j+6} - \tilde{X}_{j,\hat{i}_j+5} \leq 2\sqrt{2}c_s\varepsilon\sqrt{m_j} | Y_l)\mathbb{1}\{\tilde{j} \geq j+1\} +$$
$$\mathbb{E}_s(\tilde{X}_{j,\hat{i}_j-6} - \tilde{X}_{j,\hat{i}_j-5} \leq 2\sqrt{2}c_s\varepsilon\sqrt{m_j} | Y_l)\mathbb{1}\{\tilde{j} \geq j+1\}\big)$$
$$\leq \sum_{j=1}^{j^*-4} m_j \mathbb{E}_l\big(2\Phi(2 - \frac{\rho_m(\varepsilon; f)}{\rho_z(\varepsilon; f)} m_j \frac{\sqrt{m_j}}{c_s\sqrt{2}\varepsilon})\mathbb{1}\{\tilde{j} \geq j+1\}\big)$$
$$\leq \sum_{j=1}^{j^*-4} 2^{j^*-j}\frac{\rho_z(\varepsilon; f)}{4} \times 2\Phi(2 - \frac{2^{\frac{3}{2}(j^*-j-3)}}{2c_s})\mathbb{E}_l\left(\mathbb{1}\{\tilde{j} \geq j+1\}\right)$$
$$\leq \sum_{j=1}^{j^*-4} 2^{j^*-j}\frac{\rho_z(\varepsilon; f)}{4} \times 2\Phi(2 - 2^{\frac{3}{2}(j^*-j-4)}\sqrt{2/3})$$
$$\leq 8\rho_z(\varepsilon; f)\big(\Phi(2 - \sqrt{2/3}) + 2\Phi(2 - 4/\sqrt{3}) + 4\Phi(2 - 8 \times \sqrt{2/3})\frac{1}{1 - 0.008}\big).$$

Plugging the bounds for η_1 and η_2 back to Equation (C.73) gives

$$\mathbb{E}(m_{\hat{j}}) < 17.5\rho_z(\varepsilon; f). \tag{C.74}$$

Therefore,

$$\mathbb{E}(CI_{z,\alpha}) \leq (24 \times 2^{K_\alpha} - 3) \times \mathbb{E}(m_{\hat{j}}) \leq (24 \times 2^{K_\alpha} - 3) \times 17.5\rho_z(\varepsilon; f). \tag{C.75}$$

Further, by Theorem 2.1, Proposition 2.1, and Proposition 2.2, we have $L_{z,\alpha}(\varepsilon; f) \geq b_\alpha\omega_z(\varepsilon/3; f) \geq b_\alpha\rho_z(\varepsilon; f)/3$ when $0 < \alpha < 0.3$, which gives the statement.

□

C.9. Proof of Theorem 3.3. We introduce two quantities associated with (Y_l, Y_s) induced error and Y_e induced error.

$$\hat{f} = \frac{1}{m_{\hat{j}}}\int_{t_{\hat{i}_{\hat{j}}+\Delta-1}}^{t_{\hat{i}_{\hat{j}}+\Delta}} f(t)dt, \qquad \hat{\mathfrak{z}} = \frac{1}{m_{\hat{j}}}(W_3(t_{\hat{i}_{\hat{j}}+\Delta}) - W_3(t_{\hat{i}_{\hat{j}}+\Delta-1})),$$

where

$$\Delta = 2\left(\mathbb{1}\{\tilde{X}_{\hat{j},\hat{i}_{\hat{j}}+6} - \tilde{X}_{j,\hat{i}_{\hat{j}}+5} \leq 2\sigma_j\} - \mathbb{1}\{\tilde{X}_{\hat{j},\hat{i}_{\hat{j}}-6} - \tilde{X}_{j,\hat{i}_{\hat{j}}-5} \leq 2\sigma_j\}\right).$$

Clearly, $\hat{f}$ only depends on (Y_l, Y_s) and $\hat{\mathfrak{Z}}|(Y_l, Y_s)$ is independent with (Y_l, Y_s).

Therefore, we have:

(C.76)

$$\begin{aligned}&\mathbb{E}_{l,s,e}\big((\hat{M} - M(f))^2\big) = \mathbb{E}_{l,s,e}\big((\hat{f} - M(f))^2 + \hat{\mathfrak{Z}}^2 + 2\hat{\mathfrak{Z}}(\hat{f} - M(f))\big)\\ &\overset{(a)}{=} \mathbb{E}_{l,s}\big((\hat{f} - M(f))^2 + \frac{3\varepsilon^2}{m_{\hat{j}}}\big) \leq \mathbb{E}_{l,s}\big((\hat{f} - M(f))^2\big) + \frac{24\varepsilon^2}{\rho_z(\varepsilon; f)}\mathbb{E}\big(2^{\hat{j}-j^*}\big).\end{aligned}$$

Step (a) follows from taking conditional expectation on (Y_l, Y_s) and the mutual independence between Y_l, Y_s and Y_e.

For the second term of the right hand side of Inequality (C.76), we have the following lemma that we will prove later.

LEMMA C.13.

$$\mathbb{E}(2^{\hat{j}-j^*}) < \frac{35}{8} \leq \frac{35}{4}\frac{\rho_z(\varepsilon; f)\rho_m(\varepsilon; f)^2}{\varepsilon^2}. \tag{C.77}$$

For the first term of the right hand side of Inequality (C.76), we have

$$\begin{aligned}&\mathbb{E}_{l,s}\big((\hat{f} - M(f))^2\big)\\ &= \mathbb{E}_{l,s}\big((\hat{f} - M(f))^2\mathbb{1}\{\tilde{j} \leq \hat{j}\}\big) + \mathbb{E}_{l,s}\big((\hat{f} - M(f))^2\mathbb{1}\{\tilde{j} > \hat{j}\}\big).\end{aligned} \tag{C.78}$$

To bound the first term in Equation (C.78), on the event $\{\tilde{j} \leq \hat{j}\}$, we have

(C.79)

$$\begin{aligned}(\hat{f} - M(f))^2 &\leq \big((\hat{f} - \mu_{\hat{j},\hat{i}_{\hat{j}}})_+ + (\mu_{\hat{j},\hat{i}_{\hat{j}}} - M(f))\big)^2\\ &\leq 2\big(\hat{f} - \mu_{\hat{j},\hat{i}_{\hat{j}}}\big)_+^2 + 2\big(\mu_{\hat{j},\hat{i}_{\hat{j}}} - M(f)\big)^2\\ &\leq 2\big(\hat{f} - \mu_{\hat{j},\hat{i}_{\hat{j}}}\big)_+^2 + \frac{8}{3}\big(\mu_{\hat{j},\hat{i}_{\hat{j}}} - \mu_{\tilde{j},\hat{i}_{\tilde{j}}}\big)_+^2 + 8\big(\mu_{\tilde{j},\hat{i}_{\tilde{j}}} - M(f)\big)^2.\end{aligned}$$

Therefore, going back to Inequality (C.78), we have

(C.80)

$$\begin{aligned}&\mathbb{E}_{l,s}\big((\hat{f} - M(f))^2\big)\\ &\leq 2\mathbb{E}_{l,s}\Big(\big((\hat{f} - \mu_{\hat{j},\hat{i}_{\hat{j}}})_+\big)^2\mathbb{1}\{\tilde{j} \leq \hat{j}\}\Big) + \mathbb{E}_{l,s}\big((\hat{f} - M(f))^2\mathbb{1}\{\tilde{j} > \hat{j}\}\big)\\ &\quad + 2\mathbb{E}_{l,s}\left(\frac{4}{3}\Big(\big(\mu_{\hat{j},\hat{i}_{\hat{j}}} - \mu_{\tilde{j},\hat{i}_{\tilde{j}}}\big)_+\Big)^2\mathbb{1}\{\tilde{j} \leq \hat{j}\}\right) + 2\mathbb{E}_{l,s}\left(4\big(\mu_{\tilde{j},\hat{i}_{\tilde{j}}} - M(f)\big)^2\mathbb{1}\{\tilde{j} \leq \hat{j}\}\right).\end{aligned}$$

To bound each term in Inequality (C.80), we introduce and prove the following proposition.

PROPOSITION C.3.

$$\mathbb{E}_{l,s}\big((\hat{f}-M(f))^2\mathbb{1}\{\tilde{j}>\hat{j}\}\big) < 15949\rho_m(\varepsilon;f)^2, \tag{C.81}$$

$$\mathbb{E}_{l,s}\Big(((\hat{f}-\mu_{\hat{j},\hat{i}_{\hat{j}}})_+)^2\mathbb{1}\{\tilde{j}\le\hat{j}\}\Big) < 13064\rho_m(\varepsilon;f)^2, \tag{C.82}$$

$$\mathbb{E}_{l,s}\left(\left(\mu_{\tilde{j},\hat{i}_{\tilde{j}}}-M(f)\right)^2\mathbb{1}\{\tilde{j}\le\hat{j}\}\right) < 3104\rho_m(\varepsilon;f)^2, \tag{C.83}$$

$$\mathbb{E}_{l,s}\left(\left(\left(\mu_{\hat{j},\hat{i}_{\hat{j}}}-\mu_{\tilde{j},\hat{i}_{\tilde{j}}}\right)_+\right)^2\mathbb{1}\{\tilde{j}\le\hat{j}\}\right) < 50857\rho_m(\varepsilon;f)^2. \tag{C.84}$$

By applying Proposition C.3 to Inequality (C.80) and using Lemma C.13, followed by plugging in Inequality (C.76), we arrive at the statement of the theorem. Now we are left with proving Proposition C.3 and Lemma C.13. Before we proceed, we introduce and prove the following lemma, which makes the equation $\mathbb{E}(\mathfrak{Y})=\mathbb{E}(\mathfrak{Y}\sum_{j\ge1}\mathbb{1}\{\hat{j}=j\})$ holds for any random variable $\mathfrak{Y}$.

LEMMA C.14. $P(\hat{j}<\infty)=1$.

PROOF. To prove this, we only need to prove $\lim_{j\to\infty}P(\hat{j}>j)=0$. Suppose $j\ge j^*+3$. For $j_1\ge j^*+2$,

$$\min\{\mu_{j_1,\hat{i}_{j_1}+6}-\mu_{j_1,\hat{i}_{j_1}+5},\mu_{j_1,\hat{i}_{j_1}-6}-\mu_{j_1,\hat{i}_{j_1}-5}\} < 13.5m_{j_1}\frac{\rho_m(\varepsilon;f)}{\rho_z(\varepsilon;f)}.$$

Note that we have inequalities

$$3\varepsilon^2\ge\rho_z(\varepsilon;f)\rho_m(\varepsilon;f)^2\ge\frac{1}{2}\varepsilon^2,\ \text{and}\ \rho_z(\varepsilon;f)\ge 4m_{j^*}, \tag{C.85}$$

which gives

$$\min\{\mu_{j_1,\hat{i}_{j_1}+6}-\mu_{j_1,\hat{i}_{j_1}+5},\mu_{j_1,\hat{i}_{j_1}-6}-\mu_{j_1,\hat{i}_{j_1}-5}\}m_{j_1}/(c_s\sqrt{2m_{j_1}\varepsilon^2})\le 13.5\cdot 2^{\frac{-3(j_1-j^*+2)}{2}}.$$

Therefore,

$$\begin{aligned}
&P(\hat{j}>j)\\
&=\mathbb{E}_l\left(\mathbb{E}_s(\Pi_{j_1=j^*+2}^{j-1}\mathbb{1}\{\min\{\tilde{X}_{j_1,\hat{i}_{j_1}+6}-\tilde{X}_{j_1,\hat{i}_{j_1}+5},\tilde{X}_{j_1,\hat{i}_{j_1}-6}-\tilde{X}_{j_1,\hat{i}_{j_1}-5}\}>2\sigma_{j_1}\}|Y_l)\right)\\
&\le\mathbb{E}_l\left(\Pi_{j_1=j^*+2}^{j-1}\Phi(-2+13.5\cdot 2^{\frac{-3(j_1-j^*+2)}{2}})\right) < \Phi(-1.85)^{j-j^*-2}.
\end{aligned} \tag{C.86}$$

Therefore, $\lim_{j\to\infty}P(\hat{j}>j)\le\lim_{j\to\infty}\Phi(-1.85)^{j-j^*-2}=0$.

□

Continuing with the proof of the Proposition C.3, we have the following lemmas that we will prove in the Section D (page 117, 122, 123 and 125).

LEMMA C.15.

$$\begin{aligned}&\mathbb{E}_{l,s}\big((\hat{f}-M(f))^2\mathbb{1}\{\tilde{j}>\hat{j}\}\big)\\ \le&(7680V+2)\rho_m(\varepsilon;f)^2+78V\rho_m(\varepsilon;f)^2+\frac{1}{16}\rho_m(\varepsilon;f)^2,\end{aligned} \tag{C.87}$$

where $V=\sup_{x\ge0}x^2\Phi(2-x)$.

LEMMA C.16.

$$\mathbb{E}_{l,s}\Big(\big((\hat{f}-\mu_{\hat{j},\hat{i}_{\hat{j}}})_+\big)^2\mathbb{1}\{\tilde{j}\le\hat{j}\}\Big)\le 6355.2V\rho_m(\varepsilon;f)^2, \tag{C.88}$$

where $V=\sup_{x\ge0}x^2\Phi(2-x)$.

LEMMA C.17.

$$\mathbb{E}_{l,s}\left(\big(\mu_{\tilde{j},\hat{i}_{\tilde{j}}}-M(f)\big)^2\mathbb{1}\{\tilde{j}\le\hat{j}\}\right)<3\big(2^8+2^8\tfrac{\Phi(-1.85)}{(1-\Phi(-2+\frac{1}{12}))^2}\big)\rho_m(\varepsilon;f)^2(23\tfrac{1}{8})Q, \tag{C.89}$$

where $Q=\sup_{x\ge0}x^2\Phi(-x)$.

LEMMA C.18.

$$\mathbb{E}_{l,s}\left(\Big(\big(\mu_{\hat{j},\hat{i}_{\hat{j}}}-\mu_{\tilde{j},\hat{i}_{\tilde{j}}}\big)_+\Big)^2\mathbb{1}\{\tilde{j}\le\hat{j}\}\right)\le 277075Q\rho_m(\varepsilon;f)^2+23850.1Q\rho_m(\varepsilon;f)^2, \tag{C.90}$$

where $Q=\sup_{x\ge0}x^2\Phi(-x)$.

These four lemmas combined with Lemma C.3 give the statement of Proposition C.3.

Finally we will prove Lemma C.13.

PROOF OF LEMMA C.13. Splitting the entire probability space into smaller events gives

$$\begin{aligned}\mathbb{E}\big(2^{\hat{j}-j^*}\big)&=\mathbb{E}(2^{\hat{j}-j^*}\mathbb{1}\{\hat{j}\le j^*+2\})+\mathbb{E}(\{2^{\hat{j}-j^*}\mathbb{1}\{\hat{j}\ge j^*+3\})\\ &\le 4+\mathbb{E}\big(2^{\hat{j}-j^*}\mathbb{1}\{\hat{j}\ge j^*+3\}\big)\\ &=4+\mathbb{E}_{l,s}\Big(\sum_{j_1\ge j^*+3}2^{j_1-j^*}\mathbb{1}\{\hat{j}=j_1,t_{\hat{j},\hat{i}_{\hat{j}}}\le\frac{7\rho_m(\varepsilon;f)}{16}+Z(f)\}\\ &\qquad+\sum_{j_1\ge j^*+3}2^{j_1-j^*}\mathbb{1}\{\hat{j}=j_1,t_{\hat{j},\hat{i}_{\hat{j}}}>\frac{7\rho_m(\varepsilon;f)}{16}+Z(f)\}\Big).\end{aligned} \tag{C.91}$$

Now we will bound the second term and the third term in the Inequality (C.91). Without loss of generality, we can assume

$$\sup\{t > Z(f) : f(t) \le \rho_m(\varepsilon; f) + M(f)\} = \rho_z(\varepsilon; f) + Z(f),$$

because otherwise the following would hold:

$$\min\{t < Z(f) : f(t) \le \rho_m(\varepsilon; f) + M(f)\} = Z(f) - \rho_z(\varepsilon; f),$$

for which one only need to flip everything around with $Z(f)$ being the center. Then for the second term, simplifying event, taking conditional expectation and calculating that gives

(C.92)

$$\begin{aligned}
&\mathbb{E}_{l,s}\big(\sum_{j_1 \ge j^*+3} 2^{j_1-j^*}\mathbb{1}\{\hat{j} = j_1, t_{\hat{j},\hat{i}_{\hat{j}}} \le \frac{7\rho_m(\varepsilon; f)}{16} + Z(f)\}\big)\\
&= \mathbb{E}_{l,s}\Big[\sum_{j_1 \ge j^*+3} 2^{j_1-j^*}\mathbb{1}\{\hat{j} = j_1, t_{j_1,\hat{i}_{j_1}} \le \frac{7\rho_m(\varepsilon; f)}{16} + Z(f),\\
&\qquad \forall j^*+2 \le j \le j_1 - 1, \mathcal{E}_{j,\hat{i}_j+6}\frac{1}{\sqrt{2}c_s\varepsilon} \ge 2 - \frac{\sqrt{m_j}}{\sqrt{2}c_s\varepsilon}(\mu_{j,\hat{i}_j+6} - \mu_{j,\hat{i}_j+5})\}\Big]\\
&\le \sum_{j_1 \ge j^*+3} 2^{j_1-j^*}\mathbb{E}_l\Big[\mathbb{E}_s\big(\mathbb{1}\{t_{j_1,\hat{i}_{j_1}} \le \frac{7\rho_m(\varepsilon; f)}{16} + Z(f),\\
&\qquad \forall j^*+2 \le j \le j_1 - 1, \mathcal{E}_{j,\hat{i}_j+6}\frac{1}{\sqrt{2}c_s\varepsilon} \ge 2 - \frac{\sqrt{m_j}}{\sqrt{2}c_s\varepsilon}(\mu_{j,\hat{i}_j+6} - \mu_{j,\hat{i}_j+5})\}|Y_l\big)\Big]\\
&\le \sum_{j_1 \ge j^*+3} 2^{j_1-j^*}\mathbb{E}_l\Big[\mathbb{1}\{t_{j_1,\hat{i}_{j_1}} \le \frac{7\rho_m(\varepsilon; f)}{16} + Z(f)\}\\
&\mathbb{E}_s\big(\mathbb{1}\{\forall j^*+2 \le j \le j_1 - 1, \frac{\mathcal{E}_{j,\hat{i}_j+6}}{\sqrt{2}c_s\varepsilon} \ge 2 - \frac{\sqrt{m_j}}{\sqrt{2}c_s\varepsilon}\frac{\rho_m(\varepsilon; f)(\frac{7}{16}\rho_z(\varepsilon; f) + 6m_j)}{\rho_z(\varepsilon; f)}\}|Y_l\big)\Big],\\
&\le \sum_{j_1 \ge j^*+3} 2^{j_1-j^*}\mathbb{E}_l\Big[\mathbb{1}\{t_{j_1,\hat{i}_{j_1}} \le \frac{7\rho_m(\varepsilon; f)}{16} + Z(f)\}\\
&\qquad \Pi_{j=j^*+2}^{j_1-1}\Phi(-2 + \frac{\rho_m(\varepsilon; f)\sqrt{\rho_z(\varepsilon; f)}}{\varepsilon}\frac{2^{\frac{j^*-j-2}{2}}}{\sqrt{2}c_s}(\frac{7}{16} + 6 * 2^{j^*-j-2}))\Big]\\
&\le \sum_{j_1 \ge j^*+3} 2^{j_1-j^*}\mathbb{E}_l\Big[\mathbb{1}\{t_{j_1,\hat{i}_{j_1}} \le \frac{7\rho_m(\varepsilon; f)}{16} + Z(f)\}]\\
&\qquad \Pi_{j=j^*+2}^{j_1-1}\Phi(-2 + \frac{2^{\frac{j^*-j-2}{2}}}{\sqrt{2}}(\frac{7}{16} + 6 * 2^{j^*-j-2}))\\
&\le \sum_{j_1 \ge j^*+3} 2^{j_1-j^*}\mathbb{E}_l\Big[\mathbb{1}\{t_{j_1,\hat{i}_{j_1}} \le \frac{7\rho_m(\varepsilon; f)}{16} + Z(f)\}\Big]\Phi(-1.8)^{j_1-j^*-2}.
\end{aligned}$$

Now we go to the third term in the Inequality (C.91).

(C.93)

$$\begin{aligned}
&\mathbb{E}_{l,s}\big(\sum_{j_1\geq j^*+3} 2^{j_1-j^*}\mathbb{1}\{\hat{j}=j_1, t_{\hat{j},\hat{i}_{\hat{j}}} > \frac{7\rho_m(\varepsilon;f)}{16}+Z(f)\}\big)\\
&=\sum_{j_1\geq j^*+3} 2^{j_1-j^*}\mathbb{E}_{l,s}\big(\mathbb{1}\{\hat{j}=j_1, t_{j_1,\hat{i}_{j_1}} > \frac{7\rho_m(\varepsilon;f)}{16}+Z(f)\}\big)\\
&=\sum_{j_1\geq j^*+3} 2^{j_1-j^*}\mathbb{E}_{l,s}\big(\mathbb{1}\{\hat{j}=j_1, t_{j_1,\hat{i}_{j_1}} > \frac{7\rho_m(\varepsilon;f)}{16}+Z(f)\},\\
&\qquad\qquad \forall j^*+2\leq j\leq j_1-1, -\mathcal{E}_{j,\hat{i}_j-5}\geq 2-\frac{\sqrt{m_j}}{c_s}(\mu_{j,\hat{i}_j-6}-\mu_{j,\hat{i}_j-5})\}\big)\\
&\leq \sum_{j_1\geq j^*+3} 2^{j_1-j^*}\mathbb{E}_l\Big[\mathbb{1}\{t_{j_1,\hat{i}_{j_1}} > \frac{7\rho_m(\varepsilon;f)}{16}+Z(f)\}\\
&\qquad\qquad \mathbb{E}_s\big(\mathbb{1}\{\forall j^*+2\leq j\leq j_1-1, -\mathcal{E}_{j,\hat{i}_j-5}\frac{1}{\sqrt{m_j}c_s\varepsilon}\geq 2\}|Y_l\big)\Big]\\
&\leq \sum_{j_1\geq j^*+3} 2^{j_1-j^*}\mathbb{E}_l\Big[\mathbb{1}\{t_{j_1,\hat{i}_{j_1}} > \frac{7\rho_m(\varepsilon;f)}{16}+Z(f)\}\Big]\Phi(-2)^{j_1-j^*-2}.
\end{aligned}$$

Plugging Inequality (C.92) and Inequality (C.93) back to Inequality (C.91) gives

(C.94)

$$\begin{aligned}
&\mathbb{E}\big(2^{\hat{j}-j^*}\big)\\
&\leq 4+\Bigg(\sum_{j_1\geq j^*+3} 2^{j_1-j^*}\mathbb{E}_l\Big[\mathbb{1}\{t_{j_1,\hat{i}_{j_1}} \leq \frac{7\rho_m(\varepsilon;f)}{16}+Z(f)\}\Big]\Phi(-1.8)^{j_1-j^*-2}\\
&\quad+\sum_{j_1\geq j^*+3} 2^{j_1-j^*}\mathbb{E}_l\Big[\mathbb{1}\{t_{j_1,\hat{i}_{j_1}} > \frac{7\rho_m(\varepsilon;f)}{16}+Z(f)\}\Big]\Phi(-2)^{j_1-j^*-2}\Bigg)\\
&\leq 4+\sum_{j_1\geq j^*+3} 2^{j_1-j^*}\Phi(-1.8)^{j_1-j^*-2}=4+8\Phi(-1.8)*\frac{1}{1-2\Phi(-1.8)}\\
&<\frac{35}{8}.
\end{aligned}$$

Therefore, we have

$$\mathbb{E}\big(2^{\hat{j}-j^*}\big)\leq\frac{35}{8}\leq\frac{35}{4}\frac{\rho_m(\varepsilon;f)^2\rho_z(\varepsilon;f)}{\varepsilon}.$$

□

C.10. Proof of Theorem 3.4. We will prove the following two lemmas separately, which give rise to the theorem.

LEMMA C.19 (Coverage of the Confidence Interval for the Minimum). *For any* $0 < \alpha < 1$*, the confidence interval* $CI_{m,\alpha}$ *given in* (3.10) *is a* $1-\alpha$ *confidence interval.*

LEMMA C.20 (Length of the Confidence Interval for the Minimum). *For* $0 < \alpha < 1$*, the expected length of the confidence interval given in* (3.10) *satisfies*

$$\mathbb{E}(|f_{hi} - f_{lo}|) \le c_{m,\alpha}\rho_m(\varepsilon; f), \text{ for all } f \in \mathcal{F},$$

where $c_{m,\alpha}$ *is a constant depending only on* α*.*

Further, when $0 < \alpha < 0.3$*, we have*

$$\mathbb{E}(|f_{hi} - f_{lo}|) \le c_{m,\alpha}\rho_m(\varepsilon; f) \le C_{m,\alpha}L_{m,\alpha}(\varepsilon; f), \text{ for all } f \in \mathcal{F},$$

where $C_{m,\alpha}$ *is an absolute constant depending only on* α*.*

PROOF OF LEMMA C.19. Define five events:

$$\begin{aligned}
&E = \{Z(f) \notin [t_{(\hat{j}-K_{\frac{\alpha}{4}}-1)_+, \hat{i}_{(\hat{j}-K_{\frac{\alpha}{4}}-1)_+}-5}, t_{(\hat{j}-K_{\frac{\alpha}{4}}-1)_+, \hat{i}_{(\hat{j}-K_{\frac{\alpha}{4}}-1)_+}+4}]\} \\
&E_1 = \{\hat{j} \ge j^w + K_{\frac{\alpha}{4}} + 1\} \\
&F = \{\hat{j} \le j^* - 2 - \tilde{K}_{\frac{\alpha}{4}}\} \\
&G = \{f_{hi} < M(f)\} \\
&H = \{f_{lo} > M(f)\}.
\end{aligned} \tag{C.95}$$

By definition $\{M(f) \in [f_{lo}, f_{hi}]\} = G^c \cap H^c$. We will bound the probabilities of the above events.

Recalling $K_\alpha = \lceil \frac{\log \alpha}{\log \Phi(-2)} \rceil$, then with Lemma C.11 we have

$$P(\hat{j} \ge j^w + K_\alpha + 1) \le \alpha,$$

so $P(E_1) \le \frac{\alpha}{4}$.

When the event $E_1^c = \{\hat{j} \le j^w + K_\alpha\}$ occurs, we have

$$Z(f) \in [t_{(\hat{j}-K_\alpha-1)_+, \hat{i}_{(\hat{j}-K_\alpha-1)_+}-5}, t_{(\hat{j}-K_\alpha-1)_+, \hat{i}_{(\hat{j}-K_\alpha-1)_+}+4}],$$

so $P(E) \le \frac{\alpha}{4}$.

To bound $P(F)$, we introduce the following lemma (proved in Section D page 130), showing that the procedure can not stop too early.

LEMMA C.21. *When $\tilde{K} \geq 4$, we have*

$$P(\hat{j} \leq j^* - 2 - \tilde{K}) \leq \Phi(-2^{\frac{3}{2}(\tilde{K}-2)-\frac{1}{2}} + 2)\frac{2}{1-\exp(-40)}.$$

Using this lemma and setting $\tilde{K}_\alpha = \max\{4,\ 2 + \lceil \log_2(2 - \Phi^{-1}(\frac{\alpha}{3}))\rceil\} > \max\{4,\ 2 + \lceil \frac{2}{3}\log_2 \max\{2 - \Phi^{-1}((1-e^{-40})\frac{\alpha}{2}), 1\} + \frac{1}{3}\rceil\}$, we can conclude that

$$P(\hat{j} \leq j^* - 2 - \tilde{K}_\alpha) \leq \alpha.$$

Therefore, we have $P(F) \leq \frac{\alpha}{4}$.

We now present two more lemmas that establish the remaining foundation of the proof. The lemmas are proved in Section D (page 130 and 131).

LEMMA C.22.

$$P(G|E^c) \leq \frac{\alpha}{4}. \tag{C.96}$$

LEMMA C.23.

$$P(H|E^c \cap F^c) \leq \frac{\alpha}{4}. \tag{C.97}$$

With these additional lemmas, we have

$$\begin{aligned}
&P(M(f) \in CI_{m,\alpha}(Y)) \geq P(E^c \cap F^c \cap G^c \cap H^c) \\
&\geq (1 - P(H|E^c \cap F^c) - P(G|E^c \cap F^c))P(E^c \cap F^c) \\
&\geq -P(H|E^c \cap F^c) + P(E^c \cap F^c) - P(G \cap E^c \cap F^c) \\
&\geq -P(H|E^c \cap F^c) + 1 - P(E) - P(F) - P(G|E^c) \\
&\geq -\frac{\alpha}{4} + 1 - \frac{\alpha}{4} - \frac{\alpha}{4} - \frac{\alpha}{4} = 1 - \alpha.
\end{aligned} \tag{C.98}$$

□

PROOF OF LEMMA C.20.

$$\begin{aligned}
&\mathbb{E}(|f_{hi} - f_{lo}|) \\
&= \mathbb{E}((S_{i_R - i_L, \frac{\alpha}{4}} c_e + z_{\frac{\alpha}{4}} c_e + \sqrt{3})\frac{\varepsilon}{\sqrt{m_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}}) \\
&< (S_{i_R - i_L, \frac{\alpha}{4}} + z_{\frac{\alpha}{4}} + \sqrt{3})\frac{2^{\frac{\tilde{K}_{\frac{\alpha}{4}}}{2}} c_e \varepsilon}{\sqrt{m_{j^*}}}\mathbb{E}(2^{\frac{1}{2}(\hat{j}-j^*)}) \\
&\leq (S_{i_R - i_L, \frac{\alpha}{4}} + z_{\frac{\alpha}{4}} + \sqrt{3})2^{\frac{\tilde{K}_{\frac{\alpha}{4}}}{2}} c_e \cdot 4\rho_m(\varepsilon; f)\mathbb{E}(2^{\frac{1}{2}(\hat{j}-j^*)}).
\end{aligned} \tag{C.99}$$

Similarly to the way we bound variance in Theorem 3.3, we have

$$\begin{aligned}&\mathbb{E}(2^{\frac{1}{2}(\hat{j}-j^*)})\\&\le 2\mathbb{E}(\mathbb{1}\{\hat{j}\le j^*+2\})+\mathbb{E}(2^{\frac{1}{2}(\hat{j}-j^*)}\mathbb{1}\{\hat{j}\ge j^*+3\})\\&\le 2+2\sqrt{2}\Phi(-1.85)\frac{1}{1-2\Phi(-1.85)}\\&< 2.16.\end{aligned} \tag{C.100}$$

According to the definition of $S_{i_R-i_L,\frac{\alpha}{4}}$, $S_{i_R-i_L,\frac{\alpha}{4}}$ is decided by the following

$$(1-\Phi(-S_{i_R-i_L,\frac{\alpha}{4}}))^{i_R-i_L}=1-\frac{\alpha}{4}. \tag{C.101}$$

Therefore,

$$S_{i_R-i_L,\frac{\alpha}{4}}=-\Phi^{-1}(1-(1-\frac{\alpha}{4})^{\frac{1}{i_R-i_L}}). \tag{C.102}$$

Furthermore, we have

$$i_R-i_L=9\times 2\times 2^{\tilde{K}_{\frac{\alpha}{4}}}\times 2^{K_{\frac{\alpha}{4}}}, \tag{C.103}$$

so we know that $(S_{i_R-i_L,\frac{\alpha}{4}}+z_{\frac{\alpha}{4}}+\sqrt{3})2^{\frac{\tilde{K}_{\frac{\alpha}{4}}}{2}}c_e$ only depend on α. Therefore,

$$\mathbb{E}(|f_{hi}-f_{lo}|)\le c_{m,\alpha}\rho_m(\varepsilon;f). \tag{C.104}$$

Since for $0<\alpha<0.3$, using Theorem 2.1 and Proposition 2.2, we have

$$\rho_m(\varepsilon;f)\le 3\rho_m(\varepsilon/3;f)\le\frac{3}{b_\alpha}L_{m,\alpha}(\varepsilon;f),$$

which gives our statement.

□

C.11. Analysis of Lower Bounds of the Benchmarks in Regression Setting. To establish the optimality of the procedures, we need to analyze the lower bounds of the benchmarks. Compared with the white noise model, we will incur an additional discretization error.

This discretization error is caused by the fact that a set of convex functions can have the same values as f on the grid points (i.e., $x_0.x_1,\dots,x_n$). This fact implies that when we only look at the observations, this set of functions are equivalent. We denote this set of functions by $\mathcal{G}_n(f)$:

$$\mathcal{G}_n(f)=\{g\in\mathcal{F}:g(x_i)=f(x_i),\ \text{for all } 0\le i\le n\}. \tag{C.105}$$

However, functions in $\mathcal{G}_n(f)$ can have difference minimizers and minimums, giving rise to discretization errors for $Z(f)$ and $M(f)$ defined as

$$\mathfrak{D}_z(n, f) = \max\{Z(g) : g \in \mathcal{G}_n(f)\} - \min\{Z(g) : g \in \mathcal{G}_n(f)\}, \tag{C.106a}$$

$$\mathfrak{D}_m(n, f) = \max\{M(g) : g \in \mathcal{G}_n(f)\} - \min\{M(g) : g \in \mathcal{G}_n(f)\}. \tag{C.106b}$$

It is easy to see that $0 \le \mathfrak{D}_z(n, f) < \frac{2}{n}$ and any value in $[0, \frac{2}{n})$ can be taken by $\mathfrak{D}_z(n, f)$ for some $f \in \mathcal{F}$.

The lower bounds for the benchmarks are given as follows.

PROPOSITION C.4. *Let $\tilde{R}_{z,n}(\sigma; f)$, $\tilde{R}_{m,n}(\sigma; f)$, $\tilde{L}_{z,\alpha,n}(\sigma; f)$, $\tilde{L}_{m,\alpha,n}(\sigma; f)$ be defined in Equation (4.2). Let $\mathcal{G}_n(f)$ be defined in Equation (C.105). Let $\mathfrak{D}_z(n, f)$ and $\mathfrak{D}_m(n, f)$ be defined in Equation (C.106). Suppose $0 < \alpha < 0.3$. Then there exist constants $\tilde{C}_z, \tilde{C}_m, \tilde{C}_{z,\alpha}, \tilde{C}_{m,\alpha} > 0$ such that for all $f \in \mathcal{F}$,*

(C.107)

$$\tilde{R}_{z,n}(\sigma; f) \ge \tilde{C}_z \sup_{g \in \mathcal{G}_n(f)} \rho_z(\frac{\sigma}{\sqrt{n}}; g) \left(1 \wedge \sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}}; g)}\right) \vee \frac{1}{4}\mathfrak{D}_z(n, f),$$

$$\tilde{R}_{m,n}(\sigma; f) \ge \tilde{C}_m \sup_{g \in \mathcal{G}_n(f)} \rho_m(\frac{\sigma}{\sqrt{n}}; g) \left(1 \wedge \sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}}; g)}\right) \vee \frac{1}{4}\mathfrak{D}_m(n, f),$$

$$\tilde{L}_{z,\alpha,n}(\sigma; f) \ge \tilde{C}_{z,\alpha} \sup_{g \in \mathcal{G}_n(f)} \rho_z(\frac{\sigma}{\sqrt{n}}; g) \left(1 \wedge \sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}}; g)}\right) \vee \frac{(1-2\alpha)}{2}\mathfrak{D}_z(n, f),$$

$$\tilde{L}_{m,\alpha,n}(\sigma; f) \ge \tilde{C}_{m,\alpha} \sup_{g \in \mathcal{G}_n(f)} \rho_m(\frac{\sigma}{\sqrt{n}}; g) \left(1 \wedge \sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}}; g)}\right) \vee \frac{(1-2\alpha)}{2}\mathfrak{D}_m(n, f).$$

Compared with the lower bounds in the white noise model, the lower bounds in the regression model include additional discretization errors, which do not vanish with the noise level $\sigma \to 0$ for fixed n and f.

PROOF OF PROPOSITION C.4. Similar to white noise model. The probability density under truth f is:

$$p(y_0, \cdots y_n | f) = \Pi_{i=0}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - f(x_i))^2}{2\sigma^2}\right).$$

Hence the likelihood ratio is

$$\frac{p(y_0, \cdots y_n | f)}{p(y_0, \cdots y_n | g)} = \exp\left(\frac{\sum_{i=0}^n (f(x_i) - g(x_i))(2y_i - f(x_i) - g(x_i))}{2\sigma^2}\right).$$

Let $\theta_1 = 1$ denote the truth being f, and $\theta_2 = -1$ denote the truth being g. Suppose $\hat{\theta}$ is an estimator of θ. Then we know that $\frac{\sum_{i=0}^n (f(x_i) - g(x_i))(y_i - \frac{1}{2}f(x_i) - \frac{1}{2}g(x_i))}{\sigma^2}$ is a sufficient statistic for θ. We further standardize this statistic by $l_n(f, g) =$

$\sqrt{\sum_{i=0}^{n}\frac{1}{n}(f(x_i)-g(x_i))^2}$ and σ, which results in the following sufficient statistic for θ:

$$\check{W}=\frac{\sum_{i=0}^{n}(f(x_i)-g(x_i))(y_i-\frac{1}{2}f(x_i)-\frac{1}{2}g(x_i))}{l_n(f,g)\sqrt{n}\sigma}\sim N(\theta\frac{l_n(f,g)}{2\frac{\sigma}{\sqrt{n}}},1).$$

Letting $\hat{\theta}=\frac{2\hat{Z}-(Z(f)+Z(g))}{Z(f)-Z(g)}$ gives

$$\mathbb{E}_f(|\hat{Z}-Z(f)|)=|Z(f)-Z(g)|\mathbb{E}_{\theta=1}(\frac{1}{2}|\hat{\theta}-\theta|),$$
$$\mathbb{E}_g(|\hat{Z}-Z(g)|)=|Z(f)-Z(g)|\mathbb{E}_{\theta=-1}(\frac{1}{2}|\hat{\theta}-\theta|).$$

Therefore, similar arguments as in the white noise model give

(C.108)
$$\tilde{R}_{z,n}(\sigma;f)\geq\underbrace{\sup\{|Z(g)-Z(f)|:g\in\mathcal{F},l_n(f,g)\leq\sigma/\sqrt{n}\}}_{\mathfrak{w}_z(\sigma/\sqrt{n};f)}\Phi(-0.5).$$

For minimum, similar procedure shows that

(C.109)
$$\tilde{R}_{m,n}(\sigma;f)\geq\underbrace{\sup\{|M(g)-M(f)|:g\in\mathcal{F},l_n(f,g)\leq\sigma/\sqrt{n}\}}_{\mathfrak{w}_m(\sigma/\sqrt{n};f)}\Phi(-0.5).$$

For confidence interval with $0<\alpha<0.3$, using similar arguments as in the white noise model, we have that for $CI\in\mathcal{I}_{z,\alpha,n}(\{f,g\})$,

$$\mathbb{E}_f L(CI)\geq|Z(f)-Z(g)|(1-2\alpha-TV(P_{f,n},P_{g,n}))$$
$$\geq|Z(f)-Z(g)|(1-2\alpha-\sqrt{\chi^2(P_{f,n},P_{g,n})}),$$

where $P_{f,n}$ is the distribution of the regression model with $n+1$ observations corresponding to f.

Further, elementary calculation of chi-square divergence gives

$$\chi^2(P_{f,n},P_{g,n})=$$
$$\int\exp(\frac{\sum_{i=0}^{n}(f(x_i)-g(x_i))(2y_i-f(x_i)-g(x_i))}{\sigma^2})p(y_0,\cdots y_n|g)\mathrm{d}y_0\mathrm{d}y_1\cdots\mathrm{d}y_n-1$$
$$=\exp(\frac{l_n(f,g)^2}{\sigma^2/n})-1.$$

Picking $g\in\mathcal{F}$ such that $l_n(f,g)\leq\frac{1}{3}\frac{\sigma}{\sqrt{n}}$ gives $\mathbb{E}_f L(CI)\geq(0.6-2\alpha)|Z(f)-Z(g)|$. Therefore,

(C.110)
$$\tilde{L}_{z,\alpha,n}(\sigma;f)\geq(0.6-2\alpha)\underbrace{\sup\{|Z(g)-Z(f)|:g\in\mathcal{F},l_n(f,g)\leq\frac{1}{3}\sigma/\sqrt{n}\}}_{\mathfrak{w}_z(\frac{1}{3}\sigma/\sqrt{n};f)}.$$

Similarly, we have
(C.111)

$$\tilde{L}_{m,\alpha,n}(\sigma; f) \geq (0.6 - 2\alpha) \underbrace{\sup\{|M(g) - M(f)| : g \in \mathcal{F}, l_n(f, g) \leq \frac{1}{3}\sigma/\sqrt{n}\}}_{\mathfrak{w}_m(\frac{1}{3}\sigma/\sqrt{n}; f)}.$$

Therefore, it remains to find $\mathfrak{w}_z(\sigma/\sqrt{n}; f)$ and $\mathfrak{w}_m(\sigma/\sqrt{n}; f)$, which are analogies of continuity moduli in white noise model.

We have the following lemma that constructs convex functions g such that $l_n(f, g) \leq \sigma/\sqrt{n}$. We will use these functions to calculate lower bounds of $\mathfrak{w}_z(\sigma/\sqrt{n}; f)$ and $\mathfrak{w}_m(\sigma/\sqrt{n}; f)$. The proof of this lemma is deffered to Section D (page 131).

LEMMA C.24. *Suppose $h \in \mathcal{G}_n(f)$, where $\mathcal{G}_n(f)$ is defined in Equation (C.105). If $\rho_z(\frac{\sigma}{\sqrt{6n}}; h) \geq 1/2n$, let $g_{n,\sigma,h}(t) = \max\{h(t), M(h) + \rho_m(\frac{\sigma}{\sqrt{6n}}; h)\}$. Then we have*

$$l_n(f, g_{n,\sigma,h}) \leq \sigma^2/n.$$

If $\rho_z(\frac{\sigma}{\sqrt{6n}}; h) < 1/2n$, let

$$g_{n,\sigma,h}(t) = \max\{h(t), M(h) + \rho_m(\frac{\sigma}{\sqrt{6n}}; h)\sqrt{2n\rho_z(\frac{\sigma}{\sqrt{6n}}; h)}\}.$$

Then we have

$$l_n(f, g_{n,\sigma,h}) \leq \sigma^2/n.$$

Let $t_l(h) = \inf\{t \in [0,1] : g_{n,\sigma,h}(t) > h(t)\}$, and $t_r(h) = \sup\{t \in [0,1] : g_{n,\sigma,h}(t) > h(t)\}$, where we will omit h when there are no ambiguities. Clearly,

$$t_r - t_l \geq \rho_z(\frac{\sigma}{\sqrt{6n}}; h)(1 \wedge \sqrt{2n\rho_z(\frac{\sigma}{\sqrt{6n}}; h)}),$$
$$M(g_{n,\sigma,h}) \geq \min\{\rho_m(\frac{\sigma}{\sqrt{6n}}; h), \rho_m(\frac{\sigma}{\sqrt{6n}}; h)\sqrt{2n\rho_z(\frac{\sigma}{\sqrt{6n}}; h)}\} + M(h).$$

Similar arguments as in Proposition 2.2 give that for any $\delta > 0$, there exist $g_{n,\sigma,h,\delta,l}$, $g_{n,\sigma,h,\delta,r} \in \mathcal{F}$, such that

$$l_n(f, g_{n,\sigma,h,\delta,l}) \leq \sigma^2/n, \qquad l_n(f, g_{n,\sigma,h,\delta,r}) \leq \sigma^2/n,$$
$$Z(g_{n,\sigma,h,\delta,l}) \leq t_l + \delta, \qquad Z(g_{n,\sigma,h,\delta,r}) \geq t_r - \delta, \text{ and}$$
$$M(g_{n,\sigma,h,\delta,r}) = M(g_{n,\sigma,h,\delta,l}) \geq \rho_m(\frac{\sigma}{\sqrt{6n}}; h) \min\{1, \sqrt{2n\rho_z(\frac{\sigma}{\sqrt{6n}}; h)}\} + M(h) - \delta.$$

Then we have the following lower bounds for $\mathfrak{w}_z(\sigma/\sqrt{n};f)$, $\mathfrak{w}_m(\sigma/\sqrt{n};f)$, $\mathfrak{w}_z(\frac{1}{3}\sigma/\sqrt{n};f)$, and $\mathfrak{w}_m(\frac{1}{3}\sigma/\sqrt{n};f)$.

$$\begin{aligned}
&\mathfrak{w}_z(\sigma/\sqrt{n};f)\\
&\quad = \sup\{|Z(g)-Z(f)| : l_n(f,g)\le \sigma/\sqrt{n}, g\in\mathcal{F}\}\\
&\quad \ge \sup_{h\in\mathcal{G}_n(f)} \frac{1}{2}\lim_{\delta\to 0^+}(Z(g_{n,\sigma,h,\delta,r})-Z(g_{n,\sigma,h,\delta,l})) = \sup_{h\in\mathcal{G}_n(f)}\frac{1}{2}(t_r-t_l)\\
&\quad \ge \frac{1}{2}\sup_{h\in\mathcal{G}_n(f)} \rho_z(\frac{\sigma}{\sqrt{6n}};h)(1\wedge\sqrt{2n\rho_z(\frac{\sigma}{\sqrt{6n}};h)})\\
&\quad \ge \frac{1}{2}\sup_{h\in\mathcal{G}_n(f)} 54^{-\frac{1}{4}}\rho_z(\frac{\sigma}{\sqrt{n}};h)\left(1\wedge\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};h)}\right),\\
&\mathfrak{w}_z(\sigma/3\sqrt{n};f)\\
&\quad = \sup\{|Z(g)-Z(f)| : l_n(f,g)\le \sigma/3\sqrt{n}, g\in\mathcal{F}\}\\
&\quad \ge \sup_{h\in\mathcal{G}_n(f)} \frac{1}{2}\lim_{\delta\to 0^+}(Z(g_{n,\sigma/3,h,\delta,r})-Z(g_{n,\sigma/3,h,\delta,l})) = \sup_{h\in\mathcal{G}_n(f)}\frac{1}{2}(t_r-t_l)\\
&\quad \ge \frac{1}{2}\sup_{h\in\mathcal{G}_n(f)} \rho_z(\frac{\sigma}{3\sqrt{6n}};h)(1\wedge\sqrt{2n\rho_z(\frac{\sigma}{3\sqrt{6n}};h)})\\
&\quad \ge \frac{1}{2}\sup_{h\in\mathcal{G}_n(f)} \frac{1}{9}6^{-\frac{1}{4}}\rho_z(\frac{\sigma}{\sqrt{n}};h)\left(1\wedge\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};h)}\right),\\
&\mathfrak{w}_m(\sigma/\sqrt{n};f)\\
&\quad = \sup\{|M(g)-M(f)| : l_n(f,g)\le \sigma/\sqrt{n}, g\in\mathcal{F}\}\\
&\quad \ge \frac{1}{2}\min\{\rho_m(\frac{\sigma}{\sqrt{6n}};h), \rho_m(\frac{\sigma}{\sqrt{6n}};h)\sqrt{2n\rho_z(\frac{\sigma}{\sqrt{6n}};h)}\}\\
&\quad \ge \frac{1}{2}54^{-\frac{1}{4}}\min\{\rho_m(\frac{\sigma}{\sqrt{n}};h), \rho_m(\frac{\sigma}{\sqrt{n}};h)\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};h)}\}, \text{ and}\\
&\mathfrak{w}_m(\sigma/3\sqrt{n};f)\\
&\quad = \sup\{|M(g)-M(f)| : l_n(f,g)\le \sigma/3\sqrt{n}, g\in\mathcal{F}\}\\
&\quad \ge \frac{1}{2}\min\{\rho_m(\frac{\sigma}{3\sqrt{6n}};h), \rho_m(\frac{\sigma}{3\sqrt{6n}};h)\sqrt{2n\rho_z(\frac{\sigma}{3\sqrt{6n}};h)}\}\\
&\quad \ge \frac{1}{18}6^{-\frac{1}{4}}\min\{\rho_m(\frac{\sigma}{\sqrt{n}};h), \rho_m(\frac{\sigma}{\sqrt{n}};h)\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};h)}\}.
\end{aligned}$$

Returning to Inequalities (C.108),(C.109), (C.110), (C.111), we have

$$\tilde{R}_{z,n}(\sigma; f) \geq \frac{1}{2}\Phi(-0.5)54^{-\frac{1}{4}} \sup_{h\in\mathcal{G}_n(f)} \rho_z(\frac{\sigma}{\sqrt{n}}; h)\left(1 \wedge \sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}}; h)}\right),$$

$$\tilde{R}_{m,n}(\sigma; f) \geq \frac{1}{2}\Phi(-0.5)54^{-\frac{1}{4}} \sup_{h\in\mathcal{G}_n(f)} \rho_z(\frac{\sigma}{\sqrt{n}}; h)\left(1 \wedge \sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}}; h)}\right),$$

$$\tilde{L}_{z,\alpha,n}(\sigma; f) \geq \frac{1}{2}(0.6 - 2\alpha)\frac{1}{9}6^{-\frac{1}{4}} \sup_{h\in\mathcal{G}_n(f)} \rho_z(\frac{\sigma}{\sqrt{n}}; h)\left(1 \wedge \sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}}; h)}\right),$$

$$\tilde{L}_{m,\alpha,n}(\sigma; f) \geq (0.6 - 2\alpha)\frac{1}{18}6^{-\frac{1}{4}} \sup_{h\in\mathcal{G}_n(f)} \rho_m(\frac{\sigma}{\sqrt{n}}; h)\left(1 \wedge \sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}}; h)}\right).$$

Now we turn to the discretization error. For any $g \in \mathcal{G}_n(f)$, we have $\frac{dP_{f,n}}{dP_{g,n}}(y_0, y_1, \cdots, y_n) = 1$ for all $(y_0, y_1, \cdots, y_n) \in \mathcal{R}^n$. Therefore, for any estimator $\hat{Z}$, we have

$$\begin{aligned}\mathbb{E}_g|\hat{Z} - Z(g)| + \mathbb{E}_f|\hat{Z} - Z(f)| &= \mathbb{E}_f\left(|\hat{Z} - Z(g)| + |\hat{Z} - Z(f)|\right)\\ &\geq \mathbb{E}_f|Z(f) - Z(g)| = |Z(f) - Z(g)|.\end{aligned}$$

Hence we have

$$\tilde{R}_{z,n}(\sigma; f) \geq \frac{1}{2}\sup_{g\in\mathcal{G}_n(f)} |Z(f) - Z(g)| \geq \frac{1}{4}\mathfrak{D}_z(n, f).$$

Similarly, we have $\tilde{R}_{m,n}(\sigma; f) \geq \frac{1}{4}\mathfrak{D}_m(n, f)$. For the confidence interval, we have that for any $g \in \mathcal{G}_n(f)$, and for any $CI \in I_{z,\alpha,n}(\{f, g\})$,

$$\begin{aligned}\mathbb{E}_f L(CI) &\geq \left(1 - P_f(Z(f) \notin CI) - P_f(Z(g) \notin CI)\right)_+ |Z(f) - Z(g)|\\ &\geq (1 - 2\alpha)|Z(f) - Z(g)|.\end{aligned}$$

Hence we have

$$\tilde{L}_{z,\alpha,n}(\sigma; f) \geq (1 - 2\alpha)\cdot\frac{1}{2}\mathfrak{D}_z(n, f).$$

Similarly, we have $\tilde{L}_{m,\alpha,n}(\sigma; f) \geq (1 - 2\alpha)\cdot\frac{1}{2}\mathfrak{D}_m(n, f)$.

□

C.12. Proof of Theorem 4.1. With the lower bound in Proposition C.4, we only need to prove the following two propositions to prove the theorem.

PROPOSITION C.5. *For $\hat{Z}$ defined in (4.5), we have*

$$\mathbb{E}(|\hat{Z} - Z(f)|) \leq \check{C}_1\rho_z(\frac{\sigma}{\sqrt{n}}; f) + \frac{2}{n}. \tag{C.113}$$

Proposition C.6. *For $\hat{Z}$ defined in (4.5), if* $\sup_{h\in\mathcal{G}_n(f)}\{\rho_z(\frac{\sigma}{\sqrt{n}};h)\} < \frac{1}{2n}$, *we have*

$$\mathbb{E}(|\hat{Z}-Z(f)|) \le \check{C}_2 \sup_{h\in\mathcal{G}_n(f)} \rho_z(\frac{\sigma}{\sqrt{n}};h)\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};h)} + \mathfrak{D}_z(n,f). \tag{C.114}$$

The statement of the theorem follows from letting $C_1 = \frac{\sqrt{2}\check{C}_1+4+\check{C}_2}{\tilde{C}_z} + 4$, where $\tilde{C}_z$ is defined in (C.107).

Now we proceed with proving the Propositions.

Proof of Proposition C.5.

$$\begin{aligned}\mathbb{E}(|\hat{Z}-Z(f)|) &= \mathbb{E}(\mathbb{1}\{\hat{\mathfrak{j}} < \tilde{\mathfrak{j}}\}\big|\hat{Z}-Z(f)\big|) + \mathbb{E}(\mathbb{1}\{\hat{\mathfrak{j}} \ge \tilde{\mathfrak{j}}\}\big|\hat{Z}-Z(f)\big|)\\ &\le \mathbb{E}(\mathbb{1}\{\hat{\mathfrak{j}} < \tilde{\mathfrak{j}}\}1.5m_{\hat{\mathfrak{j}}}) + \mathbb{E}(\mathbb{1}\{\hat{\mathfrak{j}} \ge \tilde{\mathfrak{j}}\}\big|\hat{Z}-Z(f)\big|)\end{aligned}$$

To bound the two terms, we give two lemmas below, the proofs of the lemmas are in Section D (page 132, 133).

Lemma C.25.

$$\mathbb{E}(\mathbb{1}\{\hat{\mathfrak{j}} < \tilde{\mathfrak{j}}\}1.5m_{\hat{\mathfrak{j}}}) \le c_{z1}\rho_z(\frac{\sigma}{\sqrt{n}};f) + \frac{1.5}{n}\mathbb{1}\{J \le \mathfrak{j}^*-3\}. \tag{C.115}$$

Lemma C.26.

$$\mathbb{E}(\mathbb{1}\{\hat{\mathfrak{j}} \ge \tilde{\mathfrak{j}}\}\big|\hat{Z}-Z(f)\big|) \le c_{z2}\rho_z(\frac{\sigma}{\sqrt{n}};f). \tag{C.116}$$

Therefore,

(C.117)

$$\mathbb{E}(|\hat{Z}-Z(f)|) \le (c_{z1}+c_{z2})\rho_z(\frac{\sigma}{\sqrt{n}};f) + \frac{1.5}{n}\mathbb{1}\{J \le \mathfrak{j}^*-3\} \le \check{C}_1\rho_z(\frac{\sigma}{\sqrt{n}};f) + \frac{1.5}{n}.$$

□

Proof of Proposition C.6. since $\sup_{h\in\mathcal{G}_n(f)}\{\rho_z(\frac{\sigma}{\sqrt{n}};h)\} < \frac{1}{2n}$, we know that $|\{i : f(x_i) = \min\{f(x_k) : 0 \le k \le n\}\}| = 1$. Suppose $i_{min} \in \{i : f(x_i) = \min\{f(x_k) : 0 \le k \le n\}\}$. Let $\tilde{h}$ be the piece wise linear function such that $\tilde{h}(x_i) = f(x_i)$ for all $0 \le i \le n$, and $\tilde{h}$ is linear on all the sub-intervals $[k/n, k+1/n]$, for $0 \le k \le n-1$. It is clear that $Z(\tilde{h}) = x_{i_{min}}$.

Then we have

$$\begin{aligned}\mathbb{E}(|\hat{Z}-Z(f)|) &\le \mathbb{E}(|\hat{Z}-Z(\tilde{h})|) + |Z(\tilde{h})-Z(f)| \le \mathbb{E}(|\hat{Z}-Z(\tilde{h})|) + \mathfrak{D}_z(n,f)\\ &= \mathbb{E}(\mathbb{1}\{\check{j} < \infty\}|\hat{Z}-Z(\tilde{h})|) + \mathbb{E}(\mathbb{1}\{\check{j} = \infty\}|\hat{Z}-Z(\tilde{h})|) + \mathfrak{D}_z(n,f).\end{aligned}$$

Splitting the first and second terms by the $\{\hat{\mathfrak{j}} < \tilde{\mathfrak{j}}\}$ and $\{\hat{\mathfrak{j}} \geq \tilde{\mathfrak{j}}\}$ gives

$$\begin{aligned}&\mathbb{E}(\mathbb{1}\{\check{j} < \infty\}|\hat{Z} - Z(\tilde{h})|)\\ &= \mathbb{E}(\mathbb{1}\{\check{j} < \infty\}\mathbb{1}\{\check{j} < \tilde{\mathfrak{j}}\}|\hat{Z} - Z(\tilde{h})|) + \mathbb{E}(\mathbb{1}\{\check{j} < \infty\}\mathbb{1}\{\check{j} \geq \tilde{\mathfrak{j}}\}|\hat{Z} - Z(\tilde{h})|),\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}(\mathbb{1}\{\check{j} = \infty\}|\hat{Z} - Z(\tilde{h})|)\\ &= \mathbb{E}(\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{\hat{\mathfrak{j}} < \tilde{\mathfrak{j}}\}|\hat{Z} - Z(\tilde{h})|) + \mathbb{E}(\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{\hat{\mathfrak{j}} \geq \tilde{\mathfrak{j}}\}|\hat{Z} - Z(\tilde{h})|)\\ &\leq \frac{5}{n}\Phi(-\frac{\rho_m(\frac{\sigma}{\sqrt{n}};\tilde{h})}{\sqrt{3}\sigma n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}) + \mathbb{E}(\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{\hat{\mathfrak{j}} \geq \tilde{\mathfrak{j}}\}|\hat{Z} - Z(\tilde{h})|)\\ &\leq \check{c}_{z0}\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})} + \mathbb{E}(\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{\hat{\mathfrak{j}} \geq \tilde{\mathfrak{j}}\}|\hat{Z} - Z(\tilde{h})|).\end{aligned}$$

Therefore,

(C.118)
$$\begin{aligned}&\mathbb{E}(|\hat{Z} - Z(f)|) \leq \check{c}_{z0}\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})} + \mathbb{E}(\mathbb{1}\{\check{j} < \infty\}\mathbb{1}\{\check{j} < \tilde{\mathfrak{j}}\}|\hat{Z} - Z(\tilde{h})|)\\ &+ \mathbb{E}(\mathbb{1}\{\check{j} < \infty\}\mathbb{1}\{\check{j} \geq \tilde{\mathfrak{j}}\}|\hat{Z} - Z(\tilde{h})|) + \mathbb{E}(\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{\hat{\mathfrak{j}} \geq \tilde{\mathfrak{j}}\}|\hat{Z} - Z(\tilde{h})|) + \mathfrak{D}_z(n, f)\\ &= \mathbb{E}(\mathbb{1}\{\check{j} < \infty\}\mathbb{1}\{\check{j} < \tilde{\mathfrak{j}}\}|\hat{Z} - Z(\tilde{h})|) + \mathbb{E}(\mathbb{1}\{\hat{\mathfrak{j}} \geq \tilde{\mathfrak{j}}\}|\hat{Z} - Z(\tilde{h})|) + \mathfrak{D}_z(n, f)\\ &+ \check{c}_{z0}\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}.\end{aligned}$$

Finally, with the help of the following lemmas (proved in Section D, page 134, 134), we prove the proposition.

Lemma C.27.

$$\mathbb{E}(\mathbb{1}\{\check{j} < \infty\}\mathbb{1}\{\check{j} < \tilde{\mathfrak{j}}\}|\hat{Z} - Z(\tilde{h})|) \leq \check{c}_{z1}\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})} \tag{C.119}$$

Lemma C.28.

$$\mathbb{E}(\mathbb{1}\{\hat{\mathfrak{j}} \geq \tilde{\mathfrak{j}}\}|\hat{Z} - Z(\tilde{h})|) \leq \check{c}_{z2}\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})} \tag{C.120}$$

□

C.13. Proof of Theorem 4.2. With Proposition C.4, we only need to prove following three lemmas to prove the theorem.

LEMMA C.29 (length of the confidence interval for minimizer).

$$\mathbb{E}_f L(\mathtt{CI}_{z,\alpha}(Y)) < \tilde{C}_{2,\alpha}(C_0\rho_z(\frac{\sigma}{\sqrt{n}}; f) + \frac{1}{n}).$$

LEMMA C.30. *When* $\sup_{h\in\mathcal{G}_n(f)}\{\rho_z(\frac{\sigma}{\sqrt{n}}; h)\} < \frac{1}{2n}$ *, we have*

$$\mathbb{E}_f L(\mathtt{CI}_{z,\alpha}(Y)) < \check{C}_{2,\alpha} \sup_{h\in\mathcal{G}_n(f)} \rho_z(\frac{\sigma}{\sqrt{n}}; h)\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}}; h)} + 2\mathfrak{D}_z(n, f)$$

LEMMA C.31 (coverage of the confidence interval for minimizer).

$$P(Z(f) \in \mathtt{CI}_{z,\alpha}(Y)) \geq 1 - \alpha.$$

Let $C_{2,\alpha} = \max\{\frac{\tilde{C}_{2,\alpha}(C_0+2)\sqrt{2}}{\tilde{C}_{z,\alpha}}, \frac{\check{C}_{2,\alpha}}{\tilde{C}_{z,\alpha}} + \frac{4}{1-2\alpha}\}$, then we have the statement of the theorem.

PROOF OF LEMMA C.29.

$$\begin{aligned}
&\mathbb{E}_f L(\mathtt{CI}_{z,\alpha}(Y)) \\
&\leq 24 \times 2^{K\frac{\alpha}{2}} \cdot \mathbb{E}(\frac{2^{J-\hat{\mathtt{j}}}}{n}) \\
&= 24 \times 2^{K\frac{\alpha}{2}} \frac{2^J}{n}\mathbb{E}(\sum_{j=1}^{\mathtt{j}^*-1} 2^{-j}\mathbb{1}\{\hat{\mathtt{j}} = j\} + \sum_{j=\mathtt{j}^*}^{\infty} 2^{-j}\mathbb{1}\{\hat{\mathtt{j}} = j\}) \\
&\leq 24 \times 2^{K\frac{\alpha}{2}} \frac{2^J}{n}\Big(\sum_{j=1}^{\mathtt{j}^*-1} 2^{-j}\mathbb{E}(\mathbb{1}\{\hat{\mathtt{j}} = j, \tilde{\mathtt{j}} > j\} + \mathbb{1}\{\hat{\mathtt{j}} = j, \tilde{\mathtt{j}} \leq j\}) + 2^{-\mathtt{j}^*}\Big)
\end{aligned} \tag{C.121}$$

To bound the first two terms, we will introduce two lemmas. The proofs of the lemmas are given at Section D (page 134 and 135).

LEMMA C.32.

$$\sum_{j=1}^{\mathtt{j}^*-1} \mathbb{E}(2^{-j}\mathbb{1}\{\hat{\mathtt{j}} = j, \tilde{\mathtt{j}} > j\}) \leq 2^{-\mathtt{j}^*}c_{z3} + 2^{-J}\mathbb{1}\{J \leq j^* - 1\}. \tag{C.122}$$

LEMMA C.33.

$$\sum_{j=1}^{\mathtt{j}^*-1} \mathbb{E}(2^{-j}\mathbb{1}\{\hat{\mathtt{j}} = j, \tilde{\mathtt{j}} \leq j\}) \leq 2^{-\mathtt{j}^*}c_{z4}. \tag{C.123}$$

With these lemmas, we have

$$\begin{aligned}
&\mathbb{E}_f L(\mathtt{CI}_{z,\alpha}(Y)) \\
&\leq 24 \times 2^{K_{\frac{\alpha}{2}}} \left(\frac{2^{J-\mathtt{j}^*}}{n}(c_{z4} + c_{z3} + 1) + \frac{1}{n}\mathbb{1}\{J \leq \mathtt{j}^* - 1\} \right) \\
&\leq \tilde{C}_{2,\alpha}(C_0 \rho_z(\frac{\sigma}{\sqrt{n}}; f) + \frac{1}{n}\mathbb{1}\{J \leq \mathtt{j}^* - 1\}),
\end{aligned} \tag{C.124}$$

where $\tilde{C}_{2,\alpha} = 24 \times 2^{K_{\alpha/2}}$, $C_0 = \frac{c_{z3}+c_{z4}+1}{4}$. □

Proof of Lemma C.30. To prove the lemma, we introduce the following lemmas while postponing their proofs.

Lemma C.34. *When* $\sup_{h \in \mathcal{G}_n(f)}\{\rho_z(\frac{\sigma}{\sqrt{n}}; h)\} < \frac{1}{2n}$,

$$\mathbb{E}(\mathbb{1}\{\check{j} < \infty\} L(\mathtt{CI}_{z,\alpha}(Y))) \leq \check{c}_{1,\alpha} \sup_{h \in \mathcal{G}_n(f)} \rho_z(\frac{\sigma}{\sqrt{n}}; h)\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}}; h)}. \tag{C.125}$$

Lemma C.35. *When* $\sup_{h \in \mathcal{G}_n(f)}\{\rho_z(\frac{\sigma}{\sqrt{n}}; h)\} < \frac{1}{2n}$,

$$\mathbb{E}(\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{t_{hi} - t_{lo} \geq \tfrac{3}{n}\} L(\mathtt{CI}_{z,\alpha}(Y))) \leq \check{c}_{2,\alpha} \sup_{h \in \mathcal{G}_n(f)} \rho_z(\tfrac{\sigma}{\sqrt{n}}; h)\sqrt{n\rho_z(\tfrac{\sigma}{\sqrt{n}}; h)}. \tag{C.126}$$

Lemma C.36. *When* $\sup_{h \in \mathcal{G}_n(f)}\{\rho_z(\frac{\sigma}{\sqrt{n}}; h)\} < \frac{1}{2n}$,

$$\begin{aligned}
&\mathbb{E}(\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{t_{hi} - t_{lo} < \frac{3}{n}\} L(\mathtt{CI}_{z,\alpha}(Y))) \\
&\leq \check{c}_{3,\alpha} \sup_{h \in \mathcal{G}_n(f)} \rho_z(\frac{\sigma}{\sqrt{n}}; h)\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}}; h)} + 2\mathfrak{D}_z(n, f).
\end{aligned} \tag{C.127}$$

With these lemmas, we have the statement of Lemma C.30.

The proofs of the lemmas are in Section D (page 135, 137 and 138). Here we point out the common thing that will be used in all these proofs.

When $\sup_{h \in \mathcal{G}_n(f)}\{\rho_z(\frac{\sigma}{\sqrt{n}}; h)\} < \frac{1}{2n}$, we know that $|\{k : f(x_k) = \min\{f(x_i) : 0 \leq i \leq n\}\}| = 1$, we denote this unique element to be i_m.

Let $\tilde{h}$ be the piece wise linear function such that $\tilde{h}(x_i) = f(x_i)$ for all $0 \leq i \leq n$, and $\tilde{h}$ is linear on all the sub-intervals $[k/n, k + 1/n]$, for $0 \leq k \leq n-1$. Then we know that $\rho_z(\frac{\sigma}{\sqrt{n}}; \tilde{h}) < \frac{1}{2n}$.

Suppose $Y_{e,1} = \{y_{e,i} + \sqrt{3}\sigma z_{3,i} : (L-1) \vee 0 \leq i \leq (U+1) \wedge n\}$, $Y_{e,2} = \{y_{e,i} - \sqrt{3}\sigma z_{3,i} : (L-1) \vee 0 \leq i \leq (U+1) \wedge n\}$. Then we know that $\mathtt{Y}_l$,$\mathtt{Y}_s$,$Y_{e,1}$,$Y_{e,2}$ are independent.

□

Proof of Lemma C.31. For clarity of the main idea of the proof, we postpone the proofs of the supporting lemmas to Section D.

With a bit abuse of notation, define the following events:

$$\begin{aligned} E =&\Bigg\{Z(f) \in \Big[\left(\hat{\mathtt{i}}_{\hat{\mathtt{j}}} - (6\cdot 2^{K_{\alpha/2}+1} - 2) - 1\right)\frac{2^{J-\hat{\mathtt{j}}}}{n} - \frac{1}{2n}, \\ &\qquad \left(\hat{\mathtt{i}}_{\hat{\mathtt{j}}} + (6\cdot 2^{K_{\alpha/2}+1} - 2)\right)\frac{2^{J-\hat{\mathtt{j}}}}{n} - \frac{1}{2n}\Big] \cap [0,1]\Bigg\} \\ F_1 =&\Bigg\{i_l \le \min\Big\{i : f(x_i) = \min\{f(x_k) : 0 \le k \le n\}\Big\}\Bigg\} \\ F_2 =&\Bigg\{i_r + 1 \ge \max\Big\{i : f(x_i) = \min\{f(x_k) : 0 \le k \le n\}\Big\}\Bigg\}. \end{aligned} \tag{C.128}$$

For $\mathtt{j}^{\mathtt{w}}$ defined in equation (C.5), we have the following lemma (proved in Section D, on page 146).

Lemma C.37. *For $K \ge 1$,*

$$\Phi(\hat{\mathtt{j}} \ge \mathtt{j}^{\mathtt{w}} + K + 1) \le \Phi(-2)^K. \tag{C.129}$$

Therefore, with this lemma, we have

$$P(E^c) \le P(\left|\mathtt{i}^*_{\hat{\mathtt{j}}} - \hat{\mathtt{i}}_{\hat{\mathtt{j}}}\right| > 6\cdot 2^{K_{\frac{\alpha}{2}}+1} - 2) \le P(\mathbb{1}\{\hat{\mathtt{j}} > \mathtt{j}^{\mathtt{w}} + K_{\alpha/2}\}) \le \frac{\alpha}{2}. \tag{C.130}$$

Therefore,

(C.131)

$$\begin{aligned} &P(Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)) \\ &= \mathbb{E}(\mathbb{1}\{Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{E\}) + \mathbb{E}(\mathbb{1}\{Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{E^c\}) \\ &\le \mathbb{E}(\mathbb{1}\{Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{E\}\mathbb{1}\{\check{j} < \infty\}) + \\ &\qquad \mathbb{E}(\mathbb{1}\{Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{E\}\mathbb{1}\{\check{j} = \infty\}) + \frac{\alpha}{2} \\ &= \mathbb{E}(\mathbb{1}\{Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{E\}\mathbb{1}\{\check{j} = \infty\}) + \frac{\alpha}{2} \\ &\le \mathbb{E}\Big(\mathbb{1}\{Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{E\}\mathbb{1}\{\check{j} = \infty\}\big(\mathbb{1}\{F_1 \cap F_2\} + \mathbb{1}\{F_1^c\} + \mathbb{1}\{F_2^c\}\big)\Big) + \frac{\alpha}{2} \\ &\le \mathbb{E}\Big(\mathbb{1}\{Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{E\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{F_1 \cap F_2\}\Big) \\ &\qquad + \mathbb{E}(\mathbb{1}\{E\}\mathbb{1}\{\check{j} = \infty\}(\mathbb{1}\{F_1^c\} + \mathbb{1}\{F_2^c\})) + \frac{\alpha}{2}. \end{aligned}$$

We introduce the following lemma, which is proved in Section D on page 146.

Lemma C.38.

$$\mathbb{E}(\mathbb{1}\{E\}\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{F_1^c\}) \le \alpha_1, \mathbb{E}(\mathbb{1}\{E\}\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{F_2^c\}) \le \alpha_1. \tag{C.132}$$

Therefore

$$P(Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)) \le \mathbb{E}\Big(\mathbb{1}\{Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{E\}\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{F_1 \cap F_2\}\Big) + \frac{\alpha}{2} + 2\alpha_1. \tag{C.133}$$

The only term needs analysis it the first term on the right hand side. Note that the entire probability space is the union of the following three disjoint events.

$$\{(i_l - U)(i_r - L + 1) = 0\},$$
$$\{(i_l - U)(i_r - L + 1) \neq 0\} \cap \{i_{hi} - i_{lo} \le 2, 0 < i_{lo}, i_{hi} < n\},$$
$$\{(i_l - U)(i_r - L + 1) \neq 0\} \cap \{i_{hi} - i_{lo} \ge 3 \text{ or } (i_{hi} - n)i_{lo} = 0\}.$$

Further, on the event $E \cap F_1 \cap F_2 \cap \{\check{j} = \infty\} \cap \{(i_l - U)(i_r - L + 1) \neq 0\} \cap \{i_{hi} - i_{lo} \ge 3 \text{ or } (i_{hi} - n)i_{lo} = 0\}$, $Z(f) \in \mathtt{CI}_{z,\alpha}(Y)$. The first term on the right hand side of Inequality (C.133) therefore simplifies to

$$\begin{aligned}&\mathbb{E}\Big(\mathbb{1}\{Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{E\}\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{F_1 \cap F_2\}\mathbb{1}\{(i_l - U)(i_r - L + 1) = 0\}\Big)\\ &+ \mathbb{E}\Big(\mathbb{1}\{Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{E\}\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{F_1 \cap F_2\}\mathbb{1}\{(i_l - U)(i_r - L + 1) \neq 0\}\\ &\qquad \mathbb{1}\{i_{hi} - i_{lo} \le 2, 0 < i_{lo}, i_{hi} < n\}\Big).\end{aligned} \tag{C.134}$$

We have the following lemmas, which are proved in Section D on page 147 and 147.

Lemma C.39.

$$\begin{aligned}&\mathbb{E}\Big(\mathbb{1}\{Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{E\}\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{F_1 \cap F_2\}\mathbb{1}\{(i_l - U)(i_r - L + 1) = 0\}\Big)\\ &\le 3\alpha_2\mathbb{E}\left(\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{F_1 \cap F_2\}\mathbb{1}\{(i_l - U)(i_r - L + 1) = 0\}\right).\end{aligned} \tag{C.135}$$

Lemma C.40.

$$\begin{aligned}&\mathbb{E}\Big(\mathbb{1}\{Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{E\}\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{F_1 \cap F_2\}\mathbb{1}\{(i_l - U)(i_r - L + 1) \neq 0\}\\ &\qquad \mathbb{1}\{i_{hi} - i_{lo} \le 2, 0 < i_{lo}, i_{hi} < n\}\Big)\\ &\le 6\alpha_2 P\Big(\mathbb{1}\{E\}\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{F_1 \cap F_2\}\mathbb{1}\{(i_l - U)(i_r - L + 1) \neq 0\}\\ &\qquad \mathbb{1}\{i_{hi} - i_{lo} \le 2, 0 < i_{lo}, i_{hi} < n\}\Big).\end{aligned} \tag{C.136}$$

With these two lemmas, we finally have

$$P(Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)) \leq 6\alpha_2 + \frac{\alpha}{2} + 2\alpha_1 \leq \alpha \tag{C.137}$$

□

C.14. Proof of Theorem 4.3. With the lower bound in Proposition C.4, we only need to prove the following two propositions to prove the theorem.

Proposition C.7.

$$\mathbb{E}(|\hat{M} - M(f)|) \leq \check{C}_{3,0}\rho_m(\frac{\sigma}{\sqrt{n}}; f) + \sqrt{2}\left(\min\{f(x_i) : 0 \leq i \leq n\} - M(f)\right). \tag{C.138}$$

Proposition C.8. *When* $\sup_{h\in\mathcal{G}_n(f)}\{\rho_z(\frac{\sigma}{\sqrt{n}}; h)\} < \frac{1}{2n}$, *we have*

$$\begin{aligned}\mathbb{E}(|\hat{M} - M(f)|) \leq \check{C}_3 \sup_{h\in\mathcal{G}_n(f)} \rho_m(\frac{\sigma}{\sqrt{n}}; h)\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}}; h)} \\ + \sqrt{2}\left(\min\{f(x_i) : 0 \leq i \leq n\} - M(f)\right).\end{aligned} \tag{C.139}$$

Let $C_3 = \frac{\sqrt{2}\check{C}_{3,0}+\check{C}_3}{\check{C}_m} + 4\sqrt{2}$ gives the statement of Theorem 4.3.

Proof of Proposition C.7. We have

$$\mathbb{E}((\hat{M} - M(f))^2) = \mathbb{E}((\hat{M} - M(f))^2\mathbb{1}\{\check{j} < \infty\} + (\hat{M} - M(f))^2\mathbb{1}\{\check{j} = \infty\}). \tag{C.140}$$

For the first term we have

$$\begin{aligned}&\mathbb{E}((\hat{M} - M(f))^2\mathbb{1}\{\check{j} < \infty\}) = \mathbb{E}\left(\left((\hat{\mathtt{f}} - M(f)) + \mathfrak{E}_{\check{j},\tilde{\mathtt{i}}_{\check{j}},e}\frac{1}{2^{J-\check{j}}}\right)^2\mathbb{1}\{\check{j} < \infty\}\right) \\ &= \mathbb{E}\Big(\big((\hat{\mathtt{f}} - M(f))^2 + 2(\hat{\mathtt{f}} - M(f))\mathfrak{E}_{\check{j},\tilde{\mathtt{i}}_{\check{j}},e}\frac{1}{2^{J-\check{j}}} + (\mathfrak{E}_{\check{j},\tilde{\mathtt{i}}_{\check{j}},e}\frac{1}{2^{J-\check{j}}})^2\big)\mathbb{1}\{\check{j} < \infty\}\Big) \\ &= \mathbb{E}\left((\hat{\mathtt{f}} - M(f))^2\mathbb{1}\{\check{j} < \infty\}\right) + \mathbb{E}\left((\mathfrak{E}_{\check{j},\tilde{\mathtt{i}}_{\check{j}},e}\frac{1}{2^{J-\check{j}}})^2\mathbb{1}\{\check{j} < \infty\}\right).\end{aligned} \tag{C.141}$$

We introduce following two lemmas (proved in Section D on page 149 and 149) to bound the two terms.

Lemma C.41.

$$\mathbb{E}((\mathfrak{E}_{\check{j},\tilde{\mathtt{i}}_{\check{j}},e}\frac{1}{2^{J-\check{j}}})^2\mathbb{1}\{\check{j} < \infty\}) \leq c_{m1}\rho_m(\frac{\sigma}{\sqrt{n}}; f)^2. \tag{C.142}$$

LEMMA C.42.

$$\mathbb{E}((\hat{\mathtt{f}} - M(f))^2 \mathbb{1}\{\check{j} < \infty\}) \le c_{m2}\rho_m(\frac{\sigma}{\sqrt{n}}; f)^2. \tag{C.143}$$

For the second term in Equation (C.140), let

$$\begin{aligned} \mathbf{i} &= \underset{\hat{\mathtt{i}}_J - 2 \le i \le \hat{\mathtt{i}}_J + 2}{\arg\min} f(x_{i-1}), \\ f_i &= f(x_{i-1}), \\ \delta_i &= y_{e,i-1} - f(x_{i-1}), \\ \eta &= \min\{\delta_i : \hat{\mathtt{i}}_J - 2 \le i \le \hat{\mathtt{i}}_J + 2\}, \end{aligned} \tag{C.144}$$

then we know $\mathbb{E}(\eta | \hat{\mathtt{i}}_J) \le 0$, and we have

$$\begin{aligned} &\mathbb{E}\big((\hat{M} - M(f))^2 \mathbb{1}\{\check{j} = \infty\}\big) \\ &\le \mathbb{E}\big((f_{\mathbf{i}} - M(f) + \delta_{\mathbf{i}})^2 \mathbb{1}\{\check{j} = \infty\} \mathbb{1}\{\hat{M} > M(f)\}\big) \\ &\quad + \mathbb{E}\big((f_{\mathbf{i}} - M(f) + \eta)^2 \mathbb{1}\{\check{j} = \infty\} \mathbb{1}\{\hat{M} < M(f)\}\big) \\ &\le 2\mathbb{E}\big((f_{\mathbf{i}} - M(f))^2 \mathbb{1}\{\check{j} = \infty\}\big) + 2\gamma_e^2\sigma^2\mathbb{E}(\mathbb{1}\{\check{j} = \infty\}) \\ &\quad + \mathbb{E}\big(\mathbb{E}(\eta^2 \mathbb{1}\{\eta < 0\} | \mathtt{Y}_l, \mathtt{Y}_s) \mathbb{1}\{\check{j} = \infty\}\big) \\ &\le 2\mathbb{E}\big((f_{\mathbf{i}} - M(f))^2 \mathbb{1}\{\check{j} = \infty\}\big) + 2\gamma_e^2\sigma^2\mathbb{E}(\mathbb{1}\{\check{j} = \infty\}) + \sigma^2\gamma_e^2 Q_2 \mathbb{E}(\mathbb{1}\{\check{j} = \infty\}), \end{aligned} \tag{C.145}$$

where $Q_2 = \int_0^\infty x^2 5\Phi(x)^4 \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) dx \le \frac{5}{2}$.

To bound it we have the following lemmas, which are proved in Section D on page 159 and 160.

LEMMA C.43.

$$\begin{aligned} &\mathbb{E}\big((f_{\mathbf{i}} - M(f))^2 \mathbb{1}\{\check{j} = \infty\}\big) \\ &\le c_{m6}\rho_m(\frac{\sigma}{\sqrt{n}}; f)^2 + (\min\{f(x_i) : 0 \le i \le n\} - M(f))^2 \end{aligned} \tag{C.146}$$

LEMMA C.44.

$$\sigma^2 \mathbb{E}(\mathbb{1}\{\check{j} = \infty\}) \le 32\rho_m(\frac{\sigma}{\sqrt{n}}; f)^2 \tag{C.147}$$

Combining them together, we have

$$\begin{aligned} &\mathbb{E}((\hat{M} - M(f))^2) \\ &\le (c_{m1} + c_{m2} + 144\gamma_e^2 + 2c_{m_6})\rho_m(\frac{\sigma}{\sqrt{n}}; f)^2 + 2\left(\min\{f(x_i) : 0 \le i \le n\} - M(f)\right)^2 \\ &\le C_{3,0}\rho_m(\frac{\sigma}{\sqrt{n}}; f)^2 + 2\left(\min\{f(x_i) : 0 \le i \le n\} - M(f)\right)^2. \end{aligned} \tag{C.148}$$

Therefore,

$$\mathbb{E}(|\hat{M}-M(f)|) \leq \sqrt{\mathbb{E}((\hat{M}-M(f))^2)} \\ \leq \check{C}_{3,0}\rho_m(\frac{\sigma}{\sqrt{n}};f)+\sqrt{2}\left(\min\{f(x_i):0\leq i\leq n\}-M(f)\right). \tag{C.149}$$

□

Proof of Proposition C.8. Since we have

$$\sup_{h\in\mathcal{G}_n(f)}\rho_m(\frac{\sigma}{\sqrt{n}};h)\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};h)} \geq \sqrt{n}\frac{1}{\sqrt{2}}\frac{\sigma}{\sqrt{n}}=\frac{\sigma}{\sqrt{2}}, \tag{C.150}$$

we only need to prove that

$$\mathbb{E}(|\hat{M}-M(f)|) \leq \check{c}_{m1}\sigma+\sqrt{2}\left(\min\{f(x_i):0\leq i\leq n\}-M(f)\right). \tag{C.151}$$

We recycle all the notation in the proof of Proposition C.7, especially in Equation (C.144) and (D.103).

Similar to the proof of Proposition of C.7, we have

$$\begin{aligned} &\mathbb{E}((\hat{M}-M(f))^2) \\ &=\mathbb{E}((\hat{\mathfrak{f}}-M(f))^2\mathbb{1}\{\check{j}<\infty\})+\mathbb{E}((\mathfrak{E}_{\check{j},\tilde{\mathbf{i}}_{\check{j}},e}\frac{1}{2^{J-\check{j}}})^2\mathbb{1}\{\check{j}<\infty\})+ \\ &\quad 2\mathbb{E}\big((f_{\mathbf{i}}-M(f))^2\mathbb{1}\{\check{j}=\infty\}\big)+2\gamma_e^2\sigma^2\mathbb{E}(\mathbb{1}\{\check{j}=\infty\})+\sigma^2\gamma_e^2Q_2\mathbb{E}(\mathbb{1}\{\check{j}=\infty\}), \end{aligned} \tag{C.152}$$

where $Q_2=\int_0^\infty x^2\cdot 5\Phi(x)^4\frac{1}{\sqrt{2\pi}}\exp{(-\frac{x^2}{2})}dx\leq\frac{5}{2}$.

Since we have

$$\mathbb{E}((\mathfrak{E}_{\check{j},\tilde{\mathbf{i}}_{\check{j}},e}\frac{1}{2^{J-\check{j}}})^2\mathbb{1}\{\check{j}<\infty\})=\mathbb{E}(\frac{\sigma^2}{2^{J-\check{j}}}\mathbb{1}\{\check{j}<\infty\})\leq\sigma^2, \tag{C.153}$$

we are only left with bounding the terms: $\mathbb{E}((\hat{\mathfrak{f}}-M(f))^2\mathbb{1}\{\check{j}<\infty\})$, $\mathbb{E}((f_{\mathbf{i}}-M(f))^2\mathbb{1}\{\check{j}=\infty\})$.

We have the following two lemmas, which are proved in Section D on page 160 and 163.

Lemma C.45.

$$\mathbb{E}((\hat{\mathfrak{f}}-M(f))^2\mathbb{1}\{\check{j}<\infty\})\leq\check{c}_{m2}^2\sigma^2. \tag{C.154}$$

Lemma C.46.

$$\mathbb{E}((f_{\mathbf{i}}-M(f))^2\mathbb{1}\{\check{j}=\infty\})\leq\check{c}_{m3}^2\sigma^2+\left(\min\{f(x_i):0\leq i\leq n\}-M(f)\right)^2. \tag{C.155}$$

With these lemmas, we know that

$$\begin{aligned}
&\mathbb{E}((\hat{M}-M(f))^2) \\
&\le (\check{c}_{m2}^2+1+2\check{c}_{m3}^2+2\gamma_e^2+\gamma_e^2Q_2)\sigma^2+2\left(\min\{f(x_i):0\le i\le n\}-M(f)\right)^2 .
\end{aligned} \tag{C.156}$$

Therefore, we have

$$\begin{aligned}
&\mathbb{E}(\big|\hat{M}-M(f)\big|) \\
&\le \sqrt{\check{c}_{m2}^2+1+2\check{c}_{m3}^2+2\gamma_e^2+\gamma_e^2Q_2}\,\sigma+\sqrt{2}\big(\min\{f(x_i):0\le i\le n\}-M(f)\big) \\
&= \check{C}_3\frac{\sigma}{\sqrt{2}}+\sqrt{2}\big(\min\{f(x_i):0\le i\le n\}-M(f)\big).
\end{aligned} \tag{C.157}$$

□

C.15. Proof of Theorem 4.4. With Proposition C.4, we prove the theorem by proving the following lemmas.

LEMMA C.47 (length of the confidence interval for minimum 0).

$$\mathbb{E}_f L(\texttt{CI}_{m,\alpha}(Y)) \le \check{C}_{4,\alpha}\rho_m(\frac{\sigma}{\sqrt{n}};f)+\sqrt{2}\left(\min\{f(x_i):i=0,1,\cdots,n\}-\check{h}\right),$$

where $\check{h}=\inf\{M(g):g\in\mathcal{F},$ *and* $g(x_i)=f(x_i),i=0,1,\cdots,n\}$.

LEMMA C.48 (length of the confidence interval for minimum 1). *When* $\sup_{h\in\mathcal{G}_n(f)}\{\rho_z(\frac{\sigma}{\sqrt{n}};h)\}<\frac{1}{2n}$, *we have*

$$\mathbb{E}_f L(\texttt{CI}_{m,\alpha}(Y)) \le \check{C}_{5,\alpha}\sigma+\sqrt{2}\left(\min\{f(x_i):i=0,1,\cdots,n\}-\check{h}\right),$$

where $\check{h}=\min\{M(g):g\in\mathcal{F},$ *and* $g(x_i)=f(x_i),i=0,1,\cdots,n\}$.

Note that we always have

$$\sup_{h\in\mathcal{G}_n(f)}\rho_m(\frac{\sigma}{\sqrt{n}};h)\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};h)}\ge\frac{\sigma}{\sqrt{2}},$$

hence with these two lemmas we know that

$$\begin{aligned}
&\mathbb{E}_f L(\texttt{CI}_{m,\alpha}(Y)) \\
&\le (\sqrt{2}\check{C}_{4,\alpha}+\sqrt{2}\check{C}_{5,\alpha})\sup_{h\in G_n(f)}\rho_m(\frac{\sigma}{\sqrt{n}};h)\left(1\wedge\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};h)}\right)+\sqrt{2}\mathfrak{D}_m(n,f).
\end{aligned} \tag{C.158}$$

When $0 < \alpha < 0.3$, letting

$$C_{4,\alpha} = \frac{\sqrt{2}\check{C}_{4,\alpha} + \sqrt{2}\check{C}_{5,\alpha}}{\tilde{C}_{m,\alpha}} + \frac{2\sqrt{2}}{1-2\alpha} \tag{C.159}$$

gives the statement with respect to the expected length.

LEMMA C.49 (coverage of the confidence interval for minimum).

$$P(M(f) \in \mathtt{CI}_{m,\alpha}(Y)) \geq 1 - \alpha.$$

PROOF OF LEMMA C.47.

$$I_{hi} - I_{lo} + 1 \leq 2 + 9 \cdot 2^{j_l - j_s} \leq 2 + 9 \cdot 2^{K_{\frac{\alpha}{4}} + \tilde{K}_{\frac{\alpha}{4}} + 1}. \tag{C.160}$$

Therefore,

$$S_{I_{hi} - I_{lo} + 1, \frac{\alpha}{4}} \leq -\Phi^{-1}\left(\frac{\alpha}{4(2 + 9 \cdot 2^{K_{\frac{\alpha}{4}} + \tilde{K}_{\frac{\alpha}{4}} + 1})}\right). \tag{C.161}$$

$$\begin{aligned} \mathbb{E}_f L(\mathtt{CI}_{m,\alpha}(Y)) \leq &\left(S_{I_{hi} - I_{lo} + 1, \frac{\alpha}{4}} - \Phi^{-1}(\frac{\alpha}{4}) + \sqrt{3}\right)\gamma_e \mathbb{E}(\frac{\sigma}{\sqrt{2^{J-j_l}}}) \\ &+ \mathbb{E}\left((\hat{\mathtt{f}}_1 - z_{\alpha/4}\frac{\sqrt{3}\sigma}{\sqrt{2^{J-j_l}}} - \frac{\sqrt{3}\sigma}{\sqrt{2^{J-j_l}}} - \mathtt{f}_{lo})_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\} \right). \end{aligned} \tag{C.162}$$

We first bound the first term. Note that $\frac{\sigma}{\sqrt{2^{J-\mathtt{j}^*}}} = \frac{\frac{\sigma}{\sqrt{n}}}{\sqrt{m_j^*}} < \frac{\frac{\sigma}{\sqrt{n}}}{\sqrt{\frac{1}{8}\rho_z(\frac{\sigma}{\sqrt{n}};f)}} \leq 4\rho_m(\frac{\sigma}{\sqrt{n}};f)$. Therefore, we have

$$\begin{aligned} &\mathbb{E}(\frac{\sigma}{\sqrt{2^{J-j_l}}}) \leq \mathbb{1}\{\mathtt{j}^* + 2 \geq J\}\sigma + \\ &\quad \mathbb{1}\{\mathtt{j}^* + 3 \leq J\}\left(\mathbb{E}(\frac{\sigma}{\sqrt{2^{J-\hat{\mathtt{j}}-\tilde{K}_{\frac{\alpha}{4}}}}}\mathbb{1}\{\hat{\mathtt{j}} \leq J - \tilde{K}_{\frac{\alpha}{4}}\}) + \sigma\mathbb{E}(\mathbb{1}\{\hat{\mathtt{j}} > J - \tilde{K}_{\frac{\alpha}{4}}\})\right) \\ &\leq \mathbb{1}\{\mathtt{j}^* + 3 \leq J\}\Big(\frac{\sigma}{\sqrt{2^{J-\mathtt{j}^*-\tilde{K}_{\frac{\alpha}{4}}-3}}} + \sum_{j=\mathtt{j}^*+3}^{J}\frac{\sigma}{\sqrt{2^{J-j-1-\tilde{K}_{\frac{\alpha}{4}}}}}\Phi(-2+\frac{1}{6})^{j-\mathtt{j}^*-2} \\ &\qquad + \frac{\sigma}{\sqrt{2^{J-\mathtt{j}^*}}}\sqrt{2^{J-\mathtt{j}^*}}\Phi(-2+\frac{1}{6})^{(J-1-\tilde{K}_{\frac{\alpha}{4}}-\mathtt{j}^*)_+}\Big) \\ &\qquad + \mathbb{1}\{\frac{1}{n} > \frac{\rho_z(\frac{\sigma}{\sqrt{n}};f)}{32}\}\sqrt{n}\sqrt{2\rho_z(\frac{\sigma}{\sqrt{n}};f)}\rho_m(\frac{\sigma}{\sqrt{n}};f) \\ &\leq 2^{1+\frac{\tilde{K}_{\frac{\alpha}{4}}}{2}}\tilde{C}_4\rho_m(\frac{\sigma}{\sqrt{n}};f) + 8\rho_m(\frac{\sigma}{\sqrt{n}};f)\bar{C}_{1,\alpha}. \end{aligned} \tag{C.163}$$

Now we turn to the second term,

$$
\begin{aligned}
&\mathbb{E}\Big((\hat{\mathtt{f}}_1 - z_{\alpha/4}\frac{\sqrt{3}\sigma}{\sqrt{2^{J-j_l}}} - \frac{\sqrt{3}\sigma}{\sqrt{2^{J-j_l}}} - \mathtt{f}_{lo})_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\Big) \\
&\le \mathbb{E}\left((\hat{\mathtt{f}}_1 - \mathtt{f}_{lo})_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right) \\
&\le \mathbb{E}\left((\hat{\mathtt{f}}_1 - M(f))_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right) + \mathbb{E}\left((M(f) - \mathtt{f}_{lo})_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right) \\
&\le \mathbb{E}\left((\hat{M} - M(f))_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right) + \mathbb{E}\left((M(f) - \mathtt{f}_{lo})_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right),
\end{aligned}
$$

where $\hat{M}$ is defined in Equation (4.9).

Then according to Proposition C.7, we have

$$
\begin{aligned}
&\mathbb{E}\left((\hat{M} - M(f))_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right) \\
&\quad \le \check{C}_{3,0}\rho_m(\frac{\sigma}{\sqrt{n}}; f) + \sqrt{2}\left(\min\{f(x_i) : 0 \le i \le n\} - M(f)\right).
\end{aligned}
$$

Now we turn to the term $\mathbb{E}\left((M(f) - \mathtt{f}_{lo})_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right)$. We begin by defining two sequences of linear functions: $\{\tilde{v}_{l,k} : 1 \le k \le n-1\}$ and $\{\tilde{v}_{r,k} : 0 \le k \le n-2\}$. For $1 \le k \le n-1$, define linear functions

$$
\tilde{v}_{l,k} : t \mapsto \frac{f(x_k) - f(x_{k-1})}{1/n}(t - x_k) + f(x_k). \tag{C.164}
$$

For $0 \le k \le n-2$, define linear functions

$$
\tilde{v}_{r,k} : t \mapsto \frac{f(x_{k+2}) - f(x_{k+1})}{1/n}(t - x_{k+1}) + f(x_{k+1}). \tag{C.165}
$$

Recall that $\mathcal{G}_n(f)$, as defined in Equation (C.105), is the set of convex functions that take the same values as f on grid points. Now we define $\tilde{h}(k)$ for $0 \le k \le n-1$ as follows

$$
\tilde{h}(k) = \begin{cases} \min_{t \in [x_0, x_1]} \tilde{v}_{r,0}(t), & k = 0 \\ \min_{t \in [x_k, x_{k+1}]} \max\{\tilde{v}_{l,k}(t), \tilde{v}_{r,k}(t)\}, & 1 \le k \le n-2 \\ \min_{t \in [\frac{n-1}{n}, 1]} \tilde{v}_{r,n-1}(t), & k = n-1 \end{cases}. \tag{C.166}
$$

It is easy to check that for $0 \le k \le n-1$, the possible values of $\inf_{t \in [x_k, x_{k+1}]} g(t)$ for $g \in \mathcal{G}_n(f)$ are as follows.

$$
\{\inf_{t \in [x_k, x_{k+1}]} g(t) : g \in \mathcal{G}_n(f)\} = [\tilde{h}(k), \max\{f(x_k), f(x_{k+1})\}]. \tag{C.167}
$$

Now we know that

$$\max\{M(g) : g \in \mathcal{G}_n(f)\} = \min\{f(x_i) : 0 \le i \le n\},$$
$$\min\{M(g) : g \in \mathcal{G}_n(f)\} = \min\{\tilde{h}(i) : 0 \le i \le n-1\}. \tag{C.168}$$

Denote $\check{h} = \min\{\tilde{h}(i) : 0 \le i \le n-1\}$, and then we have

$$\begin{aligned}
&\mathbb{E}\left((M(f) - \mathtt{f}_{lo})_+ \mathbb{1}\{\hat{\jmath} + \tilde{K}_{\alpha/4} > J\}\right) \\
&\le (M(f) - \check{h}) + \mathbb{E}\left((\check{h} - \mathtt{f}_{lo})_+ \mathbb{1}\{\hat{\jmath} + \tilde{K}_{\alpha/4} > J\}\right) \\
&\le (M(f) - \check{h}) + \sum_{i=I_{lo}-1}^{I_{hi}-2} \mathbb{E}\left((\tilde{h}(i) - h(i))_+ \mathbb{1}\{\hat{\jmath} + \tilde{K}_{\alpha/4} > J\}\right).
\end{aligned} \tag{C.169}$$

Recall the definition of δ_i in Equation (C.144): $\delta_i = y_{e,i-1} - f(x_{i-1})$. For $(I_{lo}-1) \vee 1 \le i \le (I_{hi}-2) \wedge (n-2)$, we have

$$\begin{aligned}
&\mathbb{E}\left((\tilde{h}(i) - h(i))_+ \mathbb{1}\{\hat{\jmath} + \tilde{K}_{\alpha/4} > J\}\right) \\
&\le \mathbb{E}\Bigg(\Big(\min_{t\in[x_i,x_{i+1}]} \max\{\tilde{v}_{l,i}(t), \tilde{v}_{r,i}(t)\} - \\
&\quad \min_{t\in[x_i,x_{i+1}]} \max\{\tilde{v}_{l,i}(t) + (\delta_{i+1} - \delta_i - 2H)n(t - x_i) + \delta_{i+1} - H, \\
&\quad \tilde{v}_{r,i}(t) + (\delta_{i+2} - \delta_{i+3} - 2H)n(x_{i+1} - t) + \delta_{i+2} - H\}\Big)_+ \mathbb{1}\{\hat{\jmath} + \tilde{K}_{\alpha/4} > J\}\Bigg) \\
&\le P(\hat{\jmath} + \tilde{K}_{\alpha/4} > J)\left(\mathbb{E}\left(2|\delta_{i+1}| + |\delta_i| + 2|\delta_{i+2}| + |\delta_{i+3}|\right) + 3H\right) \\
&\le \left(6 \cdot \gamma_e \sigma \sqrt{\frac{2}{\pi}} + 3\gamma_e S_{I_{hi}-I_{lo}+3,\frac{1}{8}}\sigma\right) P(\hat{\jmath} + \tilde{K}_{\alpha/4} > J) \overset{(a)}{\le} \bar{C}_{2,\alpha}\rho_m(\frac{\sigma}{\sqrt{n}}; f).
\end{aligned} \tag{C.170}$$

Step (a) follows from $\sigma\mathbb{E}(\mathbb{1}\{\hat{\jmath} + \tilde{K}_{\alpha/4} > J\}) < \mathbb{E}(\frac{\sigma}{\sqrt{2^{J-j_l}}})$, and Inequality (C.163).

When $I_{lo} = 1$,

$$\begin{aligned}
&\mathbb{E}\left(\left(\tilde{h}(0) - h(0)\right)_+ \mathbb{1}\{\hat{\jmath} + \tilde{K}_{\alpha/4} > J\}\right) \\
&\le \mathbb{E}\Bigg(\left(\min_{t\in[0,1/n]} \tilde{v}_{r,0}(t) - \min_{t\in[0,1/n]} (\tilde{v}_{r,0}(t) + (\delta_3 - \delta_2 + 2H)n(t - x_1) + \delta_2 - H)\right)_+ \\
&\quad \mathbb{1}\{\hat{\jmath} + \tilde{K}_{\alpha/4} > J\}\Bigg) \le P(\hat{\jmath} + \tilde{K}_{\alpha/4} > J)(3H + 3\gamma_e\sigma\sqrt{\frac{2}{\pi}}) < \bar{C}_{2,\alpha}\rho_m(\frac{\sigma}{\sqrt{n}}; f).
\end{aligned} \tag{C.171}$$

When $I_{hi} - 2 = n - 1$,

(C.172)
$$\mathbb{E}\left(\left(\tilde{h}(n-1) - h(n-1)\right)_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right)$$
$$\leq \mathbb{E}\Bigg(\mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}$$
$$\left(\min_{t\in[\frac{n-1}{n},1]} \tilde{v}_{l,n-1}(t) - \min_{t\in[\frac{n-1}{n},1]} (\tilde{v}_{l,n-1}(t) + (\delta_n - \delta_{n-1} - 2H)n(t - x_{n-1}) + \delta_n - H)\right)_+\Bigg)$$
$$\leq P(\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J)(3H + 3\gamma_e\sigma\sqrt{\frac{2}{\pi}}) < \bar{C}_{2,\alpha}\rho_m(\frac{\sigma}{\sqrt{n}}; f).$$

Going back to Inequality (C.169), we have

(C.173)
$$\mathbb{E}\left((M(f) - \mathtt{f}_{lo})_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right) \leq (I_{hi} - I_{lo})\bar{C}_{2,\alpha}\rho_m(\frac{\sigma}{\sqrt{n}}; f) + (M(f) - \check{h}).$$

Combing all the terms together, we have

(C.174)
$$\mathbb{E}_f L(CI_{m,\alpha}(Y)) \leq \check{C}_{4,\alpha}\rho_m(\frac{\sigma}{\sqrt{n}}; f) + \sqrt{2}\left(\min\{f(x_i) : 0 \leq i \leq n\} - \check{h}\right).$$

□

Proof of Lemma C.48. The proof of this lemma is very similar to that of lemma C.47. For simplicity, we will omit the parts that are the same and only point out the places that are different.

Similar to Inequality (C.162), we have

(C.175)
$$\mathbb{E}_f L(\mathtt{CI}_{m,\alpha}(Y))$$
$$\leq \left(S_{I_{hi}-I_{lo}+1,\frac{\alpha}{4}} - \Phi^{-1}(\frac{\alpha}{4}) + \sqrt{3}\right)\gamma_e \mathbb{E}(\frac{\sigma}{\sqrt{2^{J-j_l}}})$$
$$+ \mathbb{E}\left((\hat{\mathtt{f}}_1 - z_{\alpha/4}\frac{\sqrt{3}\sigma}{\sqrt{2^{J-j_l}}} - \frac{\sqrt{3}\sigma}{\sqrt{2^{J-j_l}}} - \mathtt{f}_{lo})_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right)$$
$$\leq \left(S_{I_{hi}-I_{lo}+1,\frac{\alpha}{4}} - \Phi^{-1}(\frac{\alpha}{4}) + \sqrt{3}\right)\gamma_e\sigma +$$
$$\mathbb{E}\left((\hat{\mathtt{f}}_1 - M(f))_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right) + \mathbb{E}\left((M(f) - \mathtt{f}_{lo})_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right).$$

For the second term, according to the definition of $\hat{\mathtt{f}}_1$ and Proposition C.8,

we have

(C.176)

$$\mathbb{E}\left((\hat{\mathtt{f}}_1 - M(f))_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right)$$
$$\leq \mathbb{E}\left((\hat{M} - M(f))_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right)$$
$$\leq \check{C}_3 \sup_{h\in\mathcal{G}_n(f)} \rho_m(\frac{\sigma}{\sqrt{n}}; h)\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}}; h)} + \sqrt{2}\left(\min\{f(x_i) : 0 \leq i \leq n\} - M(f)\right),$$

where $\hat{M}$ is defined in (4.9).

For $\mathbb{E}\left((M(f) - \mathtt{f}_{lo})_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right)$, according to the arguments in the proof of Lemma C.47, we have

$$\begin{aligned}
&\mathbb{E}\left((M(f) - \mathtt{f}_{lo})_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right) \\
&\leq (M(f) - \check{h}) + \sum_{i=I_{lo}-1}^{I_{hi}-2} \mathbb{E}\left((\tilde{h}(i) - h(i))_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right).
\end{aligned} \tag{C.177}$$

For $(I_{lo} - 1) \vee 1 \leq i \leq (I_{hi} - 2) \wedge (n - 2)$, we have

$$\begin{aligned}
&\mathbb{E}\left((\tilde{h}(i) - h(i))_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right) \\
&\leq \left(6 \cdot \gamma_e\sqrt{\frac{2}{\pi}} + 3\gamma_e S_{I_{hi}-I_{lo}+3,\frac{1}{8}}\right) P(\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J)\sigma.
\end{aligned} \tag{C.178}$$

When $I_{lo} = 1$,

$$\begin{aligned}
&\mathbb{E}\left(\left(\tilde{h}(0) - h(0)\right)_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right) \\
&\leq P(\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J)(3H + 3\gamma_e\sigma\sqrt{\frac{2}{\pi}}).
\end{aligned} \tag{C.179}$$

When $I_{hi} - 2 = n - 1$,

$$\begin{aligned}
&\mathbb{E}\left(\left(\tilde{h}(n-1) - h(n-1)\right)_+ \mathbb{1}\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J\}\right) \\
&\leq P(\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} > J)(3H + 3\gamma_e\sigma\sqrt{\frac{2}{\pi}}).
\end{aligned} \tag{C.180}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\left((M(f)-\mathtt{f}_{lo})_+\mathbb{1}\{\hat{\mathtt{j}}+\tilde{K}_{\alpha/4}>J\}\right)\\
&\le (I_{hi}-I_{lo})\left(6\cdot\gamma_e\sqrt{\frac{2}{\pi}}+3\gamma_e S_{I_{hi}-I_{lo}+3,\frac{1}{8}}\right)P(\hat{\mathtt{j}}+\tilde{K}_{\alpha/4}>J)\sigma\\
&\qquad +(M(f)-\check{h}).
\end{aligned}\tag{C.181}$$

Hence

$$\mathbb{E}_f L(\mathtt{CI}_{m,\alpha}(Y))\le \check{C}_{5,\alpha}\sigma+\sqrt{2}\left(\min\{f(x_i):i=0,1,\cdots,n\}-\check{h}\right).\tag{C.182}$$

□

Proof of Lemma C.49. Similar to the proof of lemma C.19, define the following events:

$$\begin{aligned}
&\mathtt{E}=\{Z(f)\notin[\frac{2^{J-j_l}(I_{lo}-1)}{n},\frac{2^{J-j_l}I_{hi}-1}{n}]\cap[0,1]\}\\
&\mathtt{E}_1=\{\check{j}\ge \mathtt{j}^{\mathtt{w}}+K_{\frac{\alpha}{4}}+1,\text{ and } \mathtt{j}^{\mathtt{w}}+K_{\frac{\alpha}{4}}+1\le J\}\\
&\mathtt{F}=\{\check{j}\le \mathtt{j}^*-2-\tilde{K}_{\frac{\alpha}{4}}\}\\
&\mathtt{G}=\{\mathtt{f}_{hi}<M(f)\}\\
&\mathtt{H}=\{\mathtt{f}_{lo}>M(f)\}.
\end{aligned}\tag{C.183}$$

Then we know that

$$\mathtt{E}_1^c\subset\mathtt{E}^c.\tag{C.184}$$

So we have

$$\{M(f)\in\mathtt{CI}_{m,\alpha}(Y)\}\supset \mathtt{E}^c\cap\mathtt{F}^c\cap\mathtt{G}^c\cap\mathtt{H}^c\supset\mathtt{E}_1^c\cap\mathtt{F}^c\cap\mathtt{G}^c\cap\mathtt{H}^c.\tag{C.185}$$

Then we have

$$\begin{aligned}
&P(M(f)\in\mathtt{CI}_{m,\alpha}(Y))\\
&\ge P(\mathtt{E}_1^c\cap\mathtt{F}^c\cap\mathtt{G}^c\cap\mathtt{H}^c)\\
&=P(\mathtt{G}^c\cap\mathtt{H}^c|\mathtt{E}_1^c\cap\mathtt{F}^c)(1-P(\mathtt{E}_1)-P(\mathtt{F})+P(\mathtt{F}\cap\mathtt{E}_1))\\
&=(1-P(\mathtt{G}|\mathtt{E}_1^c\cap\mathtt{F}^c)-P(\mathtt{H}|\mathtt{E}_1^c\cap\mathtt{F}^c)\\
&\qquad+P(\mathtt{G}\cap\mathtt{H}|\mathtt{E}_1^c\cap\mathtt{F}^c))(1-P(\mathtt{E}_1)-P(\mathtt{F})+P(\mathtt{F}\cap\mathtt{E}_1))\\
&\ge 1-P(\mathtt{G}|\mathtt{E}_1^c\cap\mathtt{F}^c)-P(\mathtt{H}|\mathtt{E}_1^c\cap\mathtt{F}^c)-P(\mathtt{E}_1)-P(\mathtt{F}).
\end{aligned}\tag{C.186}$$

According to Lemma C.37, we have

$$P(\mathtt{E}_1) = P(\hat{\mathtt{j}} \geq \mathtt{j}^{\mathtt{W}} + K_{\frac{\alpha}{4}} + 1, \mathtt{j}^{\mathtt{W}} + K_{\frac{\alpha}{4}} + 1 \leq J) \\ \leq P(\check{j} \geq \mathtt{j}^{\mathtt{W}} + K_{\frac{\alpha}{4}} + 1, \mathtt{j}^{\mathtt{W}} + K_{\frac{\alpha}{4}} + 1 \leq J) \\ \leq \Phi(-2)^{K_{\frac{\alpha}{4}}} \leq \frac{\alpha}{4}. \tag{C.187}$$

Similar to the proof of Lemma C.19, especially the proof of Lemma C.21, which consists the proof of Lemma C.19, we have

$$P(\mathtt{F}) \leq P(\check{j} \leq \mathtt{j}^* - 2 - \tilde{K}_{\frac{\alpha}{4}}) \leq \frac{\alpha}{4}. \tag{C.188}$$

For the remaining terms in Inequality (C.186), we claim

Lemma C.50.

$$P(\mathtt{H}|\mathtt{E}_1^c \cap \mathtt{F}^c) \leq \frac{\alpha}{4}. \tag{C.189}$$

Proof. With a little abuse of notation, let A denote the event $\{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} \leq J\}$ in the proof of this lemma. Then

$$P(\mathtt{H}|\mathtt{E}_1^c \cap \mathtt{F}^c) = \\ P(\mathtt{H}|\mathtt{E}_1^c \cap \mathtt{F}^c \cap A)P(A|\mathtt{E}_1^c \cap \mathtt{F}^c) + P(\mathtt{H}|\mathtt{E}_1^c \cap \mathtt{F}^c \cap A^c)(1 - P(A|\mathtt{E}_1^c \cap \mathtt{F}^c)). \tag{C.190}$$

We start with the second term, for which we introduce another lemma.

Lemma C.51. *On event A^c, for $h(i)$ defined in Algorithm 2,*

$$P(h(i) \leq \min_{t \in [x_i, x_{i+1}]} f(t) \text{ for all } I_{lo} - 1 \leq i \leq I_{hi} - 2 \big| \mathtt{Y}_l, \mathtt{Y}_s) \geq 1 - \alpha/4.$$

Proof. We take the definition of δ_i in Equation (C.144): $\delta_i = y_{e,i-1} - f(x_{i-1})$. Since

$$P(\max\{|\delta_i| : (I_{lo} - 1) \vee 1 \leq i \leq (I_{hi} + 1) \wedge (n+1)\} > H \big| \mathtt{Y}_l, \mathtt{Y}_s) \\ \leq P(\max\{\delta_i : (I_{lo} - 1) \vee 1 \leq i \leq (I_{hi} + 1) \wedge (n+1)\} > H \big| \mathtt{Y}_l, \mathtt{Y}_s) \\ + P(-\min\{\delta_i : I_{lo} \leq i \leq I_{hi}\} > H \big| \mathtt{Y}_l, \mathtt{Y}_s) \leq \alpha/4, \tag{C.191}$$

we have that condition on $\mathtt{Y}_l, \mathtt{Y}_s$, on event A^c, the following event holds with probability at least $1 - \alpha/4$:

$$B = \{y_{e,i} - H \leq f(x_i) \text{ and } y_{e,i} + H \geq f(x_i) \text{ for all } (I_{lo} - 2)_+ \leq i \leq I_{hi} \wedge n\}.$$

On event B, for $(I_{lo}-1)\vee 1 \le i \le (I_{hi}-2)\wedge(n-2)$, consider two linear functions $\tilde{v}_{l,i} : t \mapsto \frac{f(x_i)-f(x_{i-1})}{1/n}(t-x_i)+f(x_i)$, $\tilde{v}_{r,i} : t \mapsto \frac{f(x_{i+2})-f(x_{i+1})}{1/n}(t-x_{i+1})+f(x_{i+1})$. Then for $t \in [x_i, x_{i+1}]$, $f(t) \ge \max\{\tilde{v}_{l,i}(t), \tilde{v}_{r,i}(t)\} \ge \max\{v_{l,i}(t), v_{r,i}(t)\}$, hence $h(i) \le \inf_{t\in[x_i,x_{i+1}]} f(t)$.

If $I_{lo} - 1 = 0$, suppose event B holds, then consider the linear function $\tilde{v}_{r,0} : t \mapsto \frac{f(x_2)-f(x_1)}{1/n}(t - x_1) + f(x_1)$. For $t \in [0, 1/n]$, we have that $f(t) \ge \tilde{v}_{r,0}(t) \ge v_{r,0}(t)$, hence $h(0) \le \min_{t\in[0,1/n]} f(t)$.

Similarly, if $I_{hi} - 2 = n - 1$, on event B we have that $h(n - 1) \le \min_{t\in[n-1/n,1]} f(t)$.

Therefore, on event B, $\min\{h(i) : I_{lo}-1 \le i \le I_{hi}-2\} \le \inf_{t\in[x_{I_{lo}-1},x_{I_{hi}-1}]} f(t)$.

Therefore,

$$P\left(h(i) \le \min_{t\in[x_i,x_{i+1}]} f(t) \text{ for all } I_{lo} - 1 \le i \le I_{hi} - 2 \,\middle|\, \mathtt{Y}_l, \mathtt{Y}_s\right) \ge P(B|\mathtt{Y}_l, \mathtt{Y}_s) \ge 1 - \alpha/4. \tag{C.192}$$

□

Recalling that on event $\mathtt{E}_1^c$, we have $Z(f) \in [x_{I_{lo}-1}, x_{I_{hi}-1}]$, together with Lemma C.51, we have

(C.193)

$$\begin{aligned}
&P(H \mid \mathtt{E}_1^c \cap \mathtt{F}^c \cap A^c) \\
&\le P\left(\min\{h(i) : I_{lo} - 1 \le i \le I_{hi} - 2\} > M(f) \mid \mathtt{E}_1^c \cap \mathtt{F}^c \cap A^c\right) \\
&= P\left(\min\{h(i) : I_{lo} - 1 \le i \le I_{hi} - 2\} > \min_{t\in[x_{I_{lo}-1},x_{I_{hi}-1}]} f(t) \mid \mathtt{E}_1^c \cap \mathtt{F}^c \cap A^c\right) \le \alpha/4.
\end{aligned}$$

Now we turn to the first term in Inequality (C.190).

First, we show that on event $E_1^c \cap A \cap F^c$, we have $\min_{I_{lo}\le i\le I_{hi}} ave_f(j_l, i) \le M(f)+\frac{\sqrt{3}\sigma}{\sqrt{2^{J-j_l}}}$ using the fact that $\rho_m(\frac{\sigma}{\sqrt{n}}; f) \le \frac{\sqrt{3}\sigma}{\sqrt{n}\sqrt{\rho_z(\frac{\sigma}{\sqrt{n}};f)}} \le \frac{\sqrt{3}\sigma}{\sqrt{n}\sqrt{2^{j-j^*+2}/n}}$:

$$\begin{aligned}
&\{\min_{I_{lo}\le i\le I_{hi}} ave_f(j_l, i) \le M(f) + \frac{\sqrt{3}\sigma}{\sqrt{2^{J-j_l}}}\} \cap \mathtt{E}_1^c \cap A \\
&\supset \{\min_{I_{lo}\le i\le I_{hi}} ave_f(j_l, i) \le M(f) + \rho_m(\frac{\sigma}{\sqrt{n}}; f)\} \cap \{j_l > \mathtt{j}^* - 2\} \cap \mathtt{E}_1^c \cap A \\
&\supset F^c \cap \{j_l > \mathtt{j}^* - 2\} \cap \mathtt{E}_1^c \cap \{\hat{\mathtt{j}} + \tilde{K}_{\alpha/4} \le J\} \\
&\supset \mathtt{E}_1^c \cap A \cap \mathtt{F}^c.
\end{aligned}$$

Denote $i_{min} = \arg\min_{I_{lo}\le i\le I_{hi}} ave_f(j_l, i)$. When there is more than one qualifying for i_{min}, take any one.

Therefore,

$$P(\mathtt{H}|\mathtt{E}_1^c \cap \mathtt{F}^c \cap A) \leq P(\mathfrak{E}_{j_l,i_{min},e} \geq -\Phi^{-1}(\frac{\alpha}{4})\sigma\gamma_e 2^{\frac{J-j_l}{2}}) \leq \frac{\alpha}{4}. \tag{C.194}$$

Therefore,

$$P(\mathtt{H}|\mathtt{E}_1^c \cap \mathtt{F}^c) \leq \frac{\alpha}{4}. \tag{C.195}$$

□

Similar to the arguments in proof of Lemma C.22, we have

$$P(\mathtt{G}|\mathtt{E}_1^c \cap \mathtt{F}^c) \leq \frac{\alpha}{4}. \tag{C.196}$$

Returning to the main theorem, we have,

$$P(M(f) \in \mathtt{CI}_{m,\alpha}(Y)) \geq 1 - \alpha. \tag{C.197}$$

□

APPENDIX D: PROOFS OF TECHNICAL LEMMAS

We prove all the technical lemmas in this section.

PROOF OF LEMMA C.1. The inequalities are due to

$$\frac{f(x_2)-f(x_1)}{x_2-x_1} - \frac{f(x_3)-f(x_2)}{x_3-x_2} \leq \frac{(x_3-x_1)(f(x_2)-\frac{f(x_1)(x_3-x_2)+f(x_3)(x_2-x_1)}{x_3-x_1})}{(x_2-x_1)(x_3-x_2)} \leq 0,$$

and

$$\frac{f(x_3)-f(x_1)}{x_3-x_1} = \frac{f(x_2)-f(x_1)}{x_2-x_1} \cdot \frac{x_2-x_1}{x_3-x_1} + \frac{f(x_3)-f(x_2)}{x_3-x_2} \cdot \frac{x_3-x_2}{x_3-x_1}.$$

□

PROOF OF LEMMA C.2. Let $t = x^{\frac{3}{2}}\sqrt{2/3} - 2$, then we have

$$\begin{aligned} &\frac{2x\Phi(2-(2x)^{\frac{3}{2}}\sqrt{2/3})}{x\Phi(2-\sqrt{2/3}x^{3/2})} \leq 2\frac{\int_{-\infty}^{-2\sqrt{2}t-(4\sqrt{2}-2)} \exp(-\frac{u^2}{2})du}{\int_{-\infty}^{-t} \exp(-\frac{u^2}{2})du} \\ &\leq 4\sqrt{2}\frac{\int_{-\infty}^{-2\sqrt{2}t} \exp(-\frac{u^2}{2})du \exp(-\frac{(4\sqrt{2}-2)^2}{2})}{\int_{-\infty}^{-2\sqrt{2}t} \exp(-\frac{u^2}{16})du} < 0.008. \end{aligned} \tag{D.1}$$

□

Proof of Lemma C.3. Let

$$q(x) = x^2\Phi(-x)$$

Then

$$q'(x) = x(2\Phi(-x) - \frac{x}{\sqrt{2\pi}}\exp{(-\frac{x^2}{2})}).$$

Taking further derivative, we know that $\mathrm{sign}((2\Phi(-x) - \frac{x}{\sqrt{2\pi}}\exp{(-\frac{x^2}{2})})') = \mathrm{sign}(x^2-3)$. Hence $q'(x)/x$ goes down and then goes up, its first root is the place that $q(x)$ takes maximum. Since $q'(1.19) > 0$, $q'(1.2) < 0$, we have $\sup_{x>0} q(x) \le 1.2^2\Phi(-1.19) < 0.168514 < 0.169$. Therefore $Q \le 1.2^2\Phi(-1.19) < 0.169$. Only in this proof, let $u(x) = x^2\Phi(2-x)$. We have $u'(x) = x(2\Phi(2-x) - x\frac{1}{\sqrt{2\pi}}\exp{(-\frac{(2-x)^2}{2})})$. Since $\mathrm{sign}((2\Phi(2-x) - x\frac{1}{\sqrt{2\pi}}\exp{(-\frac{(2-x)^2}{2})})') = \mathrm{sign}(x(x-2)-3)$, and $\min_{x>0} u'(x) < 0 < u'(1)$, we know $u'(x)$ has at least 1 root. And its first root (when the root is unique, its first root is its unique root) is where $u(x)$ takes maximum, since $u'(2.18) > 0, u'(2.19) < 0$, we have $u(x) \le 2.19^2\Phi(2-2.18) < 2.0555$. Hence $V < 2.0555$. □

Proof of Lemma C.7. Since we have for $t > 0$,

$$\Phi(-t) \ge \frac{1}{\sqrt{2\pi}}\frac{t}{t^2+1}\exp{(-t^2/2)}, \tag{D.2}$$

we set $t(\alpha) = \sqrt{2\log{(1/\alpha)}} - \sqrt{\log{(2\log{(1/\alpha)})}}$. So we get, for $\alpha < 0.03$,

$$\begin{aligned}\Phi(-t(\alpha)) &\ge \frac{1}{\sqrt{2\pi}}\alpha\exp{(\log{(2\log{(1/\alpha)})}\cdot(\sqrt{\frac{\exp{(1)}}{1}}-\frac{1}{2}))}\frac{t(\alpha)}{t(\alpha)^2+1} \\ &\ge \alpha\cdot(2\log{(1/\alpha)})^{1.14}\frac{1}{\sqrt{2\pi}}\frac{t(\alpha)}{t(\alpha)^2+1}.\end{aligned} \tag{D.3}$$

Further, denote $x = 2\log{(1/\alpha)}$, we have

$$\frac{t(\alpha)}{t(\alpha)^2+1}x = \frac{t(\alpha)^2}{t(\alpha)^2+1}\frac{x}{t(\alpha)} \ge \frac{t(\alpha)^2}{t(\alpha)^2+1}\sqrt{x} > 0.6\sqrt{x} > 1.58. \tag{D.4}$$

The inequalities are because of $t(\alpha) = \sqrt{x} - \sqrt{\log x}$, t increases with x when $x > 2$, and $x > 7$ when $\alpha < 0.03$.

Therefore, for $\alpha < 0.03$

$$\Phi(-t(\alpha)) \ge 0.82\alpha. \tag{D.5}$$

Therefore, for $\alpha \le 0.005$, $z_{3\alpha} \ge t(\frac{3}{0.82}\alpha)$, $z_{2.06\alpha} \ge t(\frac{2.06}{0.82}\alpha)$.
Note that for $\alpha < 0.02$, $t(\alpha) \ge \sqrt{\log(1/\alpha)} \times 0.689$.
Hence for $\alpha \le 0.005$,

$$
\begin{aligned}
z_{3\alpha} &\ge t(\frac{3}{0.82}\alpha) \ge 0.689 \times \sqrt{\log(0.82/3\alpha)} \ge 0.599\sqrt{\log(1/\alpha)}, \\
z_{2.06\alpha} &\ge t(\frac{2.06}{0.82}\alpha) \ge 0.689 \times \sqrt{\log(0.82/2.06\alpha)} \ge 0.627\sqrt{\log(1/\alpha)}.
\end{aligned} \tag{D.6}
$$

We are now left with bounding

$$\inf_{\alpha \in (0.005, 0.08]} \frac{z_{2.06\alpha}}{\sqrt{\log 1/\alpha}}.$$

Note that both $z_{2.06\alpha}$ and $\sqrt{\log 1/\alpha}$ increases with α decreasing. Therefore,

$$\inf_{\alpha \in (0.005, 0.08]} \frac{z_{2.06\alpha}}{\sqrt{\log 1/\alpha}} \ge \min_{5 \le k \le 79} \frac{z_{2.06\frac{k+1}{1000}}}{\sqrt{\log 1000/k}} \ge 0.61.$$

Therefore, for $\alpha < 0.08$, $\frac{z_{2.06\alpha}}{\sqrt{\log 1/\alpha}} \ge 0.61$.

□

Proof of Lemma C.8. In this proof, we extend the meaning of operator $\max\{\cdot,\cdot\}$ to allow function-value arguments. Suppose f and g are two functions, then $\max\{f, g\} := t \mapsto \max\{f(t), g(t)\}$.

For μ that will be specified later, define $x_l = \arg\min\{t \in [0,1] : f(t) \le M(f) + \mu\}, x_r = \arg\max\{t \in [0,1] : f(t) \ge M(f) + \mu\}$. We will construct several functions. Without loss of generality, we assume $x_r + x_l \ge 2Z(f)$. Otherwise, we construct those functions on the left side. As shown in the Figure 22, the function in bold is f, and the following points have the following coordinates:

$$F : (Z(f), M(f)) \quad A : (x_l, M(f)+\mu) \quad D : (x_r, M(f)+\mu) \quad N : (x_l, M(f)+2\mu) \tag{D.7}$$

Define four linear functions L_0, L_1, L_2, L_3:

$$
\begin{aligned}
L_0(t) &= M(f) + \mu && (AD), \\
L_1(t) &= \begin{cases} M(f) + (t - Z(f))\frac{\mu}{x_l - Z(f)}, & x_l \ne Z(f) \\ M(f), & x_l = Z(f) \end{cases} && (AF), \\
L_2(t) &= M(f) + (t - Z(f))\frac{\mu}{x_r - Z(f)} && (FD), \\
L_3(t) &= M(f) + \mu + (t - x_r)\frac{\mu}{x_l - x_r} && (ND).
\end{aligned} \tag{D.8}
$$

Define the following functions:

(D.9)

$$g_1 = \max(f, L_0),\ g_2 = \max(f, L_3),\ g_3 = \max(L_1, L_2, L_0),\ g_4 = \max(L_1, L_2).$$

Note that the above definition is valid for all $\mu > 0$ as it does not require A and D to be on the graph of f. Therefore, when μ goes from 0^+ to ∞, $\|g_1 - f\|$ and $\|g_2 - f\|$ also go from 0^+ to ∞. But note that $\|g_2 - f\|$, as a function of μ, may not be monotonic, nor continuous. As μ increases, g_2 may jump between the right and the left side. A jump can incur a sudden increase or decrease. On each small chunk, $\|g_2 - f\|$ is monotonic and continuous.

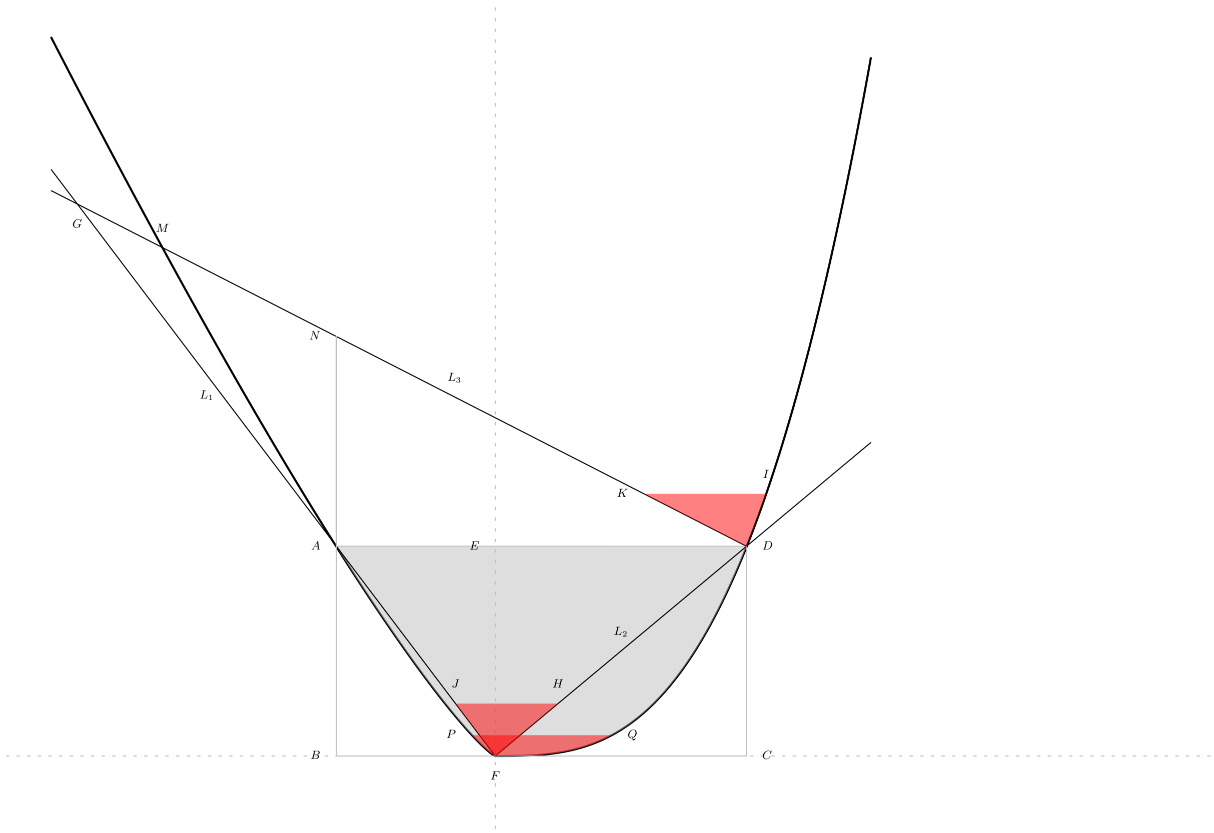


Fig 22: Illustration figure for proof of lemma C.8

Therefore, for any given $\sigma > 0$, there are two possible cases. Either $\exists \mu > 0$, s.t. $\|g_2 - f\| = \sigma$, or $\exists \mu$ such that the following three things hold.

Property 1 $x_l(\mu) + x_r(\mu) = 2Z(f)$.

Property 2 Suppose $g_{2,l}$ and $g_{2,r}$ are constructed essentially in the same way as g_2 but one on the left side ($g_{2,l}$) and one on the right side ($g_{2,r}$). Then $(\|g_{2,l} - f\| - \sigma) \cdot (\|g_{2,r} - f\| - \sigma) < 0$.

Property 3 And further, for the side ($h \in \{l, r\}$) that $\|g_{2,h} - f\| - \sigma < 0$, $\exists \mu > \tau_h > 0$ such that for any $\tau \in (0, \tau_h)$,

$$|x_h(\mu - \tau) - Z(f)| \geq \frac{|x_l(\mu - \tau) - Z(f)| + |x_r(\mu - \tau) - Z(f)|}{2}.$$

And for the other side $\tilde{h} \in \{l, r\}/\{h\}$, $\|g_{2,\tilde{h}} - f\| - \sigma > 0$, $\exists \mu > \tau_{\tilde{h}} > 0$ such that for any $\tau \in (0, \tau_{\tilde{h}})$,

$$|x_{\tilde{h}}(\mu + \tau) - Z(f)| \geq \frac{|x_l(\mu + \tau) - Z(f)| + |x_r(\mu + \tau) - Z(f)|}{2}.$$

To show the main idea more clearly, we assume for the moment that for the σ that will be chosen later, there exists a μ such that on at least one side, we have $\|g_2 - f\| = \sigma$ and use σ to denote $\|g_2 - f\|$. For the σ that does not have a corresponding μ, we will discuss it later.

Now we will introduce several inequalities with respect to $\|g_2 - f\|$, $\|g_1 - f\|$ and $\|g_3 - g_4\|$.

1.

$$\|g_2 - f\|^2 \leq 5\|g_1 - f\|^2. \tag{D.10}$$

When $Z(f) \neq x_l$, we have

$$\begin{aligned}
\|g_2 - f\|^2 &\leq \frac{1}{3}\mu^3 \frac{1}{\frac{\mu}{Z(f) - x_l} - \frac{\mu}{x_r - x_l}} + 2 \times \frac{1}{3}(x_r - x_l) \times \mu^2 + 2 \times \|g_1 - f\|^2 \\
&= \frac{1}{3}\mu^2 \frac{(Z(f) - x_l)(x_r - x_l)}{x_r - Z(f)} + \frac{2}{3}\mu^2(x_r - x_l) + 2\|g_1 - f\|^2 \\
&\leq \mu^2(x_r - x_l) + 2 \times \|g_1 - f\|^2 \leq 5\|g_1 - f\|^2.
\end{aligned}$$

Otherwise, the first term is zero, we still have Inequality (D.10).

2.

$$\|g_1 - f\|^2 \leq 3\|g_3 - g_4\|^2. \tag{D.11}$$

This follows from

$$\|g_3 - g_4\|^2 = \frac{1}{3}\mu^2(x_r - x_l) \geq \frac{1}{3}\|g_1 - f\|^2.$$

3.

$$\|g_2 - f\|^2 \leq 8\|g_3 - g_4\|^2. \tag{D.12}$$

When $x_l \neq Z(f)$, we have

$$\begin{aligned}
&\|g_2 - f\|^2 \\
&\leq \frac{1}{3}\mu^3 \frac{1}{\frac{\mu}{Z(f)-x_l} - \frac{\mu}{x_r - x_l}} + \frac{1}{3}(x_r - x_l) \times \mu^2 + \|g_1 - f\|^2 + 2 \times \mu^2 \times \frac{1}{2}(x_r - x_l) \\
&= \frac{1}{3}\mu^2 \frac{4x_r - x_l - 3Z(f)}{x_r - Z(f)}(x_r - x_l) + \|g_1 - f\|^2 \\
&\leq \frac{5}{3}\mu^2(x_r - x_l) + \|g_1 - f\|^2 \leq 8\|g_3 - g_4\|^2.
\end{aligned}$$

Otherwise, the first term is zero and Inequality (D.12) still holds.

Define linear function $g_5 = \max\{L_3, L_2\}$, then we know that

$$\rho_z(\gamma; g_2) \leq \rho_z(\gamma; g_5), \quad \forall \gamma > 0. \tag{D.13}$$

Now we will show that

$$\rho_z(\gamma; g_5) \leq 2\rho_z(\gamma; g_4), \tag{D.14}$$

for $\gamma \leq \sqrt{\frac{1}{3}\mu^2(x_r - x_l)} = \|g_3 - g_4\|$. When $\gamma \leq \|g_3 - g_4\|$, elementary calculation gives the followings:

$$\begin{aligned}
&\rho_z(\gamma; g_4) = \frac{x_r - Z(f)}{x_r - x_l}(3\frac{(x_r - x_l)^2}{\mu^2}\gamma^2)^{1/3} \geq \frac{1}{2}(3\frac{(x_r - x_l)^2}{\mu^2}\gamma^2)^{1/3}, \\
&\rho_z(\gamma; g_5) \leq (3\frac{(x_r - x_l)^2}{\mu^2}\gamma^2)^{1/3},
\end{aligned}$$

which give Inequality (D.14).

Therefore, Inequality (D.13) and Inequality (D.14) give

$$\rho_z(\gamma; g_2) \leq 2\rho_z(\gamma; g_4), \quad \forall \gamma \in (0, \|g_3 - g_4\|). \tag{D.15}$$

Further, for all $\gamma > 0$, we have

(D.16)

$$\begin{aligned}
\rho_z(\gamma; g_4) &= \left(\frac{\gamma}{\|g_3 - g_4\|}\right)^{\frac{2}{3}}(x_r - Z(f)) \leq \left(\frac{\sqrt{8}\gamma}{\|g_2 - f\|}\right)^{\frac{2}{3}}(x_r - Z(f)) \\
&= \left(\frac{\sqrt{8}\gamma}{\|g_2 - f\|}\right)^{\frac{2}{3}}|Z(g_2) - Z(f)| = \left(\frac{\sqrt{8}\gamma}{\sigma}\right)^{\frac{2}{3}}|Z(g_2) - Z(f)|.
\end{aligned}$$

Therefore, we have

$$\rho_z(\gamma; g_2) \le 4(\frac{\gamma}{\sigma})^{\frac{2}{3}}|Z(g_2) - Z(f)|, \qquad \forall \gamma \le \|g_3 - g_4\|. \tag{D.17}$$

Further we have

$$|Z(g_2) - Z(f)| = \sup\{|t - Z(f)| : g_1(t) = M(g_1)\} \ge \rho_z(\frac{1}{\sqrt{5}}\sigma; f). \tag{D.18}$$

The σ we will specify later is no smaller than $\sqrt{8}\varepsilon$, and suppose $\sigma \ge \sqrt{8}\varepsilon$ from now. This gives two consequences.

1. $|Z(g_2) - Z(f)| \ge \rho_z(\frac{1}{\sqrt{5}}\sigma; f) \ge \rho_z(\varepsilon; f)$.
2. $\|g_3 - g_4\|^2 \ge \frac{1}{8} \times 8\varepsilon^2$. By Inequality (D.15), this further implies $\rho_z(\varepsilon; g_2) \le 4(\frac{\varepsilon}{\sigma})^{\frac{2}{3}}|Z(g_2) - Z(f)|$.

As we know, for the problem of estimation $Z(h)$ with $h \in \{g_2, f\}$, the following statistic is sufficient

$$WS = \frac{\int_0^1 (g_2(t) - f(t))dY(t) - \frac{1}{2}\int_0^1 (g_2(t)^2 - f(t)^2)dt}{\varepsilon\|g_2 - f\|}, \tag{D.19}$$

and we have $WS \sim N(\theta(h)\frac{\|g_2 - f\|}{2\varepsilon}, 1)$, with $\theta(g_2) = 1, \theta(f) = -1$.

Define an event $O = \{|\hat{Z} - Z(f)| > \frac{1}{2}\rho_z(\varepsilon; f)\}$, then we have $P_f(O) \le 2c$. This is because we have $\mathbb{E}_f|\hat{Z} - Z(f)| \le c\rho_z(\varepsilon; f)$. Therefore, by arguments simliar to Neyman–Pearson lemma, we have $P_{g_2}(O) \le \Phi(\frac{\|g_2 - f\|}{\varepsilon} - \Phi^{-1}(1-2c))$. Since $|Z(g_2) - Z(f)| \ge \rho_z(\varepsilon; f)$ and $|\hat{Z} - Z(g_2)| \ge |Z(g_2) - Z(f)| - |\hat{Z} - Z(f)|$, we have the following inequalities

$$\begin{aligned} \mathbb{E}_{g_2}|\hat{Z} - Z(g_2)| &\ge \mathbb{E}_{g_2}\Big((|Z(g_2) - Z(f)| - |\hat{Z} - Z(f)|)_+\Big) \\ &\ge \mathbb{E}_{g_2}\left(\mathbb{1}\{O^c\}\left(|Z(g_2) - Z(f)| - \frac{1}{2}\rho_z(\varepsilon; f)\right)\right) \\ &\ge \Phi(\Phi^{-1}(1 - 2c) - \frac{\|g_2 - f\|}{\varepsilon})\frac{1}{2}|Z(g_2) - Z(f)|. \end{aligned} \tag{D.20}$$

For $c \le 0.0011$, let $\sigma = \Phi^{-1}(1-2c)\varepsilon$. Then $\sigma > \sqrt{8}\varepsilon$, thus $|Z(g_2) - Z(f)| \ge \rho_z(\varepsilon; f)$ and $\rho_z(\varepsilon; g_2) \le 4(\frac{\varepsilon}{\sigma})^{\frac{2}{3}}|Z(g_2) - Z(f)|$.

So we have

$$\mathbb{E}_{g_2}|\hat{Z} - Z(g_2)| \ge \frac{1}{4}|Z(g_2) - Z(f)| \ge \frac{1}{16}\Phi^{-1}(1 - 2c)^{\frac{2}{3}}\rho_z(\varepsilon; g_2). \tag{D.21}$$

Let $f_1 = g_2$, we have the result.

Now we consider the case when $\sigma = \Phi^{-1}(1-2c)\varepsilon$ does not have a corresponding μ. Then $\exists \mu > 0$ such that Property 1, Property 2 and Property 3 hold.

Without loss of generality, we assume h defined in Property 3 is r. Then Property 2 and Property 3 give that $\|g_{2,l} - f\| > \sigma$. By Property 1,

$$|Z(g_{2,r}) - Z(f)| = |Z(g_{2,l}) - Z(f)|. \tag{D.22}$$

Besides $g_{2,l}$ and $g_{2,r}$, we can construct $g_{1,l}$, $g_{3,l}$, $g_{4,l}$, $g_{5,l}$ similarly to g_1, g_3, g_4, g_5 on the left hand side, and also $g_{1,r}$, $g_{3,r}$, $g_{4,r}$, $g_{5,r}$ on the right hand side. Then we know that

$$g_{1,l} = g_{1,r} = g_1, \quad g_{3,l} = g_{3,r} = g_3, \quad g_{4,l} = g_{4,r} = g_4. \tag{D.23}$$

According to Inequality (D.12), we have $\|g_3 - g_4\|^2 \geq \frac{1}{8}\|g_{2,l} - f\|^2 \geq \varepsilon^2$. Therefore, we have

$$\begin{aligned} |Z(g_{2,r}) - Z(f)| &\overset{(i)}{=} |Z(g_{2,l}) - Z(f)| \\ &\overset{(ii)}{\geq} \left(\frac{\|g_{2,l} - f\|}{\sqrt{8}\varepsilon}\right)^{\frac{2}{3}} \rho_z(\varepsilon; g_{4,l}) \overset{(iii)}{\geq} \frac{1}{2}\Phi^{-1}(1-2c)^{\frac{2}{3}}\rho_z(\varepsilon; g_{4,l}) \\ &\overset{(iv)}{=} \frac{1}{2}\Phi^{-1}(1-2c)^{\frac{2}{3}}\rho_z(\varepsilon; g_{4,r}) \overset{(v)}{\geq} \frac{1}{4}\Phi^{-1}(1-2c)^{\frac{2}{3}}\rho_z(\varepsilon; g_{2,r}). \end{aligned} \tag{D.24}$$

Step (i) follows from Equation (D.22), step (ii) follows from Inequality (D.16), step (iii) follows from $\|g_{2,l} - f\| > \sigma = \Phi^{-1}(1-2c)\varepsilon$, step (iv) follows from Equation (D.23), and step (v) follows from Inequality (D.15).

Again, since $\sigma > \sqrt{5}\varepsilon$, we have $|Z(g_{2,r}) - Z(f)| = |Z(g_{2,l}) - Z(f)| \geq \rho_z(\varepsilon; f)$, which comes from (D.18).

Similar to the arguments in the case of g_2, we define event $O = \{|\hat{Z} - Z(f)| > \frac{1}{2}\rho_z(\varepsilon; f)\}$, then we have $P_f(O) \leq 2c$. And we have

$$\begin{aligned} \mathbb{E}_{g_{2,r}}|\hat{Z} - Z(g_{2,r})| &\geq \mathbb{E}_{g_{2,r}}\left((|Z(g_{2,r}) - Z(f)| - |\hat{Z} - Z(f)|)_+\right) \\ &\geq \mathbb{E}_{g_{2,r}}\left(\mathbb{1}\{O^c\}\left(|Z(g_{2,r}) - Z(f)| - \frac{1}{2}\rho_z(\varepsilon; f)\right)\right) \\ &\geq \Phi(\Phi^{-1}(1-2c) - \frac{\|g_{2,r} - f\|}{\varepsilon})\frac{1}{2}|Z(g_{2,r}) - Z(f)| \\ &\geq \frac{1}{16}\Phi^{-1}(1-2c)^{\frac{2}{3}}\rho_z(\varepsilon; g_{2,r}). \end{aligned} \tag{D.25}$$

We take $f_1 = g_{2,r}$ and get the statement.

□

Proof of Lemma C.9. Without loss of generality, we assume $f(Z(f)+\rho_z(\varepsilon;f)) \le M(f)+\rho_m(\varepsilon;f)$. Denote $x_l = \min\{t : f(t) \le M(f)+\rho_m(\varepsilon;f)\}$.

For $0<\delta<\frac{1}{2}\rho_z(\varepsilon;f))$, denote

$$g_\delta(t) = \max\big\{f(t),\\ M(f)+\rho_m(\varepsilon;f)+\frac{f(Z(f)+\rho_z(\varepsilon;f)-\delta)-M(f)-\rho_m(\varepsilon;f)}{\rho_z(\varepsilon;f)+Z(f)-x_l-\delta}(t-x_l)\big\}.$$

Clearly, $\|g_\delta - f\| \le \varepsilon$, $Z(g_\delta) = Z(f)+\rho_z(\varepsilon;f)-\delta$, and $\rho_z(\varepsilon;g_\delta) \le 3\rho_z(\varepsilon;f)$.

Define event O to be $O = \{|\hat{Z}-Z(f)| \ge \frac{1}{2}\rho_z(\varepsilon;f)\}$. Then $P_f(O) \le 2c$, thus $P_{g_\delta}(O) \le \Phi(1+\Phi^{-1}(2c))$.

Therefore,

$$\begin{aligned}\mathbb{E}_{g_\delta}|\hat{Z}-Z(g_\delta)| &\ge \mathbb{E}_{g_\delta}\left(\mathbb{1}\{O^c\}(|Z(f)-Z(g_\delta)|-\frac{1}{2}\rho_z(\varepsilon;f))_+\right)\\ &\ge P_{g_\delta}(O^c)(\rho_z(\varepsilon;f)-\delta-\frac{1}{2}\rho_z(\varepsilon;f))\\ &\ge (1-\Phi(1+\Phi^{-1}(2c)))(\rho_z(\varepsilon;f)-\delta-\frac{1}{2}\rho_z(\varepsilon;f))\\ &\ge (1-\Phi(1+\Phi^{-1}(2c)))\left(\frac{1}{2}-\frac{\delta}{\rho_z(\varepsilon;f)}\right)_+\rho_z(\varepsilon;f)\\ &\ge (1-\Phi(1+\Phi^{-1}(2c)))\left(\frac{1}{2}-\frac{\delta}{\rho_z(\varepsilon;f)}\right)_+\frac{\rho_z(\varepsilon;g_\delta)}{3}.\end{aligned} \tag{D.26}$$

Therefore,

$$\begin{aligned}&\sup_{\frac{1}{2}\rho_z(\varepsilon;f)>\delta>0}\frac{\mathbb{E}_{g_\delta}|\hat{Z}-Z(g_\delta)|}{\rho_z(\varepsilon;g_\delta)}\\ &\ge \limsup_{\delta\to 0^+}\frac{(1-\Phi(1+\Phi^{-1}(2c)))}{3}\left(\frac{1}{2}-\frac{\delta}{\rho_z(\varepsilon;f)}\right)_+\\ &= \frac{1}{6}\left(1-\Phi(1+\Phi^{-1}(2c))\right) > 0.1666\left(1-\Phi(1+\Phi^{-1}(2c))\right).\end{aligned} \tag{D.27}$$

Note that the inequality is strict, so we have the statement.

□

Proof of Lemma C.10. Without loss of generality, we can assume

$$t_r = \max\{t\in[0,1] : f(t) \le M(f)+\rho_m(\gamma;f)\} = Z(f)+\rho_z(\gamma;f),$$

for a $\gamma>0$ that we will specify later. Denote

$$t_l = \min\{t\in[0,1] : f(t) \le M(f)+\rho_m(\gamma;f)\} = Z(f)+\rho_z(\gamma;f).$$

It is apparent that t_r and t_l depend on γ. For $\frac{1}{4}\rho_z(\gamma; f) > \delta > 0$, define

$$g_\delta(\gamma; f) = \max\{f, M(f) + \rho_m(\gamma; f) + \frac{f(t_r - \delta) - M(f) - \rho_m(\gamma; f)}{t_r - \delta - t_l}(t - t_l)\}.$$

Therefore, we know that $\|g_\delta(\gamma; f) - f\| \leq \gamma$. We will use g to refer to $g_\delta(\gamma; f)$ when there is no ambiguity. According to the definition, we know that $\limsup_{\delta\to 0^+} \rho_m(\gamma; g) \leq \rho_m(\gamma; f)$. We will specify γ to be a quantity no smaller than ε, suppose $\gamma \geq \varepsilon$ from now.

Denote $O = \{|\hat{M} - M(f)| > \frac{1}{2}\rho_m(\varepsilon; f)\}$. Since $\mathbb{E}_f|\hat{M} - M(f)| \leq c\rho_m(\varepsilon; f)$, we have $P_f(O) \leq 2c$, then we have

$$\begin{aligned}
&\mathbb{E}_g|\hat{M} - M(g)| \\
&\geq \mathbb{E}_g \left(\mathbb{1}\{O^c\}(|M(f) - M(g)| - |\hat{M} - M(f)|)_+\right) \\
&\geq P_g(O^c)(|M(f) - M(g)| - \frac{1}{2}\rho_m(\varepsilon; f))_+ \\
&\geq \Phi(\Phi^{-1}(1 - 2c) - \frac{\gamma}{\varepsilon}) \left(|M(f) - M(g)| - \frac{1}{2}\rho_m(\varepsilon; f)\right)_+ \\
&= \Phi(\Phi^{-1}(1 - 2c) - \frac{\gamma}{\varepsilon}) \left(\rho_m(\gamma; f) - \frac{1}{2}\rho_m(\varepsilon; f) + f(t_r - \delta) - f(t_r)\right)_+ .
\end{aligned}$$

For $c \leq 0.103$, let $\gamma = \max\{\Phi^{-1}(1 - 2c)\varepsilon, \varepsilon\}$. Then $\gamma \geq \varepsilon$.

Therefore, we have

$$\begin{aligned}
&\sup_{0<\delta<\frac{1}{4}\rho_z(\gamma;f)} \frac{\mathbb{E}_g|\hat{M} - M(g)|}{\rho_m(\varepsilon; g)} \geq \limsup_{\delta\to 0^+} \frac{\mathbb{E}_g|\hat{M} - M(g)|}{\rho_m(\varepsilon; g)} \\
&\geq \limsup_{\delta\to 0^+} \frac{\Phi(z_{2c} - \max\{z_{2c}, 1\})(\left(\frac{\gamma}{\varepsilon}\right)^{\frac{2}{3}} \rho_m(\varepsilon; f) - \frac{1}{2}\rho_m(\varepsilon; f) + f(t_r - \delta) - f(t_r))_+}{\rho_m(\varepsilon; g)} \\
&\geq \frac{\Phi(z_{2c} - \max\{z_{2c}, 1\}) \left(\left(\frac{\gamma}{\varepsilon}\right)^{\frac{2}{3}} \rho_m(\varepsilon; f) - \frac{1}{2}\rho_m(\varepsilon; f)\right)}{\rho_m(\varepsilon; f)} \\
&= \Phi(z_{2c} - \max\{z_{2c}, 1\}) \left(\left(\frac{\gamma}{\varepsilon}\right)^{\frac{2}{3}} - \frac{1}{2}\right).
\end{aligned}$$

For $0.103 \geq c \geq \frac{\Phi(-1)}{2}$, we have

$$\sup_{g\in\mathcal{F}} \frac{\mathbb{E}_g|\hat{M} - M(g)|}{\rho_m(\varepsilon; g)} \geq \frac{\Phi(z_{2c} - 1)}{2} > 0.214362. \tag{D.28}$$

For $c < \frac{\Phi(-1)}{2}$, we have

$$\sup_{g\in\mathcal{F}} \frac{\mathbb{E}_g|\hat{M} - M(g)|}{\rho_m(\varepsilon; g)} \geq \frac{1}{2}\left(z_{2c}^{\frac{2}{3}} - \frac{1}{2}\right) > \frac{z_{2c}^{\frac{2}{3}}}{4}. \tag{D.29}$$

Note that for both cases, the inequality is strict, so we have the statement. $\square$

Proof of Lemma C.15. Without loss of generality, we assume

$$\sup\{t > Z(f) : f(t) \leq \rho_m(\varepsilon; f) + M(f)\} = \rho_z(\varepsilon; f) + Z(f). \tag{D.30}$$

Splitting the entire probability space into $\{\hat{j} \leq j^* + 1\}$ and $\{\hat{j} \geq j^* + 2\}$ gives

$$\begin{aligned}
&\mathbb{E}_{l,s}\big((\hat{f} - M(f))^2 \mathbb{1}\{\tilde{j} > \hat{j}\}\big) \\
&= \underbrace{\mathbb{E}_{l,s}\big(\sum_{j_1=2}^{j^*+1} (\hat{f} - M(f))^2 \mathbb{1}\{\hat{j} = j_1, \tilde{j} \geq j_1 + 1\}\big)}_{\eta_1} \\
&\quad + \underbrace{\mathbb{E}_{l,s}\big(\sum_{j_1=j^*+2}^{\infty} (\hat{f} - M(f))^2 \mathbb{1}\{\hat{j} = j_1, \tilde{j} \geq j_1 + 1\}\big)}_{\eta_2}
\end{aligned} \tag{D.31}$$

We have the following bounds that we will prove separately

$$\eta_1 \leq (7680V + 2)\rho_m(\varepsilon; f)^2, \qquad \eta_2 \leq (78V + \frac{1}{16})\rho_m(\varepsilon; f)^2, \tag{D.32}$$

which gives the statement of the lemma.

Proof of bound on η_1 in Inequality (D.32). Splitting the entire probability space by the value of Δ (i.e., $\Delta = 2$, $\Delta = -2$, $\Delta = 0$) gives

$$\begin{aligned}
\eta_1 =& \mathbb{E}_{l,s}\big(\sum_{j_1=2}^{j^*+1} (\hat{f} - M(f))^2 \mathbb{1}\{\tilde{j} \geq j_1 + 1, \hat{j} = j_1\}\big) \\
\leq& \sum_{j_1=2}^{j^*+1} \mathbb{E}_{l,s}\big((\mu_{j_1,\hat{i}_{j_1}+2} - M(f))^2 \mathbb{1}\{\tilde{j} \geq j_1 + 1, \hat{j} = j_1, \Delta = 2\}\big) \\
&+ \sum_{j_1=2}^{j^*+1} \mathbb{E}_{l,s}\big((\mu_{j_1,\hat{i}_{j_1}-2} - M(f))^2 \mathbb{1}\{\tilde{j} \geq j_1 + 1, \hat{j} = j_1, \Delta = -2\}\big) \\
&+ \underbrace{\sum_{j_1=2}^{j^*+1} \mathbb{E}_{l,s}\big((\mu_{j_1,\hat{i}_{j_1}} - M(f))^2 \mathbb{1}\{\tilde{j} \geq j_1 + 1, \hat{j} = j_1, \Delta = 0\}\big)}_{\xi}.
\end{aligned} \tag{D.33}$$

By convexity of f, we can further bound ξ in Inequality (D.33) by

$$\xi \leq \sum_{j_1=2}^{j^*+1} \mathbb{E}_{l,s}\Big(\frac{(\mu_{j_1,\hat{i}_{j_1}+2} - M(f))^2 + (\mu_{j_1,\hat{i}_{j_1}-2} - M(f))^2}{2}\mathbb{1}\{\tilde{j} \geq j_1+1, \hat{j} = j_1, \Delta = 0\}\Big).$$

Plugging this bound of ξ back into Inequality (D.33) gives

$$\begin{aligned}\eta_1 \leq & \sum_{j_1=2}^{j^*+1} \underbrace{\mathbb{E}_{l,s}\big((\mu_{j_1,\hat{i}_{j_1}+2} - M(f))^2 \mathbb{1}\{\tilde{j} \geq j_1+1, \hat{j} = j_1, \Delta \in \{0,2\}\}\big)}_{\kappa_1(j_1)} \\ & + \sum_{j_1=2}^{j^*+1} \underbrace{\mathbb{E}_{l,s}\big((\mu_{j_1,\hat{i}_{j_1}-2} - M(f))^2 \mathbb{1}\{\tilde{j} \geq j_1+1, \hat{j} = j_1, \Delta \in \{-2,0\}\}\big)}_{\kappa_2(j_1)}.\end{aligned} \tag{D.34}$$

Now we will bound $\kappa_1(j_1)$ and $\kappa_1(j_2)$ for $j_1 \leq j^*$. Before we bound for general $j_1 \leq j^*$, we list two special cases for $\kappa_1(j_1)$. By assumption (D.30), for $j_1 = j^*$ and $j_1 = j^*+1$ in the first term, we have

$$\kappa_1(j_1) \leq \rho_m(\varepsilon; f)^2 2^{2j^*-2j_1}. \tag{D.35}$$

For general $j_1 \leq j^*$, simplifying the event $\{\Delta \in \{0,2\}\}$ and taking conditional expectation with respect to Y_l gives

$$\begin{aligned}\kappa_1(j_1) \leq & \mathbb{E}_{l,s}\big((\mu_{j_1,\hat{i}_{j_1}+2} - M(f))^2 \mathbb{1}\{\tilde{j} \geq j_1+1, \frac{\tilde{X}_{j_1,\hat{i}_{j_1}+6} - \tilde{X}_{j_1,\hat{i}_{j_1}+5}}{\sqrt{2}c_s\sqrt{m_{j_1}}\varepsilon} \leq 2\}\big) \\ \leq & \mathbb{E}_l\Bigg((\mu_{j_1,\hat{i}_{j_1}+2} - M(f))^2 \mathbb{1}\{\tilde{j} \geq j_1+1\} \\ & \quad \mathbb{E}_s\big(\mathbb{1}\{\mathcal{E}_{j_1,\hat{i}_{j_1}+6}\frac{1}{\sqrt{2}c_s\varepsilon} \leq 2 - \mu_{j_1,\hat{i}_{j_1}+6}\frac{\sqrt{m_{j_1}}}{\sqrt{2}c_s\varepsilon} + \mu_{j_1,\hat{i}_{j_1}+5}\frac{\sqrt{m_{j_1}}}{\sqrt{2}c_s\varepsilon}|Y_l\big)\Bigg), \\ \kappa_2(j_2) \leq & \mathbb{E}_{l,s}\big(\sum_{j_1=2}^{j^*+1}(\mu_{j_1,\hat{i}_{j_1}-2} - M(f))^2 \mathbb{1}\{\tilde{j} \geq j_1+1, \frac{\tilde{X}_{j_1,\hat{i}_{j_1}-6} - \tilde{X}_{j_1,\hat{i}_{j_1}-5}}{\sqrt{2}c_s\sqrt{m_{j_1}}\varepsilon} \leq 2\}\big) \\ \leq & \mathbb{E}_l\Bigg((\mu_{j_1,\hat{i}_{j_1}-2} - M(f))^2 \mathbb{1}\{\tilde{j} \geq j_1+1\} \\ & \quad \mathbb{E}_s\big(\mathbb{1}\{-\mathcal{E}_{j_1,\hat{i}_{j_1}-5}\frac{1}{\sqrt{2}c_s\varepsilon} \leq 2 - \mu_{j_1,\hat{i}_{j_1}-6}\frac{\sqrt{m_{j_1}}}{\sqrt{2}c_s\varepsilon} + \mu_{j_1,\hat{i}_{j_1}-5}\frac{\sqrt{m_{j_1}}}{\sqrt{2}c_s\varepsilon}|Y_l\big)\Bigg).\end{aligned}$$

Now we will bound $\mu_{j_1,\hat{i}_{j_1}-6} - \mu_{j_1,\hat{i}_{j_1}-5}$ by an expression of $\mu_{j_1,\hat{i}_{j_1}-2} - M(f)$. As we have $|\hat{i}_{j_1} - i^*_{j_1}| \le 1$, we have $i^*_{j_1} - 3 \le \hat{i}_{j_1} - 2 \le i^*_{j_1} - 1$. We have

$$\mu_{j_1,\hat{i}_{j_1}-6} - \mu_{j_1,\hat{i}_{j_1}-5} \ge m_{j_1}\frac{f(t_{j_1,\hat{i}_{j_1}-6}) - M(f)}{t_{j_1,\hat{i}_{j_1}-6} - Z(f)} \ge m_{j_1}\frac{f(t_{j_1,\hat{i}_{j_1}-3}) - M(f)}{t_{j_1,\hat{i}_{j_1}-3} - Z(f)}$$
$$\ge m_{j_1}\frac{\mu_{j_1,\hat{i}_{j_1}-2} - M(f)}{4m_{j_1}} \ge \frac{1}{4}(\mu_{j_1,\hat{i}_{j_1}-2} - M(f)).$$

Similarly we have

$$\mu_{j_1,\hat{i}_{j_1}+6} - \mu_{j_1,\hat{i}_{j_1}+5} \ge \frac{1}{4}(\mu_{j_1,\hat{i}_{j_1}+2} - M(f)).$$

Now we plug these bounds back to bound $\kappa_2(j_1)$ and $\kappa_1(j_1)$. Since similar analysis goes for both $\kappa_2(j_1)$ and $\kappa_1(j_1)$, we only showcase the analysis for $\kappa_2(j_1)$ in detail.

$$\begin{aligned}
&\kappa_2(j_1) \\
\le&\mathbb{E}_l\big((\mu_{j_1,\hat{i}_{j_1}-2} - M(f))^2\mathbb{1}\{\tilde{j} \ge j_1+1\} \\
&\mathbb{E}_s\big(\mathbb{1}\{-\mathcal{E}_{j_1,\hat{i}_{j_1}-5}\frac{1}{\sqrt{2}c_s\varepsilon} \le 2 - \frac{1}{4}(\mu_{j_1,\hat{i}_{j_1}-2} - M(f))\frac{\sqrt{m_{j_1}}}{\sqrt{2}c_s\varepsilon}|Y_l)\big) \\
=&\mathbb{E}_l\big((\mu_{j_1,\hat{i}_{j_1}-2} - M(f))^2\mathbb{1}\{\tilde{j} \ge j_1+1\}\Phi(2 - (\mu_{j_1,\hat{i}_{j_1}-2} - M(f))2^{\frac{j^*-j_1-4}{2}}\frac{\sqrt{m_{j^*}}}{\sqrt{2}c_s\varepsilon})\big) \\
\le&\mathbb{E}_l\big(2^{4+j_1-j^*}\frac{2c_s^2\varepsilon^2}{m_{j^*}}\mathbb{1}\{\tilde{j} \ge j_1+1\} \\
&[2^{\frac{j^*-j_1-4}{2}}\frac{\sqrt{m_{j^*}}}{2\sqrt{2}c_s\varepsilon}(\mu_{j_1,\hat{i}_{j_1}-2} - M(f))]^2\Phi(2 - (\mu_{j_1,\hat{i}_{j_1}-2} - M(f))2^{\frac{j^*-j_1-4}{2}}\frac{\sqrt{m_{j^*}}}{2\sqrt{2}c_s\varepsilon})\big) \\
\le&2^{4+j_1-j^*}\frac{2c_s^2\varepsilon^2}{m_{j^*}}V,
\end{aligned} \tag{D.36}$$

where $V = \sup_{x\ge 0} x^2\Phi(2-x)$.

Similarly,

$$\kappa_1(j_1) \le 2^{4+j_1-j^*}\frac{2c_s^2\varepsilon^2}{m_{j^*}}V. \tag{D.37}$$

Plugging the bounds of $\kappa_1(j_1)$ and $\kappa_2(j_1)$ (Inequality (D.35), (D.37), (D.36)) back to Inequality (D.34) gives

(D.38)
$$
\begin{aligned}
\eta_1 &\le \sum_{j=1}^{j^*-1}(\kappa_1(j_1)+\kappa_2(j_1)) + \sum_{j=j^*}^{j^*+1}\kappa_2(j_1) + \sum_{j=j^*}^{j^*+1}\kappa_1(j_1)\\
&\le (3\times 2^{10}V + 3\times 2^{9}V + 3\times 2^{10}V + \frac{5}{4})\rho_m(\varepsilon;f)^2 \le (7680V+2)\rho_m(\varepsilon;f)^2.
\end{aligned}
$$

Proof of bound on η_2 in Inequality (D.32). Similar to the proof of bound on η_1, we split the entire probability space by the value of Δ, simplify the events, take conditional expectations, and calculate them to arrive at the bound. Details are as follows.

$$
\begin{aligned}
&\mathbb{E}_{l,s}\big(\sum_{j_1=j^*+2}^{\infty}(\hat f - M(f))^2\mathbb{1}\{\hat j = j_1, \tilde j \ge j_1+1\}\big)\\
&\le \mathbb{E}_{l,s}\big(\sum_{j_1=j^*+2}^{\infty}(\mu_{j_1,\hat i_{j_1}+2} - M(f))^2\mathbb{1}\{\tilde j \ge j_1+1, \hat j = j_1,\\
&\qquad \frac{\tilde X_{j_1,\hat i_{j_1}+6} - \tilde X_{j_1,\hat i_{j_1}+5}}{\sqrt{2}c_s\sqrt{m_{j_1}}\varepsilon} \le 2, \frac{\tilde X_{j_1,\hat i_{j_1}-6} - \tilde X_{j_1,\hat i_{j_1}-5}}{\sqrt{2}c_s\sqrt{m_{j_1}}\varepsilon} > 2\}\big)\\
&\quad + \mathbb{E}_{l,s}\big(\sum_{j_1=j^*+2}^{\infty}(\mu_{j_1,\hat i_{j_1}-2} - M(f))^2\mathbb{1}\{\tilde j \ge j_1+1, \hat j = j_1,\\
&\qquad \frac{\tilde X_{j_1,\hat i_{j_1}-6} - \tilde X_{j_1,\hat i_{j_1}-5}}{\sqrt{2}c_s\sqrt{m_{j_1}}\varepsilon} \le 2, \frac{\tilde X_{j_1,\hat i_{j_1}+6} - \tilde X_{j_1,\hat i_{j_1}+5}}{\sqrt{2}c_s\sqrt{m_{j_1}}\varepsilon} > 2\}\big)\\
&\quad + \mathbb{E}_{l,s}\big(\sum_{j_1=j^*+2}^{\infty}(\mu_{j_1,\hat i_{j_1}} - M(f))^2\mathbb{1}\{\tilde j \ge j_1+1, \hat j = j_1,\\
&\qquad \frac{\tilde X_{j_1,\hat i_{j_1}-6} - \tilde X_{j_1,\hat i_{j_1}-5}}{\sqrt{2}c_s\sqrt{m_{j_1}}\varepsilon} \le 2, \frac{\tilde X_{j_1,\hat i_{j_1}+6} - \tilde X_{j_1,\hat i_{j_1}+5}}{\sqrt{2}c_s\sqrt{m_{j_1}}\varepsilon} \le 2\}\big)\\
&\le \frac{1}{16}\rho_m(\varepsilon;f)^2 + \mathbb{E}_{l,s}\big(\sum_{j_1=j^*+2}^{\infty}(\mu_{j_1,\hat i_{j_1}-2} - M(f))^2\mathbb{1}\{\tilde j \ge j_1+1,\\
&\quad \frac{\tilde X_{j_1,\hat i_{j_1}-6} - \tilde X_{j_1,\hat i_{j_1}-5}}{\sqrt{2}c_s\sqrt{m_j}\varepsilon} \le 2, \forall j^*+1 \le j \le j_1, \frac{\tilde X_{j,\hat i_j+6} - \tilde X_{j,\hat i_j+5}}{\sqrt{2}c_s\sqrt{m_j}\varepsilon} > 2\}\big)\\
&\quad + \mathbb{E}_{l,s}\big(\sum_{j_1=j^*+2}^{\infty}(\mu_{j_1,\hat i_{j_1}} - M(f))^2\mathbb{1}\{\tilde j \ge j_1+1, \hat j = j_1, \frac{\tilde X_{j_1,\hat i_{j_1}-6} - \tilde X_{j_1,\hat i_{j_1}-5}}{\sqrt{2}c_s\sqrt{m_{j_1}}\varepsilon} \le 2,\\
&\quad \frac{\tilde X_{j_1,\hat i_{j_1}+6} - \tilde X_{j_1,\hat i_{j_1}+5}}{\sqrt{2}c_s\sqrt{m_{j_1}}\varepsilon} \le 2, \frac{\tilde X_{j,\hat i_j+6} - \tilde X_{j,\hat i_j+5}}{\sqrt{2}c_s\sqrt{m_j}\varepsilon} > 2, \forall j^*+1 \le j \le j_1-1\}\big)
\end{aligned}
$$

$$
\begin{aligned}
&\leq \mathbb{E}_l\Big(\sum_{j_1=j^*+2}^{\infty}(\mu_{j_1,\hat{i}_{j_1}-2}-M(f))^2\mathbb{1}\{\tilde{j}\geq j_1+1\}\mathbb{E}_s\big(\mathbb{1}\{\forall j^*+1\leq j\leq j_1,\\
&\quad \frac{\tilde{X}_{j,\hat{i}_j+6}-\tilde{X}_{j,\hat{i}_j+5}}{\sqrt{2}c_s\sqrt{m_j}\varepsilon}>2,\frac{\tilde{X}_{j_1,\hat{i}_{j_1}-6}-\tilde{X}_{j_1,\hat{i}_{j_1}-5}}{\sqrt{2}c_s\sqrt{m_{j_1}}\varepsilon}\leq 2\}|Y_l\big)\Big)+\frac{1}{16}\rho_m(\varepsilon;f)^2\\
&\quad+\mathbb{E}_l\Big(\sum_{j_1=j^*+2}^{\infty}(\mu_{j_1,\hat{i}_{j_1}}-M(f))^2\mathbb{1}\{\tilde{j}\geq j_1+1\}\mathbb{E}_s\big(\mathbb{1}\{\frac{\tilde{X}_{j_1,\hat{i}_{j_1}-6}-\tilde{X}_{j_1,\hat{i}_{j_1}-5}}{\sqrt{2}c_s\sqrt{m_{j_1}}\varepsilon}\leq 2,\\
&\quad \forall j^*+1\leq j\leq j_1-1,\frac{\tilde{X}_{j,\hat{i}_j+6}-\tilde{X}_{j,\hat{i}_j+5}}{\sqrt{2}c_s\sqrt{m_j}\varepsilon}>2,\frac{\tilde{X}_{j_1,\hat{i}_{j_1}+6}-\tilde{X}_{j_1,\hat{i}_{j_1}+5}}{\sqrt{2}c_s\sqrt{m_{j_1}}\varepsilon}\leq 2\}|Y_l\big)\Big)\\
&\leq \frac{1}{16}\rho_m(\varepsilon;f)^2+\mathbb{E}_l\Big(\sum_{j_1=j^*+2}^{\infty}(\mu_{j_1,\hat{i}_{j_1}-2}-M(f))^2\mathbb{1}\{\tilde{j}\geq j_1+1\}\\
&\quad \big[\Pi_{j=j^*+1}^{j_1}\Phi(-2+\frac{(\mu_{j,\hat{i}_j+6}-\mu_{j,\hat{i}_j+5})\sqrt{m_j}}{\sqrt{2}c_s\varepsilon})\big]\Phi(2-\frac{\sqrt{m_{j_1}}(\mu_{j_1,\hat{i}_{j_1}-6}-\mu_{j_1,\hat{i}_{j_1}-5})}{\sqrt{2}c_s\varepsilon})\Big)\\
&\quad+\mathbb{E}_l\Big(\sum_{j_1=j^*+2}^{\infty}(\mu_{j_1,\hat{i}_{j_1}}-M(f))^2\mathbb{1}\{\tilde{j}\geq j_1+1\}\Phi(2-\frac{\sqrt{m_{j_1}}(\mu_{j_1,\hat{i}_{j_1}-6}-\mu_{j_1,\hat{i}_{j_1}-5})}{\sqrt{2}c_s\varepsilon})\\
&\quad \Phi(2-\frac{\sqrt{m_{j_1}}(\mu_{j_1,\hat{i}_{j_1}+6}-\mu_{j_1,\hat{i}_{j_1}+5})}{\sqrt{2}c_s\varepsilon})\big[\Pi_{j=j^*+1}^{j_1-1}\Phi(-2+\frac{(\mu_{j,\hat{i}_j+6}-\mu_{j,\hat{i}_j+5})\sqrt{m_j}}{\sqrt{2}c_s\varepsilon})\big]\Big)\\
&\leq \frac{1}{16}\rho_m(\varepsilon;f)^2+\mathbb{E}_l\Big(\sum_{j_1=j^*+2}^{\infty}(\mu_{j_1,\hat{i}_{j_1}-2}-M(f))^2\mathbb{1}\{\tilde{j}\geq j_1+1\}\\
&\quad \big[\Pi_{j=j^*+1}^{j_1}\Phi(-2+\frac{\frac{\rho_m(\varepsilon;f)}{\rho_z(\varepsilon;f)}8m_j\sqrt{m_j}}{\sqrt{2}c_s\varepsilon})\big]\Phi(2-\frac{\sqrt{m_{j_1}}\frac{\mu_{j_1,\hat{i}_{j_1}-2}-M(f)}{4}}{\sqrt{2}c_s\varepsilon})\Big)\\
&\quad+\mathbb{E}_l\Big(\sum_{j_1=j^*+2}^{\infty}(\mu_{j_1,\hat{i}_{j_1}}-M(f))^2\mathbb{1}\{\tilde{j}\geq j_1+1\}\Phi(2-\frac{\sqrt{m_{j_1}}\frac{\mu_{j_1,\hat{i}_{j_1}}-M(f)}{2}}{\sqrt{2}c_s\varepsilon}))\\
&\quad \big[\Pi_{j=j^*+1}^{j_1-1}\Phi(-2+\frac{\frac{\rho_m(\varepsilon;f)}{\rho_z(\varepsilon;f)}8m_j\sqrt{m_j}}{\sqrt{2}c_s\varepsilon})\big]\Big)\\
&\leq \frac{1}{16}\rho_m(\varepsilon;f)^2+\mathbb{E}_l\Big(\sum_{j_1=j^*+2}^{\infty}\mathbb{1}\{\tilde{j}\geq j_1+1\}V\frac{32c_s^2\varepsilon^2}{m_{j_1}}\Phi(-1.75)^{j_1-j^*}\Big)\\
&\quad+\mathbb{E}_l\Big(\sum_{j_1=j^*+2}^{\infty}\mathbb{1}\{\tilde{j}\geq j_1+1\}V\frac{8c_s^2\varepsilon^2}{m_{j_1}}\Phi(-1.75)^{j_1-j^*-1}\Big)\\
&\leq \frac{1}{16}\rho_m(\varepsilon;f)^2\\
&\quad+\mathbb{E}_l\Big(\sum_{j_1=j^*+2}^{\infty}\mathbb{1}\{\tilde{j}\geq j_1+1\}V\times 32\times 3\times 8\frac{\varepsilon^2}{\rho_z(\varepsilon;f)}\times 2^{j_1-j^*}\Phi(-1.75)^{j_1-j^*}\Big)\\
&\quad+\mathbb{E}_l\Big(\sum_{j_1=j^*+2}^{\infty}\mathbb{1}\{\tilde{j}\geq j_1+1\}V\times 24\times 8\frac{\varepsilon^2}{\rho_z(\varepsilon;f)}\times 2^{j_1-j^*}\Phi(-1.75)^{j_1-j^*-1}\Big)\\
&< \frac{1}{16}\rho_m(\varepsilon;f)^2+\rho_m(\varepsilon;f)^2V\times 78.
\end{aligned}
$$

□

Proof of Lemma C.16. Note that $(\hat{f}-\mu_{\hat{j},\hat{i}_{\hat{j}}})_+$ takes non-zero value only when $\Delta \neq 0$ and $\mu_{\hat{j},\hat{i}_{\hat{j}}+\Delta} > \mu_{\hat{j},\hat{i}_{\hat{j}}}$. We split the entire probability space by the value of Δ, and then by the value of $\hat{j}$ and $\tilde{j}$. Then, we further take conditional expectation with respect to Y_l. Repetitive usage of convexity and careful calculation give the statement. The details are as follows.

$$
\begin{aligned}
&\mathbb{E}_{l,s}\Big(\big((\hat{f}-\mu_{\hat{j},\hat{i}_{\hat{j}}})_+\big)^2\mathbb{1}\{\tilde{j}\leq\hat{j}\}\Big) = \mathbb{E}_{l,s}\Big((\hat{f}-\mu_{\hat{j},\hat{i}_{\hat{j}}})^2\mathbb{1}\{\tilde{j}\leq\hat{j},\hat{f}>\mu_{\hat{j},\hat{i}_{\hat{j}}}\}\Big)\\
&\leq \mathbb{E}_{l,s}\Big((\mu_{\hat{j},\hat{i}_{\hat{j}}+2}-\mu_{\hat{j},\hat{i}_{\hat{j}}})^2\mathbb{1}\{\tilde{j}\leq\hat{j},\mu_{\hat{j},\hat{i}_{\hat{j}}+2}>\mu_{\hat{j},\hat{i}_{\hat{j}}},\frac{\tilde{X}_{\hat{j},\hat{i}_{\hat{j}}+6}-\tilde{X}_{\hat{j},\hat{i}_{\hat{j}}+5}}{\sqrt{2}\sqrt{m_{\hat{j}}}c_s\varepsilon}\leq 2,\\
&\qquad \frac{\tilde{X}_{\hat{j},\hat{i}_{\hat{j}-6}}-\tilde{X}_{\hat{j},\hat{i}_{\hat{j}-5}}}{\sqrt{2}\sqrt{m_{\hat{j}}}c_s\varepsilon}>2 \text{ if } \hat{i}_{\hat{j}-6}\geq 1\}\Big)\\
&\qquad +\mathbb{E}_{l,s}\Big((\mu_{\hat{j},\hat{i}_{\hat{j}}-2}-\mu_{\hat{j},\hat{i}_{\hat{j}}})^2\mathbb{1}\{\tilde{j}\leq\hat{j},\mu_{\hat{j},\hat{i}_{\hat{j}}-2}>\mu_{\hat{j},\hat{i}_{\hat{j}}},\frac{\tilde{X}_{\hat{j},\hat{i}_{\hat{j}}-6}-\tilde{X}_{\hat{j},\hat{i}_{\hat{j}}-5}}{\sqrt{2}\sqrt{m_{\hat{j}}}c_s\varepsilon}\leq 2,\\
&\qquad \frac{\tilde{X}_{\hat{j},\hat{i}_{\hat{j}+6}}-\tilde{X}_{\hat{j},\hat{i}_{\hat{j}+5}}}{\sqrt{2}\sqrt{m_{\hat{j}}}c_s\varepsilon}>2 \text{ if } t_{\hat{j},\hat{i}_{\hat{j}}+6}\leq 1\}\Big)\\
&\leq \sum_{j_1=2}^{\infty}\sum_{j_2=j_1}^{\infty}\Bigg(\mathbb{E}_{l,s}\Big((\mu_{j_2,\hat{i}_{j_2}+2}-\mu_{j_2,\hat{i}_{j_2}})^2\mathbb{1}\{\tilde{j}=j_1,\hat{j}=j_2,\frac{\tilde{X}_{j_2,\hat{i}_{j_2}+6}-\tilde{X}_{j_2,\hat{i}_{j_2}+5}}{\sqrt{2}\sqrt{m_{j_2}}c_s\varepsilon}\leq 2,\\
&\forall j^*+2\leq j\leq j_2-1,\frac{\tilde{X}_{j,\hat{i}_j+6}-\tilde{X}_{j,\hat{i}_j+5}}{\sqrt{2}\sqrt{m_j}c_s\varepsilon}>2,\frac{\tilde{X}_{j,\hat{i}_j-6}-\tilde{X}_{j,\hat{i}_j-5}}{\sqrt{2}\sqrt{m_j}c_s\varepsilon}>2,\mu_{j_2,\hat{i}_{j_2}+2}>\mu_{j_2,\hat{i}_{j_2}}\}\Big)\\
&\quad +\mathbb{E}_{l,s}\Big((\mu_{j_2,\hat{i}_{j_2}-2}-\mu_{j_2,\hat{i}_{j_2}})^2\mathbb{1}\{\tilde{j}=j_1,\hat{j}=j_2,\frac{\tilde{X}_{j_2,\hat{i}_{j_2}-6}-\tilde{X}_{j_2,\hat{i}_{j_2}-5}}{\sqrt{2}\sqrt{m_{j_2}}c_s\varepsilon}\leq 2,\\
&\forall j^*+2\leq j\leq j_2-1,\frac{\tilde{X}_{j,\hat{i}_j+6}-\tilde{X}_{j,\hat{i}_j+5}}{\sqrt{2}\sqrt{m_j}c_s\varepsilon}>2,\frac{\tilde{X}_{j,\hat{i}_j-6}-\tilde{X}_{j,\hat{i}_j-5}}{\sqrt{2}\sqrt{m_j}c_s\varepsilon}>2,\mu_{j_2,\hat{i}_{j_2}-2}>\mu_{j_2,\hat{i}_{j_2}}\}\Big)\Bigg)\\
&\leq \sum_{\substack{j_1\geq 2\\ j_2\geq j_1}}\Bigg(\mathbb{E}_l\Big((\mu_{j_2,\hat{i}_{j_2}+2}-\mu_{j_2,\hat{i}_{j_2}})^2\mathbb{1}\{\tilde{j}=j_1,\mu_{j_2,\hat{i}_{j_2}+2}>\mu_{j_2,\hat{i}_{j_2}}\}\Phi(2-\frac{\mu_{j_2,\hat{i}_{j_2}+2}-\mu_{j_2,\hat{i}_{j_2}}}{2}\frac{\sqrt{m_{j_2}}}{\sqrt{2}c_s\varepsilon})\\
&\qquad \Pi_{j=j^*+2}^{j_2-1}\max\{\Phi(-2),\Phi(-2+(\frac{7}{16}+\frac{6m_j}{\rho_z(\varepsilon;f)})\rho_m(\varepsilon;f)\frac{\sqrt{m_j}}{\sqrt{2}c_s\varepsilon})\}\Big)\\
&\quad +\mathbb{E}_l\Big((\mu_{j_2,\hat{i}_{j_2}-2}-\mu_{j_2,\hat{i}_{j_2}})^2\mathbb{1}\{\tilde{j}=j_1,\mu_{j_2,\hat{i}_{j_2}-2}>\mu_{j_2,\hat{i}_{j_2}}\}\Phi(2-\frac{\mu_{j_2,\hat{i}_{j_2}-2}-\mu_{j_2,\hat{i}_{j_2}}}{2}\frac{\sqrt{m_{j_2}}}{\sqrt{2}c_s\varepsilon})\\
&\qquad \Pi_{j=j^*+2}^{j_2-1}\max\{\Phi(-2),\Phi(-2+(\frac{7}{16}+\frac{6m_j}{\rho_z(\varepsilon;f)})\rho_m(\varepsilon;f)\frac{\sqrt{m_j}}{\sqrt{2}c_s\varepsilon})\}\Big)\Bigg).\\
&\leq \sum_{j_1=2}^{\infty}\sum_{j_2=j_1}^{\infty} 2\times\mathbb{E}_l\Big(\mathbb{1}\{\tilde{j}=j_1\}\frac{8c_s^2\varepsilon^2}{m_{j_2}}V\Phi(-1.85)^{(j_2-j^*-2)_+}\Big)\\
&\leq \sum_{j_1=2}^{\infty}\sum_{j_2=j_1}^{\infty} 2\times\mathbb{E}_l\Big(\mathbb{1}\{\tilde{j}=j_1\}\times 8c_s^2\times 2^{j_2-j^*+4}\rho_m(\varepsilon;f)^2V\Phi(-1.85)^{(j_2-j^*-2)_+}\Big)
\end{aligned}
$$

$$\leq \sum_{j_1=2}^{\infty}\sum_{j_2=j_1}^{\infty}\mathbb{E}_l\Big(\mathbb{1}\{\tilde{j}=j_1\}\Big)\times 2^{10}\times 3\times \rho_m(\varepsilon;f)^2 V 2^{j_2-j^*-2}\Phi(-1.85)^{(j_2-j^*-2)_+}$$

$$\leq \sum_{j_1=2}^{\infty}\mathbb{E}_l\Big(\mathbb{1}\{\tilde{j}=j_1\}\Big)\times 2^{10}\times 3\times \rho_m(\varepsilon;f)^2 V(2\times \mathbb{1}\{j_1\leq j^*+2\}+\frac{2\Phi(-1.85)}{1-2\Phi(-1.85)})$$

$$\leq 2^{11}\times 3\rho_m(\varepsilon;f)^2 V\times P(\tilde{j}\leq j^*+2)+2^{11}\times 3\times \Phi(-1.85)\frac{\rho_m(\varepsilon;f)^2 V}{1-2\Phi(-1.85)}$$

$$\leq 6355.2V\rho_m(\varepsilon;f)^2$$

□

Proof of Lemma C.17. We introduce the shorthand $Op(j_2)$ as the set of all possible $\hat{i}_{j_2}$ values when $j_2=\tilde{j}$. More precisely, $Op(j_2)=\{i^*_{j_2}-4, i^*_{j_2}-3, i^*_{j_2}-2, i^*_{j_2}+2, i^*_{j_2}+3, i^*_{j_2}+4\}$. By the definition of $\tilde{j}$, it is easy to verify that $\hat{i}_{\tilde{j}}\in Op(\tilde{j})$. We introduce another shorthand $\hat{i}_{\tilde{j}}\in Op(\tilde{j})$. Without loss of generality, we assume

$$\sup\{t>Z(f): f(t)\leq \rho_m(\varepsilon;f)+M(f)\}=\rho_z(\varepsilon;f)+Z(f).$$

We split the entire probability space by the value of $\hat{j}$, $\tilde{j}$, $\hat{i}_{\tilde{j}}$, and take conditional expectation on Y_l. Simplifying the events, repetitive usage of convexity, and careful calculation give the statement. Details are as follows.

$$\mathbb{E}_{l,s}\big((\mu_{\tilde{j},\hat{i}_{\tilde{j}}}-M(f))^2\mathbb{1}\{\tilde{j}\leq\hat{j}\}\big)$$

$$=\sum_{j_2=2}^{\infty}\sum_{j_1=j_2}^{\infty}\mathbb{E}_{l,s}\big((\mu_{j_2,\hat{i}_{j_2}}-M(f))^2\mathbb{1}\{\tilde{j}=j_2,\hat{j}=j_1\}\big)$$

$$=\sum_{j_2=2}^{\infty}\sum_{j_1=j_2}^{\infty}\sum_{i\in Op(j_2)}\mathbb{E}_{l,s}\big((\mu_{j_2,i}-M(f))^2\mathbb{1}\{\tilde{j}=j_2,\hat{j}=j_1,\hat{i}_{j_2}=i\}\big)$$

$$\leq\sum_{j_2=2}^{\infty}\sum_{j_1=j_2}^{\infty}\sum_{i\in Op(j_2)}\mathbb{E}_l\Big((\mu_{j_2,i}-M(f))^2\mathbb{1}\{\tilde{j}=j_2,\hat{i}_{j_2}=i\}\mathbb{E}_s\big(\mathbb{1}\{\hat{j}=j_1\}|Y_l\big)\Big)$$

$$\leq\sum_{j_2=2}^{\infty}\sum_{j_1=j_2}^{\infty}\sum_{i\in Op(j_2)}$$

$$\mathbb{E}_l\Big((\mu_{j_2,i}-M(f))^2\mathbb{1}\{\tilde{j}=j_2,\hat{i}_{j_2}=i\}\Big(\mathbb{E}_s\big(\mathbb{1}\{\hat{j}=j_1\}|Y_l\big)\mathbb{1}\{j_1\leq j^*+2\}+$$

$$\mathbb{1}\{j_1\geq j^*+3\}\Pi_{j=j^*+2}^{j_1-1}\max\{\Phi(-2),\Phi(-2+(\frac{7}{16}+\frac{6m_j}{\rho_z(\varepsilon;f)})\rho_m(\varepsilon;f)\frac{\sqrt{m_j}}{\sqrt{2}c_s\varepsilon})\}\Big)\Big)$$

$$
\begin{aligned}
&\le \sum_{j_2=2}^{\infty}\sum_{i\in Op(j_2)}\mathbb{E}_l\Big((\mu_{j_2,i}-M(f))^2\mathbb{1}\{\tilde{j}=j_2,\hat{i}_{j_2}=i\}\big(\mathbb{1}\{j_2\le j^*+2\}+\\
&\qquad\qquad \mathbb{1}\{j_2\ge j^*+3\}\Phi(-1.85)\frac{\Phi(-2+\frac{1}{12})^{j_2-j^*-3}}{1-\Phi(-2+\frac{1}{12})})\Big)\\
&\le \sum_{j_2=2}^{\infty}\sum_{i\in Op(j_2)}\mathbb{E}_l\Big((\mu_{j_2,i}-M(f))^2\mathbb{1}\{X_{j_2,i}\le X_{j_2,i^*_{j_2}+Ind(j_2,i)}\}\big(\mathbb{1}\{j_2\le j^*+2\}+\\
&\qquad\qquad \mathbb{1}\{j_2\ge j^*+3\}\Phi(-1.85)\frac{\Phi(-2+\frac{1}{12})^{j_2-j^*-3}}{1-\Phi(-2+\frac{1}{12})})\Big)\\
&= \sum_{j_2=2}^{\infty}\big(\mathbb{1}\{j_2\le j^*+2\}+\mathbb{1}\{j_2\ge j^*+3\}\Phi(-1.85)\frac{\Phi(-2+\frac{1}{12})^{j_2-j^*-3}}{1-\Phi(-2+\frac{1}{12})})\\
&\qquad \sum_{i\in Op(j_2)}(\mu_{j_2,i}-M(f))^2\Phi(\frac{\mu_{j_2,i^*_{j_2}+Ind(j_2,i)}-\mu_{j_2,i}}{\sqrt{2}c_l\varepsilon}\sqrt{m_{j_2}})\\
&\overset{(a)}{\le} \sum_{j_2=2}^{\infty}\big(\mathbb{1}\{j_2\le j^*+2\}+\mathbb{1}\{j_2\ge j^*+3\}\Phi(-1.85)\frac{\Phi(-2+\frac{1}{12})^{j_2-j^*-3}}{1-\Phi(-2+\frac{1}{12})})\\
&\qquad \sum_{i\in Op(j_2)}(\mu_{j_2,i}-M(f))^2\Phi(-(\mu_{j_2,i}-M(f))\frac{|i-i^*_{j_2}|-1}{|i-i^*_{j_2}|+\frac{1}{2}}\frac{\sqrt{m_{j_2}}}{\sqrt{2}c_l\varepsilon})\\
&\le \sum_{j_2=2}^{\infty}\big(\mathbb{1}\{j_2\le j^*+2\}+\mathbb{1}\{j_2\ge j^*+3\}\Phi(-1.85)\frac{\Phi(-2+\frac{1}{12})^{j_2-j^*-3}}{1-\Phi(-2+\frac{1}{12})})\\
&\qquad \sum_{i\in Op(j_2)}\frac{2c_l^2\varepsilon^2}{m_{j_2}}(\frac{|i-i^*_{j_2}|+\frac{1}{2}}{|i-i^*_{j_2}|-1})^2Q\\
&< \sum_{j_2=2}^{\infty}\big(\mathbb{1}\{j_2\le j^*+2\}+\mathbb{1}\{j_2\ge j^*+3\}\Phi(-1.85)\frac{\Phi(-2+\frac{1}{12})^{j_2-j^*-3}}{1-\Phi(-2+\frac{1}{12})})\times\\
&\quad 3\times 2^{4+j_2-j^*}\rho_m(\varepsilon;f)^2(23\frac{1}{8})Q\times 2\\
&< 3\times(2^8+2^8\frac{\Phi(-1.85)}{(1-\Phi(-2+\frac{1}{12}))^2})\rho_m(\varepsilon;f)^2(23\frac{1}{8})Q,
\end{aligned}
$$

where

$$Q=\sup_{x\ge 0}x^2\Phi(-x).$$

Step (a) follows from the following reasoning. Without loss of generality, we can assume $i\ge i^*_{j_2}+2$. This assumption and conveixty give that $\mu_{j_2,i}\ge f(t_{j_2,i}-\frac{1}{2})>\mu_{j_2,i^*_{j_2}+1}>M(f)$, that $f(t_{j_2,i}-\frac{1}{2})\ge f(t_{j_2,i}+x)$ for $x\in[0,1]$, and that $\frac{f(t_{j_2,i}-\frac{1}{2})-f(t_{j_2,i^*_{j_2}}+x)}{t_{j_2,i}-\frac{1}{2}-(t_{j_2,i^*_{j_2}}+x)} \geq \frac{f(t_{j_2,i}-\frac{1}{2})-M(f)}{t_{j_2,i}-\frac{1}{2}-Z(f)}$. Consequently,

$$\frac{\mu_{j_2,i}-\mu_{j_2,i^*_{j_2}+1}}{\mu_{j_2,i}-M(f)} \geq \frac{f(t_{j_2,i}-\frac{1}{2})-\mu_{j_2,i^*_{j_2}+1}}{f(t_{j_2,i}-\frac{1}{2})-M(f)} \geq \int_{[0,1]} \frac{t_{j_2,i}-\frac{1}{2}-t_{j_2,i^*_{j_2}}-x}{t_{j_2,i}-\frac{1}{2}-Z(f)}dx$$
$$= \frac{|i-i^*_{j_2}|-1}{t_{j_2,i}-\frac{1}{2}-Z(f)} \geq \frac{|i-i^*_{j_2}|-1}{|i-i^*_{j_2}|+\frac{1}{2}}.$$

□

Proof of Lemma C.18. First, with a bit of abuse of notation, define the events A_r, B_r, C_r, D_r to be the following (they only mean events but not constants in this proof):

$$\begin{aligned}
A_r &= \{\omega : \hat{i}_{\tilde{j}+r} < i^*_{\tilde{j}+r}, t_{\tilde{j}+r+1,\hat{i}_{\tilde{j}+r+1}} = t_{\tilde{j}+r,\hat{i}_{\tilde{j}+r}} - m_{\tilde{j}+r}\} \\
&\quad \cup \{\omega : \hat{i}_{\tilde{j}+r} > i^*_{\tilde{j}+r}, t_{\tilde{j}+r+1,\hat{i}_{\tilde{j}+r+1}} = t_{\tilde{j}+r,\hat{i}_{\tilde{j}+r}} + m_{\tilde{j}+r+1}\} \\
B_r &= \{\omega : \hat{i}_{\tilde{j}+r} < i^*_{\tilde{j}+r}, t_{\tilde{j}+r+1,\hat{i}_{\tilde{j}+r+1}} = t_{\tilde{j}+r,\hat{i}_{\tilde{j}+r}} - m_{\tilde{j}+r+1}\} \\
&\quad \cup \{\omega : \hat{i}_{\tilde{j}+r} > i^*_{\tilde{j}+r}, t_{\tilde{j}+r+1,\hat{i}_{\tilde{j}+r+1}} = t_{\tilde{j}+r,\hat{i}_{\tilde{j}+r}}\} \\
C_r &= \{\omega : \hat{i}_{\tilde{j}+r} < i^*_{\tilde{j}+r}, t_{\tilde{j}+r+1,\hat{i}_{\tilde{j}+r+1}} = t_{\tilde{j}+r,\hat{i}_{\tilde{j}+r}}\} \\
&\quad \cup \{\omega : \hat{i}_{\tilde{j}+r} > i^*_{\tilde{j}+r}, t_{\tilde{j}+r+1,\hat{i}_{\tilde{j}+r+1}} = t_{\tilde{j}+r,\hat{i}_{\tilde{j}+r}} - m_{\tilde{j}+r+1}\} \\
D_r &= \{\omega : \hat{i}_{\tilde{j}+r} < i^*_{\tilde{j}+r}, t_{\tilde{j}+r+1,\hat{i}_{\tilde{j}+r+1}} = t_{\tilde{j}+r,\hat{i}_{\tilde{j}+r}} + m_{\tilde{j}+r+1}\} \\
&\quad \cup \{\omega : \hat{i}_{\tilde{j}+r} > i^*_{\tilde{j}+r}, t_{\tilde{j}+r+1,\hat{i}_{\tilde{j}+r+1}} = t_{\tilde{j}+r,\hat{i}_{\tilde{j}+r}} - m_{\tilde{j}+r}\}
\end{aligned} \tag{D.39}$$

Basically, these events indicate which interval the localization procedure picks at the step $\tilde{j}+r+1$, and from the highest average to the lowest average is A to D. These sets of notation for events are only used in this proof, and in the proof of other theorems, the same notation can denote different things.

Still, without loss of generality, we assume

$$\sup\{t > Z(f) : f(t) \leq \rho_m(\varepsilon; f) + M(f)\} = \rho_z(\varepsilon; f) + Z(f).$$

Note that $(\mu_{\hat{j},\hat{i}_{\hat{j}}} - \mu_{\tilde{j},\hat{i}_{\tilde{j}}})_+$ is non-zero only when $\mu_{\hat{j},\hat{i}_{\hat{j}}} > \mu_{\tilde{j},\hat{i}_{\tilde{j}}}$. We split the entire probability space by events A_r, B_r, C_r, D_r and the values of $\hat{j}$ and $\tilde{j}$. We remove the ones that $\mu_{\hat{j},\hat{i}_{\hat{j}}} \leq \mu_{\tilde{j},\hat{i}_{\tilde{j}}}$ and only keep the ones that $\mu_{\hat{j},\hat{i}_{\hat{j}}} > \mu_{\tilde{j},\hat{i}_{\tilde{j}}}$ is possible. We further write $(\mu_{\hat{j},\hat{i}_{\hat{j}}} - \mu_{\tilde{j},\hat{i}_{\tilde{j}}})_+$ as a summation of $(\mu_{j+1,\hat{i}_{j+1}} - \mu_{j,\hat{i}_j})_+$ for $\tilde{j} \leq j < \hat{j}$ and bound this sum. Details are as follows.

(D.40)

$$
\begin{aligned}
&\mathbb{E}_{l,s}\Big(\big((\mu_{\hat{j},\hat{i}_{\hat{j}}}-\mu_{\tilde{j},\hat{i}_{\tilde{j}}})_+\big)^2\mathbb{1}\{\tilde{j}\le\hat{j}\}\Big)\\
&=\mathbb{E}_{l,s}\Big(\big((\mu_{\hat{j},\hat{i}_{\hat{j}}}-\mu_{\tilde{j},\hat{i}_{\tilde{j}}})_+\big)^2\mathbb{1}\{\tilde{j}\le\hat{j}-1\}\Big)\\
&=\mathbb{E}_{l,s}\Big(\big((\mu_{\hat{j},\hat{i}_{\hat{j}}}-\mu_{\tilde{j},\hat{i}_{\tilde{j}}})_+\big)^2\mathbb{1}\{\tilde{j}\le\hat{j}-1,A_0\cup B_0\cup(C_0\cap(A_1\cup B_1))\}\Big)\\
&=\mathbb{E}_{l,s}\Big(\big((\mu_{\hat{j},\hat{i}_{\hat{j}}}-\mu_{\tilde{j},\hat{i}_{\tilde{j}}})_+\big)^2\mathbb{1}\{\tilde{j}\le\hat{j}-1,A_0\cup(B_0\cap D_1^c)\cup(B_0\cap D_1\cap\{\hat{j}=\tilde{j}+1\})\}\Big)\\
&\quad+\mathbb{E}_{l,s}\Big(\big((\mu_{\hat{j},\hat{i}_{\hat{j}}}-\mu_{\tilde{j},\hat{i}_{\tilde{j}}})_+\big)^2\mathbb{1}\{\tilde{j}\le\hat{j}-2,C_0\cap A_1\}\Big)\\
&\quad+\mathbb{E}_{l,s}\Big(\big((\mu_{\hat{j},\hat{i}_{\hat{j}}}-\mu_{\tilde{j},\hat{i}_{\tilde{j}}})_+\big)^2\mathbb{1}\{\tilde{j}\le\hat{j}-3,(C_0\cap B_1)\cup(B_0\cap D_1)\}\Big)\\
&\le\sum_{j_2=2}^{\infty}\sum_{j_1=j_2+1}^{\infty}\mathbb{E}_{l,s}\Big(\big(\sum_{j=j_2}^{j_1-1}(\mu_{j+1,\hat{i}_{j+1}}-\mu_{j,\hat{i}_j})_+\big)^2\\
&\qquad\qquad\mathbb{1}\{\hat{j}=j_1,\tilde{j}=j_2,A_0\cup(B_0\cap D_1^c)\cup(B_0\cap D_1\cap\{j_1=j_2+1\})\}\Big)\\
&\quad+\sum_{j_2=2}^{\infty}\sum_{j_1=j_2+2}^{\infty}\mathbb{E}_{l,s}\Big(\big(\sum_{j=j_2}^{j_1-1}(\mu_{j+1,\hat{i}_{j+1}}-\mu_{j,\hat{i}_j})_+\big)^2\mathbb{1}\{\hat{j}=j_1,\tilde{j}=j_2,C_0\cap A_1\}\Big)\\
&\quad+\mathbb{E}_{l,s}\Big(\big((\mu_{\hat{j},\hat{i}_{\hat{j}}}-\mu_{\tilde{j},\hat{i}_{\tilde{j}}})_+\big)^2\mathbb{1}\{\hat{j}\ge\tilde{j}+3,(C_0\cap B_1)\cup(B_0\cap D_1)\}\Big)\\
&\le\underbrace{\sum_{j_2=2}^{\infty}\sum_{j_1=j_2+1}^{\infty}\mathbb{E}_{l,s}\Big(2\sum_{j=j_2}^{j_1-1}2^{j-j_2}\big((\mu_{j+1,\hat{i}_{j+1}}-\mu_{j,\hat{i}_j})_+\big)^2
\mathbb{1}\{\hat{j}=j_1,\tilde{j}=j_2,A_0\cup(B_0\cap D_1^c)\cup(B_0\cap D_1\cap\{j_1=j_2+1\})\}\Big)}_{\kappa_1}\\
&\quad+\underbrace{\sum_{j_2=2}^{\infty}\sum_{j_1=j_2+2}^{\infty}\mathbb{E}_{l,s}\Big(2\sum_{j=j_2+1}^{j_1-1}2^{j-j_2-1}\big((\mu_{j+1,\hat{i}_{j+1}}-\mu_{j,\hat{i}_j})_+\big)^2\mathbb{1}\{\hat{j}=j_1,\tilde{j}=j_2,C_0\cap A_1\}\Big)}_{\kappa_2}\\
&\quad+\underbrace{\mathbb{E}_{l,s}\Big(\big((\mu_{\hat{j},\hat{i}_{\hat{j}}}-\mu_{\tilde{j},\hat{i}_{\tilde{j}}})_+\big)^2\mathbb{1}\{\hat{j}\ge\tilde{j}+3,(C_0\cap B_1)\cup(B_0\cap D_1)\}\Big)}_{\kappa_3}.
\end{aligned}
$$

We will bound $\kappa_1+\kappa_2$ and κ_3 in Inequality (D.40) separately. We start with $\kappa_1+\kappa_2$ and introduce the shorthand $\delta_0=\mathbb{1}\{j_1=j_2+1\},\delta=\mathbb{1}\{j=j_2\}$ that we will only use in bounding $\kappa_1+\kappa_2$. Changing the order of summation, taking conditional expectation with respect to Y_l, and elementary calculation give the following.

$$\begin{aligned}
&\kappa_1 + \kappa_2 \\
&= \sum_{j_2=2}^{\infty} \sum_{j=j_2}^{\infty} 2^{j+1-j_2} \sum_{j_1=j+1}^{\infty} \mathbb{E}_{l,s}\Big((\mu_{j+1,\hat{i}_{j+1}} - \mu_{j,\hat{i}_j})^2 \mathbb{1}\{\mu_{j+1,\hat{i}_{j+1}} > \mu_{j,\hat{i}_j}\} \mathbb{1}\{\hat{j} = j_1\} \\
&\qquad \Big(\mathbb{1}\{\tilde{j} = j_2, A_0 \cup (B_0 \cap D_1^c)\} + \mathbb{1}\{\tilde{j} = j_2, j_1 = j_2 + 1, j = j_2, B_0 \cap D_1\} \Big) \Big) \\
&\quad + \sum_{j_2=2}^{\infty} \sum_{j=j_2+1}^{\infty} 2^{j-j_2} \sum_{j_1=j+1}^{\infty} \mathbb{E}_{l,s}\Big((\mu_{j+1,\hat{i}_{j+1}} - \mu_{j,\hat{i}_j})^2 \mathbb{1}\{\mu_{j+1,\hat{i}_{j+1}} > \mu_{j,\hat{i}_j}\} \\
&\qquad\qquad \mathbb{1}\{\tilde{j} = j_2, C_0 \cap A_1\} \mathbb{1}\{\hat{j} = j_1\} \Big) \\
&\le \sum_{j_2=2}^{\infty} \sum_{j=j_2}^{\infty} 2^{j+1-j_2} \mathbb{E}_l\Big((\mu_{j+1,\hat{i}_{j+1}} - \mu_{j,\hat{i}_j})^2 \mathbb{1}\{\mu_{j+1,\hat{i}_{j+1}} > \mu_{j,\hat{i}_j}\} \\
&\qquad\qquad \sum_{j_1=j+1}^{\infty} \Phi(-1.85)^{(j_2-j^*-\delta_0)_+} \Phi(-2)^{(j_1-j_2-2)_+} \\
&\qquad \Big(\mathbb{1}\{\tilde{j} = j_2, A_0 \cup (B_0 \cap D_1^c)\} + \mathbb{1}\{\tilde{j} = j_2, B_0 \cap D_1, j_1 = j_2 + 1, j = j_2\} \Big) \Big) \\
&\quad + \sum_{j_2=2}^{\infty} \sum_{j=j_2+1}^{\infty} 2^{j-j_2} \mathbb{E}_l\Big((\mu_{j+1,\hat{i}_{j+1}} - \mu_{j,\hat{i}_j})^2 \mathbb{1}\{\mu_{j+1,\hat{i}_{j+1}} > \mu_{j,\hat{i}_j}\} \mathbb{1}\{\tilde{j} = j_2, C_0 \cap A_1\} \\
&\qquad\qquad \sum_{j_1=j+1}^{\infty} \Phi(-1.85)^{(j_2-j^*)_+} \Phi(-2)^{(j_1-j_2-2)_+} \Big) \\
&\le \sum_{j_2=2}^{\infty} \sum_{j=j_2}^{\infty} \Phi(-1.85)^{(j_2-j^*-\delta)_+} 2^{j+1-j_2} \mathbb{E}_l\Big((\mu_{j+1,\hat{i}_{j+1}} - \mu_{j,\hat{i}_j})^2 \mathbb{1}\{\mu_{j+1,\hat{i}_{j+1}} > \mu_{j,\hat{i}_j}\} \mathbb{1}\{\tilde{j} = j_2\} \\
&\quad \big(\mathbb{1}\{j = j_2, A_0 \cup B_0\}(1 + \frac{1}{1 - \Phi(-2)}) + \mathbb{1}\{j \ge j_2 + 1, A_0 \cup (B_0 \cap D_1^c)\} \frac{\Phi(-2)^{j-j_2-1}}{1 - \Phi(-2)} \big) \Big) \\
&\quad + \sum_{j_2=2}^{\infty} \Phi(-1.85)^{(j_2-j^*)_+} \sum_{j=j_2+1}^{\infty} 2^{j-j_2} \mathbb{E}_l\Big((\mu_{j+1,\hat{i}_{j+1}} - \mu_{j,\hat{i}_j})^2 \\
&\qquad \mathbb{1}\{\mu_{j+1,\hat{i}_{j+1}} > \mu_{j,\hat{i}_j}\} \mathbb{1}\{\tilde{j} = j_2, C_0 \cap A_1\} \Big) \big(\Phi(-2)^{j-j_2-1} \frac{1}{1 - \Phi(-2)} \big).
\end{aligned}$$

We will further split the probability space by the sequence given in localization procedure. Define the set $C(j,k,k+1)$ to be the set of pairs (i_1, i_2) such that $P(\hat{i}_{k+1} = i_2, \hat{i}_k = i_1 | \tilde{j} = j) > 0$. Clearly, $|C(j,k,k+1)| \le$

$\min\{10\times 2^{k-j}\times 4, 6\times 4^{k+1-j}\}$. Continuing with the bound, we have

$$\kappa_1+\kappa_2 \le \sum_{j_2=2}^{\infty}\sum_{j=j_2}^{\infty}\Phi(-1.85)^{(j_2-j^*-\delta)_+}\cdot 2^{j+1-j_2}\sum_{(i_1,i_2)\in C(j_2,j,j+1)}$$
$$\mathbb{E}_l\Big((\mu_{j+1,i_2}-\mu_{j,i_1})^2\mathbb{1}\{\mu_{j+1,i_2}>\mu_{j,i_1}\}\mathbb{1}\{\tilde{j}=j_2, A_0\cup B_0, \hat{i}_j=i_1, \hat{i}_{j+1}=i_2\}\Big)$$
$$\big(\mathbb{1}\{j=j_2\}(1+\frac{1}{1-\Phi(-2)})+\mathbb{1}\{j\ge j_2+1\}\frac{\Phi(-2)^{j-j_2-1}}{1-\Phi(-2)}\big)$$
$$+\sum_{j_2=2}^{\infty}\Phi(-1.85)^{(j_2-j^*)_+}\sum_{j=j_2+1}^{\infty}2^{j-j_2}\sum_{(i_1,i_2)\in C(j_2,j,j+1)}$$
$$\mathbb{E}_l\Big((\mu_{j+1,i_2}-\mu_{j,i_1})^2\mathbb{1}\{\mu_{j+1,i_2}>\mu_{j,i_1}\}\mathbb{1}\{\tilde{j}=j_2, \hat{i}_{j+1}=i_2, \hat{i}_j=i_1, C_0\cap A_1\}\Big)$$
$$\big(\Phi(-2)^{j-j_2-1}\frac{1}{1-\Phi(-2)}\big)$$
$$\le \sum_{j_2=2}^{\infty}\sum_{j=j_2}^{\infty}\Phi(-1.85)^{(j_2-j^*-\delta)_+}\cdot 2^{j+1-j_2}\sum_{(i_1,i_2)\in C(j_2,j,j+1)}\frac{2c_l^2\varepsilon^2}{m_{j+1}}Q\mathbb{1}\{\mu_{j+1,i_2}>\mu_{j,i_1}\}$$
$$\big(\mathbb{1}\{j=j_2\}(1+\frac{1}{1-\Phi(-2)})+\mathbb{1}\{j\ge j_2+1\}\frac{\Phi(-2)^{j-j_2-1}}{1-\Phi(-2)}\big)$$
$$+\sum_{j_2=2}^{\infty}\Phi(-1.85)^{(j_2-j^*)_+}\sum_{j=j_2+1}^{\infty}2^{j-j_2}\sum_{(i_1,i_2)\in C(j_2,j,j+1)}\frac{2c_l^2\varepsilon^2}{m_{j+1}}Q\mathbb{1}\{\mu_{j+1,i_2}>\mu_{j,i_1}\}$$
$$\big(\Phi(-2)^{j-j_2-1}\frac{1}{1-\Phi(-2)}\big)$$
$$\le \sum_{j_2=2}^{\infty}\sum_{j=j_2}^{\infty}\Phi(-1.85)^{(j_2-j^*-\delta)_+}\cdot 2^{j+1-j_2}\times\min\{10\times 2^{j-j_2}\times 2, 6\times 4^{j-j_2}\times 2\}\frac{2c_l^2\varepsilon^2}{m_{j+1}}Q$$
$$\big(\mathbb{1}\{j=j_2\}(1+\frac{1}{1-\Phi(-2)})+\mathbb{1}\{j\ge j_2+1\}\frac{\Phi(-2)^{j-j_2-1}}{1-\Phi(-2)}\big)$$
$$+\sum_{j_2=2}^{\infty}\Phi(-1.85)^{(j_2-j^*)_+}\sum_{j=j_2+1}^{\infty}2^{j-j_2}\times\min\{10\times 2^{j-j_2}\times 2, 6\times 4^{j-j_2}\times 2\}\frac{2c_l^2\varepsilon^2}{m_{j+1}}Q$$
$$\big(\Phi(-2)^{j-j_2-1}\frac{1}{1-\Phi(-2)}\big)$$
$$=\frac{c_l^2Q\varepsilon^2}{m_{j^*}}\sum_{j_2=2}^{\infty}2^{j_2+3-j^*}\times(12\times(1+\frac{1}{1-\Phi(-2)})\times\Phi(-1.85)^{(j_2-j^*-1)_+}+$$
$$\Phi(-1.85)^{(j_2-j^*)_+}\times 160\times\frac{1}{1-\Phi(-2)}\times\frac{1}{1-8\Phi(-2)})$$
$$+\frac{c_l^2Q\varepsilon^2}{m_{j^*}}\sum_{j_2=2}^{\infty}\Phi(-1.85)^{(j_2-j^*)_+}2^{7+j_2-j^*}\times 5\times\frac{1}{1-\Phi(-2)}\times\frac{1}{1-8\Phi(-2)}$$
$$<\frac{c_l^2Q\varepsilon^2}{m_{j^*}}2790.303\times(\frac{1}{1-2\Phi(-1.85)}+2-1)\le Q\times 277075\rho_m(\varepsilon;f)^2.$$

Now we will turn to κ_3 in Inequality (D.40). Analysis similar to bounding $\kappa_1 + \kappa_2$ gives

$$
\begin{aligned}
\kappa_3 &= \mathbb{E}_{l,s}\Big(\big((\mu_{\hat{j},\hat{i}_{\hat{j}}} - \mu_{\tilde{j},\hat{i}_{\tilde{j}}})_+\big)^2 \mathbb{1}\{\hat{j} \geq \tilde{j}+3, (C_0 \cap B_1) \cup (B_0 \cap D_1)\}\Big) \\
&\leq \sum_{j_2=2}^{\infty} \sum_{j_1=j_2+3}^{\infty} \mathbb{E}_{l,s}\Big(\big(\sum_{j=j_2+2}^{j_1-1} (\mu_{j+1,\hat{i}_{j+1}} - \mu_{j,\hat{i}_j})_+\big)^2 \\
&\qquad \mathbb{1}\{\hat{j} = j_1, \tilde{j} = j_2, (C_0 \cap B_1) \cup (B_0 \cap D_1)\}\Big) \\
&\leq \sum_{j_2=2}^{\infty} \sum_{j_1=j_2+3}^{\infty} \mathbb{E}_{l,s}\Big(2 \sum_{j=j_2+2}^{j_1-1} 2^{j-j_2-2}\big((\mu_{j+1,\hat{i}_{j+1}} - \mu_{j,\hat{i}_j})_+\big)^2 \\
&\qquad \mathbb{1}\{\hat{j} = j_1, \tilde{j} = j_2, (C_0 \cap B_1) \cup (B_0 \cap D_1)\}\Big) \\
&\leq \sum_{j_2=2}^{\infty} \mathbb{E}_{l}\Big(2 \sum_{j=j_2+2}^{\infty} 2^{j-j_2-2}\big((\mu_{j+1,\hat{i}_{j+1}} - \mu_{j,\hat{i}_j})_+\big)^2 \\
&\qquad \mathbb{1}\{\tilde{j} = j_2, (C_0 \cap B_1) \cup (B_0 \cap D_1)\} \times \Phi(-1.85)^{(j_2+1-j^*)_+} \frac{\Phi(-2)^{j-j_2-2}}{1-\Phi(-2)}\Big) \\
&\overset{(a)}{\leq} \sum_{j_2=2}^{\infty} \sum_{j=j_2+2}^{\infty} 2^{j-j_2-1}(2 \cdot 3 \cdot 2^{j-j_2-2} \cdot 2) \frac{2c_l^2\varepsilon^2}{m_{j+1}} Q \Phi(-1.85)^{(j_2+1-j^*)_+} \frac{\Phi(-2)^{j-j_2-2}}{1-\Phi(-2)} \\
&= \frac{c_l^2\varepsilon^2}{m_{j^*}} Q \sum_{j_2=2}^{\infty} \frac{192}{1-\Phi(-2)} \times 2^{j_2+1-j^*} \times \Phi(-1.85)^{(j_2+1-j^*)_+} \frac{1}{1-8\Phi(-2)} \\
&\leq \frac{c_l^2\varepsilon^2}{m_{j^*}} Q \frac{192}{1-\Phi(-2)} \times \Big(\frac{1}{1-2\Phi(-1.85)} + 2 - 1\Big) \frac{1}{1-8\Phi(-2)} \\
&\leq 48Q \times \frac{192}{1-\Phi(-2)} \times \Big(\frac{1}{1-2\Phi(-1.85)} + 1\Big) \frac{1}{1-8\Phi(-2)} \rho_m(\varepsilon; f)^2 \\
&\leq 23850.1 \rho_m(\varepsilon; f)^2 Q.
\end{aligned}
$$

Step (a) follows from the fact that the number of possible pairs of $(\hat{i}_j, \hat{i}_{j+1})$ such that $(C_0 \cap B_1) \cup (B_0 \cap D_1)$, $\mu_{j+1,\hat{i}_{j+1}} > \mu_{j,\hat{i}_j}$, $\tilde{j} = j_2$, $j \geq j_2 + 2$, and $\mu_{\hat{j},\hat{i}_{\hat{j}}} > \mu_{\tilde{j},\hat{i}_{\tilde{j}}}$ is at most $2 \times 3 \times 2^{j-(j_2+2)} \times 2$. Plugging the bounds of $\kappa_1 + \kappa_2$ and κ_3 back to Inequalit (D.40) gives

$$
\begin{aligned}
&\mathbb{E}_{l,s}\Big(\big((\mu_{\hat{j},\hat{i}_{\hat{j}}} - \mu_{\tilde{j},\hat{i}_{\tilde{j}}})_+\big)^2 \mathbb{1}\{\tilde{j} \leq \hat{j}\}\Big) \\
&\leq Q \times 277075 \times \rho_m(\varepsilon; f)^2 + Q \times 23850.1 \times \rho_m(\varepsilon; f)^2.
\end{aligned} \tag{D.41}
$$

□

PROOF OF LEMMA C.21.

(D.42)

$$
\begin{aligned}
&P(\hat{j} \le j^* - 2 - \tilde{K}) \\
&\le P(\hat{j} \le j^* - 2 - \tilde{K}, |\hat{i}_{\hat{j}} - i_{\hat{j}^*}| \le 4) + P(\hat{j} \le j^* - 2 - \tilde{K}, |\hat{i}_{\hat{j}} - i^*_{\hat{j}}| \ge 5) \\
&\le \sum_{j=1}^{j^*-2-\tilde{K}} P(|\hat{i}_j - i_{j^*}| \le 4, X_{\hat{i}_j+6} - X_{\hat{i}_j+5} \le 2c_s\sqrt{2}\varepsilon) + \\
&\qquad P(|\hat{i}_j - i_{j^*}| \le 4, X_{\hat{i}_j-6} - X_{\hat{i}_j-5} \le 2c_s\sqrt{2}\varepsilon) + P(|\hat{i}_{j-1} - i^*_{j-1}| \ge 2) \\
&\le \sum_{j=1}^{j^*-2-\tilde{K}} \\
&2\Phi(2 - (\frac{m_j}{\rho_z(\varepsilon; f)})^{\frac{3}{2}} \frac{\rho_m(\varepsilon; f)\sqrt{\rho_z(\varepsilon; f)}}{\sqrt{2}c_s\varepsilon}) + 2\Phi(-(\frac{m_{j-1}}{\rho_z(\varepsilon; f)})^{\frac{3}{2}} \frac{\rho_m(\varepsilon; f)\sqrt{\rho_z(\varepsilon; f)}}{\sqrt{2}c_s\varepsilon}) \\
&+ 2\Phi(-2(\frac{m_{j-1}}{\rho_z(\varepsilon; f)})^{\frac{3}{2}} \frac{\rho_m(\varepsilon; f)\sqrt{\rho_z(\varepsilon; f)}}{\sqrt{2}c_s\varepsilon}) + 2\Phi(-3(\frac{m_{j-1}}{\rho_z(\varepsilon; f)})^{\frac{3}{2}} \frac{\rho_m(\varepsilon; f)\sqrt{\rho_z(\varepsilon; f)}}{\sqrt{2}c_s\varepsilon}) \\
&< 2\sum_{j=1}^{j^*-2-\tilde{K}} \Big(\Phi(2 - 2^{\frac{3}{2}(j^*-j-4)-\frac{1}{2}}) + \Phi(-2^{\frac{3}{2}(j^*-j-3)-\frac{1}{2}}) + \Phi(-2^{\frac{3}{2}(j^*-j-3)+\frac{1}{2}}) \\
&\qquad + \Phi(-3 \times 2^{\frac{3}{2}(j^*-j-3)-\frac{1}{2}})\Big) \\
&\le 2\sum_{k=\tilde{K}}^{\infty} \Big(\Phi(2 - 2^{\frac{3}{2}(k-2)-\frac{1}{2}}) + \Phi(-2^{\frac{3}{2}(k-1)-\frac{1}{2}}) + \Phi(-2^{\frac{3}{2}k-1}) \\
&\qquad + \Phi(-3 \times 2^{\frac{3}{2}(k-1)-\frac{1}{2}})\Big) \\
&\le 2(\Phi(2 - 2^{\frac{3}{2}(\tilde{K}-2)-\frac{1}{2}}) \frac{1 + 3\exp(-44)}{1 - \exp(-44)}) \\
&\le \frac{2}{1 - \exp(-40)} \Phi(2 - 2^{\frac{3}{2}(\tilde{K}-2)-\frac{1}{2}}).
\end{aligned}
$$

The last three equations use the fact that $\Phi(-2\sqrt{2}x) \le 2\sqrt{2}\exp(-\frac{7x^2}{2})\Phi(-x)$, for $x > 0$.

□

PROOF OF LEMMA C.22. For the ease of expression, we define $\tilde{\mathcal{E}}_{j,i} = \frac{1}{\sqrt{m_j}}(Y_3(t_{j,i}) - Y_3(t_{j,i-1}) - \int_{t_{j,i-1}}^{t_{j,i}} f(x)dx)$. Then $\tilde{\mathcal{E}}_{j,i} \overset{i.i.d}{\sim} N(0, \varepsilon^2 c_e^2), i = 1, 2, \cdots, 2^j$. Recall that

$$E^c = \{Z(f) \in [t_{(\hat{j}-K_{\frac{\alpha}{4}}-1)_+, \hat{i}_{(\hat{j}-K_{\frac{\alpha}{4}}-1)_+}-5}, t_{(\hat{j}-K_{\frac{\alpha}{4}}-1)_+, \hat{i}_{(\hat{j}-K_{\frac{\alpha}{4}}-1)_+}+4}]\},$$

we have

(D.43)

$$P\Big(G\Big|E^c\Big) = P\Big(\hat{f}_1 + S_{i_R-i_L,\frac{\alpha}{4}}\frac{c_e\varepsilon}{\sqrt{m_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}} < M(f)\Big|E^c\Big)$$

$$\leq P\Big(M(f) + \frac{1}{\sqrt{m_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}}\min_{i_L<i\leq i_R}\tilde{\mathcal{E}}_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}},i} + S_{i_R-i_L,\frac{\alpha}{4}}\frac{c_e\varepsilon}{\sqrt{m_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}} < M(f)\Big|E^c\Big)$$

$$\leq \frac{\mathbb{E}\left(\mathbb{E}\Big(\mathbb{1}\{\min_{i_L<i\leq i_R}\tilde{\mathcal{E}}_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}},i} + c_e\varepsilon S_{i_R-i_L,\frac{\alpha}{4}} < 0\}\Big|Y_l, Y_s\Big)\mathbb{1}\{E^c\}\right)}{\mathbb{E}\Big(\mathbb{1}\{E^c\}\Big)}$$

$$\leq \frac{\mathbb{E}\Big(\frac{\alpha}{4}\mathbb{1}\{E^c\}\Big)}{\mathbb{E}\Big(\mathbb{1}\{E^c\}\Big)} = \frac{\alpha}{4}.$$

□

Proof of Lemma C.23.

(D.44)

$$P(H|E^c\cap F^c)$$

$$\leq P(\hat{f}_1 + \Phi^{-1}(\frac{\alpha}{4})\frac{c_e\varepsilon}{\sqrt{m_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}} - \frac{\sqrt{3}\varepsilon}{\sqrt{m_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}} > M(f)|E^c\cap F^c)$$

$$\leq P(\int_{t_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}},i^*_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}}^{t_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}},i^*_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}+1}} f(x)\frac{1}{m_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}dx + \frac{1}{\sqrt{m_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}}\tilde{\mathcal{E}}_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}},i^*_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}+1} +$$

$$\Phi^{-1}(\frac{\alpha}{4})\frac{c_e\varepsilon}{\sqrt{m_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}} - \frac{\sqrt{3}\varepsilon}{\sqrt{m_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}} > M(f)|E^c\cap F^c)$$

$$\leq P(\tilde{\mathcal{E}}_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}},i^*_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}\frac{1}{c_e} + \Phi^{-1}(\frac{\alpha}{4})\varepsilon + \rho_m(\varepsilon;f)\sqrt{m_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}} - \sqrt{3}\varepsilon > 0|E^c\cap F^c)$$

$$\leq P(\tilde{\mathcal{E}}_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}i^*_{\hat{j}+\tilde{K}_{\frac{\alpha}{4}}}}\frac{1}{c_e} + \Phi^{-1}(\frac{\alpha}{4})\varepsilon + \rho_m(\varepsilon;f)\sqrt{\frac{1}{2}\rho_z(\varepsilon;f)} - \sqrt{3}\varepsilon > 0|E^c\cap F^c)$$

$$\leq \frac{\alpha}{4}.$$

□

Proof of Lemma C.24. Let $i_l = \min\{i : g_{n,\sigma,h}(x_i) > f(x_i)\}$, $i_r = \max\{i : g_{n,\sigma,h}(x_i) > f(x_i)\}$.

We will first prove the lemma for the case $\rho_z(\frac{\sigma}{\sqrt{6n}}; h) \geq 1/2n$.

When $\{i : g_{n,\sigma,h}(x_i) > f(x_i)\} = \emptyset$, the lemma holds naturally.

When $i_l = i_r$, let $x_l = \inf\{x : g_{n,\sigma,h}(x) > h(x)\}$, $x_r = \sup\{x : g_{n,\sigma,h}(x) > h(x)\}$, then we have

$$\begin{aligned}\frac{\sigma^2}{6n} \geq \|h - g_{n,\sigma,h}\|_2^2 \geq \frac{1}{3}(x_r - x_l)\rho_m(\frac{\sigma}{\sqrt{6n}}; h)^2 \geq \frac{1}{6}\frac{\rho_m(\frac{\sigma}{\sqrt{6n}}; h)^2}{n} \\ \geq \frac{1}{6} l_n(h, g_{n,\sigma,h})^2 = \frac{1}{6} l_n(f, g_{n,\sigma,h})^2.\end{aligned}$$

When $i_l < i_r$,

$$\begin{aligned}\frac{\sigma^2}{6n} \geq \|h - g_{n,\sigma,h}\|_2^2 \geq \sum_{k=i_l}^{i_r} \frac{1}{3}\frac{1}{2n}(h(x_k) - g_{n,\sigma,h}(x_k))^2 \geq \frac{1}{6} l_n(h, g_{n,\sigma,h})^2 \\ = \frac{1}{6} l_n(f, g_{n,\sigma,h})^2.\end{aligned}$$

Now we turn to the second case $\rho_z(\frac{\sigma}{\sqrt{6n}}; h) < 1/2n$.

Since $\rho_z(\frac{\sigma}{\sqrt{6n}}; h) < 1/2n$, then $|\{i : g_{n,\sigma,h}(x_i) > f(x_i)\}| \leq 1$. When $|\{i : g_{n,\sigma,h}(x_i) > f(x_i)\}| = 0$, the lemma holds naturally. When $|\{i : g_{n,\sigma,h}(x_i) > f(x_i)\}| = 1$, we have

$$l_n(f, g_{n,\sigma,h})^2 = l_n(h, g_{n,\sigma,h})^2 \leq \frac{1}{n}\rho_m(\frac{\sigma}{\sqrt{6n}}; h)^2 \cdot 2n\rho_z(\frac{\sigma}{\sqrt{6n}}; h) \leq \sigma^2.$$

□

Proof of Lemma C.25.

(D.45)

$$\begin{aligned}&\mathbb{E}(\mathbb{1}\{\hat{\jmath} < \tilde{\jmath}\}1.5m_{\hat{\jmath}}) \\ &\leq \mathbb{E}(\mathbb{1}\{\hat{\jmath} < \tilde{\jmath}\}1.5m_{\hat{\jmath}}\mathbb{1}\{\hat{\jmath} \leq \mathrm{j}^* - 3\}) + \mathbb{E}(\mathbb{1}\{\hat{\jmath} < \tilde{\jmath}\}1.5m_{\hat{\jmath}}\mathbb{1}\{\hat{\jmath} \geq \mathrm{j}^* - 2\}) \\ &\leq 1.5\mathbb{E}(\mathbb{1}\{\hat{\jmath} < \tilde{\jmath}\}m_{\hat{\jmath}}\mathbb{1}\{\hat{\jmath} \leq \mathrm{j}^* - 3\}) + 1.5 \times \rho_z(\frac{\sigma}{\sqrt{n}}; f) \\ &\leq 1.5\left(\sum_{j=0}^{(\mathrm{j}^*-3)\wedge(J-1)} \mathbb{E}(\mathbb{1}\{\hat{\jmath} = j, \tilde{\jmath} > j\}m_{\hat{\jmath}}) + \frac{1}{n}\mathbb{1}\{J \leq \mathrm{j}^* - 3\}\right) + 1.5 \times \rho_z(\frac{\sigma}{\sqrt{n}}; f)\end{aligned}$$

Also we have

(D.46)
$$\sum_{j=0}^{(\mathtt{j}^*-3)\wedge(J-1)} \mathbb{E}(\mathbb{1}\{\hat{\mathtt{j}}=j,\tilde{\mathtt{j}}>j\}m_{\tilde{\mathtt{j}}})$$
$$\leq \sum_{j=0}^{(\mathtt{j}^*-3)\wedge(J-1)} m_j\Big(\mathbb{E}(\mathbb{1}\{\tilde{\mathtt{j}}>j, Y_{j,\hat{\mathtt{i}}_j+6,s}-Y_{j,\hat{\mathtt{i}}_j+5,s}\leq \gamma_s 2\sqrt{2}\sqrt{2^{J-j}}\sigma\})$$
$$+\mathbb{E}(\mathbb{1}\{\tilde{\mathtt{j}}>j, Y_{j,\hat{\mathtt{i}}_j-6,s}-Y_{j,\hat{\mathtt{i}}_j-5,s}\leq \gamma_s 2\sqrt{2}\sqrt{2^{J-j}}\sigma\})\Big)$$
$$\leq \sum_{j=0}^{(\mathtt{j}^*-3)\wedge(J-1)} m_j\mathbb{E}\Big(\mathbb{1}\{\tilde{\mathtt{j}}>j, \frac{\sqrt{2^{J-j}}}{\gamma_s\sqrt{2}\sigma}\big(ave_f(j,\hat{\mathtt{i}}_j+6)-ave_f(j,\hat{\mathtt{i}}_j+5)\big)\leq$$
$$\frac{\big(\mathfrak{E}_{j,\hat{\mathtt{i}}_j+5,s}-\mathfrak{E}_{j,\hat{\mathtt{i}}_j+6,s}\big)}{\sqrt{2}\sqrt{2^{J-j}}\gamma_s\sigma}+2\}\Big)+m_j\mathbb{E}\Big(\mathbb{1}\{\tilde{\mathtt{j}}>j,$$
$$\frac{\sqrt{2^{J-j}}}{\gamma_s\sqrt{2}\sigma}\big(ave_f(j,\hat{\mathtt{i}}_j-6)-ave_f(j,\hat{\mathtt{i}}_j-5)\big)\leq \frac{\big(\mathfrak{E}_{j,\hat{\mathtt{i}}_j-5,s}-\mathfrak{E}_{j,\hat{\mathtt{i}}_j-6,s}\big)}{\gamma_s\sqrt{2}\sigma\sqrt{2^{J-j}}}+2\}\Big)$$
$$\leq \sum_{j=0}^{(\mathtt{j}^*-3)\wedge(J-1)} m_j\mathbb{E}\Big(\mathbb{1}\{\tilde{\mathtt{j}}>j\}\Phi(2-\frac{\rho_m(\frac{\sigma}{\sqrt{n}};f)}{\rho_z(\frac{\sigma}{\sqrt{n}};f)}m_j^{\frac{3}{2}}\frac{\sqrt{n}}{\gamma_s\sqrt{2}\sigma})\Big)\times 2$$
$$\leq \sum_{j=0}^{(\mathtt{j}^*-3)\wedge(J-1)} \rho_z(\frac{\sigma}{\sqrt{n}};f)2^{\mathtt{j}^*-j-2}\mathbb{E}(\mathbb{1}\{\tilde{\mathtt{j}}>j\})\times 2\Phi(2-\frac{1}{2\gamma_s}2^{\frac{3}{2}(\mathtt{j}^*-j-3)})$$
$$\leq c_{z0}\rho_z(\frac{\sigma}{\sqrt{n}};f).$$

Therefore,

$$\mathbb{E}(\mathbb{1}\{\hat{\mathtt{j}}<\tilde{\mathtt{j}}\}1.5m_{\tilde{\mathtt{j}}})\leq c_{z1}\rho_z(\frac{\sigma}{\sqrt{n}};f)+\frac{3}{2}\frac{1}{n}\mathbb{1}\{J\leq \mathtt{j}^*-3\}. \tag{D.47}$$

□

Proof of Lemma C.26.

$$\mathbb{E}(\mathbb{1}\{\hat{\mathtt{j}}\geq\tilde{\mathtt{j}}\}\big|\hat{Z}-Z(f)\big|)\leq \mathbb{E}(\mathbb{1}\{\hat{\mathtt{j}}\geq\tilde{\mathtt{j}}\}6m_{\tilde{\mathtt{j}}})$$
$$\leq 6\sum_{j=3}^{(j^*-3)\wedge J} \rho_z(\frac{\sigma}{\sqrt{n}};f)2^{\mathtt{j}^*-j-2}\Phi(-\frac{\rho_m(\frac{\sigma}{\sqrt{n}};f)}{\rho_z(\frac{\sigma}{\sqrt{n}};f)}m_{\mathtt{j}^*}^{\frac{3}{2}}2^{\frac{3}{2}(\mathtt{j}^*-j)}\frac{\sqrt{n}}{\gamma_l\sigma\sqrt{2}}) \tag{D.48}$$
$$+6\mathbb{1}\{J\geq j^*-2\}\rho_z(\frac{\sigma}{\sqrt{n}};f)\leq c_{z2}\rho_z(\frac{\sigma}{\sqrt{n}};f)$$

□

PROOF OF LEMMA C.27.

$$
\begin{aligned}
&\mathbb{E}(\mathbb{1}\{\check{j} < \infty\}\mathbb{1}\{\check{j} < \tilde{\mathrm{j}}\}|\hat{Z} - Z(\tilde{h})|) \\
&\le \mathbb{E}(\mathbb{1}\{\check{j} < \infty\}\mathbb{1}\{\check{j} < \tilde{\mathrm{j}}\}1.5 m_{\check{j}}) \\
&\le \sum_{j=3}^{J} 1.5\frac{2^{J-j}}{n}\cdot 2\Phi(2 - \frac{\rho_m(\frac{\sigma}{\sqrt{n}};\tilde{h})}{\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\frac{2^{J-j}}{n}2^{J-j}\frac{1}{\gamma_s\sigma\sqrt{2}\sqrt{2^{J-j}}}) \\
&\le \sum_{j=3}^{J} 3\frac{2^{J-j}}{n}\cdot \Phi(2 - \frac{\frac{1}{\sqrt{2n}}\sigma}{\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}}\frac{2^{\frac{3}{2}(J-j)}}{n}\frac{1}{\sqrt{2}\sigma\gamma_s}) \\
&= \sum_{j=3}^{J} 3\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\frac{2^{J-j}}{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\cdot \Phi(2 - (\frac{2^{J-j}}{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})})^{\frac{3}{2}}\frac{1}{2\gamma_s}) \\
&\le \sum_{j=3}^{J} 3\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}2^{\frac{j-J}{2}}\cdot 2\gamma_s\check{C} \\
&\le \rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\check{c}_{z1},
\end{aligned} \tag{D.49}
$$

where $\check{C} = \sup_{x>0} x\Phi(2-x)$. □

PROOF OF LEMMA C.28.

$$
\begin{aligned}
&\mathbb{E}(\mathbb{1}\{\hat{\mathrm{j}} \ge \tilde{\mathrm{j}}\}|\hat{Z} - Z(\tilde{h})|) \le \mathbb{E}(\mathbb{1}\{\hat{\mathrm{j}} \ge \tilde{\mathrm{j}}\}6 m_{\tilde{\mathrm{j}}}) \\
&\le \sum_{j=1}^{J} 6\frac{2^{J-j}}{n}\cdot 6\Phi(-\frac{2^{J-j}\frac{\rho_m(\frac{\sigma}{\sqrt{n}};\tilde{h})}{\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\frac{2^{J-j}}{n}}{\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j}}}) \le \sum_{j=1}^{J} 6\frac{2^{J-j}}{n}\cdot 6\Phi(-\frac{(\frac{2^{J-j}}{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})})^{\frac{3}{2}}}{2\gamma_s}) \\
&\le \rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\check{c}_{z2}
\end{aligned} \tag{D.50}
$$

□

Proof of Lemma C.32.

(D.51)

$$\sum_{j=1}^{\mathtt{j}^*-1}\mathbb{E}(2^{-j}\mathbb{1}\{\hat{\mathtt{j}}=j,\tilde{\mathtt{j}}>j\}) \quad\leq\quad 2^{-J}\mathbb{1}\{J\leq j^*-1\}+$$
$$\sum_{j=1}^{\min\{\mathtt{j}^*-1,J\}}\mathbb{E}\Big(2^{-j}\big(\mathbb{1}\{Y_{j,\hat{\mathtt{i}}_j+6,s}-Y_{j,\hat{\mathtt{i}}_j+5,s}\leq 2\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j}}\}$$
$$+\mathbb{1}\{Y_{j,\hat{\mathtt{i}}_j-6}-Y_{j,\hat{\mathtt{i}}_j-5}\leq 2\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j}}\}\big)\mathbb{1}\{\tilde{\mathtt{j}}>j\}\Big)$$
$$\leq\sum_{j=1}^{\min\{\mathtt{j}^*-1,J\}}2^{-j+1}\Phi(2-\frac{\frac{\rho_m(\frac{\sigma}{\sqrt{n}};f)}{\rho_z(\frac{\sigma}{\sqrt{n}};f)}\frac{2^{J-j}}{n}2^{J-j}}{\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j}}})\mathbb{E}(\mathbb{1}\{\tilde{\mathtt{j}}>j\})$$
$$+2^{-J}\mathbb{1}\{J\leq j^*-1\}$$
$$\leq\sum_{j=1}^{\min\{\mathtt{j}^*-1,J\}}2^{-\mathtt{j}^*}\cdot 2^{(\mathtt{j}^*-j)+1}\Phi(2-\frac{1}{2\gamma_s}2^{\frac{3}{2}(\mathtt{j}^*-j-3)})+2^{-J}\mathbb{1}\{J\leq j^*-1\}$$
$$\leq c_{z3}2^{-\mathtt{j}^*}+2^{-J}\mathbb{1}\{J\leq j^*-1\}.$$

□

Proof of Lemma C.33.

$$\sum_{j=1}^{\mathtt{j}^*-1}\mathbb{E}(2^{-j}\mathbb{1}\{\hat{\mathtt{j}}=j,\tilde{\mathtt{j}}\leq j\})\leq\sum_{j=1}^{(\mathtt{j}^*-3)\wedge(J-1)}\mathbb{E}(2^{-j}\mathbb{1}\{\tilde{\mathtt{j}}=j\})$$
$$\leq\sum_{j=1}^{(\mathtt{j}^*-3)\wedge(J-1)}2^{-j}\cdot 6\Phi(-\frac{\frac{\rho_m(\frac{\sigma}{\sqrt{n}};f)}{\rho_z(\frac{\sigma}{\sqrt{n}};f)}\cdot\frac{2^{J-j}}{n}2^{J-j}}{\sqrt{2}\gamma_l\sigma\sqrt{2^{J-j}}})$$
(D.52)
$$\leq\sum_{j=1}^{(\mathtt{j}^*-3)\wedge(J-1)}2^{-j}\cdot 6\Phi(-\frac{1}{2\gamma_l}2^{\frac{3}{2}(\mathtt{j}^*-j-3)})$$
$$\leq 2^{-\mathtt{j}^*}\sum_{j=1}^{\infty}6\cdot 2^{j}\Phi(-\frac{1}{2\gamma_l}2^{\frac{3}{2}(j-3)})\leq 2^{-\mathtt{j}^*}c_{z4}.$$

□

PROOF OF LEMMA C.34.

(D.53)

$$
\begin{aligned}
&\mathbb{E}(\mathbb{1}\{\check{j} < \infty\}L(\mathtt{CI}_{z,\alpha}(Y))) \leq 12 \cdot 2^{K_{\alpha/2}+1}\mathbb{E}(\frac{2^{J-\hat{\jmath}}}{n}\mathbb{1}\{\check{j} < \infty\}) \\
&= 12 \cdot 2^{K_{\alpha/2}+1}\sum_{j=3}^{J}\mathbb{E}(\mathbb{1}\{\check{j} = j\})\frac{2^{J-j}}{n} \\
&= 12 \cdot 2^{K_{\alpha/2}+1}\times \\
&\qquad \left(\sum_{j=3}^{J}\mathbb{E}(\mathbb{1}\{\check{j} = j\}\mathbb{1}\{\tilde{\jmath} \leq j\})\frac{2^{J-j}}{n} + \sum_{j=3}^{J}\mathbb{E}(\mathbb{1}\{\check{j} = j\}\mathbb{1}\{\tilde{\jmath} > j\})\frac{2^{J-j}}{n}\right).
\end{aligned}
$$

We bound the two terms separately and we start with the first term.

(D.54)

$$
\begin{aligned}
&\sum_{j=3}^{J}\mathbb{E}(\mathbb{1}\{\check{j} = j\}\mathbb{1}\{\tilde{\jmath} \leq j\})\frac{2^{J-j}}{n} \leq \sum_{j=3}^{J}\mathbb{E}\left(\mathbb{1}\{\check{j} = j\}\mathbb{1}\{\tilde{\jmath} \leq j\}\frac{2^{J-\tilde{\jmath}}}{n}\right) \\
&\leq \sum_{j=1}^{J}\mathbb{E}(\mathbb{1}\{\tilde{\jmath} = j\})\frac{2^{J-j}}{n} \leq \sum_{j=1}^{J}\frac{2^{J-j}}{n}6\Phi(-\frac{\rho_m(\frac{\sigma}{\sqrt{n}};\tilde{h})}{\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\frac{2^{J-j}}{n}\frac{2^{J-j}}{\sqrt{2^{J-j}}\sqrt{2}\gamma_l\sigma}) \\
&\leq \sum_{j=1}^{J}\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}2^{\frac{j-J}{2}}\left(\frac{2^{J-j}}{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\right)^{\frac{3}{2}} \cdot 6 \cdot \Phi(-(\frac{2^{J-j}}{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})})^{\frac{3}{2}}\frac{1}{2\gamma_l}) \\
&\leq 6\frac{1}{1-\sqrt{1/2}}\check{C}\cdot 2\gamma_l \cdot \sup_{h\in\mathcal{G}_n(f)}\rho_z(\frac{\sigma}{\sqrt{n}};h)\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};h)},
\end{aligned}
$$

where $\check{C} = \sup_{t>0} t\Phi(-t)$.

For the second term, we have

(D.55)

$$
\begin{aligned}
&\sum_{j=3}^{J}\mathbb{E}(\mathbb{1}\{\check{j} = j\}\mathbb{1}\{\tilde{\jmath} > j\})\frac{2^{J-j}}{n} \\
&\leq \sum_{j=3}^{J}\frac{2^{J-j}}{n}2\Phi(2-\frac{\rho_m(\frac{\sigma}{\sqrt{n}};\tilde{h})}{\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\frac{2^{J-j}}{n}\frac{2^{J-j}}{\sqrt{2^{J-j}}\sqrt{2}\gamma_s\sigma}) \\
&\leq \sum_{j=3}^{J}2\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}2^{\frac{j-J}{2}}\left(\frac{2^{J-j}}{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\right)^{\frac{3}{2}}\Phi(2-(\frac{2^{J-j}}{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})})^{\frac{3}{2}}\frac{1}{2\gamma_s}) \\
&\leq 2\cdot 2\gamma_s\cdot\check{Q}\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\frac{1}{1-\sqrt{1/2}},
\end{aligned}
$$

where $\check{Q} = \sup_{t>0} t\Phi(2-t)$.

Let $\check{c}_{1,\alpha} = (6\frac{1}{1-\sqrt{1/2}}2\gamma_l \check{C} + 4\gamma_s \cdot \check{Q}\frac{1}{1-\sqrt{1/2}}) \cdot 12 \cdot 2^{K_{\alpha/2}+1}$ gives the statement of Lemma C.34.

□

Proof of Lemma C.35. When $2 \leq i_m \leq n-2$, $t_{hi} - t_{lo} \geq \frac{3}{n}$ implies that $i_l \leq i_m - 1$ or $i_r \geq i_m$. When $i_m \leq 1$, $t_{hi} - t_{lo} \geq \frac{3}{n}$ implies that $i_r \geq i_m$. When $i_m \geq n-1$, $t_{hi} - t_{lo} \geq \frac{3}{n}$ implies that $i_l \leq i_m - 1$. Therefore, we have

(D.56)

$$\begin{aligned}
&\mathbb{E}(\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{t_{hi} - t_{lo} \geq \frac{3}{n}\}L(\mathtt{CI}_{z,\alpha}(Y))) \\
&\leq \frac{12 \cdot 2^{K_{\alpha/2}+1}}{n}\mathbb{E}\Big(\mathbb{1}\{i_l \leq i_m - 1\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{i_m \geq 2\} \\
&\qquad + \mathbb{1}\{i_r \geq i_m\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{i_m \leq n-2\}\Big) \\
&= \frac{12 \cdot 2^{K_{\alpha/2}+1}}{n}\mathbb{E}\Big(\mathbb{1}\{i_l \leq i_m - 1\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{U \leq i_m - 1, i_m \geq 2\} + \\
&\qquad \mathbb{1}\{i_l \leq i_m - 1\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{U \geq i_m, i_m \geq 2\} \\
&\qquad + \mathbb{1}\{i_r \geq i_m\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{L \geq i_m + 1, i_m \leq n-2\} \\
&\qquad + \mathbb{1}\{i_r \geq i_m\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{L \leq i_m, i_m \leq n-2\}\Big).
\end{aligned}$$

Since $\{U \leq i_m - 1, i_m \geq 2\} \cup \{L \geq i_m + 1, i_m \leq n-2\}$ implies that $\tilde{\mathrm{j}} < J$, and $\{U \leq i_m - 1\} \cap \{L \geq i_m + 1\} = \emptyset$, we have

(D.57)

$$\begin{aligned}
&\mathbb{E}\Big(\mathbb{1}\{i_l \leq i_m - 1\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{U \leq i_m - 1\} + \mathbb{1}\{i_r \geq i_m\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{L \geq i_m + 1\}\Big) \\
&\leq \mathbb{E}(\mathbb{1}\{\tilde{\mathrm{j}} < J\}) = \sum_{j=1}^{J-1} P(\tilde{\mathrm{j}} = j) \leq \sum_{j=1}^{J-1} \Phi(-\frac{\rho_m(\frac{\sigma}{\sqrt{n}}; \tilde{h})}{\rho_z(\frac{\sigma}{\sqrt{n}}; \tilde{h})}\frac{2^{J-j}}{n}\frac{2^{J-j}}{\sqrt{2^{J-j}}\sqrt{2}\gamma_l \sigma}) \\
&\leq \sum_{j=1}^{J-1} \Phi(-\left(\frac{2^{J-j}}{n\rho_z(\frac{\sigma}{\sqrt{n}}; \tilde{h})}\right)^{\frac{3}{2}}\frac{1}{2\gamma_l}) \\
&\leq n\rho_z(\frac{\sigma}{\sqrt{n}}; \tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}}; \tilde{h})}\frac{1}{1-\sqrt{\frac{1}{8}}}2\gamma_l \check{C},
\end{aligned}$$

where $\check{C} = \sup_{t>0} t\Phi(-t)$.

And for $\mathbb{E}(\mathbb{1}\{i_l \leq i_m - 1\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{U \geq i_m, i_m \geq 2\} + \mathbb{1}\{i_r \geq i_m\}\mathbb{1}\{\check{j} =$

$\infty\}\mathbb{1}\{L \le i_m, i_m \le n-2\})$, we have

$$\begin{aligned}
&\mathbb{E}\Big(\mathbb{1}\{i_l \le i_m - 1\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{U \ge i_m, i_m \ge 2\} \\
&\qquad + \mathbb{1}\{i_r \ge i_m\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{L \le i_m, i_m \le n-2\}\Big) \\
&= \mathbb{E}\Big(\mathbb{E}(\mathbb{1}\{i_l \le i_m - 1\}|\mathtt{Y}_l, \mathtt{Y}_s)\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{U \ge i_m, i_m \ge 2\} \\
&\qquad + \mathbb{E}(\mathbb{1}\{i_r \ge i_m\}|\mathtt{Y}_l, \mathtt{Y}_s)\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{L \le i_m, i_m \le n-2\}\Big) \\
&\le 2\mathbb{E}(U-L)\Phi(-\frac{\rho_m(\frac{\sigma}{\sqrt{n}};\tilde{h})}{\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\frac{1}{n2\sqrt{3}\sigma} + z_{\alpha_1}) \\
&\le 2\mathbb{E}(U-L)\Phi(-\left(\frac{1}{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\right)^{\frac{3}{2}}\frac{1}{\sqrt{24}} + z_{\alpha_1}) \\
&\le n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\sqrt{24}\cdot 2\cdot \check{Q}_2 \times 24 \times 2^{K_{\alpha/2}},
\end{aligned} \tag{D.58}$$

where $\check{Q}_2 = \sup_{t>0} t\Phi(z_{\alpha_1} - t)$.

Therefore,

$$\begin{aligned}
&\mathbb{E}(\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{t_{hi} - t_{lo} \ge \frac{3}{n}\}L(\mathtt{CI}_{z,\alpha}(Y))) \\
&\le \check{c}_{2,\alpha}\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})} \\
&\le \check{c}_{2,\alpha}\sup_{h\in\mathcal{G}_n(f)}\rho_z(\frac{\sigma}{\sqrt{n}};h)\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};h)}.
\end{aligned} \tag{D.59}$$

□

Proof of Lemma C.36. Note that when $0 < t_{hi} - t_{lo} < \frac{3}{n}$, one of the following holds: $i_l = n = U = i_r + 1$, $i_r = -1 = L - 1 = i_l - 1$, $L + 1 \le i_l = i_r + 1 \le U - 1$, $i_l = L = i_r$, $i_r = U - 1 = i_l$. We denote event

$$\begin{aligned}
H_1 =& \{i_l = n = U = i_r + 1\} \cup \{i_r = -1 = L - 1 = i_l - 1\} \cup \{L + 1 \le i_l = i_r + 1 \le U - 1\}, \\
H_2 =& \{i_l = L = i_r\} \cup \{i_r = U - 1 = i_l\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}(\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{t_{hi} - t_{lo} < \frac{3}{n}\}L(\mathtt{CI}_{z,\alpha}(Y))) \\
&= \mathbb{E}\left(\mathbb{1}\{\check{j} = \infty\}L(\mathtt{CI}_{z,\alpha}(Y))\mathbb{1}\{H_1\}\right) + \mathbb{E}\left(\mathbb{1}\{\check{j} = \infty\}L(\mathtt{CI}_{z,\alpha}(Y))\mathbb{1}\{H_2\}\right).
\end{aligned} \tag{D.60}$$

We start with the second term

(D.61)

$$
\begin{aligned}
&\mathbb{E}(\mathbb{1}\{\check{j}=\infty\}L(\mathtt{CI}_{z,\alpha}(Y))\mathbb{1}\{H_2\}) \\
&\leq \mathbb{E}\Big(\mathbb{1}\{\check{j}=\infty\}(t_{hi}-t_{lo})\big(\mathbb{1}\{i_m\leq L-1\}+\mathbb{1}\{i_m\geq L\}\mathbb{1}\{i_l=L=i_r\} \\
&\qquad +\mathbb{1}\{i_m\geq U+1\}+\mathbb{1}\{i_m\leq U\}\mathbb{1}\{i_r=U-1=i_l\}\big)\Big) \\
&\leq \frac{2}{n}\Big(\sum_{j=1}^{J-1}6\Phi\Big(-\frac{\rho_m(\frac{\sigma}{\sqrt{n}};\tilde{h})}{\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\frac{2^{J-j}}{n}\frac{2^{J-j}}{\sqrt{2^{J-j}}\gamma_l\sigma\sqrt{2}}\Big)+2\Phi\Big(-\frac{\rho_m(\frac{\sigma}{\sqrt{n}};\tilde{h})}{\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\frac{1}{n}\frac{1}{2\sqrt{3}\sigma}+z_{\alpha_1}\Big)\Big) \\
&\leq \rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}(24\gamma_l 2^{-\frac{3}{2}}\frac{1}{1-\sqrt{\frac{1}{8}}}\check{C}+4\cdot 2\sqrt{6}\check{Q}_2),
\end{aligned}
$$

where $\check{C}=\sup_{t>0}t\Phi(-t)$, $\check{Q}_2=\sup_{t>0}t\Phi(z_{\alpha_1}-t)$.

Now we turn to the first term and split based on $\{i_l=i_m\}$ and $\{i_l\neq i_m\}$ as follows

(D.62)

$$
\begin{aligned}
&\mathbb{E}(\mathbb{1}\{\check{j}=\infty\}L(\mathtt{CI}_{z,\alpha}(Y))\mathbb{1}\{H_1\}) \\
\leq&\mathbb{E}\Big(\mathbb{1}\{\check{j}=\infty\}(t_{hi}-t_{lo})\big(\mathbb{1}\{i_l=n=U=i_r+1\} \\
&\qquad +\mathbb{1}\{i_r=-1=L-1=i_l-1\}+\mathbb{1}\{L+1\leq i_l=i_r+1\leq U-1\}\big)\Big) \\
\leq&\underbrace{\mathbb{E}\Big(\mathbb{1}\{\check{j}=\infty\}(t_{hi}-t_{lo})\mathbb{1}\{i_l=i_m\}\big(\mathbb{1}\{i_l=n=U=i_r+1\} +\mathbb{1}\{i_r=-1=L-1=i_l-1\}+\mathbb{1}\{L+1\leq i_l=i_r+1\leq U-1\}\big)\Big)}_{\kappa_1} \\
&+\underbrace{\mathbb{E}\Big(\mathbb{1}\{\check{j}=\infty\}(t_{hi}-t_{lo})\mathbb{1}\{i_l\neq i_m\}\big(\mathbb{1}\{i_l=n=U=i_r+1\} +\mathbb{1}\{i_r=-1=L-1=i_l-1\}+\mathbb{1}\{L+1\leq i_l=i_r+1\leq U-1\}\big)\Big)}_{\kappa_2}.
\end{aligned}
$$

We will bound κ_1 and κ_2 in Inequality (D.62) separately, we start with κ_2. Note that the event $H_1\cap\{i_l\neq i_m\}$ is a subset of $\{i_m\notin[L,U]\}\cup\{i_m\in[L,U],i_l\neq i_m,i_l=i_r+1\}$. This fact gives

(D.63)

$$
\begin{aligned}
\kappa_2=&\mathbb{E}(\mathbb{1}\{\check{j}=\infty\}(t_{hi}-t_{lo})\mathbb{1}\{H_1\}\mathbb{1}\{i_l\neq i_m\}) \\
\leq&\mathbb{E}(\mathbb{1}\{\check{j}=\infty\}(t_{hi}-t_{lo})\mathbb{1}\{i_m\notin[L,U]\}) \\
&+\mathbb{E}(\mathbb{1}\{\check{j}=\infty\}(t_{hi}-t_{lo})\mathbb{1}\{i_m\in[L,U]\}\mathbb{1}\{i_l\neq i_m\}\mathbb{1}\{i_l=i_r+1\}).
\end{aligned}
$$

To bound the two expectations, note two facts. One is that $i_m \notin [L, U]$ implies $\tilde{\mathrm{j}} < J$. Another is that $\{\check{j} = \infty, i_m \in [L, U], i_l \neq i_m, i_l = i_r + 1\}$ is a subset of $\{\check{j} = \infty, i_m \in [L, U], i_l \geq i_m + 1\} \cup \{\check{j} = \infty, i_m \in [L, U], i_r + 1 \leq i_m - 1\}$. This two facts give that

(D.64)

$$\begin{aligned}
\kappa_2 \leq & \frac{3}{n}\Bigg(\sum_{j=1}^{J-1} 6\Phi(-(\frac{2^{J-j}}{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})})^{\frac{3}{2}}\frac{1}{2\gamma_l}) + \\
& \quad 2\mathbb{E}(\Phi(-\frac{\rho_m(\frac{\sigma}{\sqrt{n}};\tilde{h})}{\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\frac{1}{n\sqrt{12}\sigma})\mathbb{1}\{\check{j} = \infty, i_m \in [L, U]\})\Bigg) \\
\leq & \frac{3}{n}\left(\sum_{j=1}^{J-1} 6\Phi(-(\frac{2^{J-j}}{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})})^{\frac{3}{2}}\frac{1}{2\gamma_l}) + 2\Phi(-(\frac{1}{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})})^{\frac{3}{2}}\frac{1}{\sqrt{24}})\right) \\
\leq & \rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})} \cdot 3 \cdot (12\gamma_l + 2\sqrt{24})\check{C}.
\end{aligned}$$

Now we turn to κ_1 in Inequality (D.62). We discuss the four settings: $i_m = 0$, $i_m = n$, $2 \leq i_m \leq n-2$, $(i_m - 1)(i_m - n + 1) = 0$.

Note that under the case $(i_m - 1)(i_m - n + 1) = 0$, we have $\mathfrak{D}_z(n, f) \geq \frac{1}{n}$. Therefore, in this case

(D.65)

$$\begin{aligned}
\kappa_1 = & \mathbb{E}\Big(\mathbb{1}\{\check{j} = \infty\}(t_{hi} - t_{lo})\mathbb{1}\{i_l = i_m\}\big(\mathbb{1}\{i_l = n = U = i_r + 1\} \\
& \quad + \mathbb{1}\{i_r = -1 = L - 1 = i_l - 1\} + \mathbb{1}\{L + 1 \leq i_l = i_r + 1 \leq U - 1\}\big)\Big) \\
\leq & \frac{2}{n} \leq 2\mathfrak{D}_z(n, f).
\end{aligned}$$

Now we turn to the cases $(i_m - 1)(i_m - n + 1) \neq 0$. Note that under the event $H_1 = \{i_l = n = U = i_r + 1\} \cup \{i_r = -1 = L - 1 = i_l - 1\} \cup \{L + 1 \leq i_l = i_r + 1 \leq U - 1\}$, we have $t_{lo} \leq i_l/n \leq t_{hi}$. Therefore, under the case $(i_m - 1)(i_m - n + 1) \neq 0$, we can split $t_{hi} - t_{lo}$ into two non-negative parts: $t_{hi} - i_l/n$ and $i_l/n - t_{lo}$. This gives an alternative form of κ_1:

$$\begin{aligned}
\kappa_1 = & \underbrace{\mathbb{E}\Big(\mathbb{1}\{\check{j} = \infty\}(t_{hi} - i_l/n)\mathbb{1}\{i_l = i_m\}\mathbb{1}\{H_1\}\Big)}_{\varphi_1} \\
& + \underbrace{\mathbb{E}\Big(\mathbb{1}\{\check{j} = \infty\}(i_l/n - t_{lo})\mathbb{1}\{i_l = i_m\}\mathbb{1}\{H_1\}\Big)}_{\varphi_2}.
\end{aligned} \tag{D.66}$$

Due to the symmetric nature of the procedure, the case $(i_m-1)(i_m-n+1) \neq 0$, and the event $\{i_l = i_m\} \cap \{i_l = n = U = i_r + 1\} \cup \{i_r = -1 = L-1 = i_l - 1\} \cup \{L+1 \leq i_l = i_r + 1 \leq U-1\}$, we only need to bound the first term (φ_1), and the second term (φ_2) shares the similar (symmetric) bound.

(D.67)

$$\begin{aligned}\varphi_1 &= \mathbb{E}\Big(\mathbb{1}\{\check{j}=\infty\}(t_{hi}-i_l/n)\mathbb{1}\{i_l=i_m\}\big(\mathbb{1}\{i_l=n=U=i_r+1\} \\ &\qquad + \mathbb{1}\{i_r=-1=L-1=i_l-1\} + \mathbb{1}\{L+1\leq i_l=i_r+1\leq U-1\}\big)\Big) \\ &= \mathbb{E}\Big(\mathbb{1}\{\check{j}=\infty\}(t_{hi}-i_l/n)\mathbb{1}\{i_l=i_m\} \\ &\qquad \big(\mathbb{1}\{i_r=-1=L-1=i_l-1\} + \mathbb{1}\{L+1\leq i_l=i_r+1\leq U-1\}\big)\Big).\end{aligned}$$

We further simplify φ_1 by analyzing the event $H_3 = \{\check{j} = \infty, i_l = i_m\} \cap (\{i_r = -1 = L-1 = i_l - 1\} \cup \{L+1 \leq i_l = i_r + 1 \leq U-1\})$. Note that H_3, under the case $(i_m - 1)(i_m - n + 1) \neq 0$ can be alternatively written as

$$\begin{aligned}H_3 = \{\check{j}=\infty\} \cap \{i_l = i_r+1, i_l \leq U-1, i_l = i_m\} \cap \\ \{(i_m-1)(i_m-n+1)\neq 0\} \cap \{i_l = L = 0, \text{ or } i_l \geq L+1\}.\end{aligned} \tag{D.68}$$

This event is non-empty only when $i_m \leq n-2$. This means when $i_m = n$, $\varphi_1 = 0$. Therefore, we only need to bound φ_1 under the case $\{i_m \leq n-2, i_m \neq 1\}$. Now we will simplify $t_{hi} - i_l/n$ under event H_3 under this case. From now on, this case is taken as default. Note that when $\{L+1 \leq i_l = i_r+1 \leq U-1\}$, we have that $i_{hi} = i_l + 1$, $i_{lo} = i_l - 1$. When $\{i_r = -1 = L-1 = i_l - 1\}$, i_{hi} is not defined in the algorithm. We define $i_{hi} = i_l + 1$ for the event H_3 under the case $\{i_m \leq n-2, i_m \neq 1\}$. Clearly, this definition is consistent with the ones already defined in the algorithm. Now we introduce two quantities:

$$t_{raw}(i) = \frac{y_{e,i-1} - y_{e,i} - \sqrt{3}\sigma(z_{3,i-1} - z_{3,i}) + 2\sqrt{6}\sigma z_{\alpha_2}}{n(y_{e,i+1} - y_{e,i} - \sqrt{3}\sigma(z_{3,i+1} - z_{3,i}) + 2\sqrt{6}\sigma z_{\alpha_2})}, \tag{D.69}$$

$$q(i) = n(y_{e,i+1} - y_{e,i} - \sqrt{3}\sigma(z_{3,i+1} - z_{3,i}) + 2\sqrt{6}\sigma z_{\alpha_2}). \tag{D.70}$$

Under event H_3 under the case $\{i_m \leq n-2, i_m \neq 1\}$, t_{hi} can be expressed as

$$t_{hi} = \begin{cases} \Big((t_{raw}(i_{hi}) + \frac{i_{hi}}{n}) \vee \frac{i_{hi}-1}{n}\Big) \wedge \frac{i_{hi}}{n}, & q(i_{hi}) > 0 \\ i_l/n = t_{lo}, & q(i_{hi}) \leq 0 \end{cases}. \tag{D.71}$$

Note that under the event H_3 under the case $\{i_m \le n-2, i_m \ne 1\}$, we can further simplify $t_{hi} - i_l/n$ using $i_m = i_l \le n-2$ and $i_{hi} = i_l + 1$ as follows

$$t_{hi} - i_l/n = \begin{cases} (t_{raw}(i_m+1) + \frac{1}{n})_+ \wedge \frac{1}{n}, & q(i_m+1) > 0 \\ 0, & q(i_m+1) \le 0 \end{cases}. \tag{D.72}$$

Note that $(t_{raw}(i_m+1), q(i_m+1))$ is independent with $(\mathtt{Y}_l, \mathtt{Y}_s, \mathtt{Y}_{e,1})$. Plugging in the simplified $(t_{hi} - i_l/n)$ in Equation (D.72) (under event H_3 under case $\{i_m \le n-2, i_m \ne 1\}$) to φ_1 in Equation (D.67) and taking conditional expectation with respect to $(\mathtt{Y}_l, \mathtt{Y}_s, \mathtt{Y}_{e,1})$ give that

$$\begin{aligned} \varphi_1 &= \mathbb{E}((t_{hi} - i_l/n)\mathbb{1}\{H_3\}) \\ &= \mathbb{E}\left(\left(\left(t_{raw}(i_m+1) + \frac{1}{n}\right)_+ \wedge \frac{1}{n}\right)\mathbb{1}\{H_3\}\mathbb{1}\{q(i_m+1) > 0\}\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(\left(\left(t_{raw}(i_m+1) + \frac{1}{n}\right)_+ \wedge \frac{1}{n}\right)\mathbb{1}\{H_3\}\mathbb{1}\{q(i_m+1) > 0\}\middle| \mathtt{Y}_l, \mathtt{Y}_s, \mathtt{Y}_{e,1}\right)\right) \\ &= \mathbb{E}\left(\left(\left(t_{raw}(i_m+1) + \frac{1}{n}\right)_+ \wedge \frac{1}{n}\right)\mathbb{1}\{q(i_m+1) > 0\}\right) P(H_3). \end{aligned} \tag{D.73}$$

Now we introduce the shorthand for the error terms $\zeta_i = y_{e,i} - f(x_i) - \sqrt{3}\sigma z_{3,i}$. Clearly, $\{\frac{\zeta_i}{\sqrt{6}\sigma}\} \overset{i.i.d.}{\sim} N(0,1)$, and

$$t_{raw}(i_m+1) + \frac{1}{n} = \frac{f(x_{i_m}) - 2f(x_{i_m+1}) + f(x_{i_m+2}) + \zeta_{i_m} - 2\zeta_{i_m+1} + \zeta_{i_m+2} + 4\sqrt{6}\sigma z_{\alpha_2}}{n(f(x_{i_m+2}) - f(x_{i_m+1}) + \zeta_{i_m+2} - \zeta_{i_m+1} + 2\sqrt{6}\sigma z_{\alpha_2})}.$$

Therefore, when we, with a bit abuse of the notation, denote the event A_0 only in this proof to be the following event:

$$\begin{aligned} A_0 = \Big\{ \zeta_{i_m+2} &\ge -\frac{f(x_{i_m+2}) - f(x_{i_m+1})}{6} - \sqrt{6}\sigma z_{\alpha_2}, \\ \zeta_{i_m+1} &\le \frac{f(x_{i_m+2}) - f(x_{i_m+1})}{6} + \sqrt{6}\sigma z_{\alpha_2}, \\ \zeta_{i_m} &\ge -\frac{f(x_{i_m+2}) - f(x_{i_m+1})}{6} - \sqrt{6}\sigma z_{\alpha_2} \Big\} \end{aligned} \tag{D.74}$$

, we have, on event A_0,

$$t_{raw}(i_m+1) + \frac{1}{n} \ge -\frac{1}{n},$$

$$f(x_{i_m+2}) - f(x_{i_m+1}) + \zeta_{i_m+2} - \zeta_{i_m+1} + 2\sqrt{6}\sigma z_{\alpha_2} \ge \frac{2(f(x_{i_m+2}) - f(x_{i_m+1}))}{3}.$$

With a bit abuse of notation, denote event B only in this proof to be

(D.75)
$$B = \{\zeta_{i_m} - 2\zeta_{i_m+1} + \zeta_{i_m+2} + f(x_{i_m}) - 2f(x_{i_m+1}) + f(x_{i_m+2}) + 4\sqrt{6}\sigma z_{\alpha_2} \geq 0\}.$$

Then on $B^c \cap A_0$, $t_{raw}(i_m+1) + \frac{1}{n} < 0$; on $B \cap A_0$, $t_{raw}(i_m+1) + \frac{1}{n} \geq 0$.

Further, we have

(D.76)
$$\begin{aligned} P(A_0^c) &\leq P(\zeta_{i_m+2} < -\frac{f(x_{i_m+2}) - f(x_{i_m+1})}{6} - \sqrt{6}\sigma z_{\alpha_2}) \\ &\quad + P(\zeta_{i_m+1} > \frac{f(x_{i_m+2}) - f(x_{i_m+1})}{6} + \sqrt{6}\sigma z_{\alpha_2}) \\ &\quad + P(\zeta_{i_m} < -\frac{f(x_{i_m+2}) - f(x_{i_m+1})}{6} - \sqrt{6}\sigma z_{\alpha_2}) \\ &= 3\Phi(-\frac{f(x_{i_m+2}) - f(x_{i_m+1})}{6\sqrt{6}\sigma} - z_{\alpha_2}) \leq 3\Phi(-\left(\frac{1}{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\right)^{\frac{3}{2}} \frac{1}{6\sqrt{12}} - z_{\alpha_2}) \\ &\leq n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}18\sqrt{12}\check{Q}_3, \end{aligned}$$

where $\check{Q}_3 = \sup_{x>0} x\Phi(-x - z_{\alpha_2})$.

Therefore, going back to Inequality (D.73) and spliting the entire probability space by A_0 and B give

(D.77)
$$\begin{aligned} \varphi_1 &\leq \mathbb{E}\Bigg(\left(\left(t_{raw}(i_m+1) + \frac{1}{n}\right) \wedge \frac{1}{n}\right) \mathbb{1}\{q(i_m+1) > 0, -\frac{1}{n} \leq t_{raw}(i_m+1)\}\Bigg) \\ &= \mathbb{E}\Bigg(\left(\left(t_{raw}(i_m+1) + \frac{1}{n}\right) \wedge \frac{1}{n}\right) \mathbb{1}\{q(i_m+1) > 0, -\frac{1}{n} \leq t_{raw}(i_m+1)\} \\ &\qquad (\mathbb{1}\{A_0 \cap B\} + \mathbb{1}\{A_0 \cap B^c\} + \mathbb{1}\{A_0^c\})\Bigg) \\ &\leq \mathbb{E}\Bigg(\left(t_{raw}(i_m+1) + \frac{1}{n}\right) \mathbb{1}\{A_0 \cap B\}\Bigg) + \frac{1}{n}P(A_0^c) \\ &\leq \mathbb{E}\Bigg(\left(t_{raw}(i_m+1) + \frac{1}{n}\right) \mathbb{1}\{A_0 \cap B\}\Bigg) + \rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}18\sqrt{12}\check{Q}_3. \end{aligned}$$

Further, convexity gives that

$$\sup\{Z(h) : h(x_i) = f(x_i), 0 \leq i \leq n\} - \frac{i_m}{n} = \frac{f(x_{i_m}) - f(x_{i_m+1})}{n(f(x_{i_m+2}) - f(x_{i_m+1}))} + \frac{1}{n}.$$

Therefore, we have

(D.78)
$$\mathbb{E}\left(\left(t_{raw}(i_m+1)+\frac{1}{n}\right)\mathbb{1}\{A_0\cap B\}\right)=\sup\{Z(h):h(x_i)=f(x_i),0\le i\le n\}$$
$$-\frac{i_m}{n}+\mathbb{E}\left(\left(t_{raw}(i_m+1)-\frac{f(x_{i_m})-f(x_{i_m+1})}{n(f(x_{i_m+2})-f(x_{i_m+1}))}\right)\mathbb{1}\{A_0\cap B\}\right).$$

Further, on event A_0, we have
(D.79)
$$t_{raw}(i_m+1)-\frac{f(x_{i_m})-f(x_{i_m+1})}{n(f(x_{i_m+2})-f(x_{i_m+1}))}=$$
$$\frac{\zeta_{i_m}(f(x_{i_m+2})-f(x_{i_m+1}))+\zeta_{i_m+1}(f(x_{i_m})-f(x_{i_m+2}))}{n(f(x_{i_m+2})-f(x_{i_m+1})+\zeta_{i_m+2}-\zeta_{i_m+1}+2\sqrt{6}\sigma z_{\alpha_2})(f(x_{i_m+2})-f(x_{i_m+1}))}$$
$$+\frac{\zeta_{i_m+2}(f(x_{i_m+1})-f(x_{i_m}))+2\sqrt{6}\sigma z_{\alpha_2}\,(f(x_{i_m+2})-f(x_{i_m}))}{n(f(x_{i_m+2})-f(x_{i_m+1})+\zeta_{i_m+2}-\zeta_{i_m+1}+2\sqrt{6}\sigma z_{\alpha_2})(f(x_{i_m+2})-f(x_{i_m+1}))},$$
$$\le\Big(|\zeta_{i_m}|(f(x_{i_m+2})-f(x_{i_m+1}))+|\zeta_{i_m+1}|(f(x_{i_m})-f(x_{i_m+2}))$$
$$+|\zeta_{i_m+2}|(f(x_{i_m+1})-f(x_{i_m}))+2\sqrt{6}\sigma z_{\alpha_2}\,(f(x_{i_m+2})-f(x_{i_m}))\Big)$$
$$\frac{1}{\frac{2}{3}n\left(f(x_{i_m+2})-f(x_{i_m+1})\right)^2}$$
$$\le\sqrt{6}\sigma\frac{3}{2n}\Big(\Big|\frac{\zeta_{i_m}}{\sqrt{6}\sigma}\Big|+2\Big|\frac{\zeta_{i_m+1}}{\sqrt{6}\sigma}\Big|+\Big|\frac{\zeta_{i_m+2}}{\sqrt{6}\sigma}\Big|+4z_{\alpha_2}\Big)\frac{1}{f(x_{i_m+2})-f(x_{i_m+1})}.$$

Therefore,
(D.80)
$$\mathbb{E}\left(\left(t_{raw}(i_m+1)-\frac{f(x_{i_m})-f(x_{i_m+1})}{n(f(x_{i_m+2})-f(x_{i_m+1}))}\right)\mathbb{1}\{A_0\cap B\}\right)$$
$$\le\mathbb{E}\left(\sqrt{6}\sigma\frac{3}{2n}\Big(\Big|\frac{\zeta_{i_m}}{\sqrt{6}\sigma}\Big|+2\Big|\frac{\zeta_{i_m+1}}{\sqrt{6}\sigma}\Big|+\Big|\frac{\zeta_{i_m+2}}{\sqrt{6}\sigma}\Big|+4z_{\alpha_2}\Big)\frac{1}{f(x_{i_m+2})-f(x_{i_m+1})}\mathbb{1}\{A_0\cap B\}\right)$$
$$\le\mathbb{E}\left(\sqrt{6}\sigma\frac{3}{2n}\Big(\Big|\frac{\zeta_{i_m}}{\sqrt{6}\sigma}\Big|+2\Big|\frac{\zeta_{i_m+1}}{\sqrt{6}\sigma}\Big|+\Big|\frac{\zeta_{i_m+2}}{\sqrt{6}\sigma}\Big|+4z_{\alpha_2}\Big)\frac{1}{f(x_{i_m+2})-f(x_{i_m+1})}\right)$$
$$\le\sqrt{6}\sigma\frac{3}{2n}(4\check{Q}_4+4z_{\alpha_2})\frac{1}{f(x_{i_m+2})-f(x_{i_m+1})}$$
$$\le\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\sqrt{12}(6\check{Q}_4+6z_{\alpha_2}),$$

where $\check{Q}_4=\int_{-\infty}^{\infty}|x|\frac{1}{\sqrt{2\pi}}\exp{(-x^2/2)}dx$.

Going back to Equation (D.67), we have

$$\varphi_1 \leq \rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\sqrt{12}(6\check{Q}_4 + 6z_{\alpha_2} + 18\check{Q}_3) + \sup\{Z(h) : h(x_i) = f(x_i), 0 \leq i \leq n\} - \frac{i_m}{n}. \tag{D.81}$$

Therefore, under the case $(i_m - n + 1)(i_m) \neq 0$, φ_1 is bounded.
Similarly, for the φ_2 in Equation (D.66), we have

$$\begin{aligned}\varphi_2 = \mathbb{E}\Big(&\mathbb{1}\{\check{j} = \infty\}(i_l/n - t_{lo})\mathbb{1}\{i_l = i_m\}\big(\mathbb{1}\{i_l = n = U = i_r + 1\} \\ &+ \mathbb{1}\{i_r = -1 = L - 1 = i_l - 1\} + \mathbb{1}\{L + 1 \leq i_l = i_r + 1 \leq U - 1\}\big)\Big) \\ \leq{}& \rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}\sqrt{12}(6\check{Q}_4 + 6z_{\alpha_2} + 18\check{Q}_3) \\ &+ \frac{i_m}{n} - \inf\{Z(h) : h(x_i) = f(x_i), 0 \leq i \leq n\}.\end{aligned} \tag{D.82}$$

Therefore, under the case $(i_m - n + 1)(i_m) \neq 0$,

$$\begin{aligned}\kappa_1 = \mathbb{E}\Big(&\mathbb{1}\{\check{j} = \infty\}(t_{hi} - t_{lo})\mathbb{1}\{i_l = i_m\}\big(\mathbb{1}\{i_l = n = U = i_r + 1\} \\ &+ \mathbb{1}\{i_r = -1 = L - 1 = i_l - 1\} + \mathbb{1}\{L + 1 \leq i_l = i_r + 1 \leq U - 1\}\big)\Big) \\ \leq{}& \rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}2\sqrt{12}(6\check{Q}_4 + 6z_{\alpha_2} + 18\check{Q}_3) + \\ &\sup\{Z(h) : h(x_i) = f(x_i), 0 \leq i \leq n\} - \inf\{Z(h) : h(x_i) = f(x_i), 0 \leq i \leq n\} \\ ={}& \rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};\tilde{h})}2\sqrt{12}(6\check{Q}_4 + 6z_{\alpha_2} + 18\check{Q}_3) + \mathfrak{D}_z(n, f).\end{aligned} \tag{D.83}$$

All the cases analyzed, and all the terms added up (Inequatliy (D.83), Inequality (D.65), Inequality (D.64), Inequality (D.62), Inequality (D.61)) give the statement

$$\begin{aligned}&\mathbb{E}(\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{t_{hi} - t_{lo} < \frac{3}{n}\}L(\mathtt{CI}_{z,\alpha}(Y))) \\ &\leq \check{c}_{3,\alpha} \sup_{h \in \mathcal{G}_n(f)} \rho_z(\frac{\sigma}{\sqrt{n}};h)\sqrt{n\rho_z(\frac{\sigma}{\sqrt{n}};h)} + 2\mathfrak{D}_z(n, f).\end{aligned} \tag{D.84}$$

□

PROOF OF LEMMA C.37.

(D.85)
$$\begin{aligned}
&P(\hat{\mathtt{j}} \geq \mathtt{j}^{\mathtt{w}} + K + 1) = \mathbb{E}(\mathbb{1}\{\hat{\mathtt{j}} \geq \mathtt{j}^{\mathtt{w}} + K + 1\}\mathbb{1}\{\mathtt{j}^{\mathtt{w}} < \infty\}) \\
&\leq \mathbb{E}(\mathbb{1}\{\forall \mathtt{j}^{\mathtt{w}} + 1 \leq j \leq \mathtt{j}^{\mathtt{w}} + K, \\
&\min\{Y_{j,\hat{\mathtt{i}}_j - 6,s} - Y_{j,\hat{\mathtt{i}}_j - 5,s}, Y_{j,\hat{\mathtt{i}}_j + 6,s} - Y_{j,\hat{\mathtt{i}}_j + 5,s}\} > 2\gamma_s\sqrt{2}\sigma\sqrt{2^{J-j}}\}\mathbb{1}\{\mathtt{j}^{\mathtt{w}} < \infty\}) \\
&\leq \Phi(-2)^K \mathbb{E}(\mathbb{1}\{\mathtt{j}^{\mathtt{w}} < \infty\}) \leq \Phi(-2)^K.
\end{aligned}$$

The second inequality is by taking conditional expectation on the localization copy of the observation (i.e. $\mathtt{Y}_l$), and the fact that for the iteration steps j such that $\mathtt{j}^{\mathtt{w}} + 1 \leq j \leq \mathtt{j}^{\mathtt{w}} + K$ the target interval is more than 6 blocks away from the estimated one. □

PROOF OF LEMMA C.38. Given the symmetric nature of our procedure, we only need to prove

$$\mathbb{E}(\mathbb{1}\{E\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{F_1^c\}) \leq \alpha_1. \tag{D.86}$$

Note that, when $\check{j} = \infty$, $E = \{Z(f) \in [\big(\hat{\mathtt{i}}_{\hat{\mathtt{j}}} - (6 \cdot 2^{K_{\alpha/2}+1} - 2) - 1\big)\frac{2^{J-\hat{\mathtt{j}}}}{n} - \frac{1}{2n}, \big(\hat{\mathtt{i}}_{\hat{\mathtt{j}}} + (6 \cdot 2^{K_{\alpha/2}+1} - 2)\big)\frac{2^{J-\hat{\mathtt{j}}}}{n} - \frac{1}{2n}] \cap [0,1]\} \subset \{\frac{\hat{\mathtt{i}}_{\hat{\mathtt{j}}} - (6 \cdot 2^{K_{\alpha/2}+1} - 2) - 2}{n} < Z(f) < \frac{\hat{\mathtt{i}}_{\hat{\mathtt{j}}} + 6 \cdot 2^{K_{\alpha/2}+1} - 2}{n}\}$

Let $L_0 = \hat{\mathtt{i}}_{\hat{\mathtt{j}}} - (6 \cdot 2^{K_{\alpha/2}+1} - 2) - 2$, $U_0 = \hat{\mathtt{i}}_{\hat{\mathtt{j}}} + 6 \cdot 2^{K_{\alpha/2}+1} - 2$. Hence we know that when $L_0 \geq 1$, $L = L_0 - 1$; when $U_0 \leq n - 1$, $U = U_0 + 1$.

Let $i_m = \min\{k : f(x_k) = \min\{f(x_i) : 0 \leq i \leq n\}\}$. Then we know that, on E, $L_0 \leq i_m \leq U_0$. And also $i_m = n$ implies F_1, hence we only needs to consider the case $i_m \leq n - 1$ to compute F_1^c. And $\{i_m \leq n - 1\} \cap \{L_0 \leq i_m \leq U_0\}$ implies that $i_m < U$.

We also know that $\{y_{e,i} + \sqrt{3}\sigma z_{3,i} : 0 \leq i \leq n\}$, $\{y_{e,i} - \sqrt{3}\sigma z_{3,i} : 0 \leq i \leq n\}$, $\{y_{s,i} : 0 \leq i \leq n\}$, $\{y_{l,i} : 0 \leq i \leq n\}$ are independent random variables.

Therefore,

(D.87)
$$\begin{aligned}
&\mathbb{E}(\mathbb{1}\{E\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{F_1^c\}) \leq \mathbb{E}\Bigg(\mathbb{E}\Big(\mathbb{1}\{E\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{i_m < U\} \\
&\mathbb{1}\{y_{e,i_m} + \sqrt{3}\sigma z_{3,i_m} - (y_{e,i_m+1} + \sqrt{3}\sigma z_{3,i_m+1}) > 2\sqrt{3}\sigma z_{\alpha_1}\}\Big|\mathtt{Y}_s, \mathtt{Y}_l\Big)\Bigg) \leq \alpha_1.
\end{aligned}$$

□

Proof of Lemma C.39. The event $E\cap\{\check{j}=\infty\}\cap F_1\cap F_2\cap\{(i_l-U)(i_r-L+1)=0\}$ is the union of the following four events.

$$\begin{aligned}G_1&=E\cap\{\check{j}=\infty\}\cap F_1\cap F_2\cap\{i_l=U,U\neq n\},\\G_2&=E\cap\{\check{j}=\infty\}\cap F_1\cap F_2\cap\{i_l=U,U=n\},\\G_3&=E\cap\{\check{j}=\infty\}\cap F_1\cap F_2\cap\{i_r=L-1,L=0\},\\G_4&=E\cap\{\check{j}=\infty\}\cap F_1\cap F_2\cap\{i_r=L-1,L\neq 0\}.\end{aligned}\tag{D.88}$$

Since $\{U\neq n\}\cap\{\check{j}=\infty\}$ means $U_0\le U-1\le n-2$; and on $E\cap\{\check{j}=\infty\}\cap F_1\cap F_2$ we have $i_l\le\min\{k:f(x_k)=\min\{f(x_i)\}\}$ and $\min\{k:f(x_k)=\min\{f(x_i)\}\}\le U_0$, we know that $G_1=\emptyset$. Similarly, we have $G_4=\emptyset$. Also, on $E\cap\{\check{j}=\infty\}\cap F_1\cap F_2$, we know that $i_l\le i_r+1$, hence we have $G_2\cap G_3=\emptyset$.

Also, on G_2, we know that $f(x_n)=\min\{f(x_i)\}$ and $f(x_k)>\min\{f(x_i):0\le i\le n\}$ for all $k\le n-1$, which implies that $Z(f)\ge\frac{f(x_{n-1})-f(x_n)}{n(f(x_{n-2})-f(x_{n-1}))}+\frac{n-1}{n}$.

Suppose $Y_{e,1}=\{y_{e,i}+\sqrt{3}\sigma z_{3,i}:(L-1)\vee 0\le i\le(U+1)\wedge n\}$, $Y_{e,2}=\{y_{e,i}-\sqrt{3}\sigma z_{3,i}:(L-1)\vee 0\le i\le(U+1)\wedge n\}$. Then we know that $\mathtt{Y}_l$,$\mathtt{Y}_s$,$Y_{e,1}$,$Y_{e,2}$ are independent.

If we denote $\kappa_{i,1}=y_{e,i}+\sqrt{3}\sigma z_{3,i}-f(x_i)$, $\kappa_{i,2}=y_{e,i}-\sqrt{3}\sigma z_{3,i}-f(x_i)$, then we know that on G_2 when we further have $\kappa_{n,2}\ge-\sqrt{6}\sigma z_{\alpha_2}$, $\kappa_{n-1,2}\le\sqrt{6}\sigma z_{\alpha_2}$, $\kappa_{n-2,2}\ge-\sqrt{6}\sigma z_{\alpha_2}$, then $t_{lo}\le Z(f)$. Further, $t_{hi}\ge Z(f)$ trivially holds on G_2.

We have similar analysis for G_3. Hence we know that

$$\begin{aligned}&\mathbb{E}\Big(\mathbb{1}\{Z(f)\notin\mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{E\}\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{F_1\cap F_2\}\mathbb{1}\{(i_l-U)(i_r-L+1)=0\}\Big)\\&=\mathbb{E}\Big(\mathbb{1}\{Z(f)\notin\mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{G_2\}\Big)+\mathbb{E}\Big(\mathbb{1}\{Z(f)\notin\mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{G_3\}\Big)\\&=\mathbb{E}\Big(\mathbb{E}(\mathbb{1}\{t_{lo}>Z(f)\}|\mathtt{Y}_l,\mathtt{Y}_s,Y_{e,1})\mathbb{1}\{G_2\}\Big)\\&\quad+\mathbb{E}\Big(\mathbb{E}(\mathbb{1}\{t_{hi}<Z(f)\}|\mathtt{Y}_l,\mathtt{Y}_s,Y_{e,1})\mathbb{1}\{G_1\}\Big)\\&\le 3\alpha_2P(G_2)+3\alpha_2P(G_3)\\&\le 3\alpha_2P(\mathbb{1}\{E\}\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{F_1\cap F_2\}\mathbb{1}\{(i_l-U)(i_r-L+1)=0\}).\end{aligned}$$

□

PROOF OF LEMMA C.40.

$$\begin{aligned}
&\mathbb{E}\Big(\mathbb{1}\{Z(f) \notin \mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{E\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{F_1 \cap F_2\}\mathbb{1}\{(i_l - U)(i_r - L + 1) \neq 0\} \\
&\qquad \mathbb{1}\{i_{hi} - i_{lo} \le 2, 0 < i_{lo}, i_{hi} < n\}\Big) \\
&\le \mathbb{E}\Big(\mathbb{1}\{Z(f) > t_{hi}\}\mathbb{1}\{E\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{F_1 \cap F_2\}\mathbb{1}\{(i_l - U)(i_r - L + 1) \neq 0\} \\
&\qquad \mathbb{1}\{i_{hi} - i_{lo} \le 2, 0 < i_{lo}, i_{hi} < n\}\Big) \\
&\quad + \mathbb{E}\Big(\mathbb{1}\{Z(f) < t_{lo}\}\mathbb{1}\{E\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{F_1 \cap F_2\}\mathbb{1}\{(i_l - U)(i_r - L + 1) \neq 0\} \\
&\qquad \mathbb{1}\{i_{hi} - i_{lo} \le 2, 0 < i_{lo}, i_{hi} < n\}\Big).
\end{aligned}$$

The symmetric nature of the procedure means the bound for the first term also applied to the second. We show that for the first term.

Suppose $Y_{e,1} = \{y_{e,i} + \sqrt{3}\sigma z_{3,i} : (L-1) \vee 0 \le i \le (U+1) \wedge n\}$, $Y_{e,2} = \{y_{e,i} - \sqrt{3}\sigma z_{3,i} : (L-1) \vee 0 \le i \le (U+1) \wedge n\}$. Then we know that $\mathtt{Y}_l$,$\mathtt{Y}_s$,$Y_{e,1}$,$Y_{e,2}$ are independent.

On the event $E \cap \{\check{j} = \infty\} \cap F_1 \cap F_2 \cap \{(i_l - U)(i_r - L + 1) \neq 0\} \cap \{i_{hi} - i_{lo} \le 2, 0 < i_{lo}, i_{hi} < n\}$, we know that $|\{k : f(x_k) = \min\{f(x_i) : 0 \le i \le n\}\}| = 1$, we denote this unique element to be i_m. Also, when this event is not empty, we know that $2 \le i_m \le n - 2$. Hence we know that $Z(f) \le \frac{f(x_{i_m}) - f(x_{i_m+1})}{(f(x_{i_m+2}) - f(x_{i_m+1}))/\frac{1}{n}} + \frac{i_m+1}{n}$. If we denote $\kappa_{i,1} = y_{e,i} + \sqrt{3}\sigma z_{3,i} - f(x_i)$, $\kappa_{i,2} = y_{e,i} - \sqrt{3}\sigma z_{3,i} - f(x_i)$, then we know that on event $E \cap \{\check{j} = \infty\} \cap F_1 \cap F_2 \cap \{(i_l - U)(i_r - L + 1) \neq 0\} \cap \{i_{hi} - i_{lo} \le 2, 0 < i_{lo}, i_{hi} < n\}$, if we further have $\kappa_{i_m+2,2} \ge -\sqrt{6}\sigma z_{\alpha_2}, \kappa_{i_m+1,2} \le \sqrt{6}\sigma z_{\alpha_2}, \kappa_{i_m,2} \ge -\sqrt{6}\sigma z_{\alpha_2}$, then $Z(f) \le t_{hi}$.

$$\begin{aligned}
&\mathbb{E}\Big(\mathbb{1}\{Z(f) > t_{hi}\}\mathbb{1}\{E\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{F_1 \cap F_2\}\mathbb{1}\{(i_l - U)(i_r - L + 1) \neq 0\} \\
&\qquad \mathbb{1}\{i_{hi} - i_{lo} \le 2, 0 < i_{lo}, i_{hi} < n\}\Big) \\
&= \mathbb{E}\Big(\mathbb{E}\big(\mathbb{1}\{Z(f) > t_{hi}\}|\mathtt{Y}_l, \mathtt{Y}_s, Y_{e,1}\big)\mathbb{1}\{E\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{F_1 \cap F_2\}\mathbb{1}\{(i_l - U)(i_r - L + 1) \neq 0\} \\
&\quad \mathbb{1}\{i_{hi} - i_{lo} \le 2, 0 < i_{lo}, i_{hi} < n\}\mathbb{1}\{i_{lo} + 1 = i_{hi} - 1 = i_m\}\Big) \\
&\le \mathbb{E}\Big(\mathbb{E}\big(\mathbb{1}\{\kappa_{i_m+2,2} < -\sqrt{6}\sigma z_{\alpha_2} \text{ or } \kappa_{i_m+1,2} > \sqrt{6}\sigma z_{\alpha_2} \text{ or } \kappa_{i_m,2} < -\sqrt{6}\sigma z_{\alpha_2}\}|\mathtt{Y}_l, \mathtt{Y}_s, Y_{e,1}\big) \\
&\quad \mathbb{1}\{E\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{F_1 \cap F_2\}\mathbb{1}\{(i_l - U)(i_r - L + 1) \neq 0\} \\
&\quad \mathbb{1}\{i_{hi} - i_{lo} \le 2, 0 < i_{lo}, i_{hi} < n\}\mathbb{1}\{i_{lo} + 1 = i_{hi} - 1 = i_m\}\Big) \\
&\le 3\alpha_2\mathbb{E}(\mathbb{1}\{E\}\mathbb{1}\{\check{j} = \infty\}\mathbb{1}\{F_1 \cap F_2\}\mathbb{1}\{(i_l - U)(i_r - L + 1) \neq 0\} \\
&\quad \mathbb{1}\{i_{hi} - i_{lo} \le 2, 0 < i_{lo}, i_{hi} < n\}\mathbb{1}\{i_{lo} + 1 = i_{hi} - 1 = i_m\}).
\end{aligned}$$

Therefore,

(D.89)

$$\mathbb{E}\Big(\mathbb{1}\{Z(f)\notin \mathtt{CI}_{z,\alpha}(Y)\}\mathbb{1}\{E\}\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{F_1\cap F_2\}\mathbb{1}\{(i_l-U)(i_r-L+1)\neq 0\}$$
$$\mathbb{1}\{i_{hi}-i_{lo}\leq 2, 0<i_{lo},i_{hi}<n\}\Big)$$
$$\leq 6\alpha_2\mathbb{E}\Big(\mathbb{1}\{E\}\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{F_1\cap F_2\}\mathbb{1}\{(i_l-U)(i_r-L+1)\neq 0\}$$
$$\mathbb{1}\{i_{hi}-i_{lo}\leq 2, 0<i_{lo},i_{hi}<n\}\mathbb{1}\{i_{lo}+1=i_{hi}-1=i_m\}\Big).$$

□

Proof of Lemma C.41.

(D.90)

$$\mathbb{E}((\mathfrak{E}_{\check{j},\tilde{\mathtt{i}}_{\check{j}},e}\frac{1}{2^{J-\check{j}}})^2\mathbb{1}\{\check{j}<\infty\})=\mathbb{E}(\frac{1}{2^{J-\check{j}}}\sigma^2\gamma_e^2\mathbb{1}\{\check{j}<\infty\})$$
$$=\sigma^2\gamma_e^2 2^{\mathtt{j}^*-J}(\sum_{j=1}^{\mathtt{j}^*+2}\mathbb{E}(2^{-\mathtt{j}^*+j}\mathbb{1}\{\check{j}=j\})+\sum_{j=\mathtt{j}^*+3}^{\infty}\mathbb{E}(2^{-\mathtt{j}^*+j}\mathbb{1}\{\check{j}=j\}))$$
$$\leq\sigma^2\gamma_e^2 2^{\mathtt{j}^*-J}\Big(4+\sum_{j=\mathtt{j}^*+3}^{\infty}2^{-\mathtt{j}^*+j}\Phi(-2+\frac{\frac{13}{16}\rho_m(\frac{\sigma}{\sqrt{n}};f)\sqrt{2^{J-\mathtt{j}^*-2}}}{\sigma\gamma_s\sqrt{2}})$$
$$\Phi(-2+\frac{\frac{13}{32}\rho_m(\frac{\sigma}{\sqrt{n}};f)\sqrt{2^{J-\mathtt{j}^*-3}}}{\sigma\gamma_s\sqrt{2}})^{(j-\mathtt{j}^*-3)_+}\Big)$$
$$\leq\sigma^2\gamma_e^2 2^{\mathtt{j}^*-J}(4+\sum_{j=\mathtt{j}^*+3}^{\infty}2^{-\mathtt{j}^*+j}\Phi(-2+\frac{13\sqrt{3}}{\gamma_s 16\sqrt{2}}2^{\frac{-4}{2}})\Phi(-2+\frac{13\sqrt{3}}{\gamma_s 32\sqrt{2}}2^{\frac{-5}{2}})^{(j-\mathtt{j}^*-3)_+}$$
$$\leq\sigma^2\gamma_e^2 2^{\mathtt{j}^*-J}(4+\frac{8\Phi(-2+\frac{13\sqrt{3}}{\gamma_s 16\sqrt{2}}2^{\frac{-4}{2}})}{1-2\Phi(-2+\frac{13\sqrt{3}}{\gamma_s 32\sqrt{2}}2^{\frac{-5}{2}})})$$
$$\leq 2n\rho_m(\frac{\sigma}{\sqrt{n}};f)^2\rho_z(\frac{\sigma}{\sqrt{n}};f)\gamma_e^2\frac{8}{n\rho_z(\frac{\sigma}{\sqrt{n}};f)}(4+8\frac{\Phi(-2+\frac{13\sqrt{3}}{\gamma_s 16\sqrt{2}}2^{\frac{-4}{2}})}{1-2\Phi(-2+\frac{13\sqrt{3}}{\gamma_s 32\sqrt{2}}2^{\frac{-5}{2}})})$$
$$=c_{m1}\rho_m(\frac{\sigma}{\sqrt{n}};f)^2.$$

□

PROOF OF LEMMA C.42.

(D.91)
$$
\begin{aligned}
&\mathbb{E}((\hat{\mathtt{f}} - M(f))^2 \mathbb{1}\{\check{j} < \infty\}) \\
&= \mathbb{E}\Big((\hat{\mathtt{f}} - M(f))^2 \big(\mathbb{1}\{\tilde{\mathtt{j}} > \check{j}\} + \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j}\}\big) \mathbb{1}\{\check{j} < \infty\}\Big) \\
&= \sum_{j_1=2}^{\mathtt{j}^*+1} \mathbb{E}((\hat{\mathtt{f}} - M(f))^2 \mathbb{1}\{\tilde{\mathtt{j}} > \check{j} = j_1\}) + \sum_{j_1=\mathtt{j}^*+2}^{\infty} \mathbb{E}((\hat{\mathtt{f}} - M(f))^2 \mathbb{1}\{\tilde{\mathtt{j}} > \check{j} = j_1\}) \\
&\quad + \mathbb{E}\Big(\big((\hat{\mathtt{f}} - ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}))_+ + (ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}) - M(f))\big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j} = j_1\} \mathbb{1}\{\check{j} < \infty\}\Big) \\
&\le \sum_{j_1=2}^{\mathtt{j}^*+1} \mathbb{E}((\hat{\mathtt{f}} - M(f))^2 \mathbb{1}\{\tilde{\mathtt{j}} > \check{j} = j_1\}) + \sum_{j_1=\mathtt{j}^*+2}^{\infty} \mathbb{E}((\hat{\mathtt{f}} - M(f))^2 \mathbb{1}\{\tilde{\mathtt{j}} > \check{j} = j_1\}) \\
&\qquad + 2\mathbb{E}\Big(\big((\hat{\mathtt{f}} - ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}))_+\big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j}\} \mathbb{1}\{\check{j} < \infty\}\Big) + \\
&\qquad\quad 2\mathbb{E}\Big(\big(ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}) - M(f)\big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j}\} \mathbb{1}\{\check{j} < \infty\}\Big).
\end{aligned}
$$

We have following four lemmas to bound each term respectively.

LEMMA D.1.

$$
\sum_{j_1=2}^{\mathtt{j}^*+1} \mathbb{E}((\hat{\mathtt{f}} - M(f))^2 \mathbb{1}\{\tilde{\mathtt{j}} > \check{j} = j_1\}) \le c_{m3} \rho_m(\frac{\sigma}{\sqrt{n}}; f)^2. \tag{D.92}
$$

LEMMA D.2.

$$
\sum_{j_1=\mathtt{j}^*+2}^{\infty} \mathbb{E}((\hat{\mathtt{f}} - M(f))^2 \mathbb{1}\{\tilde{\mathtt{j}} > \check{j} = j_1\}) \le c_{m4} \rho_m(\frac{\sigma}{\sqrt{n}}; f)^2. \tag{D.93}
$$

LEMMA D.3.

$$
\mathbb{E}(((\hat{\mathtt{f}} - ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}))_+)^2 \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j}\} \mathbb{1}\{\check{j} < \infty\}) \le c_{m5} \rho_m(\frac{\sigma}{\sqrt{n}}; f)^2. \tag{D.94}
$$

LEMMA D.4.

$$
\mathbb{E}((ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}) - M(f))^2 \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j}\} \mathbb{1}\{\check{j} < \infty\}) \le c_{m6} \rho_m(\frac{\sigma}{\sqrt{n}}; f)^2. \tag{D.95}
$$

With these four lemmas, we know that

(D.96)
$$
\mathbb{E}((\hat{\mathtt{f}} - M(f))^2 \mathbb{1}\{\check{j} < \infty\}) \le (c_{m3} + c_{m4} + 2c_{m5} + 2c_{m6}) \rho_m(\tfrac{\sigma}{\sqrt{n}}; f)^2 = c_{m2} \rho_m(\tfrac{\sigma}{\sqrt{n}}; f)^2.
$$

Now we prove these four lemmas. We assume $\check{j} < \infty$ consistently in the remaining of the proof of Lemma C.42 to avoid repeatedly writing $\mathbb{1}\{\check{j} < \infty\}$.

Proof of Lemma D.1. Similarly to the white noise model, we have

$$\sum_{j_1=2}^{\mathtt{j}^*+1} \mathbb{E}((\hat{\mathtt{f}} - M(f))^2 \mathbb{1}\{\tilde{\mathtt{j}} > \check{j} = j_1\})$$
$$\leq \sum_{j_1=2}^{\mathtt{j}^*+1} \mathbb{E}\Big(\big((ave_f(j_1, \hat{\mathtt{i}}_{j_1}+2) - M(f))^2 \mathbb{1}\{Y_{j_1,\hat{\mathtt{i}}_{j_1}+6,s} - Y_{j_1,\hat{\mathtt{i}}_{j_1}+5,s} \leq 2\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j_1}}\}$$
$$+ (ave_f(j_1, \hat{\mathtt{i}}_{j_1}-2) - M(f))^2 \mathbb{1}\{Y_{j_1,\hat{\mathtt{i}}_{j_1}+6,s} - Y_{j_1,\hat{\mathtt{i}}_{j_1}+5,s} \leq 2\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j_1}}\}\big) \mathbb{1}\{\tilde{\mathtt{j}} > j_1\}\Big)$$
$$\leq \sum_{j_1=2}^{\mathtt{j}^*+1} \big((ave_f(j_1, \hat{\mathtt{i}}_{j_1}+2) - M(f))^2 \Phi(2 - \frac{(ave_f(j_1, \hat{\mathtt{i}}_{j_1}+2) - M(f))2^{\frac{1}{2}(J-j_1)}}{3.5\sigma\gamma_s\sqrt{2}})$$
$$+ (ave_f(j_1, \hat{\mathtt{i}}_{j_1}-2) - M(f))^2 \Phi(2 - \frac{(ave_f(j_1, \hat{\mathtt{i}}_{j_1}-2) - M(f))2^{\frac{1}{2}(J-j_1)}}{3.5\sigma\gamma_s\sqrt{2}})\big)\mathbb{E}(\mathbb{1}\{\tilde{\mathtt{j}} > j_1\}),$$

Calculation shows that this is further bounded by

$$\sum_{j_1=2}^{\mathtt{j}^*+1} 2 \cdot 2^{j_1-J}(3.5\sqrt{2}\sigma\gamma_s)^2 V \leq 4 \times 2^{\mathtt{j}^*+1-J}\sigma^2 \frac{49}{2}\gamma_s^2 V$$
$$\leq 32 \times 49 \times 2\gamma_s^2 V \rho_m(\frac{\sigma}{\sqrt{n}}; f)^2 = c_{m3}\rho_m(\frac{\sigma}{\sqrt{n}}; f)^2.$$

V in the inequalities are the same as the V in the white noise model:

$$V = \max_{x>0} x^2 \Phi(2 - x).$$

□

Proof of Lemma D.2.

$$\sum_{j_1=\mathtt{j}^*+2}^{\infty} \mathbb{E}((\hat{\mathtt{f}} - M(f))^2 \mathbb{1}\{\tilde{\mathtt{j}} > \check{j} = j_1\})$$
$$\leq \sum_{j_1=\mathtt{j}^*+2}^{\infty} \mathbb{E}\Big(\Big((ave_f(j_1, \hat{\mathtt{i}}_{j_1}+2) - M(f))^2 \mathbb{1}\{Y_{j_1,\hat{\mathtt{i}}_{j_1}+6,s} - Y_{j_1,\hat{\mathtt{i}}_{j_1}+5,s} \leq 2\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j_1}}\}$$
$$\mathbb{1}\{\forall \mathtt{j}^*+1 \leq j \leq j_1-1, \min\{Y_{j,\hat{\mathtt{i}}_j+6,s} - Y_{j,\hat{\mathtt{i}}_j+5,s}, Y_{j,\hat{\mathtt{i}}_j-6,s} - Y_{j,\hat{\mathtt{i}}_j-5,s}\} > 2\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j_1}}\}$$
$$+ (ave_f(j_1, \hat{\mathtt{i}}_{j_1}-2) - M(f))^2 \mathbb{1}\{Y_{j_1,\hat{\mathtt{i}}_{j_1}+6,s} - Y_{j_1,\hat{\mathtt{i}}_{j_1}+5,s} \leq 2\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j_1}}\}$$
$$\mathbb{1}\{\forall \mathtt{j}^*+1 \leq j \leq j_1-1, \min\{Y_{j,\hat{\mathtt{i}}_j+6,s} - Y_{j,\hat{\mathtt{i}}_j+5,s}, Y_{j,\hat{\mathtt{i}}_j-6,s} - Y_{j,\hat{\mathtt{i}}_j-5,s}\} > 2\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j_1}}\}$$
$$\Big)\mathbb{1}\{\tilde{\mathtt{j}} > j_1\}\Big)$$

$$\leq \sum_{j_1=\mathtt{j}^*+2}^{\infty} \mathbb{E}\Bigg(\Big(\big(ave_f(j_1,\hat{\mathtt{i}}_{j_1}+2)-M(f)\big)^2\Phi(2-\frac{(ave_f(j_1,\hat{\mathtt{i}}_{j_1}+2)-M(f))2^{\frac{1}{2}(J-j_1)}}{3.5\sigma\gamma_s\sqrt{2}})$$
$$\Phi(-2+\frac{\rho_m(\frac{\sigma}{\sqrt{n}};f)2^{\frac{1}{2}(J-\mathtt{j}^*-1)}}{\sigma\sqrt{2}\gamma_s})\Phi(-2+\frac{\rho_m(\frac{\sigma}{\sqrt{n}};f)2^{\frac{1}{2}(J-\mathtt{j}^*-2)}}{\sigma 2\sqrt{2}\gamma_s})^{(j_1-\mathtt{j}^*-2)_+}$$
$$+(ave_f(j_1,\hat{\mathtt{i}}_{j_1}-2)-M(f))^2\Phi(2-\frac{(ave_f(j_1,\hat{\mathtt{i}}_{j_1}-2)-M(f))2^{\frac{1}{2}(J-j_1)}}{3.5\sigma\gamma_s\sqrt{2}})$$
$$\Phi(-2+\frac{\rho_m(\frac{\sigma}{\sqrt{n}};f)2^{\frac{1}{2}(J-\mathtt{j}^*-1)}}{\sigma\sqrt{2}\gamma_s})\Phi(-2+\frac{\rho_m(\frac{\sigma}{\sqrt{n}};f)2^{\frac{1}{2}(J-\mathtt{j}^*-2)}}{\sigma 2\sqrt{2}\gamma_s})^{(j_1-\mathtt{j}^*-2)_+}$$
$$\Big)\mathbb{E}(\mathbb{1}\{\tilde{\mathtt{j}}>j_1\})\Bigg).$$

Calculation shows that it can be further bounded by

$$\begin{aligned}&\sum_{j_1=\mathtt{j}^*+2}^{\infty} 2\cdot 2^{j_1-J}(3.5\sqrt{2}\sigma\gamma_s)^2 V\cdot\Phi(-2+\frac{\sqrt{3}}{4\gamma_s})\Phi(-2+\frac{\sqrt{3}}{8\sqrt{2}\gamma_s})^{(j_1-\mathtt{j}^*-2)_+}\\ &\leq 2\cdot 2^{\mathtt{j}^*+2-J}\frac{49}{2}\sigma^2\gamma_s^2 V\Phi(-2+\frac{1}{4})\frac{1}{1-2\Phi(-1.9)}\leq c_{m4}\rho_m(\frac{\sigma}{\sqrt{n}};f)^2.\end{aligned}\tag{D.97}$$

□

PROOF OF LEMMA D.3.

$$\begin{aligned}&\mathbb{E}(((\hat{\mathtt{f}}-ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}}))_+)^2\mathbb{1}\{\mathtt{j}\leq\check{j}<\infty\})\\
&=\sum_{j_2=1}^{J}\sum_{j_1=j_2}^{J}\mathbb{E}\Big(((\hat{\mathtt{f}}-ave_f(j_1,\hat{\mathtt{i}}_{j_1}))_+)^2\mathbb{1}\{\tilde{\mathtt{j}}=j_2\}\mathbb{1}\{\check{j}=j_1\}\Big)\leq\\
&\underbrace{\begin{aligned}\sum_{j_2=1}^{J}\sum_{j_1=j_2}^{J}\mathbb{E}\Big(&\mathbb{1}\{\tilde{\mathtt{j}}=j_2\}((ave_f(j_1,\hat{\mathtt{i}}_{j_1}+2)-ave_f(j_1,\hat{\mathtt{i}}_{j_1}))_+)^2\\
&\mathbb{1}\{Y_{j_1,\hat{\mathtt{i}}_{j_1+6},s}-Y_{j_1,\hat{\mathtt{i}}_{j_1}+5,s}\leq 2\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j_1}}\}\\
&\mathbb{1}\{\forall\mathtt{j}^*+2\leq j\leq j_1-1, Y_{j,\hat{\mathtt{i}}_{j-6},s}-Y_{j,\hat{\mathtt{i}}_j-5,s}>2\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j}},\\
&\qquad Y_{j,\hat{\mathtt{i}}_{j+6},s}-Y_{j,\hat{\mathtt{i}}_j+5,s}>2\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j}},\text{if exists}\}\Big)\end{aligned}}_{\kappa_1}\\
&+\underbrace{\begin{aligned}\sum_{j_2=1}^{J}\sum_{j_1=j_2}^{J}\mathbb{E}\Big(&\mathbb{1}\{\tilde{\mathtt{j}}=j_2\}((ave_f(j_1,\hat{\mathtt{i}}_{j_1}-2)-ave_f(j_1,\hat{\mathtt{i}}_{j_1}))_+)^2\\
&\mathbb{1}\{Y_{j_1,\hat{\mathtt{i}}_{j_1-6},s}-Y_{j_1,\hat{\mathtt{i}}_{j_1}-5,s}\leq 2\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j_1}}\}\\
&\mathbb{1}\{\forall\mathtt{j}^*+2\leq j\leq j_1-1, Y_{j,\hat{\mathtt{i}}_{j-6},s}-Y_{j,\hat{\mathtt{i}}_j-5,s}>2\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j}},\\
&\qquad Y_{j,\hat{\mathtt{i}}_{j+6},s}-Y_{j,\hat{\mathtt{i}}_j+5,s}>2\sqrt{2}\gamma_s\sigma\sqrt{2^{J-j}},\text{if exists}\}\Big)\end{aligned}}_{\kappa_2}.\end{aligned}\tag{D.98}$$

κ_1 and κ_2 in Inequality (D.98) are symmetric, we only need to bound κ_1 and κ_2 shares the same bound. κ_1 is upper bounded by

$$\kappa_1 \leq \sum_{j_2=1}^{J}\sum_{j_1=j_2}^{J}\mathbb{E}\Big((ave_f(j_1,\hat{\mathtt{i}}_{j_1}+2)-ave_f(j_1,\hat{\mathtt{i}}_{j_1}))^2\mathbb{1}\{ave_f(j_1,\hat{\mathtt{i}}_{j_1}+2)-ave_f(j_1,\hat{\mathtt{i}}_{j_1})>0\}$$
$$\Phi(2-\frac{\frac{(ave_f(j_1,\hat{\mathtt{i}}_{j_1}+2)-ave_f(j_1,\hat{\mathtt{i}}_{j_1}))}{2}\sqrt{2^{J-j_1}}}{\sqrt{2}\gamma_s\sigma})\Phi(-2+\frac{\frac{13}{16}\rho_m(\frac{\sigma}{\sqrt{n}};f)\sqrt{2^{J-\mathtt{j}^*-2}}}{\sigma\gamma_s\sqrt{2}})^{(j_1-\mathtt{j}^*-2)_+}\mathbb{1}\{\tilde{\mathtt{j}}=j_2\}\Big)$$
$$\leq \sum_{j_2=1}^{J}\sum_{j_1=j_2}^{J}\mathbb{E}\Big(\mathbb{1}\{ave_f(j_1,\hat{\mathtt{i}}_{j_1}+2)>ave_f(j_1,\hat{\mathtt{i}}_{j_1})\}2^{3+j_1-J}\gamma_s^2\sigma^2V\mathbb{1}\{\tilde{\mathtt{j}}=j_2\}$$
$$\Phi(-2+\frac{13\sqrt{3}}{64\sqrt{2}\gamma_s})^{(j_1-\mathtt{j}^*-2)_+}\Big)$$
$$\leq \sum_{j_2=1}^{J}\mathbb{E}(\mathbb{1}\{\tilde{\mathtt{j}}=j_2\})\gamma_s^2\sigma^2V2^{5+\mathtt{j}^*-J}(1+\frac{1}{1-2\Phi(-2+\frac{13\sqrt{3}}{64\sqrt{2}\gamma_s})}).$$

Therefore,

$$\mathbb{E}(((\hat{\mathtt{f}}-ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}}))_+)^2\mathbb{1}\{\tilde{\mathtt{j}}\leq\check{j}<\infty\})\leq c_{m5}\rho_m(\frac{\sigma}{\sqrt{n}};f)^2. \tag{D.99}$$

□

Proof of Lemma D.4. Although we take $\check{j}<\infty$ by default, it is not a key condition in this proof. We only need it to establish that $\tilde{\mathtt{j}}\leq J$ and $\tilde{\mathtt{j}}\leq\hat{\mathtt{j}}$.

$$\begin{aligned}&\mathbb{E}((ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}})-M(f))^2\mathbb{1}\{\tilde{\mathtt{j}}\leq\check{j}\})\\ &\leq 2\mathbb{E}\Big(\Big((ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}})-ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}}))_+\Big)^2\mathbb{1}\{\tilde{\mathtt{j}}\leq\check{j}\}\Big)\\ &\quad+2\mathbb{E}\Big(\Big((ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}})-M(f))_+\Big)^2\mathbb{1}\{\tilde{\mathtt{j}}\leq\check{j}\}\Big).\end{aligned} \tag{D.100}$$

Now we introduce two lemmas that we will prove later.

Lemma D.5.

$$\mathbb{E}\Big(\Big((ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}})-ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}}))_+\Big)^2\mathbb{1}\{\tilde{\mathtt{j}}\leq\check{j}\}\Big)\leq c_{m7}\rho_m(\frac{\sigma}{\sqrt{n}};f)^2. \tag{D.101}$$

Lemma D.6.

$$\mathbb{E}\Big(\Big((ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}})-M(f))_+\Big)^2\mathbb{1}\{\tilde{\mathtt{j}}\leq\check{j}\}\Big)\leq c_{m8}\rho_m(\frac{\sigma}{\sqrt{n}};f)^2. \tag{D.102}$$

With these two lemmas, we have

$$\mathbb{E}((ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}) - M(f))^2 \mathbb{1}\{\tilde{\mathtt{j}} \leq \check{j}\}) \leq 2(c_{m7} + c_{m8})\rho_m(\tfrac{\sigma}{\sqrt{n}}; f)^2 = c_{m6}\rho_m(\tfrac{\sigma}{\sqrt{n}}; f)^2.$$

□

PROOF OF LEMMA D.5 . Similar to the white noise model, we will first define the following events to describe the relative location of one iteration further compared to the current one at stage $\tilde{\mathtt{j}} + r$:

$$\begin{aligned} \tilde{A}_r =& \{\omega : \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r} < \mathtt{i}^*_{\tilde{\mathtt{j}}+r}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1} = 2\hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1} - 2\} \\ & \cup \{\omega : \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r} > \mathtt{i}^*_{\tilde{\mathtt{j}}+r}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1} = 2\hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1} + 1\}, \\ \tilde{B}_r =& \{\omega : \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r} < \mathtt{i}^*_{\tilde{\mathtt{j}}+r}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1} = 2\hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1} - 1\} \\ & \cup \{\omega : \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r} > \mathtt{i}^*_{\tilde{\mathtt{j}}+r}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1} = 2\hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1}\}, \\ \tilde{C}_r =& \{\omega : \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r} < \mathtt{i}^*_{\tilde{\mathtt{j}}+r}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1} = 2\hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1}\}\} \\ & \cup \{\omega : \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r} > \mathtt{i}^*_{\tilde{\mathtt{j}}+r}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1} = 2\hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1} - 1\}, \\ \tilde{D}_r =& \{\omega : \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r} < \mathtt{i}^*_{\tilde{\mathtt{j}}+r}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1} = 2\hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1} + 1\}\} \\ & \cup \{\omega : \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r} > \mathtt{i}^*_{\tilde{\mathtt{j}}+r}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1} = 2\hat{\mathtt{i}}_{\tilde{\mathtt{j}}+r+1} - 2\}. \end{aligned} \tag{D.103}$$

Basically, from $\tilde{A}_r$ to $\tilde{D}_r$, the average of the signal of the chosen interval are from the highest to the lowest.

Then we have

$$\begin{aligned} &\mathbb{E}\bigg(\Big(\big(ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}) - ave_f(\tilde{\mathtt{j}}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+\Big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \leq \check{j}\}\bigg) \\ &= \mathbb{E}\bigg(\Big(\big(ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}) - ave_f(\tilde{\mathtt{j}}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+\Big)^2 \mathbb{1}\{\tilde{\mathtt{j}} + 1 \leq \check{j}\}\bigg) \\ &\leq \mathbb{E}\bigg(\Big(\big(ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}) - ave_f(\tilde{\mathtt{j}}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+\Big)^2 \mathbb{1}\{\tilde{\mathtt{j}} + 1 \leq \check{j}\}(\mathbb{1}\{\tilde{C}_0 \cap \tilde{A}_1\} \\ &\qquad + \mathbb{1}\{\tilde{A}_0 \cup (\tilde{B}_0 \cap \tilde{D}_1^c) \cup (\tilde{B}_0 \cap \tilde{D}_1 \cap \{\check{j} = \tilde{\mathtt{j}} + 1\})\} \\ &\qquad + \mathbb{1}\{(\tilde{B}_0 \cap \tilde{D}_1) \cup (\tilde{C}_0 \cap \tilde{B}_1)\}\mathbb{1}\{\check{j} \geq \tilde{\mathtt{j}} + 3\})\bigg). \end{aligned} \tag{D.104}$$

We will bound the three terms separately. To simply the presentation for deriving bounds of these three terms, denote $\delta = \mathbb{1}\{j_1 = j_2 + 1\}$, $\delta_0 = \mathbb{1}\{j =$

$j_2\}$. For the second term, we have

(D.105)

$$\mathbb{E}\Big(\Big(\big(ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}})-ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+\Big)^2\mathbb{1}\{\tilde{\mathtt{j}}+1\le\check{j}\}(\mathbb{1}\{\tilde{A}_0\cup(\tilde{B}_0\cap\tilde{D}_1^c)\cup(\tilde{B}_0\cap\tilde{D}_1\cap\{\check{j}=\tilde{\mathtt{j}}+1\})\})\Big)$$

$$\le\sum_{j_2=1}^{\infty}\sum_{j_1=j_2+1}^{\infty}\mathbb{E}\Big(2\sum_{j=j_2}^{j_1-1}2^{j-j_2}\Big(\big(ave_f(j+1,\hat{\mathtt{i}}_{j+1})-ave_f(j,\hat{\mathtt{i}}_j)\big)_+\Big)^2$$

$$\mathbb{1}\{\check{j}=j_1,\tilde{\mathtt{j}}=j_2\}\big(\mathbb{1}\{\tilde{A}_0\cup(\tilde{B}_0\cap\tilde{D}_1^c)\}+\mathbb{1}\{\tilde{B}_0\cap\tilde{D}_1\}\mathbb{1}\{j_1=j_2+1\}\big)\Big)$$

$$\le\sum_{j_2=1}^{\infty}\sum_{j_1=j_2+1}^{\infty}\sum_{j=j_2}^{j_1-1}2^{j+1-j_2}\mathbb{E}\Big(\mathbb{1}\{\check{j}=j_1\}\mathbb{1}\{\tilde{\mathtt{j}}=j_2\}\big(ave_f(j+1,\hat{\mathtt{i}}_{j+1})-ave_f(j,\hat{\mathtt{i}}_j)\big)^2$$

$$\mathbb{1}\{ave_f(j+1,\hat{\mathtt{i}}_{j+1})>ave_f(j,\hat{\mathtt{i}}_j)\}\big(\mathbb{1}\{\tilde{A}_0\cup(\tilde{B}_0\cap\tilde{D}_1^c)\}+\mathbb{1}\{\tilde{B}_0\cap\tilde{D}_1\}\mathbb{1}\{j=j_2,j_1=j+1\}\big)\Big)$$

$$\le\sum_{j_2=1}^{\infty}\sum_{j=j_2}^{\infty}2^{j+1-j_2}\mathbb{E}\Big(\big(ave_f(j+1,\hat{\mathtt{i}}_{j+1})-ave_f(j,\hat{\mathtt{i}}_j)\big)^2\mathbb{1}\{ave_f(j+1,\hat{\mathtt{i}}_{j+1})>ave_f(j,\hat{\mathtt{i}}_j)\}$$

$$\mathbb{1}\{\tilde{\mathtt{j}}=j_2\}\sum_{j_1=j+1}^{\infty}\Phi(-2)^{(j_1-j_2-2)_+}\Phi(-2+\frac{\frac{13}{16}\rho_m(\frac{\sigma}{\sqrt{n}};f)\sqrt{2^{J-\mathtt{j}^*-2}}}{\gamma_s\sigma\sqrt{2}})^{(j_2-\mathtt{j}^*-\delta)_+}$$

$$\big(\mathbb{1}\{\tilde{A}_0\cup(\tilde{B}_0\cap\tilde{D}_1^c)\}+\mathbb{1}\{\tilde{B}_0\cap\tilde{D}_1\}\mathbb{1}\{j=j_2,j_1=j+1\}\big)\Big)$$

$$\le\sum_{j_2=1}^{\infty}\sum_{j=j_2}^{\infty}2^{j+1-j_2}\mathbb{E}\Big(\big(ave_f(j+1,\hat{\mathtt{i}}_{j+1})-ave_f(j,\hat{\mathtt{i}}_j)\big)^2\mathbb{1}\{ave_f(j+1,\hat{\mathtt{i}}_{j+1})>ave_f(j,\hat{\mathtt{i}}_j)\}$$

$$\mathbb{1}\{\tilde{\mathtt{j}}=j_2,\tilde{A}_0\cup\tilde{B}_0\}\big(\mathbb{1}\{j=j_2\}(1+\frac{1}{1-\Phi(-2)})+\mathbb{1}\{j\ge j_2+1\}\frac{\Phi(-2)^{j-j_2-1}}{1-\Phi(-2)})$$

$$\Phi(-2+\frac{\frac{13}{16}\rho_m(\frac{\sigma}{\sqrt{n}};f)\sqrt{2^{J-\mathtt{j}^*-2}}}{\gamma_s\sigma\sqrt{2}})^{(j_2-\mathtt{j}^*-\delta_0)_+}\Big)$$

$$\le\sum_{j_2=1}^{\infty}\sum_{j=j_2}^{\infty}2^{j+1-j_2}\big(\mathbb{1}\{j=j_2\}(1+\frac{1}{1-\Phi(-2)})+\mathbb{1}\{j\ge j_2+1\}\frac{\Phi(-2)^{j-j_2-1}}{1-\Phi(-2)})$$

$$\Phi(-2+\frac{13\sqrt{3}}{64\sqrt{2}\gamma_s})^{(j_2-\mathtt{j}^*-\delta_0)_+}$$

$$\mathbb{E}\Big(\big(ave_f(j+1,\hat{\mathtt{i}}_{j+1})-ave_f(j,\hat{\mathtt{i}}_j)\big)^2\mathbb{1}\{ave_f(j+1,\hat{\mathtt{i}}_{j+1})>ave_f(j,\hat{\mathtt{i}}_j)\}\mathbb{1}\{\tilde{\mathtt{j}}=j_2,\tilde{A}_0\cup\tilde{B}_0\}\Big)$$

Denote $\mathtt{C}(j,k)$ as the set of pairs (i_1,i_2) such that $P(\hat{\mathtt{i}}_{k+1}=i_2,\hat{\mathtt{i}}_k=i_1|\tilde{\mathtt{j}}=j)>0$ and $ave_f(j+1,i_2)>ave_f(j,i_1)$. Clearly, $\big|\mathtt{C}(j,k)\big|\le\min\{10\cdot2^{k-j}\cdot2,3\cdot4^{k+1-j}\}$.

Then we have

$$\begin{aligned}
&\mathbb{E}\Big(\big(ave_f(j+1,\hat{\mathtt{i}}_{j+1})-ave_f(j,\hat{\mathtt{i}}_j)\big)^2\mathbb{1}\{ave_f(j+1,\hat{\mathtt{i}}_{j+1})>ave_f(j,\hat{\mathtt{i}}_j)\}\mathbb{1}\{\tilde{\mathtt{j}}=j_2,\tilde{A}_0\cup\tilde{B}_0\}\Big)\\
&\le \sum_{(i_1,i_2)\in\mathtt{C}(j_2,j)}\mathbb{E}\Big(\big(ave_f(j+1,i_2)-ave_f(j,i_1)\big)^2\mathbb{1}\{\tilde{\mathtt{j}}=j_2,\tilde{A}_0\cup\tilde{B}_0\}\mathbb{1}\{\hat{\mathtt{i}}_{j+1}=i_2,\hat{\mathtt{i}}_j=i_1\}\Big)\\
&\le \sum_{(i_1,i_2)\in\mathtt{C}(j_2,j)}\mathbb{E}\Big(\big(ave_f(j+1,i_2)-ave_f(j,i_1)\big)^2\mathbb{1}\{\hat{\mathtt{i}}_{j+1}=i_2,\hat{\mathtt{i}}_j=i_1\}\Big)\\
&\le \sum_{(i_1,i_2)\in\mathtt{C}(j_2,j)}\big(ave_f(j+1,i_2)-ave_f(j,i_1)\big)^2\Phi\Big(-\frac{\big(ave_f(j+1,i_2)-ave_f(j,i_1)\big)\sqrt{2^{J-j-1}}}{\gamma_l\sigma\sqrt{2}}\Big)\\
&\le \sum_{(i_1,i_2)\in\mathtt{C}(j_2,j)}2^{j+1-J}\cdot 2\sigma^2\gamma_l^2Q\\
&\le \min\{10\cdot 2^{j-j_2}\cdot 2, 3\cdot 4^{j+1-j_2}\}2^{j+1-J}\cdot 2\sigma^2\gamma_l^2Q.
\end{aligned}$$

Still, $Q=\sup_{x>0}x^2\Phi(-x)$.

Continue with Inequality (D.105), we have

(D.106)

$$\begin{aligned}
&\mathbb{E}\Big(\big(ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}})-ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+\mathbb{1}\{\tilde{\mathtt{j}}+1\le\check{j}\}\mathbb{1}\{\tilde{A}_0\cup\tilde{B}_0\}\Big)\\
&\le \sum_{j_2=1}^{\infty}\sum_{j=j_2}^{\infty}2^{j+1-j_2}\big(\mathbb{1}\{j=j_2\}(1+\frac{1}{1-\Phi(-2)})+\mathbb{1}\{j\ge j_2+1\}\frac{\Phi(-2)^{j-j_2-1}}{1-\Phi(-2)}\big)\\
&\qquad\Phi(-2+\frac{13\sqrt{3}}{64\sqrt{2}\gamma_s})^{(j_2-\mathtt{j}^*-\delta_0)_+}\min\{10\cdot 2^{j-j_2}\cdot 2, 3\cdot 4^{j+1-j_2}\}2^{j+1-J}\cdot 2\sigma^2\gamma_l^2Q\\
&\le \sum_{j_2=1}^{\infty}\Big(24\big(1+\frac{1}{1-\Phi(-2)}\big)+\frac{4}{1-\Phi(-2)}\frac{80}{1-8\Phi(-2)}\Big)\\
&\qquad\Phi(-2+\frac{13\sqrt{3}}{64\sqrt{2}\gamma_s})^{(j_2-\mathtt{j}^*-1)_+}2^{j_2+2-J}\sigma^2\gamma_l^2Q\\
&\le 2^{\mathtt{j}^*-J}\sigma^2\gamma_l^2Q\sum_{j_2=1}^{\infty}\Big(24\big(1+\frac{1}{1-\Phi(-2)}\big)+\frac{80\frac{4}{1-\Phi(-2)}}{1-8\Phi(-2)}\Big)\\
&\qquad\Phi(-2+\frac{13\sqrt{3}}{64\sqrt{2}\gamma_s})^{(j_2-\mathtt{j}^*-1)_+}2^{j_2-\mathtt{j}^*+2}\\
&\le \frac{8\sigma^2\gamma_l^2Q}{n\rho_z(\frac{\sigma}{\sqrt{n}};f)}\tilde{c}_{m_9}\le\rho_m(\frac{\sigma}{\sqrt{n}};f)^2\cdot 16\gamma_l^2Q\tilde{c}_{m_9}=c_{m_9}\rho_m(\frac{\sigma}{\sqrt{n}};f)^2
\end{aligned}$$

Now let us turn to the third term.

$$\mathbb{E}\Big(\big(ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}})-ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+\mathbb{1}\{\tilde{\mathtt{j}}+3\le\check{j}\}\mathbb{1}\{(\tilde{B}_0\cap\tilde{D}_1)\cup(\tilde{C}_0\cap\tilde{B}_1)\}\Big)$$

$$\le\sum_{j_2=1}^{\infty}\sum_{j_1=j_2+3}^{\infty}2\sum_{j=j_2+2}^{j_1-1}\mathbb{E}\Big(2^{j-j_2-2}\Big(\big(ave_f(j+1,\hat{\mathtt{i}}_{j+1})-ave_f(j,\hat{\mathtt{i}}_j)\big)_+\Big)^2$$

$$\mathbb{1}\{\check{j}=j_1,\tilde{\mathtt{j}}=j_2\}\mathbb{1}\{(\tilde{B}_0\cap\tilde{D}_1)\cup(\tilde{C}_0\cap\tilde{B}_1)\}\Big)$$

$$\le\sum_{j_2=1}^{\infty}\sum_{j=j_2+2}^{\infty}2^{j-j_2-1}\sum_{j_1=j+1}^{\infty}\sum_{(i_1,i_2)\in\mathtt{C}(j_2,j)}\mathbb{E}\Big(\Big(ave_f(j+1,i_2)-ave_f(j,i_1)\Big)^2$$

$$\mathbb{1}\{\check{j}=j_1\}\mathbb{1}\{\tilde{\mathtt{j}}=j_2\}\mathbb{1}\{\hat{\mathtt{i}}_{j+1}=i_2,\hat{\mathtt{i}}_j=i_1\}\mathbb{1}\{(\tilde{B}_0\cap\tilde{D}_1)\cup(\tilde{C}_0\cap\tilde{B}_1)\}\Big)$$

$$\le\sum_{j_2=1}^{\infty}\sum_{j=j_2+2}^{\infty}2^{j-j_2-1}\sum_{(i_1,i_2)\in\mathtt{C}(j_2,j)}\sum_{j_1=j+1}^{\infty}\mathbb{E}\Big(\Big(ave_f(j+1,i_2)-ave_f(j,i_1)\Big)^2$$

$$\mathbb{1}\{\tilde{\mathtt{j}}=j_2\}\mathbb{1}\{\hat{\mathtt{i}}_{j+1}=i_2,\hat{\mathtt{i}}_j=i_1\}\mathbb{1}\{(\tilde{B}_0\cap\tilde{D}_1)\cup(\tilde{C}_0\cap\tilde{B}_1)\}\Big)$$

$$\Phi(-2)^{j_1-j_2-3}\Phi(-2+\frac{\frac{13}{16}\rho_m(\frac{\sigma}{\sqrt{n}};f)\sqrt{2^{J-\mathtt{j}^*-2}}}{\gamma_s\sigma\sqrt{2}})^{(j_2-\mathtt{j}^*)_+}$$

$$\le\sum_{j_2=1}^{\infty}\sum_{j=j_2+2}^{\infty}2^{j-j_2-1}\min\{20\cdot 2^{j-j_2},3\cdot 4^{j+1-j_2}\}2^{j+2-J}\sigma^2\gamma_l^2Q$$

$$\frac{\Phi(-2)^{j-j_2-2}}{1-\Phi(-2)}\Phi(-2+\frac{13\sqrt{3}}{64\sqrt{2}\gamma_s})^{(j_2-\mathtt{j}^*)_+}$$

$$=2^{\mathtt{j}^*-J}\sigma^2\sum_{j_2=1}^{\infty}2^{j_2-\mathtt{j}^*}\Phi(-2+\frac{13\sqrt{3}}{64\sqrt{2}\gamma_s})^{(j_2-\mathtt{j}^*)_+}\tilde{c}_{m10}\le c_{m10}\rho_m(\frac{\sigma}{\sqrt{n}};f)^2.$$

Finally, let us look at the first term.

$$\begin{aligned}
&\mathbb{E}\Big(\big(ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}}) - ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+ \mathbb{1}\{\tilde{\mathtt{j}}+1 \le \check{j}\}\mathbb{1}\{\tilde{C}_0 \cap \tilde{A}_1\}\Big) \\
&\le \sum_{j_2=1}^{\infty}\sum_{j=j_2+1}^{\infty} 2^{j-j_2} \sum_{j_1=j+1}^{\infty}\sum_{(i_1,i_2)\in \mathtt{C}(j_2,j)} \mathbb{E}\Big(\big(ave_f(j+1,i_2) - ave_f(j,i_1)\big)^2 \\
&\qquad \mathbb{1}\{\check{j}=j_1\}\mathbb{1}\{\tilde{\mathtt{j}}=j_2\}\mathbb{1}\{\hat{\mathtt{i}}_{j+1}=i_2,\hat{\mathtt{i}}_j=i_1\}\mathbb{1}\{(\tilde{C}_0\cap\tilde{A}_1)\}\Big) \\
&\le \sum_{j_2=1}^{\infty}\sum_{j=j_2+1}^{\infty} 2^{j-j_2}\min\{20\cdot 2^{j-j_2}, 3\cdot 4^{j+1-j_2}\}2^{j+2-J}\sigma^2\gamma_l^2 Q \\
&\qquad \frac{\Phi(-2)^{j-j_2-2}}{1-\Phi(-2)}\Phi(-2+\frac{13\sqrt{3}}{64\sqrt{2}\gamma_s})^{(j_2-\mathtt{j}^*)_+} \\
&\le 2^{\mathtt{j}^*-J}\sigma^2\tilde{c}_{m11} \le c_{m11}\rho_m(\frac{\sigma}{\sqrt{n}};f)^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\Big(\Big(\big(ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}}) - ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+\Big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j}\}\Big) \\
&\le (c_{m9}+c_{m10}+c_{m11})\rho_m(\frac{\sigma}{\sqrt{n}};f)^2 = c_{m7}\rho_m(\frac{\sigma}{\sqrt{n}};f)^2.
\end{aligned} \tag{D.107}$$

□

Proof of Lemma D.6. First, we introduce the following set:

$$IH(j) = \{\mathtt{i}_j^*-4, \mathtt{i}_j^*-3, \mathtt{i}_j^*-2, \mathtt{i}_j^*+2, \mathtt{i}_j^*+3, \mathtt{i}_j^*+4\},$$

which denotes the possible values of $\hat{\mathtt{i}}_j$ if $j = \tilde{\mathtt{j}}$.

$$
\begin{aligned}
&\mathbb{E}\Big(\Big(ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}}-M(f))\Big)^2\mathbb{1}\{\tilde{\mathtt{j}}\le\check{j}\}\Big)\\
&=\sum_{j_2=1}^{J}\sum_{j_1=j_2}^{J}\sum_{i\in IH(j_2)}\mathbb{E}\Big(\big(ave_f(j_2,i)-M(f)\big)^2\mathbb{1}\{\tilde{\mathtt{j}}=j_2,\check{j}=j_1,\hat{\mathtt{i}}_{j_2}=i\}\Big)\\
&\le\sum_{j_2=1}^{J}\sum_{i\in IH(j_2)}\sum_{j_1=j_2}^{J}\mathbb{E}\Big(\Big(ave_f(j_2,i)-M(f)\Big)^2\mathbb{1}\{\tilde{\mathtt{j}}=j_2,\hat{\mathtt{i}}_{j_2}=i\}\\
&\quad\Big(\mathbb{E}(\mathbb{1}\{\check{j}=j_1\}\big|\mathtt{Y}_l)\mathbb{1}\{j_1\le\mathtt{j}^*+2\}+\mathbb{1}\{j_1\ge j^*+3\}\Big[\\
&\quad\quad\Pi_{j=\mathtt{j}^*+2}^{j_1-1}\max\{\Phi(-2),\Phi(-2+(\frac{7}{16}+\frac{6m_j}{\rho_z(\frac{\sigma}{\sqrt{n}};f)})\rho_m(\frac{\sigma}{\sqrt{n}};f)\frac{\sqrt{2^{J-j}}}{\sqrt{2}\gamma_s\sigma})\}\Big]\Big)\Big)\\
&\le\sum_{j_2=1}^{J}\sum_{i\in IH(j_2)}\mathbb{E}\Big(\Big(ave_f(j_2,i)-M(f)\Big)^2\mathbb{1}\{\tilde{\mathtt{j}}=j_2,\hat{\mathtt{i}}_{j_2}=i\}\\
&\quad\quad\quad\Big(\mathbb{1}\{j_2\le\mathtt{j}^*+2\}+\mathbb{1}\{j_2\ge j^*+3\}\frac{\Phi(-2+\frac{13}{16}\frac{\sqrt{3}}{4\sqrt{2}\gamma_s})^{j_2-\mathtt{j}^*-2}}{1-\Phi(-2+\frac{13}{16}\frac{\sqrt{3}}{4\sqrt{2}\gamma_s})}\Big)\Big)\\
&\le\sum_{j_2=1}^{J}\sum_{i\in IH(j_2)}\Big(\mathbb{1}\{j_2\le\mathtt{j}^*+2\}+\mathbb{1}\{j_2\ge j^*+3\}\frac{\Phi(-2+\frac{1}{6})^{j_2-\mathtt{j}^*-2}}{1-\Phi(-2+\frac{1}{6})}\Big)\\
&\Big(ave_f(j_2,i)-M(f)\Big)^2\Phi(-\frac{\big(ave_f(j_2,i)-ave_f(j_2,\mathtt{i}^*_{j_2}+\mathrm{sign}(i-\mathtt{i}^*_{j_2}))\big)\sqrt{2^{J-j_2}}}{\sqrt{2}\gamma_l\sigma})\\
&\le\sum_{j_2=1}^{J}(23\frac{1}{8})2\gamma_l^2\sigma^2 2^{j_2-J}Q\frac{\Phi(-2+\frac{1}{6})^{(j_2-\mathtt{j}^*-2)_+}}{1-\Phi(-2+\frac{1}{6})}\le 2^{\mathtt{j}^*+2-J}\sigma^2\tilde{c}_{m8}\\
&\le c_{m8}\rho_m(\frac{\sigma}{\sqrt{n}};f)^2.
\end{aligned}
$$

□

Proof of Lemma C.43.

$$
\begin{aligned}
&\mathbb{E}\big((f_{\mathtt{i}}-M(f))^2\mathbb{1}\{\check{j}=\infty\}\big)\\
&=(\min\{f(x_i):0\le i\le n\}-M(f))^2\\
&\quad\quad+\mathbb{E}\big((f_{\mathtt{i}}-M(f))^2\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{\big|\hat{\mathtt{i}}_J-\mathtt{i}^*_J\big|\ge2\}\big)\\
&\le(\min\{f(x_i):0\le i\le n\}-M(f))^2\\
&\quad\quad+\mathbb{E}\big((f_{\hat{\mathtt{i}}_J}-M(f))^2\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{\big|\hat{\mathtt{i}}_J-\mathtt{i}^*_J\big|\ge2\}\big)
\end{aligned}
\tag{D.108}
$$

In the proof of Lemma D.4, all the argument using properties of $\check{j}$ only uses that $\mathtt{T}_j > 2\tilde{\sigma}_j$ for $j < \check{j}$, so for the second term, all the argument can also go through here in the case $\check{j} = \infty$. So we have

$$\mathbb{E}\big((f_{\mathtt{i}} - M(f))^2 \mathbb{1}\{\check{j} = \infty\}\big) \le c_{m6}\rho_m(\frac{\sigma}{\sqrt{n}}; f)^2 + (\min\{f(x_i) : 0 \le i \le n\} - M(f))^2$$

□

Proof of Lemma C.44.

$$\begin{aligned}
&\sigma^2 \mathbb{E}(\mathbb{1}\{\check{j} = \infty\})\\
&\le \sigma^2 \mathbb{1}\{J \le \mathtt{j}^* + 1\} + \sigma^2 \mathbb{E}(\mathbb{1}\{\check{j} = \infty\})\mathbb{1}\{J \ge \mathtt{j}^* + 2\}\\
&< \sigma^2 \frac{16}{n\rho_z(\frac{\sigma}{\sqrt{n}}; f)} \mathbb{1}\{J \le \mathtt{j}^* + 1\} + \sigma^2 \Phi(-2 + \frac{1}{6})^{J - \mathtt{j}^* - 1}\\
&\le 32\rho_m(\frac{\sigma}{\sqrt{n}}; f)^2 \mathbb{1}\{J \le \mathtt{j}^* + 1\} + \sigma^2 \frac{\frac{1}{n}}{\frac{2^{J-\mathtt{j}^*-1}}{n}} \big(2\Phi(-2 + \frac{1}{6})\big)^{J - \mathtt{j}^* - 1} \mathbb{1}\{J \ge \mathtt{j}^* + 2\}\\
&\le 32\rho_m(\frac{\sigma}{\sqrt{n}}; f)^2 \mathbb{1}\{J \le \mathtt{j}^* + 1\} + 32\rho_m(\frac{\sigma}{\sqrt{n}}; f)^2 \cdot 2\Phi(-2 + \frac{1}{6}) \mathbb{1}\{J \ge \mathtt{j}^* + 2\}\\
&\le 32\rho_m(\frac{\sigma}{\sqrt{n}}; f)^2
\end{aligned}$$

□

Proof of Lemma C.45. Similar to the proof of Lemma C.42, we bound the expectation by the sum of three terms and further bound those terms separately.

$$\begin{aligned}
\text{(D.109)}\quad \mathbb{E}((\hat{\mathtt{f}} - M(f))^2 \mathbb{1}\{\check{j} < \infty\}) &\le \sum_{j_1=2}^{J} \mathbb{E}((\hat{\mathtt{f}} - M(f))^2 \mathbb{1}\{\tilde{\mathtt{j}} > \check{j} = j_1\})\\
&\quad + 2\mathbb{E}\Big(\big((\hat{\mathtt{f}} - ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}))_+\big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j}\}\mathbb{1}\{\check{j} < \infty\}\Big)\\
&\quad + 2\mathbb{E}\Big(\big(ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}) - M(f)\big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j}\}\mathbb{1}\{\check{j} < \infty\}\Big).
\end{aligned}$$

Similar to the arguments in the proof of Lemma D.1, we have

$$\begin{aligned}
&\sum_{j_1=2}^{J} \mathbb{E}((\hat{\mathtt{f}} - M(f))^2 \mathbb{1}\{\tilde{\mathtt{j}} > \check{j} = j_1\}) \le \sum_{j_1=2}^{J} 2 \cdot 2^{j_1 - J}(3.5\sqrt{2}\sigma\gamma_s)^2 V\\
&\le 4 \cdot \frac{49}{2}\gamma_s^2 V \sigma^2 = \check{c}_{m4}\sigma^2,
\end{aligned}$$

where $V = \max_{x>0} x^2 \Phi(2 - x)$.

Similar to the arguments in the proof of Lemma D.3, we have

$$2\mathbb{E}\Big(\big((\hat{\mathtt{f}} - ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}))_+\big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j}\}\mathbb{1}\{\check{j} < \infty\}\Big) \le 2\sum_{j_2=1}^{J}\sum_{j_1=j_2}^{J}\Phi(-2 + \frac{13\sqrt{3}}{64\sqrt{2}\gamma_s})^{(j_1 - \mathtt{j}^* - 2)_+}$$

$$\mathbb{E}\Bigg(\mathbb{1}\{ave_f(j_1, \hat{\mathtt{i}}_{j_1} + 2) > ave_f(j_1, \hat{\mathtt{i}}_{j_1})\}2^{3+j_1-J}\gamma_s^2\sigma^2 V\mathbb{1}\{\tilde{\mathtt{j}} = j_2\}$$

$$+ \mathbb{1}\{ave_f(j_1, \hat{\mathtt{i}}_{j_1} - 2) > ave_f(j_1, \hat{\mathtt{i}}_{j_1})\}2^{3+j_1-J}\gamma_s^2\sigma^2 V\mathbb{1}\{\tilde{\mathtt{j}} = j_2\}\Bigg)$$

$$\le 4\sum_{j_2=1}^{J}\mathbb{E}(\mathbb{1}\{\tilde{\mathtt{j}} = j_2\})\gamma_s^2 V 2^4\sigma^2 \le 4\gamma_s^2 V 2^4 \sigma^2 = \check{c}_{m5}\sigma^2,$$

where $V = \max_{x>0} x^2\Phi(2-x)$.

For the third term, we have

$$\begin{aligned}
&2\mathbb{E}\Big(\big(ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}) - M(f)\big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j}\}\mathbb{1}\{\check{j} < \infty\}\Big) \\
&\le 4\mathbb{E}\Bigg(\Big(\big(ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}) - ave_f(\tilde{\mathtt{j}}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+\Big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j} < \infty\}\Bigg) \\
&\quad + 4\mathbb{E}\Bigg(\Big(\big(ave_f(\tilde{\mathtt{j}}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}}) - M(f)\big)_+\Big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j} < \infty\}\Bigg).
\end{aligned}$$

Now we have the following lemmas which we will prove later:

LEMMA D.7.

$$\mathbb{E}\Bigg(\Big(\big(ave_f(\check{j}, \hat{\mathtt{i}}_{\check{j}}) - ave_f(\tilde{\mathtt{j}}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+\Big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j} < \infty\}\Bigg) \le \check{c}_{m6}\sigma^2. \tag{D.110}$$

LEMMA D.8.

$$\mathbb{E}\Bigg(\Big(\big(ave_f(\tilde{\mathtt{j}}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}}) - M(f)\big)_+\Big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j} < \infty\}\Bigg) \le \check{c}_{m7}\sigma^2. \tag{D.111}$$

Now we can conclude that

$$\mathbb{E}((\hat{\mathtt{f}} - M(f))^2\mathbb{1}\{\check{j} < \infty\}) \le (\check{c}_{m4} + \check{c}_{m5} + 4\check{c}_{m6} + 4\check{c}_{m7})\sigma^2 = \check{c}_{m2}^2\sigma^2. \tag{D.112}$$

□

PROOF OF LEMMA D.7. $\tilde{\mathtt{j}} \le \check{j} < \infty$ implies $\tilde{\mathtt{j}} \le J$. Most of the arguments of Lemma D.5 are applicable here. Note that the difference between this lemma and Lemma D.5 is that we have additional assumption $j^* - 3 > J$

in this lemma. Also note that the place $\check{j}$ appears in Lemma D.5 are for invoking the followings: $\mathtt{T}_j > 2\tilde{\sigma}_j$ for $j < \check{j}$ and $\tilde{\mathtt{j}} \le \check{j} < \infty$. These also hold in this lemma. Therefore, using the arguments in the proof of Lemma D.5 and taking the notation there, we have

$$\mathbb{E}\Big(\Big(\big(ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}}) - ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+\Big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \le \check{j}\}\Big) \le \mathbb{E}\Big(\Big(\big(ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}}) - ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+\Big)^2$$
$$\mathbb{1}\{\tilde{\mathtt{j}}+1 \le \check{j}\}(\mathbb{1}\{\tilde{A}_0 \cup (\tilde{B}_0 \cap \tilde{D}_1^c) \cup (\tilde{B}_0 \cap \tilde{D}_1 \cap \{\check{j} = \tilde{\mathtt{j}}+1\})\}$$
$$+ \mathbb{1}\{(\tilde{B}_0 \cap \tilde{D}_1) \cup (\tilde{C}_0 \cap \tilde{B}_1)\}\mathbb{1}\{\check{j} \ge \tilde{\mathtt{j}}+3\} + \mathbb{1}\{\tilde{C}_0 \cap \tilde{A}_1\})\Big),$$

$$\mathbb{E}\Big(\Big(\big(ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}}) - ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+\Big)^2 \mathbb{1}\{\tilde{\mathtt{j}}+1 \le \check{j}\}\mathbb{1}\{\tilde{A}_0 \cup (\tilde{B}_0 \cap \tilde{D}_1^c) \cup (\tilde{B}_0 \cap \tilde{D}_1 \cap \{\check{j} = \tilde{\mathtt{j}}+1\})\}\Big)$$
$$\le \sum_{j_2=1}^{J}\sum_{j_1=j_2+1}^{J} \mathbb{E}\Big(2\sum_{j=j_2}^{j_1-1} 2^{j-j_2}\Big(\big(ave_f(j+1,\hat{\mathtt{i}}_{j+1}) - ave_f(j,\hat{\mathtt{i}}_j)\big)_+\Big)^2\Big)$$
$$\le \sum_{j_2=1}^{J}\Big(24\big(1+\frac{1}{1-\Phi(-2)}\big) + \frac{4}{1-\Phi(-2)}\frac{80}{1-8\Phi(-2)}\Big)\Phi(-2+\frac{13\sqrt{3}}{64\sqrt{2}\gamma_s})^{(j_2-\mathtt{j}^*-1)_+}2^{j_2+2-J}\sigma^2\gamma_l^2 Q$$
$$\le \Big(24\big(1+\frac{1}{1-\Phi(-2)}\big) + \frac{4}{1-\Phi(-2)}\frac{80}{1-8\Phi(-2)}\Big)2^3\sigma^2\gamma_l^2 Q = \check{c}_{m8}\sigma^2,$$

$$\mathbb{E}\Big(\Big(\big(ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}}) - ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+\Big)^2 \mathbb{1}\{\tilde{\mathtt{j}}+3 \le \check{j}\}\mathbb{1}\{(\tilde{B}_0 \cap \tilde{D}_1) \cup (\tilde{C}_0 \cap \tilde{B}_1)\}\Big)$$
$$\le \sum_{j_2=1}^{J-3}\sum_{j_1=j_2+3}^{J} 2\sum_{j=j_2+2}^{j_1-1} \mathbb{E}\Big(2^{j-j_2-2}\Big(\big(ave_f(j+1,\hat{\mathtt{i}}_{j+1}) - ave_f(j,\hat{\mathtt{i}}_j)\big)_+\Big)^2$$
$$\mathbb{1}\{\check{j} = j_1, \tilde{\mathtt{j}} = j_2\}\mathbb{1}\{(\tilde{B}_0 \cap \tilde{D}_1) \cup (\tilde{C}_0 \cap \tilde{B}_1)\}\Big)$$
$$\le 2^{-J}\sigma^2 \sum_{j_2=1}^{J-3} 2^{j_2}\Phi(-2+\frac{13\sqrt{3}}{64\sqrt{2}\gamma_s})^{(j_2-\mathtt{j}^*)_+}\tilde{c}_{m10} \le \check{c}_{m9}\sigma^2,$$

and

$$\mathbb{E}\Big(\Big(\big(ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}})-ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+\Big)^2\mathbb{1}\{\tilde{\mathtt{j}}+1\leq\check{j}\}\mathbb{1}\{\tilde{C}_0\cap\tilde{A}_1\}\Big)$$
$$\leq\sum_{j_2=1}^{J-2}\sum_{j=j_2+1}^{J-1}2^{j-j_2}\sum_{j_1=j+1}^{J}\sum_{(i_1,i_2)\in\mathtt{C}(j_2,j)}\mathbb{E}\Big(\Big(ave_f(j+1,i_2)-ave_f(j,i_1)\Big)^2$$
$$\mathbb{1}\{\check{j}=j_1\}\mathbb{1}\{\tilde{\mathtt{j}}=j_2\}\mathbb{1}\{\hat{\mathtt{i}}_{j+1}=i_2,\hat{\mathtt{i}}_j=i_1\}\mathbb{1}\{(\tilde{C}_0\cap\tilde{A}_1)\}\Big)$$
$$\leq\sum_{j_2=1}^{J-2}\sum_{j=j_2+1}^{J-1}2^{j-j_2}\min\{20\cdot2^{j-j_2},3\cdot4^{j+1-j_2}\}2^{j+2-J}\sigma^2\gamma_l^2Q$$
$$\frac{\Phi(-2)^{j-j_2-2}}{1-\Phi(-2)}\Phi(-2+\frac{13\sqrt{3}}{64\sqrt{2}\gamma_s})^{(j_2-\mathtt{j}^*)_+}\leq\check{c}_{m10}\sigma^2.$$

Therefore,

$$\mathbb{E}\Big(\Big(\big(ave_f(\check{j},\hat{\mathtt{i}}_{\check{j}})-ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}})\big)_+\Big)^2\mathbb{1}\{\tilde{\mathtt{j}}\leq\check{j}\}\Big)\leq(\check{c}_{m8}+\check{c}_{m9}+\check{c}_{m10})\sigma^2. \tag{D.113}$$

□

Proof of Lemma D.8. The arguments in the proof of Lemma D.6 hold, and we only need to change the last two inequalities to get the statement of this lemma.

Specifically,

$$\mathbb{E}\Big(\Big(ave_f(\tilde{\mathtt{j}},\hat{\mathtt{i}}_{\tilde{\mathtt{j}}}-M(f))\Big)^2\mathbb{1}\{\tilde{\mathtt{j}}\leq\check{j}<\infty\}\Big)\leq\sum_{j_2=1}^{J}(23\frac{1}{8})2\gamma_l^2\sigma^2 2^{j_2-J}Q\frac{\Phi(-2+\frac{1}{6})^{(j_2-\mathtt{j}^*-2)_+}}{1-\Phi(-2+\frac{1}{6})}\leq\check{c}_{m7}\sigma^2. \tag{D.114}$$

□

Proof of Lemma C.46. Similar to the arguments in Lemma C.43, we have

$$\begin{aligned}\mathbb{E}\big((f_{\mathtt{i}}-M(f))^2\mathbb{1}\{\check{j}=\infty\}\big)&=(\min\{f(x_i):0\leq i\leq n\}-M(f))^2\\&\quad+\mathbb{E}\big((f_{\mathtt{i}}-M(f))^2\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{\big|\hat{\mathtt{i}}_J-\mathtt{i}_J^*\big|\geq2\}\big)\\&\leq(\min\{f(x_i):0\leq i\leq n\}-M(f))^2\\&\quad+\mathbb{E}\big((f_{\hat{\mathtt{i}}_J}-M(f))^2\mathbb{1}\{\check{j}=\infty\}\mathbb{1}\{\big|\hat{\mathtt{i}}_J-\mathtt{i}_J^*\big|\geq2\}\big).\end{aligned} \tag{D.115}$$

In addition, note that in the proof of Lemma D.4, and Lemma D.7, Lemma D.8, all the argument using properties of $\check{j}$ only uses that $\mathtt{T}_j > 2\tilde{\sigma}_j$ for $j < \check{j}$, and $\tilde{\mathtt{j}} \leq \check{j} < \infty$. All these holds with $\check{j}$ replaced by J for the second term. Hence all the arguments go through here and we have

$$\begin{aligned}
&\mathbb{E}\big((f_{\hat{\mathtt{i}}_J} - M(f))^2 \mathbb{1}\{\check{j} = \infty\} \mathbb{1}\{\big|\hat{\mathtt{i}}_J - \mathtt{i}^*_J\big| \geq 2\}\big) \\
&\leq 2\mathbb{E}\Big(\big((ave_f(J, \hat{\mathtt{i}}_J) - ave_f(\tilde{\mathtt{j}}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}}))_+\big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \leq J < \check{j}\}\Big) + 2\mathbb{E}\Big(\big(ave_f(\tilde{\mathtt{j}}, \hat{\mathtt{i}}_{\tilde{\mathtt{j}}}) - M(f)\big)^2 \mathbb{1}\{\tilde{\mathtt{j}} \leq J\}\Big) \\
&\leq 2(\check{c}_{m6} + \check{c}_{m7})\sigma^2.
\end{aligned}$$

Therefore,

$$\mathbb{E}\big((f_{\hat{\mathtt{i}}} - M(f))^2 \mathbb{1}\{\check{j} = \infty\}\big) \leq (\min\{f(x_i) : 0 \leq i \leq n\} - M(f))^2 + 2(\check{c}_{m6} + \check{c}_{m7})\sigma^2. \tag{D.116}$$

Let $\check{c}_{m3} = \sqrt{2(\check{c}_{m6} + \check{c}_{m7})}$ gives the statement of the lemma.

□

References.

Belitser, E., Ghosal, S., and van Zanten, H. (2021). Correction note: “Optimal two-stage procedures for estimating location and size of the maximum of a multivariate regression function” Ann. Statist. 40 (2012) 2850–2876. *The Annals of Statistics*, 49(1):612 – 613.

Cai, T. T. and Guo, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *The Annals of Statistics*, 45(2):615–646.

Deng, H., Han, Q., and Sen, B. (2020). Inference for local parameters in convexity constrained models. *arXiv preprint arXiv:2006.10264*.

Ghosal, P. and Sen, B. (2017). On univariate convex regression. *Sankhya A*, 79(2):215–253.

Shoung, J.-M. and Zhang, C.-H. (2001). Least squares estimators of the mode of a unimodal regression function. *The Annals of Statistics*, 29(3):648–665.

Simonetto, A. (2021). Smooth strongly convex regression. In *2020 28th European Signal Processing Conference (EUSIPCO)*, pages 2130–2134. IEEE.

DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104
USA
E-MAIL: tcai@wharton.upenn.edu
ran1chen@wharton.upenn.edu
yuancheng.zhu@gmail.com
URL: http://www-stat.wharton.upenn.edu/∼tcai/
URL: http://ran-chen.com/