International Journal of Statistics and Probability; Vol. 5, No. 2; 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

On Consistency of Absolute Deviations Estimators of Convex Functions

Yao Luo[1] & Eunji Lim[2]

[1] Office Depot, Inc., Boca Raton, FL 33434, USA

[2] College of Business and Public Administration, Kean University, Union, NJ 07083, USA

Correspondence: Eunji Lim, College of Business and Public Administration, Kean University, Union, NJ 07083, USA. Tel: 1-908-737-4709. E-mail: elim@kean.edu

Received: January 16, 2016 Accepted: January 31, 2016 Online Published: February 10, 2016

doi:10.5539/ijsp.v5n2p1 URL: http://dx.doi.org/10.5539/ijsp.v5n2p1

Abstract

When estimating an unknown function from a data set of n observations, the function is often known to be convex. For example, the long-run average waiting time of a customer in a single server queue is known to be convex in the service rate even though there is no closed-form formula for the mean waiting time, and hence, it needs to be estimated from a data set. A computationally efficient way of finding the best fit of the convex function to the data set is to compute the least absolute deviations estimator minimizing the sum of absolute deviations over the set of convex functions. This estimator exhibits numerically preferred behavior since it can be computed faster and for a larger data sets compared to other existing methods. In this paper, we establish the validity of the least absolute deviations estimator by proving that the least absolute deviations estimator converges almost surely to the true function as n increases to infinity under modest assumptions.

Keywords: convex regression, absolute deviations estimator, consistency

1. Introduction

We study the problem of finding the best fit of an unknown convex function $f_* : [0,1]^d \to \mathbb{R}$ to a data set of n observations $(X_1, Y_1), \ldots, (X_n, Y_n)$, where

$$Y_i = f_*(X_i) + \varepsilon_i$$

for $1 \le i \le n$, the X_is are continuous $[0,1]^d$–valued independent and identically distributed (iid) random vectors and the ε_is are iid random variables with a zero median and $\mathbb{E}(|\varepsilon_1|) < \infty$.

This problem has been studied extensively for the past few decades. Hildreth (1954) proposed computing the minimizer $\tilde{g}_n : [0,1]^d \to \mathbb{R}$ of the sum of squared errors

$$\sum_{i=1}^{n} (Y_i - g(X_i))^2 /n$$

over the set of convex functions

$$C = \left\{ g : [0,1]^d \to \mathbb{R} \text{ such that } g \text{ is convex} \right\}$$

for the case when $d = 1$. Hanson & Pledger (1976) established the almost sure consistency of $\tilde{g}_n$ when $d = 1$ and Groeneboom et al. (2001) computed the rate of convergence of $\tilde{g}_n$ when $d = 1$. Kuosmanen (2008) has shown that $\tilde{g}_n$ can be computed as the solution to a quadratic programm with $(d+1)n$ decision variables and n^2 constraints when $d \ge 1$. Computation of $\tilde{g}_n$ becomes increasingly challenging when n gets large since it involves solving a quadratic program with n^2 constraints. Recently, there have been extensive studies on how to compute the best fit of an unknown convex function more efficiently. Lim & Luo (2014) suggest computing $\hat{g}_n : [0,1]^d \to \mathbb{R}$ that minimizes the sum of absolute deviations

$$\sum_{i=1}^{n} |Y_i - g(X_i)| /n$$

over C instead of the least squares estimator $\tilde{g}_n$. In fact, Lim & Lou (2014) reveals that $\hat{g}_n$ can be found by solving a linear program with $(d+3)n$ decision variables and $n^2 + 3n$ constraints. The following table compares the least absolute deviations estimator $\hat{g}_n$ to the least squares estimator $\tilde{g}_n$.

	Formulation	Number of decision variables	Number of constraints
Least absolute deviations estimator $\hat{g}_n$	Linear program	$(d+3)n$	n^2+3n
Least squares estimator $\tilde{g}_n$	Quadratic program	$(d+1)n$	n^2

Since a linear program can be solved more efficiently than a quadratic program when the other factors remain unchanged, the least absolute deviations estimator can be preferable from a computational point of view. Numerical results presented in Lim & Luo (2014) suggests that the least squares estimator $\hat{g}_n$ is computed faster and for a larger data sets than the least squares estimator $\tilde{g}_n$.

Another advantage of least absolute deviations estimators is that they can provide more robust results because they are not sensitive to outliers in the dataset (Bassett & Koenker 1978, Wagner 1959).

In this paper, we establish the strong consistency of $\hat{g}_n$ and prove that $\hat{g}_n(x)$ converges to $f_*(x)$ for any $x \in [0,1]^d$ as $n \to \infty$ with probability one. Our result will establish that $\hat{g}_n$ is a valid estimator of f_*.

This paper is organized as follows. In Section 2, we introduce some definitions. Section 3 introduces the mathematical framework for our analysis, and precisely states the main theorems (Theorems 1 and 2) of this paper. Proofs of the main results are provided in Section 4.

2. Definitions

We view $x \in \mathbb{R}^d$ as a column vector. For $x \in \mathbb{R}^d$, we write its kth component as x^k, so $x = (x^1, \ldots, x^d)^T$. We let $\|x\|_\infty = \max(|x^i| : 1 \le i \le d)$ and $\|x\| = \left((x^1)^2 + \cdots + (x^d)^2\right)^{1/2}$. For $y \in \mathbb{R}$, we write $y_+ = \max(0, y)$.

For a function $g : [0,1]^d \to \mathbb{R}$, g is differentiable at $x \in (0,1)^d$ if and only if there exists a vector $v \in \mathbb{R}^d$ with the property that

$$\lim_{z\to x}(g(z) - g(x) - v^T(z-x))/\|z-x\| = 0.$$

Such a v, if it exists, is called the gradient of g at x and is denoted by $\nabla g(x)$.

For any convex function $g : [0,1]^d \to \mathbb{R}$, a vector $\xi \in \mathbb{R}^d$ is said to be a subgradient of g at $x \in (0,1)^d$ if $g(y) \ge g(x) + \xi^T(y-x)$ for all $y \in (0,1)^d$. The set of all subgradients of g at x is called the subdifferential of g at x and is denoted by $\partial g(x)$. The subdifferential $\partial g(x)$ of a convex function $g : [0,1]^d \to \mathbb{R}$ is non–empty for any $x \in (0,1)^d$; see pp. 215–217 of Rockafella (1970).

Let $(a_n : n \ge 1)$ and $(b_n : n \ge 1)$ be sequences of real numbers. We say $a_n = O(b_n)$ if there exist positive constants c and n_0 such that $|a_n| \le c|b_n|$ for all $n \ge n_0$.

3. The Main Result

We start with Proposition 1, provided in Lim & Luo (2014), that reveals how $\hat{g}_n$ can be computed numerically.

Proposition 1 *Consider the minimization problem in the decision variables* $(g_1, \xi_1), \ldots, (g_n, \xi_n)$

$$\begin{array}{ll} \min & \displaystyle\frac{1}{n}\sum_{i=1}^n |Y_i - g_i| \\ s/t & g_j \ge g_i + \xi_i^T(X_j - X_i), \quad 1 \le i, j \le n, \end{array} \tag{1}$$

where $g_i \in \mathbb{R}$ *and* $\xi_i \in \mathbb{R}^d$ *for* $1 \le i \le n$. *Then, problem (1) has a minimizer* $(\hat{g}^1, \hat{\xi}^1), \ldots, (\hat{g}^n, \hat{\xi}^n)$ *and* $\hat{g}_n : [0,1]^d \to \mathbb{R}$, *defined by*

$$\hat{g}_n(x) \quad = \quad \max_{1\le i\le n}(\hat{g}^i + (\hat{\xi}^i)^T(x - X_i)) \tag{2}$$

for $x \in [0,1]^d$, *minimizes* φ_n *over* C.

Furthermore, problem (1) has a minimizer $(\hat{g}^1, \hat{\xi}^1), \ldots, (\hat{g}^n, \hat{\xi}^n)$ *if and only if* $(\hat{g}^1, (Y_1 - \hat{g}^1)_+, (-Y_1 + \hat{g}^1)_+, \hat{\xi}^1), \ldots, (\hat{g}^n, (Y_n - \hat{g}^n)_+, (-Y_n + \hat{g}^n)_+, \hat{\xi}^n)$ *is a solution to the following LP in the decision variables* $(g_1, p_1, m_1, \xi_1), \ldots, (g_n, p_n, m_n, \xi_n)$:

$$\begin{array}{lll} \min & \displaystyle\frac{1}{n}\sum_{i=1}^n (p_i + m_i) & \\ s/t & g_j \ge g_i + \xi_i^T(X_j - X_i), & 1 \le i, j \le n \\ & Y_i - g_i = p_i - m_i, & 1 \le i \le n \\ & p_i, m_i \ge 0, & 1 \le i \le n, \end{array} \tag{3}$$

where $g_i \in \mathbb{R}, p_i \in \mathbb{R}, m_i \in \mathbb{R}$, *and* $\xi_i \in \mathbb{R}^d$ *for* $1 \leq i \leq n$.

Throughout this paper, we will work with the set of minimizers of φ_n over C:

$$\mathcal{S}_n = \{g_n \in C : \varphi_n(g_n) \leq \varphi_n(g) \text{ for all } g \in C\}$$

for $n \geq 1$. By Proposition 1, $\mathcal{S}_n$ is nonempty for all $n \geq 1$ almost surely. Proposition 1 suggests a way of computing an element $\hat{g}_n$ in $\mathcal{S}_n$ by using (1), (2), and (3). The convex function $\hat{g}_n$ is our estimator of f_*. In order to analyze this estimator, we impose some probabilistic assumptions on the (X_i, Y_i)s. In particular, we require that:

A1. $X_1, X_2, \ldots$ is a sequence of iid $[0,1]^d$–valued random vectors having a common continuous positive density $\kappa : [0,1]^d \to \mathbb{R}$.

A2. For $i \geq 1$, $Y_i = f_*(X_i) + \varepsilon_i$. Given $X_1, X_2, \ldots$, the ε_is are iid random variables with the common cumulative distribution function F.

A3. $\mathbb{E}(|\varepsilon_1|) < \infty$, thereby implying that

$$\mathbb{E}(|\varepsilon_1|\,|X_1) = \int_{\mathbb{R}} |y| F(dy|X_1) < \infty \quad a.s.$$

A4. For each $x \in [0,1]^d$, we have $F(0|x) = 1/2$. Therefore, given $X_1, X_2, \ldots$, the ε_is have a zero median.

A5. f_* is bounded; i.e., there exists a positive constant M such that $|f_*(x)| \leq M$ for all $x \in [0,1]^d$.

We are now ready to state our main results.

Theorem 1 *Assume A1–A5 and that* $f_* \in C$. *Then for each* $0 < c < 1/2$,

$$\sup_{x\in[c,1-c]^d, \overline{g}_n \in \mathcal{S}_n} \left|\overline{g}_n(x) - f_*(x)\right| \to 0$$

as $n \to \infty$ *with probability one.*

Theorem 2 *Assume A1–A5 and that* $f_* \in C$. *If* f_* *is differentiable at* $z \in (0,1)^d$, *then*

$$\sup_{\xi\in\partial\overline{g}_n(z), \overline{g}_n \in \mathcal{S}_n} \|\xi - \nabla f_*(z)\| \to 0$$

as $n \to \infty$ *with probability one.*

Furthermore, if f_* *is differentiable on* $[c, 1-c]^d$ *for any* $0 < c \leq 1/2$,

$$\sup_{x\in[c,1-c]^d, \xi\in\partial\overline{g}_n(x), \overline{g}_n \in \mathcal{S}_n} \|\xi - \nabla f_*(x)\| \to 0$$

as $n \to \infty$ *with probability one.*

Theorems 1 and 2 justify our choice of the least absolute deviations estimator $\hat{g}_n$ as an estimator of f_*. The next section provides the proof of Theorem 1. The proof of Theorem 2 is given in the Appendix.

4. Proof of Theorem 1

Our proof of Theorem 1 can be broken down into a number of key steps.

Step 1 Since $\varphi_n(\overline{g}_n) \leq \varphi_n(f_*)$ for any $\overline{g}_n \in \mathcal{S}_n$, we must have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \overline{g}_n(X_i)\right| &\leq \frac{1}{n}\sum_{i=1}^{n}|Y_i - f_*(X_i)| \\ &= \frac{1}{n}\sum_{i=1}^{n}|f_*(X_i) + \varepsilon_i - f_*(X_i)| = \frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i|. \qquad (4)\end{aligned}$$

Step 2 Observe that, for any $\overline{g}_n \in \mathcal{S}_n$, we must have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left|\overline{g}_n(X_i)\right| &\leq \frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i| + \frac{1}{n}\sum_{i=1}^{n}|Y_i| \quad \text{by (4)} \\
&\leq \frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i| + \frac{1}{n}\sum_{i=1}^{n}|f_*(X_i)+\varepsilon_i| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i| + \frac{1}{n}\sum_{i=1}^{n}|f_*(X_i)| + \frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i|.
\end{aligned}$$

Thus,

$$\sup_{\overline{g}_n\in\mathcal{S}_n}\frac{1}{n}\sum_{i=1}^{n}\left|\overline{g}_n(X_i)\right| \leq \quad 2\mathbb{E}\,|\varepsilon_1| + \mathbb{E}\,|f_*(X_1)| + 1 \triangleq \beta < \infty$$

a.s. for n sufficiently large by A3 and the strong law of large numbers.

Step 3 We show that for any $A \subset [0,1]^d$ with a nonempty interior, there exists $\tilde{\beta}(A)$ such that

$$\sup_{\overline{g}_n\in\mathcal{S}_n}\inf_{x\in A}\left|\overline{g}_n(x) - f_*(x)\right| \leq \tilde{\beta}(A)$$

a.s. for n sufficiently large.

To fill in the details, we observe that the strong law of large numbers and A3 ensure

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}|f_*(X_i)| = \frac{1}{n}\sum_{i=1}^{n}|Y_i-\varepsilon_i| \leq \frac{1}{n}\sum_{i=1}^{n}|Y_i| + \frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i| \\
&\leq \mathbb{E}\,|Y_1| + \mathbb{E}\,|\varepsilon_1| + 1 \triangleq \tilde{\beta}
\end{aligned}$$

a.s. for n sufficiently large.

The strong law of large numbers also guarantees that

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}I(X_i\in A) \geq \mathbb{P}(X_1\in A) \quad \text{almost surely.}$$

Let

$$\begin{aligned}
B \quad = \quad & \Bigg\{\sup_{\overline{g}_n\in\mathcal{S}_n}\frac{1}{n}\sum_{i=1}^{n}\left|\overline{g}_n(X_i)\right| \leq \beta \text{ for } n \text{ sufficiently large,} \\
& \frac{1}{n}\sum_{i=1}^{n}|f_*(X_i)| \leq \tilde{\beta} \text{ for } n \text{ sufficiently large,} \\
& \text{and } \liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}I(X_i\in A) \geq \mathbb{P}(X_1\in A)\Bigg\},
\end{aligned}$$

then by Step 2 and the above arguments, we have $\mathbb{P}(B) = 1$.

Set $\tilde{\beta}(A) \triangleq (\beta + \tilde{\beta} + 1)/\mathbb{P}(X_1 \in A)$. We will prove that $\mathbb{P}(C) = 1$, where

$$C = \left\{\sup_{\overline{g}_n\in\mathcal{S}_n}\inf_{x\in A}|\hat{g}_n(x) - f_*(x)| \leq \tilde{\beta}(A) \text{ for } n \text{ sufficiently large}\right\},$$

by showing that $B \cap C^c = \emptyset$.

Suppose, on the contrary, that $\omega \in B \cap C^c$. Then for such ω, there exists $\overline{g}_n \in \mathcal{S}_n$ such that $\inf_{x\in A}\left|\overline{g}_n(x) - f_*(x)\right| > \tilde{\beta}(A)$

for infinitely many n. So, we would have

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \left|\overline{g}_n(X_i) - f_*(X_i)\right| \\
\geq\ & \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \left|\overline{g}_n(X_i) - f_*(X_i)\right| I(X_i \in A) \\
\geq\ & \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} I(X_i \in A) \liminf_{n\to\infty} \frac{\sum_{i=1}^{n} \left|\overline{g}_n(X_i) - f_*(X_i)\right| I(X_i \in A)}{\max(1, \sum_{i=1}^{n} I(X_i \in A))} \\
\geq\ & \mathbb{P}(X_1 \in A)\tilde{\beta}(A) = \beta + \tilde{\beta} + 1. \qquad (5)
\end{aligned}$$

On the other hand, we have

$$\frac{1}{n}\sum_{i=1}^{n} \left|\overline{g}_n(X_i) - f_*(X_i)\right| \quad\leq\quad \frac{1}{n}\sum_{i=1}^{n} \left|\overline{g}_n(X_i)\right| + \frac{1}{n}\sum_{i=1}^{n} |f_*(X_i)| \leq \beta + \tilde{\beta}$$

for n sufficiently large, which contradicts (5). So, we must have $B \cap C^c = \emptyset$, proving Step 3.

Step 4 We observe the following lemma whose proof is provided in the Appendix.

Lemma 1. *Let $e_0 = (0, 0, \ldots, 0)^T$ and e_i be the ith unit vector for $1 \leq i \leq d$. Let $v_* = (1/(4d), 1/d, \ldots, 1/d)$. Let A_i be defined as follows:*

$$\begin{aligned}
A_0 &= \left\{x \in [0,1]^d : \|x - e_0\| \leq \tau\right\}, \\
A_1 &= [1/2, 1] \times [0,1] \times \cdots \times [0,1] \subset [0,1]^d, \\
A_i &= \left\{x \in [0,1]^d : \|x - e_i\| \leq \tau\right\} \text{ for } 2 \leq i \leq d, \\
A_{d+1} &= \left\{x \in [0,1]^d : \|x - v_*\| \leq \tau\right\}.
\end{aligned}$$

Then there exists a positive constant τ such that for any y in A_{d+1} and x_i in A_i for $0 \leq i \leq d$, there exist nonnegative real numbers $p^0, p^1, \ldots, p^d$ summing to one such that

$$p^0 x_0 + p^1 x_1 + \cdots + p^d x_d = y$$

and that $p^1 \geq 1/(16d)$.

Step 5 We observe the following lemma:

Lemma 2. *Let u_i be the vector identical to e_i except that its first element is one minus e_i's first element for $0 \leq i \leq d$. Let $w_* = (1 - 1/(4d), 1/d, \ldots, 1/d)$. Let B_i be defined as follows:*

$$\begin{aligned}
B_0 &= \left\{x \in [0,1]^d : \|x - u_0\| \leq \tau\right\}, \\
B_1 &= [0, 1/2] \times [0,1] \times \cdots \times [0,1] \subset [0,1]^d, \\
B_i &= \left\{x \in [0,1]^d : \|x - u_i\| \leq \tau\right\} \text{ for } 2 \leq i \leq d, \\
B_{d+1} &= \left\{x \in [0,1]^d : \|x - w_*\| \leq \tau\right\}.
\end{aligned}$$

Then, there exists a positive constant τ such that for any y in B_{d+1} and x_i in B_i for $0 \leq i \leq d$, there exist nonnegative real numbers $p^0, p^1, \ldots, p^d$ summing to one such that

$$p^0 x_0 + p^1 x_1 + \cdots + p^d x_d = y$$

and that $p^1 \geq 1/(16d)$.

The proof of Lemma 2 is similar to that of Lemma 1 and is omitted.

Step 6 We observe the following lemma whose proof is given in the Appendix.

Lemma 3. *There exists a negative constant $\tilde{\gamma}$ such that*

$$\inf_{x \in [0,1]^d, \overline{g}_n \in \mathcal{S}_n} \overline{g}_n(x) \geq \tilde{\gamma}$$

a.s. for n sufficiently large.

Step 7 We observe the following lemma whose proof is given in the Appendix.

Lemma 4. *For any $c > 0$, there exists a positive constant $\tilde{\gamma}(c)$ such that*

$$\sup_{x\in\mathcal{H}_c,\overline{g}_n\in\mathcal{S}_n} \overline{g}_n(x) \leq \tilde{\gamma}(c)$$

a.s. for n sufficiently large, where $\mathcal{H}_c = [c, 1-c]^d$.

Step 8 Observe that the a.s. bound on $\left|\overline{g}_n\right|$ and $|f_*|$ uniformly in n over $\mathcal{H}_{c/2} = [c/2, 1-c/2]^d$ implies that $\overline{g}_n$ and f_* is Lipschitz over $\mathcal{H}_c = [c, 1-c]^d$ uniformly in n a.s. In particular, there exists a positive constant $\alpha(c)$ such that

$$\sup_{\overline{g}_n\in\mathcal{S}_n} \left|\overline{g}_n(x) - \overline{g}_n(y)\right| \leq \alpha(c)\|x-y\|$$

and

$$|f_*(x) - f_*(y)| \leq \alpha(c)\|x-y\|$$

for $x, y \in \mathcal{H}_c$ a.s. for n sufficiently large; see, for example, Roberts & Barberg (1974).

Step 9 Let

$$\begin{aligned} C_c \quad = \quad & \{h : \mathcal{H}_c \to \mathbb{R} \text{ such that } h \text{ is convex on } \mathcal{H}_c, \\ & |h(x)| \leq |\bar{\gamma}| + \tilde{\gamma}(c) \text{ and } |h(x) - h(y)| \leq \alpha(c)\|x-y\| \text{ for } x, y \in \mathcal{H}_c\}. \end{aligned}$$

Note that Steps 6, 7, and 8 guarantee that for each $c \geq 0$ there exists $n(c)$ such that $n \geq n(c)$ and $\overline{g}_n \in \mathcal{S}_n$ imply that $\overline{g}_n$ restricted to $\mathcal{H}_c$ belongs to C_c a.s. Furthermore, C_c is compact in the uniform metric d_c given by

$$d_c(h_1, h_2) = \sup_{x\in\mathcal{H}_c} |h_1(x) - h_2(x)|.$$

It follows that for each $\epsilon > 0$, there exists a finite collection of functions $h_1, \ldots, h_m$ in C_c such that

$$\bigcup_{i=1}^{m} \{h \in C_c : d_c(h_i, h) < \epsilon\} \supseteq C_c.$$

That is, $h_1, h_2, \ldots, h_m$ is an ϵ–net for C_c; see Theorem 6 of Bronshtein (1976).

Step 10 We observe the following lemma whose proof is given in the Appendix.

Lemma 5. *For any positive real numbers ϵ and δ and for any $z \in [0, 1]^d$, we have*

$$\sup_{\overline{g}_n\in\mathcal{S}_n} \inf_{x\in B(z,\delta)} \left(f_*(x) - \overline{g}_n(x)\right) \leq \epsilon$$

a.s. for n sufficiently large, where $B(z, \delta) \triangleq \left\{x \in [0, 1]^d : \|x - z\| \leq \delta\right\}$.

Step 11 We observe the following lemma whose proof is given in the Appendix.

Lemma 6. *For any positive real numbers ϵ and δ and for any $z \in [0, 1]^d$, we have*

$$\sup_{\overline{g}_n\in\mathcal{S}_n} \inf_{x\in B(z,\delta)} \left(\overline{g}_n(x) - f_*(x)\right) \leq \epsilon$$

a.s. for n sufficiently large, where $B(z, \delta) \triangleq \left\{x \in [0, 1]^d : \|x - z\| \leq \delta\right\}$.

Step 12 We will prove that for any $\epsilon > 0$,

$$\sup_{x\in\mathcal{H}_c,\overline{g}_n\in\mathcal{S}_n} \left(f_*(x) - \overline{g}_n(x)\right) \leq \epsilon$$

a.s. for n sufficiently large.

Take $\delta = \epsilon/(6\alpha(c))$, where $\alpha(c)$ is given as in Step 8. Since $\mathcal{H}_c$ is compact, there exist a finite number of points $y_1, \ldots, y_l$ in $\mathcal{H}_c$ such that $B_c(y_i, \delta) \triangleq \{x \in \mathcal{H}_c : \|x - y_i\| \leq \delta\}$ for $1 \leq i \leq l$ covers $\mathcal{H}_c$.

If there exists $\overline{g}_n \in \mathcal{S}_n$ such that $\sup_{x\in\mathcal{H}_c}(f_*(x) - \overline{g}_n(x)) > \epsilon$ for infinitely many n, for each of such n, there exists a point x_n in $\mathcal{H}_c$ such that

$$f_*(x_n) - \overline{g}_n(x_n) \quad > \quad \epsilon/2. \tag{6}$$

In this case, infinitely many of the x_ns will be in $B_c(y_j, \delta)$ for some j, so if we choose a subsequence $(n_k : k \geq 1)$ so that x_{n_k} is in $B_c(y_j, \delta)$ for all $k \geq 1$, then for any $x \in B_c(y_j, \delta)$ we have

$$\begin{aligned}
& f_*(x) - \overline{g}_n(x) \\
&= f_*(x) - f_*(x_{n_k}) + f_*(x_{n_k}) - \overline{g}_n(x_{n_k}) + \overline{g}_n(x_{n_k}) - \overline{g}_n(x) \\
&\geq -\epsilon/6 + \epsilon/2 - \epsilon/6 \geq \epsilon/6
\end{aligned}$$

by (6).

So, $\sup_{x\in\mathcal{H}_c}(f_*(x) - \overline{g}_n(x)) > \epsilon$ implies $\inf_{x\in B_c(x_j,\delta)}(f_*(x) - \overline{g}_n(x)) \geq \epsilon/6$ for some j, and hence,

$$\begin{aligned}
& \mathbb{P}\left(\sup_{x\in\mathcal{H}_c,\overline{g}_n\in\mathcal{S}_n}(f_*(x) - \overline{g}_n(x)) > \epsilon \text{ for infinitely many } n\right) \\
&= \sum_{j=1}^{l}\mathbb{P}\left(\sup_{\overline{g}_n\in\mathcal{S}_n}\inf_{x\in B_c(x_j,\delta)}(f_*(x) - \overline{g}_n(x)) \geq \epsilon/6 \text{ for infinitely many } n\right) \\
&= 0
\end{aligned}$$

by Step 10, proving Step 12.

Step 13 For any $\epsilon > 0$,

$$\sup_{x\in\mathcal{H}_c,\overline{g}_n\in\mathcal{S}_n}(\overline{g}_n(x) - f_*(x)) \leq \epsilon$$

a.s. for n sufficiently large.

The proof is similar to that of Step 12 (Step 11 is used instead of Step 10) and is omitted.

Step 14 Theorem 1 follows from Steps 12 and 13.

Appendix

A.1 Proof of Lemma 1

Let $y = (y^1, \ldots, y^d)$ be any point in A_{d+1} and $x_i = (x_i^1, \ldots, x_i^d)$ be any point in A_i for $0 \leq i \leq d$. We will show that there exists a nonnegative solution $p^0, p^1, \ldots, p^d$ (summing to one) to the linear system

$$p^0 x_0 + p^1 x_1 + \cdots + p^d x_d = y$$

with $p^1 \geq 1/(16d)$.

Or equivalently, we will show that there exists a nonnegative solution $p^1, \ldots, p^d$ (summing less than or equal to one) to the linear system

$$\sum_{i=1}^{d} p^i(x_i - x_0) = y - x_0. \tag{7}$$

The linear system can be reexpressed as $Fp = y - x_0$, where $p = (p^1, \ldots, p^d)^T$ and $F = (F_{ij} : 1 \leq i, j \leq d)$ is a square $d \times d$ matrix in which the ith column is $x_i - x_0$ for $1 \leq i \leq d$. Note that F is invertible for sufficiently small $\tau > 0$ because we have

$$|F_{ii}|\left|F_{jj}\right| > \left(\sum_{k=1,k\neq i}^{n} F_{ik}\right)\left(\sum_{k=1,k\neq j}^{n} F_{jk}\right)$$

for all $i \neq j$ and $1 \leq i, j \leq d$ with sufficiently small τ, and hence, Theorem V of Taussky (1949) applies. So, there exists a solution $p^1, \ldots, p^d$ to (7).

To show that $p^1, \ldots, p^d$ are nonnegative, sum less than or equal to one, and $p^1 \geq 1/(16d)$, we let $G = \left(G_{ij} : 1 \leq i, j \leq d\right)$ be a square $d \times d$ matrix in which the first column is x_1 and the ith column is e_i for $2 \leq i \leq d$. Observe that $q^1, \ldots, q^d$ defined by

$$\begin{aligned}
q^1 &= y^1/x_1^1 \\
q^i &= y^i - y^1 x_1^i/x_1^1
\end{aligned}$$

for $2 \le i \le d$ satisfy $Gq = y$, where $q = (q^1, \ldots, q^d)$. Note also that $1/(8d) \le q^1 \le 2y^1$ and $1/(8d) \le q^i \le y^i$ for $2 \le i \le d$ for τ sufficiently small.

Set $|||F||| \triangleq \sup_{||x||=1} ||Fx||$. Mapping a $d \times d$ square matrix to its inverse is continuous with respect to $|||\cdot|||$ in a neighborhood of F because F is invertible. Thus, we can make $|||F^{-1} - G^{-1}|||$ sufficiently small by making $|||F - G|||$ or τ sufficiently small. Also, $|||F^{-1}||| \le 1/|||F||| \le 1/\max_{1\le i,j\le d} |F_{ij}| \le 4$ for τ sufficiently small. So,

$$\begin{aligned} ||p - q|| &= ||F^{-1}(y - x_0) - G^{-1}y|| \\ &= ||(F^{-1} - G^{-1})y - F^{-1}x_0|| \\ &\le ||(F^{-1} - G^{-1})y|| + ||F^{-1}x_0|| \\ &\le |||F^{-1} - G^{-1}||| \cdot ||y|| + |||F^{-1}||| \cdot ||x_0|| \\ &\le |||F^{-1} - G^{-1}||| + |||F^{-1}|||\tau, \end{aligned}$$

and hence, $||p - q|| \le 1/(16d)$ for sufficiently small τ. Thus, $p^1, \ldots, p^d$ are nonnegative and sum less than or equal to one, and $p^1 \ge 1/(16d)$. Lemma 1 is proved.

A.2 Proof of Lemma 3

First, we show that

$$\inf_{x\in A_1, \overline{g}_n\in\mathcal{S}_n} \overline{g}_n(x) \ge \tilde{\gamma}$$

a.s. for n sufficiently large. Then it will follow similarly that

$$\inf_{x\in B_1, \overline{g}_n\in\mathcal{S}_n} \overline{g}_n(x) \ge \tilde{\gamma}$$

a.s. for n sufficiently large.

By Step 3, there exists a positive constant γ such that

$$\sup_{\overline{g}_n\in\mathcal{S}_n} \inf_{x\in A_i} \left|\overline{g}_n(x) - f_*(x)\right| \le \gamma$$

a.s. for all $0 \le i \le d+1$ and n sufficiently large.

Since $|f_*(x)| \le M$ for $x \in [0,1]^d$ by A5, we have

$$\sup_{\overline{g}_n\in\mathcal{S}_n} \inf_{x\in A_i} \left|\overline{g}_n(x)\right| \le M + \gamma \tag{8}$$

a.s. for all $0 \le i \le d+1$ and n sufficiently large.

Set $\tilde{\gamma} = -32d(M + \gamma + 1)$. For any $\overline{g}_n \in \mathcal{S}_n$, if $\overline{g}_n(x_1) \le \tilde{\gamma}$ for some $x_1 \in A_1$ and $\overline{g}_n(x_i) \le (M + \gamma + 1)$ for some $x_i \in A_i$ $(i = 0, 2, \ldots, d)$, then Step 4 guarantees that for any y in A_{d+1}, there exist nonnegative real numbers $p^0, p^1, \ldots, p^d$ summing to one that satisfy

$$p^0x_0 + p^1x_1 + \cdots + p^dx_d = y$$

and $p^1 \ge 1/(16d)$. So, we have

$$\begin{aligned} \overline{g}_n(y) &= \overline{g}_n(p^0x_0 + \cdots + p^dx_d) \\ &\le p^0\overline{g}_n(x_0) + p^1\overline{g}_n(x_1) + \cdots + p^d\overline{g}_n(x_d) \quad \text{because } \overline{g}_n \text{ is convex} \\ &\le \tilde{\gamma}/(16d) + (M + \gamma + 1) \quad \text{because } p^1 \ge 1/(16d) \text{ and } \overline{g}_n(x_1) \le \tilde{\gamma} \le 0 \\ &= -(M + \gamma + 1). \end{aligned}$$

So, if $\overline{g}_n(x) \le \tilde{\gamma}$ for some $x \in A_1$, then we should either have

$$\inf_{x\in A_i} \overline{g}_n(x) \ge M + \gamma + 1$$

for some $i \in \{0, 2, \ldots, d\}$ or

$$\sup_{x\in A_{d+1}} \overline{g}_n(x) \le -(M + \gamma + 1).$$

Thus,

$$\begin{aligned}
&\mathbb{P}\left(\inf_{x\in A_1,\overline{g}_n\in\mathcal{S}_n}\overline{g}_n(x)\le\tilde{\gamma}\text{ for infinitely many } n\right)\\
\le\quad&\sum_{i=0,2,\ldots,d}\mathbb{P}\left(\sup_{\overline{g}_n\in\mathcal{S}_n}\inf_{x\in A_i}\overline{g}_n(x)\ge M+\gamma+1\text{ for infinitely many } n\right)\\
&+\mathbb{P}\left(\inf_{\overline{g}_n\in\mathcal{S}_n}\sup_{x\in A_{d+1}}\overline{g}_n(x)\le -(M+\gamma+1)\text{ for infinitely many } n\right)\\
\le\quad&\sum_{i=0,2,\ldots,d}\mathbb{P}\left(\sup_{\overline{g}_n\in\mathcal{S}_n}\inf_{x\in A_i}\left|\overline{g}_n(x)\right|\ge M+\gamma+1\text{ for infinitely many } n\right)\\
&+\mathbb{P}\left(\sup_{\overline{g}_n\in\mathcal{S}_n}\inf_{x\in A_{d+1}}\left|\overline{g}_n(x)\right|\ge M+\gamma+1\text{ for infinitely many } n\right)\\
=\quad&0
\end{aligned}$$

by (8), proving Lemma 3.

A.3 Proof of Lemma 4

First we prove that there exists a positive constant $\tau(c)$ such that for any $y\in\mathcal{H}_c$ and for any $x_i\in C_i$ $(1\le i\le d)$, where $C_i=\left\{x\in[0,1]^d:\|x-e_i\|\le\tau(c)\right\}$, there exist nonnegative real numbers $p^1,\ldots,p^d$ such that $p^1x_1+\cdots+p^dx_d=y$ and that $p^i\le 1$ for $1\le i\le d$.

To fill in the details, we note that we need to show that there exists a solution $p=(p^1,\ldots,p^d)^T$ to the linear equation $Hp=y$ with $0\le p^i\le 1$ for $1\le i\le d$, where $H=(H_{ij}:1\le i,j\le d)$ is a square $d\times d$ matrix in which the ith column is x_i for $1\le i\le d$. Set $\|H\|_\infty=\max_{1\le i\le d}\sum_{j=1}^d\left|H_{ij}\right|$ and note that $\|H-I_d\|_\infty\le\tau(c)$, where I_d is the $d\times d$ identity matrix. Hence, for $\tau(c)<1/2$, H is invertible and we have

$$\|H^{-1}\|_\infty=\|(I_d+H-I_d)^{-1}\|_\infty\le(1-\|H-I_d\|_\infty)^{-1}\le 2.$$

Therefore,

$$\|p-y\|_\infty=\|H^{-1}y-y\|_\infty\le\|H^{-1}-I_d\|_\infty\|y\|_\infty\le\|H^{-1}-I_d^{-1}\|_\infty.$$

Since mapping a $d\times d$ matrix to its inverse matrix is continuous with respect to $\|\cdot\|_\infty$ in a neighborhood of H and $\|H-I_d\|_\infty\le\tau(c)$, there exists a positive number $\tau(c)$ that guarantees $\|H^{-1}-I_d^{-1}\|\le c/2$. So, for such $\tau(c)$, $\|p-y\|_\infty\le c/2$. Since $y\in[c,1-c]^d$, $p^1,\ldots,p^d$ are nonnegative and each of them is less than or equal to one.

Now we prove Step 7. For $1\le i\le d$, $r>0$, and $\overline{g}_n\in\mathcal{S}_n$,

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^n I\left(X_j\in C_i,\left|\overline{g}_n(X_j)\right|\le r\right)\\
&\quad\ge\quad\frac{1}{n}\sum_{j=1}^n I\left(X_j\in C_i\right)-\frac{1}{n}\sum_{j=1}^n I\left(X_j\in C_i,\left|\overline{g}_n(X_j)\right|>r\right).
\end{aligned}$$

However, Markov inequality and Step 2 imply that

$$\sup_{\overline{g}_n\in\mathcal{S}_n}\frac{1}{n}\sum_{j=1}^n I\left(X_j\in C_i,\left|\overline{g}_n(X_j)\right|>r\right)\quad\le\quad\sup_{\overline{g_n}\in\mathcal{S}_n}r^{-1}\frac{1}{n}\sum_{j=1}^n\left|\hat{g}_n(X_j)\right|\le\beta/r$$

a.s. for n sufficiently large. Choose r_0 so large that $\beta/r_0\le\overline{\gamma}\triangleq\min\{\mathbb{P}(X_1\in C_i):1\le i\le d\}/2$, then

$$\inf_{\overline{g}_n\in\mathcal{S}_n}\frac{1}{n}\sum_{j=1}^n I\left(X_j\in C_i,\left|\overline{g}_n(X_j)\right|\le r_0\right)\ge\overline{\gamma}$$

a.s. for n sufficiently large.

For each such n, there exists $X_{I(i)}\in C_i$ with $1\le I(i)\le n$ and $\left|\overline{g}_n(X_{I(i)})\right|\le r_0$. For each $y\in[c,1-c]^d$ and $X_{I(i)}\in C_i$ for $1\le i\le d$, there exist $p^1,\ldots,p^d$ such that

$$y=p^1X_{I(1)}+\cdots+p^dX_{I(d)}$$

and that $0 \le p^i \le 1$ for $1 \le i \le d$. So, the convexity of $\overline{g}_n$ yields

$$\overline{g}_n(y) \le p^1 \overline{g}_n(X_{I(1)}) + \cdots + p^d \overline{g}_n(X_{I(d)}) \le dr_0,$$

proving that

$$\sup_{x \in [c,1-c]^d, \overline{g}_n \mathcal{S}_n} \overline{g}_n(x) \le dr_0$$

a.s. for n sufficiently large.

A.4 Proof of Lemma 5

Let

$$C = \{ \sup_{\overline{g}_n \in \mathcal{S}_n} \inf_{x \in B(z,\delta)} \left(f_*(x) - \overline{g}_n(x) \right) \le \epsilon \text{ for } n \text{ sufficiently large} \}.$$

We will prove that $\mathbb{P}(C) = 1$ by showing that $\mathbb{P}(A \cap B \cap C^c) = \emptyset$, where A and B will be defined subsequently. (It will also be shown later that $\mathbb{P}(A) = \mathbb{P}(B) = 1$.)

Let

$$A = \left\{ \frac{1}{n} \sum_{i=1}^{n} I(X_i \in B(z,\delta), -\epsilon/2 \le \varepsilon_i \le 0) \ge \eta/2 \text{ for } n \text{ sufficiently large} \right\},$$

where $\eta \triangleq \mathbb{P}(X_1 \in B(z,\delta), -\epsilon/2 \le \varepsilon_1 \le 0)$. By the strong law of large numbers, $\mathbb{P}(A) = 1$.

On the other hand, the dominated convergence theorem guarantees that for $H_\nu = [\nu, 1-\nu]^d$,

$$\mathbb{E}(I(X_1 \notin \mathcal{H}_\nu)) \to 0$$

as $\nu \to 0$ because $I(X_1 \notin \mathcal{H}_\nu) \downarrow 0$ a.s. as $\nu \downarrow 0$. So, take ν_0 small enough so that

$$\mathbb{E}(I(X_1 \notin \mathcal{H}_{\nu_0})) \quad \le \quad \epsilon\eta/(24(M + |\tilde{\gamma}|))$$

and note that

$$\frac{1}{n} \sum_{i=1}^{n} I(X_i \notin \mathcal{H}_{\nu_0}) \le \epsilon\eta/(12(M + |\tilde{\gamma}|))$$

a.s. for n sufficiently large by the strong law of large numbers. Also, by Step 6 and A5, we have $(f_*(X_i) - \overline{g}_n(X_i) - \epsilon/2)^+ \le M + |\tilde{\gamma}|$ a.s. for n sufficiently large, so

$$\sup_{\overline{g}_n \in \mathcal{S}_n} \frac{1}{n} \sum_{i=1}^{n} (f_*(X_i) - \overline{g}_n(X_i) - \epsilon/2)^+ I(X_i \notin \mathcal{H}_{\nu_0}) \quad \le \quad \epsilon\eta/12 \tag{9}$$

a.s. for n sufficiently large.

Let $h_1, \ldots, h_m$ be an $\epsilon\eta/12$–net for $\mathcal{H}_{\nu_0}$. For each $j \in \{1, \ldots, m\}$, the strong law of large numbers guarantees that

$$\frac{1}{n} \sum_{i=1}^{n} (f_*(X_i) - h_j(X_i) - \epsilon/2)^+ (1/2 - I(\varepsilon_i \le 0)) I(X_i \in \mathcal{H}_{\nu_0}) \to 0$$

as $n \to \infty$ because the X_is and the ε_is are independent and the ε_i's have zero median. So,

$$\begin{aligned} &\max_{1 \le j \le m} \left| \frac{1}{n} \sum_{i=1}^{n} (f_*(X_i) - h_j(X_i) - \epsilon/2)^+ (1/2 - I(\varepsilon_i \le 0)) I(X_i \in \mathcal{H}_{\nu_0}) \right| \\ &\le \epsilon\eta/24 \end{aligned} \tag{10}$$

a.s. for n sufficiently large.

We let B be the set

$$\begin{aligned} \Bigg\{ &\sup_{\overline{g}_n \in \mathcal{S}_n} \frac{1}{n} \sum_{i=1}^{n} (f_*(X_i) - \overline{g}_n(X_i) - \epsilon/2)^+ I(X_i \notin \mathcal{H}_{\nu_0}) \le \epsilon\eta/12 \text{ for } n \text{ sufficiently large}, \\ &\max_{1 \le j \le m} \left| \frac{1}{n} \sum_{i=1}^{n} (f_*(X_i) - h_j(X_i) - \epsilon/2)^+ (1/2 - I(\varepsilon_i \le 0)) I(X_i \in \mathcal{H}_{\nu_0}) \right| \le \epsilon\eta/24 \\ &\quad \text{for } n \text{ sufficiently large} \Bigg\}, \end{aligned}$$

then by (9) and (10) we have $\mathbb{P}(B) = 1$.

Now, it remains to show that $A \cap B \cap C^c = \emptyset$. Suppose, on the contrary, that $\omega \in A \cap B \cap C^c$. Then for such ω, there exists $\tilde{g}_n \in \mathcal{S}_n$ such that

$$\inf_{x \in B(z,\delta)} (f_*(x) - \tilde{g}_n(x)) \quad > \quad \epsilon \tag{11}$$

for infinitely many n.

Define $k_n : [0,1]^d \to \mathbb{R}$ by $k_n = \max(f_*(x) - \epsilon/2, \tilde{g}_n(x))$ for $x \in [0,1]^d$. Since k_n is convex, we must have

$$\varphi_n(k_n) \geq \varphi_n(\tilde{g}_n),$$

or equivalently,

$$\begin{aligned} 0 \quad &\leq \varphi_n(k_n) - \varphi(\tilde{g}_n) \\ = \quad & \frac{1}{n}\sum_{i=1}^n |Y_i - k_n(X_i)| - \frac{1}{n}\sum_{i=1}^n |Y_i - \tilde{g}_n(X_i)| \\ = \quad & \frac{1}{n}\sum_{X_i \in P_n} |Y_i - k_n(X_i)| - \frac{1}{n}\sum_{X_i \in P_n} |Y_i - \tilde{g}_n(X_i)|, \end{aligned}$$

where $P_n = \left\{x \in [0,1]^d : f_*(x) - \epsilon/2 \geq \tilde{g}_n(x)\right\}$.

We denote

$$\begin{aligned} Q_{i,n} &= \{X_i \in P_n\} \cap \{\varepsilon_i + \epsilon/2 < -(f_*(X_i) - \tilde{g}_n(X_i) - \epsilon/2)\} \\ R_{i,n} &= \{X_i \in P_n\} \cap \{-(f_*(X_i) - \tilde{g}_n(X_i) - \epsilon/2) \leq \varepsilon_i + \epsilon/2 < 0\} \\ S_{i,n} &= \{X_i \in P_n\} \cap \{0 \leq \varepsilon_i + \epsilon/2\} \end{aligned}$$

for $1 \leq i \leq n$ and observe that

$$\begin{aligned} 0 \quad &\leq \quad \varphi_n(k_n) - \varphi(\tilde{g}_n) \\ &= \quad \frac{1}{n}\sum_{X_i \in P_n} |Y_i - (f_*(X_i) - \epsilon/2)| - \frac{1}{n}\sum_{X_i \in P_n} |Y_i - \tilde{g}_n(X_i)| \\ &= \quad \frac{1}{n}\sum_{X_i \in P_n} |\varepsilon_i + \epsilon/2| - \frac{1}{n}\sum_{X_i \in P_n} |\varepsilon_i + \epsilon/2 + (f_*(X_i) - \tilde{g}_n(X_i) - \epsilon/2)| \\ &= \quad \frac{1}{n}\sum_{i=1}^n (f_*(X_i) - \tilde{g}_n(X_i) - \epsilon/2)\, I(Q_{i,n}) \\ &\quad -\frac{1}{n}\sum_{i=1}^n (2\varepsilon_i + \epsilon + f_*(X_i) - \tilde{g}_n(X_i) - \epsilon/2)\, I(R_{i,n}) \\ &\quad -\frac{1}{n}\sum_{i=1}^n (f_*(X_i) - \tilde{g}_n(X_i) - \epsilon/2)\, I(S_{i,n}) \\ &= \quad -\frac{1}{n}\sum_{i=1}^n (f_*(X_i) - \tilde{g}_n(X_i) - \epsilon/2)^+ (1 - 2I(Q_{i,n})) - \frac{2}{n}\sum_{i=1}^n (\varepsilon_i + \epsilon/2)\, I(R_{i,n}) \\ &= \quad -\frac{2}{n}\sum_{i=1}^n (f_*(X_i) - \tilde{g}_n(X_i) - \epsilon/2)^+ (1/2 - I(Q_{i,n})) - \frac{2}{n}\sum_{i=1}^n (\varepsilon_i + \epsilon/2)\, I(R_{i,n}) \\ &= \quad -\frac{2}{n}\sum_{i=1}^n (f_*(X_i) - \tilde{g}_n(X_i) - \epsilon/2)^+ (I(\varepsilon_i \leq 0) - I(Q_{i,n})) \end{aligned} \tag{12}$$

$$
\begin{aligned}
& -\frac{2}{n}\sum_{i=1}^{n}\left(f_*(X_i)-\tilde{g}_n(X_i)-\epsilon/2\right)^+(1/2-I(\varepsilon_i\le 0))-\frac{2}{n}\sum_{i=1}^{n}(\varepsilon_i+\epsilon/2)\,I(R_{i,n})\\
= & -\frac{2}{n}\sum_{i=1}^{n}\left(f_*(X_i)-\tilde{g}_n(X_i)-\epsilon/2\right)^+\left(I(R_{i,n})+I(-\epsilon/2\le\varepsilon_i\le 0)\right)\\
& -\frac{2}{n}\sum_{i=1}^{n}\left(f_*(X_i)-\tilde{g}_n(X_i)-\epsilon/2\right)^+(1/2-I(\varepsilon_i\le 0))-\frac{2}{n}\sum_{i=1}^{n}(\varepsilon_i+\epsilon/2)\,I(R_{i,n})\\
= & -\frac{2}{n}\sum_{i=1}^{n}\left(f_*(X_i)-\tilde{g}_n(X_i)+\varepsilon_i\right)I(R_{i,n})\\
& -\frac{2}{n}\sum_{i=1}^{n}\left(f_*(X_i)-\tilde{g}_n(X_i)-\epsilon/2\right)^+ I(-\epsilon/2\le\varepsilon_i\le 0)\\
& -\frac{2}{n}\sum_{i=1}^{n}\left(f_*(X_i)-\tilde{g}_n(X_i)-\epsilon/2\right)^+(1/2-I(\varepsilon_i\le 0))\\
\le & -\frac{2}{n}\sum_{i=1}^{n}\left(f_*(X_i)-\tilde{g}_n(X_i)-\epsilon/2\right)^+ I(-\epsilon/2\le\varepsilon_i\le 0)\\
& -\frac{2}{n}\sum_{i=1}^{n}\left(f_*(X_i)-\tilde{g}_n(X_i)-\epsilon/2\right)^+(1/2-I(\varepsilon_i\le 0))=\mathrm{I}+\mathrm{II},\ \text{say.}
\end{aligned}
\tag{13}
$$

The last inequality follows because $f_*(X_i)-\tilde{g}_n(X_i)+\varepsilon_i\ge 0$ on $R_{i,n}$.

From (11) and the fact that $\omega\in A$, we have

$$
\mathrm{I}\le-\frac{1}{n}\sum_{i=1}^{n}\epsilon I(X_i\in B(z,\delta),-\epsilon/2\le\varepsilon_i\le 0)\le-\epsilon\eta/2 \tag{14}
$$

for infinitely many n.

On the other hand, for each $1\le j\le m$,

$$
\begin{aligned}
\mathrm{II}=&-(2/n)\sum_{i=1}^{n}\left(f_*(X_i)-\tilde{g}_n(X_i)-\epsilon/2\right)^+(1/2-I(\varepsilon_i\le 0))\\
= & -(2/n)\sum_{i=1}^{n}\left(f_*(X_i)-\tilde{g}_n(X_i)-\epsilon/2\right)^+(1/2-I(\varepsilon_i\le 0))I(X_i\notin\mathcal{H}_{\nu_0})\\
& -(2/n)\sum_{i=1}^{n}\left(f_*(X_i)-\tilde{g}_n(X_i)-\epsilon/2\right)^+(1/2-I(\varepsilon_i\le 0))I(X_i\in\mathcal{H}_{\nu_0})\\
\le & -(2/n)\sum_{i=1}^{n}\left(f_*(X_i)-\tilde{g}_n(X_i)-\epsilon/2\right)^+(1/2-I(\varepsilon_i\le 0))I(X_i\notin\mathcal{H}_{\nu_0})\\
& -(2/n)\sum_{i=1}^{n}\left(f_*(X_i)-h_j(X_i)-\epsilon/2\right)^+(1/2-I(\varepsilon_i\le 0))I(X_i\in\mathcal{H}_{\nu_0})\\
& +(2/n)\sum_{i=1}^{n}\left|h_j(X_i)-\tilde{g}_n(X_i)\right|\,|1/2-I(\varepsilon_i\le 0)|\,I(X_i\in\mathcal{H}_{\nu_0})\\
& \quad\text{because }-(a+b)^+c\le-a^+c+|b||c|\text{ for }a,b,c\in\mathbb{R}\\
\le & \ (2/n)\sum_{i=1}^{n}\left(f_*(X_i)-\tilde{g}_n(X_i)-\epsilon/2\right)^+|1/2-I(\varepsilon_i\le 0)|\,I(X_i\notin\mathcal{H}_{\nu_0})\\
& +2\max_{1\le j\le m}\left|\frac{1}{n}\sum_{X_i\in\mathcal{H}_{\nu_0}}\left(f_*(X_i)-h_j(X_i)-\epsilon/2\right)^+(1/2-I(\varepsilon_i\le 0))\right|\\
& +(2/n)\sum_{i=1}^{n}\sup_{x\in\mathcal{H}_{\nu_0}}\left|h_j(x)-\tilde{g}_n(x)\right|\,|1/2-I(\varepsilon_i\le 0)|\,I(X_i\in\mathcal{H}_{\nu_0}).
\end{aligned}
\tag{15}
$$

Since (15) holds for any $j \in \{1, \ldots, m\}$,

$$\begin{aligned}
\mathrm{II} \quad &\leq \quad (2/n) \sum_{i=1}^{n} (f_*(X_i) - \tilde{g}_n(X_i) - \epsilon/2)^+ \left|1/2 - I(\varepsilon_i \leq 0)\right| I(X_i \notin \mathcal{H}_{\nu_0}) \\
&\quad +2 \max_{1 \leq j \leq m} \left| \frac{1}{n} \sum_{X_i \in \mathcal{H}_{\nu_0}} \left(f_*(X_i) - h_j(X_i) - \epsilon/2\right)^+ (1/2 - I(\varepsilon_i \leq 0)) \right| \\
&\quad +\epsilon\eta/12 \\
&\leq \quad \epsilon\eta/12 + \epsilon\eta/12 + \epsilon\eta/12 \quad \text{because } \omega \in B \\
&= \quad \epsilon\eta/4 \qquad (16)
\end{aligned}$$

a.s. for n sufficiently large.

Combination of (13), (14), and (16) gives $0 \leq \varphi(k_n) - \varphi(\tilde{g}_n) \leq -\epsilon\eta/4$ for infinitely many n, which is a contradiction. This proves that $A \cap B \cap C^c = \emptyset$ and that $\mathbb{P}(C) = 1$.

A.5 Proof of Lemma 6

Let

$$C = \{ \sup_{\overline{g}_n \in \mathcal{S}_n} \inf_{x \in B(z,\delta)} (\overline{g}_n(x) - f_*(x)) \leq \epsilon \text{ for } n \text{ sufficiently large}\}.$$

We will prove that $\mathbb{P}(C) = 1$ by showing that $\mathbb{P}(A \cap B \cap C^c) = \emptyset$, where A and B will be defined subsequently. (It will also be shown later that $\mathbb{P}(A) = \mathbb{P}(B) = 1$.)

Let

$$A = \left\{ \frac{1}{n} \sum_{i=1}^{n} I(X_i \in B(z,\delta), 0 < \varepsilon_i < \epsilon/2) \geq \eta/2 \text{ for } n \text{ sufficiently large} \right\},$$

where $\eta \triangleq \mathbb{P}(X_1 \in B(z,\delta), 0 < \varepsilon_1 < \epsilon/2)$. By the strong law of large numbers, $\mathbb{P}(A) = 1$.

On the other hand, the strong law of large numbers and A4 ensure that

$$\frac{1}{n} \sum_{i=1}^{n} (1/2 - I(\varepsilon_i > 0)) = \frac{1}{n} \sum_{i=1}^{n} (I(\varepsilon_i \leq 0) - 1/2) \geq -\eta/16$$

a.s. for n sufficiently large. Also, similar arguments leading to (16) ensure that

$$\inf_{\overline{g}_n \in \mathcal{S}_n} \frac{1}{n} \sum_{i=1}^{n} (f_*(X_i) - \overline{g}_n(X_i) + \epsilon)^+ (1/2 - I(\varepsilon_i \leq 0)) \geq -\epsilon\eta/16$$

a.s. for n sufficiently large.

So, if we let

$$\begin{aligned}
B = \Bigg\{ &\frac{1}{n} \sum_{i=1}^{n} (1/2 - I(\varepsilon_i > 0)) \geq -\eta/16 \text{ for } n \text{ sufficiently large} \\
&\inf_{\overline{g}_n \in \mathcal{S}_n} \frac{1}{n} \sum_{i=1}^{n} (f_*(X_i) - \overline{g}_n(X_i) + \epsilon)^+ (1/2 - I(\varepsilon_i < 0)) \geq -\epsilon\eta/16 \\
&\text{for } n \text{ sufficiently large} \Big\},
\end{aligned}$$

then $\mathbb{P}(B) = 1$.

Now, it remains to show that $A \cap B \cap C^c = \emptyset$. Suppose, on the contrary, that $\omega \in A \cap B \cap C^c$. Then for such ω, there exists $\tilde{g}_n \in \mathcal{S}_n$ such that

$$\inf_{x \in B(z,\delta)} (\tilde{g}_n(x) - f_*(x)) \quad > \quad \epsilon \qquad (17)$$

for infinitely many n.

Define $k_n : [0,1]^d \to \mathbb{R}$ by $k_n(x) = \max(\tilde{g}_n(x) - \epsilon, f_*(x))$ for $x \in [0,1]^d$. Since k_n is convex, we must have

$$\varphi_n(k_n) \geq \varphi_n(\tilde{g}_n),$$

or equivalently,

$$\begin{aligned} 0 &\leq \varphi_n(k_n) - \varphi_n(\tilde{g}_n) \\ &= \frac{1}{n}\sum_{i=1}^n |Y_i - k_n(X_i)| - \frac{1}{n}\sum_{i=1}^n |Y_i - \tilde{g}_n(X_i)| \\ &= \frac{1}{n}\sum_{X_i \in P_n} |Y_i - \tilde{g}_n(X_i) + \epsilon| - \frac{1}{n}\sum_{X_i\in P_n} |Y_i - \tilde{g}_n(X_i)| \\ &\quad + \frac{1}{n}\sum_{X_i \in P_n^c} |\varepsilon_i| - \frac{1}{n}\sum_{X_i \in P_n^c} |\varepsilon_i + f_*(X_i) - \tilde{g}_n(X_i)| \\ &= \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}, \text{ say,} \end{aligned} \tag{18}$$

where $P_n = \left\{x \in [0,1]^d : \tilde{g}_n(x) - \epsilon \geq f_*(x)\right\}$.
We denote

$$\begin{aligned} Q_{i,n} &= \{X_i \in P_n\} \cap \{\varepsilon_i \geq -(f_*(X_i) - \tilde{g}_n(X_i))\} \\ R_{i,n} &= \{X_i \in P_n\} \cap \{-(f_*(X_i) - \tilde{g}_n(X_i) + \epsilon) \leq \varepsilon_i < -(f_*(X_i) - \tilde{g}_n(X_i))\} \\ S_{i,n} &= \{X_i \in P_n\} \cap \{\varepsilon_i < -(f_*(X_i) - \tilde{g}_n(X_i) + \epsilon)\} \end{aligned}$$

and observe that

$$\begin{aligned} &\mathrm{I} + \mathrm{II} = \frac{1}{n}\sum_{X_i\in P_n} |Y_i - \tilde{g}_n(X_i) + \epsilon| - \frac{1}{n}\sum_{X_i\in P_n} |Y_i - \tilde{g}_n(X_i)| \\ &= \frac{1}{n}\sum_{X_i\in P_n} |f_*(X_i) - \tilde{g}_n(X_i) + \varepsilon_i + \epsilon| - \frac{1}{n}\sum_{X_i\in P_n} |f_*(X_i) - \tilde{g}_n(X_i) + \varepsilon_i| \\ &= \frac{1}{n}\sum_{i=1}^n \epsilon I(Q_{i,n}) + \frac{1}{n}\sum_{i=1}^n (2f_*(X_i) - 2\tilde{g}_n(X_i) + 2\varepsilon_i + \epsilon)\, I(R_{i,n}) - \frac{1}{n}\sum_{i=1}^n \epsilon I(S_{i,n}) \\ &= -\frac{1}{n}\sum_{X_i\in P_n} \epsilon\left(1 - 2I(S_{i,n}^c)\right) + \frac{2}{n}\sum_{i=1}^n (f_*(X_i) - \tilde{g}_n(X_i) + \varepsilon_i)\, I(R_{i,n}) \\ &= -\frac{2}{n}\sum_{X_i\in P_n} \epsilon\left(1/2 - I(S_{i,n}^c)\right) + \frac{2}{n}\sum_{i=1}^n (f_*(X_i) - \tilde{g}_n(X_i) + \varepsilon_i)\, I(R_{i,n}) \\ &= -\frac{2}{n}\sum_{X_i\in P_n} \epsilon\left(I(\varepsilon_i > 0) - I(S_{i,n}^c)\right) - \frac{2}{n}\sum_{X_i\in P_n} \epsilon\,(1/2 - I(\varepsilon_i > 0)) \\ &\qquad + \frac{2}{n}\sum_{i=1}^n (f_*(X_i) - \tilde{g}_n(X_i) + \varepsilon_i)\, I(R_{i,n}) \\ &= -\frac{2}{n}\sum_{X_i\in P_n} \epsilon I\,(0 < \varepsilon_i < -(f_*(X_i) - \tilde{g}_n(X_i) + \epsilon)) - \frac{2}{n}\sum_{X_i\in P_n} \epsilon\,(1/2 - I(\varepsilon_i > 0)) \\ &+ \frac{2}{n}\sum_{i=1}^n (f_*(X_i) - \tilde{g}_n(X_i) + \varepsilon_i)\, I(R_{i,n}) \\ &\leq -\frac{2}{n}\sum_{X_i\in P_n} \epsilon I\,(0 < \varepsilon_i < -(f_*(X_i) - \tilde{g}_n(X_i) + \epsilon)) - \frac{2}{n}\sum_{X_i\in P_n} \epsilon\,(1/2 - I(\varepsilon_i > 0)) \\ &+ \frac{2}{n}\sum_{X_i\in P_n} (f_*(X_i) - \tilde{g}_n(X_i) + \varepsilon_i) \\ &\cdot I(-(f_*(X_i) - \tilde{g}_n(X_i) + \epsilon) \leq \varepsilon_i < -(f_*(X_i) - \tilde{g}_n(X_i) + \epsilon/2)) \end{aligned} \tag{19}$$

$$\le -\frac{2}{n}\sum_{X_i\in P_n} \epsilon I\left(0<\varepsilon_i<-\left(f_*(X_i)-\tilde g_n(X_i)+\epsilon\right)\right)-\frac{2}{n}\sum_{X_i\in P_n}\epsilon\left(1/2-I(\varepsilon_i>0)\right)$$
$$-\frac{1}{n}\sum_{X_i\in P_n}\epsilon I(-\left(f_*(X_i)-\tilde g_n(X_i)+\epsilon\right)\le\varepsilon_i<-\left(f_*(X_i)-\tilde g_n(X_i)+\epsilon/2\right))$$
$$\le -\frac{1}{n}\sum_{X_i\in P_n}\epsilon I\left(0<\varepsilon_i<-\left(f_*(X_i)-\tilde g_n(X_i)+\epsilon/2\right)\right)$$
$$-\frac{2}{n}\sum_{X_i\in P_n}\epsilon\left(1/2-I(\varepsilon_i>0)\right)$$
$$\le -\frac{1}{n}\sum_{X_i\in P_n}\epsilon I\left(0<\varepsilon_i<\epsilon/2\right)-\frac{2}{n}\sum_{X_i\in P_n}\epsilon\left(1/2-I(\varepsilon_i>0)\right). \tag{20}$$

On the other hand, to handle III + IV, we consider the cases when 1) $0\le f_*(X_i)-\tilde g_n(X_i)$ and 2) $-\epsilon<f_*(X_i)-\tilde g_n(X_i)<0$. For case 1), we consider the three cases $\varepsilon_i\le-(f_*(X_i)-\tilde g_n(X_i))$, $-(f_*(X_i)-\tilde g_n(X_i))<\varepsilon_i\le 0$, and $0<\varepsilon_i$. For case 2), we consider the three cases $\varepsilon_i\le 0$, $0<\varepsilon_i\le-(f_*(X_i)-\tilde g_n(X_i))$, and $-(f_*(X_i)-\tilde g_n(X_i))<\varepsilon_i$. This yields the following relation:

$$\text{III}+\text{IV}$$
$$=\frac{1}{n}\sum_{X_i\in P_n^c}|\varepsilon_i|-\frac{1}{n}\sum_{X_i\in P_n^c}|\varepsilon_i+f_*(X_i)-\tilde g_n(X_i)|$$
$$=-\frac{2}{n}\sum_{X_i\in P_n^c}\left(f_*(X_i)-\tilde g_n(X_i)+\varepsilon_i\right)$$
$$\cdot I\left(0\le f_*(X_i)-\tilde g_n(X_i),-\left(f_*(X_i)-\tilde g_n(X_i)\right)<\varepsilon_i\le 0\right)$$
$$-\frac{2}{n}\sum_{X_i\in P_n^c}\left(f_*(X_i)-\tilde g_n(X_i)\right)$$
$$\cdot\left((1/2)I(0\le f_*(X_i)-\tilde g_n(X_i))-I\left(0\le f_*(X_i)-\tilde g_n(X_i),\ \varepsilon_i\le 0\right)\right)$$
$$+\frac{2}{n}\sum_{X_i\in P_n^c}\left(f_*(X_i)-\tilde g_n(X_i)+\varepsilon_i\right)$$
$$\cdot I(-\epsilon<f_*(X_i)-\tilde g_n(X_i)<0,0<\varepsilon_i\le-\left(f_*(X_i)-\tilde g_n(X_i)\right))$$
$$+\frac{2}{n}\sum_{X_i\in P_n^c}\left(f_*(X_i)-\tilde g_n(X_i)\right)$$
$$\cdot\left((1/2)I\left(-\epsilon<f_*(X_i)-\tilde g_n(X_i)<0\right)-I\left(-\epsilon<f_*(X_i)-\tilde g_n(X_i)<0,\ \varepsilon_i>0\right)\right).$$

Since the first and the third terms in the above equations are always negative, we have

$$\text{III}+\text{IV}$$
$$\le -\frac{2}{n}\sum_{X_i\in P_n^c}\left(f_*(X_i)-\tilde g_n(X_i)\right)$$
$$\cdot\left((1/2)I(0\le f_*(X_i)-\tilde g_n(X_i))-I\left(0\le f_*(X_i)-\tilde g_n(X_i),\ \varepsilon_i\le 0\right)\right)$$
$$+\frac{2}{n}\sum_{X_i\in P_n^c}\left(f_*(X_i)-\tilde g_n(X_i)\right)$$
$$\cdot\left((1/2)I\left(-\epsilon<f_*(X_i)-\tilde g_n(X_i)<0\right)-I\left(-\epsilon<f_*(X_i)-\tilde g_n(X_i)<0,\ \varepsilon_i>0\right)\right)$$
$$=-\frac{2}{n}\sum_{X_i\in P_n^c}\left(f_*(X_i)-\tilde g_n(X_i)+\epsilon\right)$$
$$\cdot\left((1/2)I(0\le f_*(X_i)-\tilde g_n(X_i))-I\left(0\le f_*(X_i)-\tilde g_n(X_i),\ \varepsilon_i\le 0\right)\right)$$
$$+\frac{2}{n}\sum_{X_i\in P_n^c}\epsilon$$

$$
\begin{aligned}
&\cdot ((1/2)I(0 \le f_*(X_i) - \tilde{g}_n(X_i)) - I(0 \le f_*(X_i) - \tilde{g}_n(X_i),\ \varepsilon_i \le 0)) \\
&+ \frac{2}{n} \sum_{X_i \in P_n^c} (f_*(X_i) - \tilde{g}_n(X_i) + \epsilon) \\
&\cdot ((1/2)I(-\epsilon < f_*(X_i) - \tilde{g}_n(X_i) < 0) - I(-\epsilon < f_*(X_i) - \tilde{g}_n(X_i) < 0,\ \varepsilon_i > 0)) \\
&- \frac{2}{n} \sum_{X_i \in P_n^c} \epsilon \\
&\cdot ((1/2)I(-\epsilon < f_*(X_i) - \tilde{g}_n(X_i) < 0) - I(-\epsilon < f_*(X_i) - \tilde{g}_n(X_i) < 0,\ \varepsilon_i > 0)) \\
&= -\frac{2}{n} \sum_{X_i \in P_n^c} (f_*(X_i) - \tilde{g}_n(X_i) + \epsilon) \\
&\cdot ((1/2)I(0 \le f_*(X_i) - \tilde{g}_n(X_i)) - I(0 \le f_*(X_i) - \tilde{g}_n(X_i),\ \varepsilon_i \le 0)) \\
&- \frac{2}{n} \sum_{X_i \in P_n^c} \epsilon \\
&\cdot ((1/2)I(0 \le f_*(X_i) - \tilde{g}_n(X_i)) - I(0 \le f_*(X_i) - \tilde{g}_n(X_i),\ \varepsilon_i > 0)) \\
&- \frac{2}{n} \sum_{X_i \in P_n^c} (f_*(X_i) - \tilde{g}_n(X_i) + \epsilon) \\
&\cdot ((1/2)I(-\epsilon < f_*(X_i) - \tilde{g}_n(X_i) < 0) - I(-\epsilon < f_*(X_i) - \tilde{g}_n(X_i) < 0,\ \varepsilon_i \le 0)) \\
&- \frac{2}{n} \sum_{X_i \in P_n^c} \epsilon \\
&\cdot ((1/2)I(-\epsilon < f_*(X_i) - \tilde{g}_n(X_i) < 0) - I(-\epsilon < f_*(X_i) - \tilde{g}_n(X_i) < 0,\ \varepsilon_i > 0)).
\end{aligned}
$$

Combining the first and third terms in the above expression and combining the second and fourth terms in the above expression yield

$$
\begin{aligned}
&\mathrm{III} + \mathrm{IV} \\
&\le -\frac{2}{n} \sum_{X_i \in P_n^c} (f_*(X_i) - \tilde{g}_n(X_i) + \epsilon) \\
&\cdot ((1/2)I(-\epsilon < f_*(X_i) - \tilde{g}_n(X_i)) - I(-\epsilon < f_*(X_i) - \tilde{g}_n(X_i),\ \varepsilon_i \le 0)) \\
&- \frac{2}{n} \sum_{X_i \in P_n^c} \epsilon \\
&\cdot ((1/2)I(-\epsilon < f_*(X_i) - \tilde{g}_n(X_i)) - I(-\epsilon < f_*(X_i) - \tilde{g}_n(X_i),\ \varepsilon_i > 0)) \\
&= -\frac{2}{n} \sum_{i=1}^{n} (f_*(X_i) - \tilde{g}_n(X_i) + \epsilon)^+ (1/2 - I(\varepsilon_i \le 0)) - \frac{2}{n} \sum_{X_i \in P_n^c} \epsilon(1/2 - I(\varepsilon_i > 0)). \qquad (21)
\end{aligned}
$$

From (20) and (21), we obtain

$$
\begin{aligned}
&\mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} \\
&\le \quad -\frac{1}{n} \sum_{X_i \in P_n} \epsilon I(0 < \varepsilon_i < \epsilon/2) - \frac{2}{n} \sum_{X_i \in P_n} \epsilon (1/2 - I(\varepsilon_i > 0)) \\
&\qquad -\frac{2}{n} \sum_{i=1}^{n} (f_*(X_i) - \tilde{g}_n(X_i) + \epsilon)^+ (1/2 - I(\varepsilon_i \le 0)) \\
&\qquad -\frac{2}{n} \sum_{X_i \in P_n^c} \epsilon(1/2 - I(\varepsilon_i > 0)) \\
&= \quad -\frac{1}{n} \sum_{i=1}^{n} \epsilon I(X_i \in P_n, 0 < \varepsilon_i < \epsilon/2) - \frac{2}{n} \sum_{i=1}^{n} \epsilon (1/2 - I(\varepsilon_i > 0)) \\
&\qquad -\frac{2}{n} \sum_{i=1}^{n} (f_*(X_i) - \tilde{g}_n(X_i) + \epsilon)^+ (1/2 - I(\varepsilon_i \le 0)). \qquad (22)
\end{aligned}
$$

By (17),

$$-\frac{1}{n}\sum_{i=1}^{n}\epsilon I\left(X_i \in P_n, 0 < \varepsilon_i < \epsilon/2\right) \le -\frac{1}{n}\sum_{i=1}^{n}\epsilon I\left(X_i \in B(z,\delta), 0 < \varepsilon_i < \epsilon/2\right)$$

for infinitely many n and because $\omega \in A$,

$$-\frac{1}{n}\sum_{i=1}^{n}\epsilon I\left(X_i \in P_n, 0 < \varepsilon_i < \epsilon/2\right) \quad \le \quad -\epsilon\eta/2 \tag{23}$$

for infinitely many n. However, because $\omega \in B$,

$$-\frac{2}{n}\sum_{i=1}^{n}\epsilon\left(1/2 - I(\varepsilon_i > 0)\right) - \frac{2}{n}\sum_{i=1}^{n}\left(f_*(X_i) - \tilde{g}_n(X_i) + \epsilon\right)^+\left(1/2 - I(\varepsilon_i \le 0)\right) \le \epsilon\eta/4 \tag{24}$$

for n sufficiently large. Combination of (18), (22), (23), and (24) gives $0 \le \varphi(k_n) - \varphi(\tilde{g}_n) = \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} \le -\epsilon\eta/4$ for infinitely many n, which is a contradiction. This proves that $A \cap B \cap C^c = \emptyset$ and that $\mathbb{P}(C) = 1$.

A.6 Proof of Theorem 2

It suffices to prove the second part of Theorem 2. The first part of Theorem 2 can be justified similarly to the second part. Suppose that f_* is differentiable on $[c, 1-c]^d$. Take $c_0 < c$ and let

$$A = \left\{\sup_{x\in[c_0,1-c_0]^d, \overline{g}_n\in\mathcal{S}_n}\left|\overline{g}_n(x) - f_*(x)\right| \to 0 \text{ as } n \to \infty\right\},$$

then $\mathbb{P}(A) = 1$ by Theorem 1. We will show that $\mathbb{P}(B) = 1$, where

$$B = \left\{\sup_{x\in[c,1-c]^d, \xi\in\partial\overline{g}_n(x), \overline{g}_n\in\mathcal{S}_n}\|\xi - \nabla f_*(x)\| \to 0 \text{ as } n \to \infty\right\},$$

by proving that $A \cap B^c = \emptyset$.Suppose, on the contrary, that $\omega \in A \cap B^c$ exists. For such an ω, there exists $\epsilon > 0$, $x_n \in [c, 1-c]^d$, $\tilde{g}_n \in \mathcal{S}_n$ and $\xi_n \in \partial\tilde{g}_n(x_n)$ such that

$$\|\xi_n - \nabla f_*(x_n)\| > \epsilon$$

for infinitely many n. Furthermore, there exists an index $i \in \{1, \ldots, d\}$ such that

$$\left|e_i^T\xi_n - e_i^T\nabla f_*(x_n)\right| \quad > \quad \epsilon/d \tag{25}$$

for infinitely many n, where e_i is the ith unit vector. Equation (25) implies that either

$$e_i^T\xi_n \quad > \quad e_i^T\nabla f_*(x_n) + \epsilon/d \tag{26}$$

or

$$e_i^T\xi_n \quad < \quad e_i^T\nabla f_*(x_n) - \epsilon/d \tag{27}$$

holds. We first consider the case where (26) holds. Since $[c, 1-c]^d$ is compact, there exists a subsequence $(x_{n_k} : 1 \le k)$ that converges to a point x_0 in $[c, 1-c]^d$. Passing to subsequences if necessary, for any $\lambda > 0$ small enough that $x_0 + \lambda e_i \in [(c + c_0)/2, 1 - (c + c_0)/2]^d$, we have $x_n + \lambda e_i \in [c_0, 1 - c_0]^d$ for all sufficiently large n and

$$e_i^T\xi_n \quad \le \quad \left(\tilde{g}_n(x_n + \lambda e_i) - \tilde{g}_n(x_n)\right)/\lambda. \tag{28}$$

Since $\omega \in A$ and the $\tilde{g}_n$s are continuous on $[c_0, 1-c_0]^d$, $\tilde{g}_n(x_n + \lambda e_i)$ tends to $f_*(x_0 + \lambda e_i)$ and $\tilde{g}_n(x_n)$ tends to $f_*(x_0)$ as $n \to \infty$. By Theorem 25.5 on p. 246 of [**?**], ∇f_* is continuous on $[c, 1-c]^d$, and hence, $\nabla f_*(x_n)$ tends to $\nabla f_*(x_0)$. Therefore,

$$\begin{aligned}
e_i^T\nabla f_*(x_0) + \epsilon/d \quad &= \quad \lim_{n\to\infty} e_i^T\nabla f_*(x_n) + \epsilon/d \\
&\le \quad \limsup_{n\to\infty} e_i^T\xi_n \quad \text{by (26)} \\
&\le \quad \lim_{n\to\infty}\left(\tilde{g}_n(x_n + \lambda e_i) - \tilde{g}_n(x_n)\right)/\lambda \quad \text{by (28)} \\
&= \quad \left(f_*(x_0 + \lambda e_i) - f_*(x_0)\right)/\lambda.
\end{aligned} \tag{29}$$

This is supposed to hold for every sufficiently small $\lambda > 0$. However

$$e_i^T \nabla f_*(x_0) = \lim_{\lambda \downarrow 0} \left(f_*(x_0 + \lambda e_i) - f_*(x_0) \right) / \lambda,$$

which contradicts (29). Similar arguments can be applied to reach a contradiction in the case of (27). Hence, Theorem 2 is proved.

References

Bassett, G., & Koenker, R. (1978). Asymptotic theory of least absolute error regression. *J. Amer. Statist. Assoc., 363*, 618–622. http://dx.doi.org/10.1080/01621459.1978.10480065

Bronshtein, E. M. (1976). ϵ-entropy of convex sets and functions. *Siberian Math. J., 17,* 393–398.

Groeneboom, P., Jongbloed, G., & Wellner, J. A. (2001). Estimation of a convex function: characterizations and asymptotic theory. *Ann. Statist., 29*, 1653–1698.

Hanson, D. L., & Pledger, G. (1976). Consistency in concave regression. *Ann. Statist., 4*, 1038–1050. http://dx.doi.org/10.1214/aos/1176343640

Hildreth, C. (1954). Point estimates of ordinates of concave functions. *J. Amer. Statist. Assoc., 49*, 598–619 http://dx.doi.org/10.1080/016

Kuosmanen, T. (2008). Representation theorem for convex nonparametric least squares. *Econometrics Journal, 11*, 308–325. http://dx.doi.org/10.1111/j.1368-423X.2008.00239.x

Lim, E., & Luo, Y. (2014). On a least absolute deviations estimator of a convex function. *In Proceedings of the 2014 Winter Simulation Conference,* 2682–2691. http://dx.doi.org/10.1109/wsc.2014.7020112

Roberts, A. W., & Varberg, D. E. (1974). Another proof that convex functions are locally Lipschitz. *Amer. Math. Monthly, 81*, 1014–1016. http://dx.doi.org/10.2307/2319313

Rockafellar, R. T. (1970). *Convex analysis.* Princeton, NJ: Princeton University Press. http://dx.doi.org/10.1515/9781400873173

Taussky. O. (1949). A recurring theorem on determinants. *Amer. Math. Monthly, 56*, 672–676. http://dx.doi.org/10.2307/2305561

Wagner, H. M. (1959). Linear programming techniques for regression analysis. *J. Amer. Statist. Assoc., 285*, 206–212. http://dx.doi.org/10.1080/01621459.1959.10501506

Weber, R. R. (1983). A note on waiting times in single server queues. *Operations Research, 31*, 950–951. http://dx.doi.org/10.1287/opre.31.5.950

Copyrights