On Consistency of Absolute Deviations Estimators of Convex Functions

Yao Luo & Eunji Lim

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 \(f^* : [0, 1]^d \rightarrow \mathbb{R}\) to a data set of \(n\) observations \((X_1, Y_1), \ldots, (X_n, Y_n)\), where

\[ Y_i = f^*(X_i) + \epsilon_i \]

for \(1 \leq i \leq n\), the \(X_i\)s are continuous \([0, 1]^d\)-valued independent and identically distributed (iid) random vectors and the \(\epsilon_i\)s are iid random variables with a zero median and \(\mathbb{E}(|\epsilon_1|) < \infty\).

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

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

over the set of convex functions

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

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 program with \((d + 1)n\) decision variables and \(n^2\) constraints when \(d \geq 1\). 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 \(\hat{g}_n : [0, 1]^d \rightarrow \mathbb{R}\) 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 \(\tilde{g}_n\).
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 is computed faster and for a larger data sets than the least squares estimator $\hat{\beta}$.

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 $\hat{\beta}$ and prove that $\hat{\beta}(x)$ converges to $f_\epsilon(x)$ for any $x \in [0,1]^d$ as $n \to \infty$ with probability one. Our result will establish that $\hat{\beta}$ is a valid estimator of $f_\epsilon$.

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 $k$th component as $x^k$, so $x = (x^1, \ldots, x^d)^T$. We let $\|x\|_\infty = \max(|x^1| : 1 \leq i \leq d)$ and $\|x\|_2 = (\sum_{i=1}^{d} x^i)^{1/2}$. 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

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

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 $g(y) \geq g(x) + \xi^T(y - x)$ 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 $(a_n : n \geq 1)$ and $(b_n : n \geq 1)$ be sequences of real numbers. We say $a_n = O(b_n)$ if there exist positive constants $c$ and $n_0$ such that $|a_n| \leq c|b_n|$ for all $n \geq n_0$.

3. The Main Result

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

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

$$\min \frac{1}{n} \sum_{i=1}^{n} |Y_i - g_i|$$

s.t. $g_j \geq g_i + \xi_i^T(X_j - X_i), \quad 1 \leq i, j \leq n,$

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

$$\hat{\beta}(x) = \max_{1 \leq i \leq n} (\hat{g}^i + (\hat{\xi}^i)^T(x - X_i))$$

for $x \in [0,1]^d$, minimizes $\varphi_\epsilon$ 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 $(g_1, p_1, m_1, \xi_1), \ldots, (g_n, p_n, m_n, \xi_n)$:

$$\min \frac{1}{n} \sum_{i=1}^{n} (p_i + m_i)$$

s.t. $g_j \geq g_i + \xi_i^T(X_j - X_i), \quad 1 \leq i, j \leq n,$

$Y_i - g_i = p_i - m_i, \quad 1 \leq i \leq n,$

$p_i, m_i \geq 0, \quad 1 \leq i \leq n,$
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 \( \varphi_n \) over \( C: \)

\[
S_n = \{ g_n \in C : \varphi_n(g_n) \leq \varphi_n(g) \text{ for all } g \in C \}
\]

for \( n \geq 1 \). By Proposition 1, \( S_n \) is nonempty for all \( n \geq 1 \) almost surely. Proposition 1 suggests a way of computing an element \( \hat{g}_n \) in \( S_n \) 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 \( \varepsilon_i\)s are iid random variables with the common cumulative distribution function \( F \).

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

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

A4. For each \( x \in [0,1]^d, \) we have \( F(0|x) = 1/2 \). Therefore, given \( X_1, X_2, \ldots \), the \( \varepsilon_i\)s 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 \( 0 < c < 1/2, \)

\[
\sup_{x \in [c, 1-c]^d, \xi, \epsilon \in S_n} \left| \bar{g}_n(x) - f_*(x) \right| \to 0
\]

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

\[
\sup_{\xi \in \bar{g}_n(z), \epsilon \in S_n} \| \xi - \nabla f_*(z) \| \to 0
\]

as \( n \to \infty \) with probability one.

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

\[
\sup_{x \in [c, 1-c]^d, \xi \in \bar{g}_n(x), \epsilon \in S_n} \| \xi - \nabla f_*(x) \| \to 0
\]

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 \( \varphi_n(\bar{g}_n) \leq \varphi_n(f_*) \) for any \( \bar{g}_n \in S_n \), we must have

\[
\frac{1}{n} \sum_{i=1}^{n} |Y_i - \bar{g}_n(X_i)| \leq \frac{1}{n} \sum_{i=1}^{n} |Y_i - f_*(X_i)|
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} |f_*(X_i) + \varepsilon_i - f_*(X_i)| = \frac{1}{n} \sum_{i=1}^{n} |\varepsilon_i|.
\]

(4)
Step 2 Observe that, for any $\bar{g}_n \in S_n$, we must have
\[
\frac{1}{n} \sum_{i=1}^{n} |\bar{g}_n(X_i)| \leq \frac{1}{n} \sum_{i=1}^{n} |\epsilon_i| + \frac{1}{n} \sum_{i=1}^{n} |Y_i| \quad \text{by (4)}
\]
\[
\leq \frac{1}{n} \sum_{i=1}^{n} |\epsilon_i| + \frac{1}{n} \sum_{i=1}^{n} |f_e(X_i) + \epsilon_i|
\]
\[
\leq \frac{1}{n} \sum_{i=1}^{n} |\epsilon_i| + \frac{1}{n} \sum_{i=1}^{n} |f_e(X_i)| + \frac{n}{n} \sum_{i=1}^{n} |\epsilon_i|.
\]
Thus,
\[
\sup_{\bar{g}_n \in S_n} \frac{1}{n} \sum_{i=1}^{n} |\bar{g}_n(X_i)| \leq 2E|\epsilon_1| + E|f_e(X_1)| + 1 \triangleq \beta < \infty
\]
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
\[
\sup_{\bar{g}_n \in S_n} \inf_{x \in A} \left| \hat{g}_n(x) - f_e(x) \right| \leq \tilde{\beta}(A)
\]
a.s. for $n$ sufficiently large.

To fill in the details, we observe that the strong law of large numbers and A3 ensure
\[
\frac{1}{n} \sum_{i=1}^{n} |f_e(X_i)| = \frac{1}{n} \sum_{i=1}^{n} |Y_i - \epsilon_i| \leq \frac{1}{n} \sum_{i=1}^{n} |Y_i| + \frac{1}{n} \sum_{i=1}^{n} |\epsilon_i|
\]
\[
\leq E|Y_1| + E|\epsilon_1| + 1 \triangleq \tilde{\beta}
\]
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
\[
B = \left\{ \sup_{\bar{g}_n \in S_n} \frac{1}{n} \sum_{i=1}^{n} |\bar{g}_n(X_i)| \leq \beta \text{ for } n \text{ sufficiently large,} \right. \]
\[
\frac{1}{n} \sum_{i=1}^{n} |f_e(X_i)| \leq \tilde{\beta} \text{ for } n \text{ sufficiently large,}
\]
\[
\left. \text{and } \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} I(X_i \in A) \geq \mathbb{P}(X_1 \in A) \right\},
\]
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
\[
C = \left\{ \sup_{\bar{g}_n \in S_n} \inf_{x \in A} \left| \hat{g}_n(x) - f_e(x) \right| \leq \tilde{\beta}(A) \text{ for } n \text{ sufficiently large} \right\},
\]
by showing that $B \cap C^c = \emptyset$.

Suppose, on the contrary, that $\omega \in B \cap C^c$. Then for such $\omega$, there exists $\bar{g}_n \in S_n$ such that $\inf_{x \in A} \left| \hat{g}_n(x) - f_e(x) \right| > \tilde{\beta}(A)$.
for infinitely many \( n \). So, we would have

\[
\liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} |\overline{f}_n(X_i) - f_*(X_i)| \\
\geq \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} |\overline{f}_n(X_i) - f_*(X_i)| I(X_i \in A) \\
\geq \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} I(X_i \in A) \liminf_{n \to \infty} \frac{\sum_{i=1}^{n} |\overline{f}_n(X_i) - f_*(X_i)| I(X_i \in A)}{\max(1, \sum_{i=1}^{n} I(X_i \in A))} \\
\geq \mathbb{P}(X_1 \in A)\beta(A) = \beta + \tilde{\beta} + 1.
\]

On the other hand, we have

\[
\frac{1}{n} \sum_{i=1}^{n} |\overline{f}_n(X_i) - f_*(X_i)| \leq \frac{1}{n} \sum_{i=1}^{n} |\overline{f}_n(X_i)| + \frac{1}{n} \sum_{i=1}^{n} |f_*(X_i)| \leq \beta + \tilde{\beta}
\]

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 \( i \)th unit vector for \( 1 \leq i \leq d \). Let \( v_* = (1/(4d), 1/d, \ldots, 1/d) \). Let \( A_i \) be defined as follows:

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

Then there exists a positive constant \( \tau \) 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:

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

Then, there exists a positive constant \( \tau \) 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

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

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

$$
\sup_{x \in \mathcal{H}, \tilde{g}_n \in \mathcal{S}_n} \tilde{g}_n(x) \leq \tilde{\gamma}(c)
$$
a.s. for $n$ sufficiently large, where $\mathcal{H} = [c, 1-c]^d$.

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

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

and

$$
|f_x(x) - f_y(x)| \leq \alpha(c)|x-y|
$$

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

Step 9 Let

$$
C_c = \{ h : \mathcal{H} \to \mathbb{R} \text{ such that } h \text{ is convex on } \mathcal{H},
$$

$$
|h(x)| \leq |\tilde{\gamma}| + \tilde{\gamma}(c) \text{ and } |h(x) - h(y)| \leq \alpha(c)|x-y| \text{ for } x, y \in \mathcal{H} \}.
$$

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 $\tilde{g}_n \in \mathcal{S}_n$ imply that $\tilde{g}_n$ restricted to $\mathcal{H}$ belongs to $C_c$ a.s. Furthermore, $C_c$ is compact in the uniform metric $d_c$ given by

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

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 $\epsilon$–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 $\epsilon$ and $\delta$ and for any $z \in [0, 1]^d$, we have

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

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

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

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

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

$$
\sup_{x \in \mathcal{H}, \tilde{g}_n \in \mathcal{S}_n} \left( f_x(x) - \tilde{g}_n(x) \right) \leq \epsilon
$$
a.s. for $n$ sufficiently large.

Take $\delta = \epsilon/(6\alpha(c))$, where $\alpha(c)$ is given as in Step 8. Since $\mathcal{H}$ is compact, there exist a finite number of points $y_1, \ldots, y_l$ in $\mathcal{H}$ such that $B_c(y_i, \delta) \triangleq \{ x \in \mathcal{H} : \|x - y_i\| \leq \delta \}$ for $1 \leq i \leq l$ covers $\mathcal{H}$. 
If there exists \( \overrightarrow{s}_n \in S_n \) such that \( \sup_{x \in \mathcal{F}} (f(x) - \overrightarrow{s}_n(x)) > \epsilon \) for infinitely many \( n \), for each of such \( n \), there exists a point \( x_n \) in \( \mathcal{H}_n \) such that

\[
f(x_n) - \overrightarrow{s}_n(x_n) > \epsilon/2. \tag{6}
\]

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

\[
f(x) - f(x_{n_k}) + f(x_{n_k}) - \overrightarrow{s}_n(x_{n_k}) + \overrightarrow{s}_n(x_{n_k}) - \overrightarrow{s}_n(x) \geq -\epsilon/6 + \epsilon/2 - \epsilon/6 \geq \epsilon/6
\]

by (6). So, \( \sup_{x \in \mathcal{H}} (f(x) - \overrightarrow{s}_n(x)) > \epsilon \) implies \( \inf_{x \in B_r(x_j, \delta)} (f(x) - \overrightarrow{s}_n(x)) \geq \epsilon/6 \) for some \( j \), and hence,

\[
\mathbb{P}\left( \sup_{x \in \mathcal{H}, \overrightarrow{s}_n \in S_n} (f(x) - \overrightarrow{s}_n(x)) > \epsilon \text{ for infinitely many } n \right) = \sum_{j=1}^{\infty} \mathbb{P}\left( \sup_{x \in \mathcal{H}, \overrightarrow{s}_n \in S_n} \inf_{x \in B_r(x_j, \delta)} (f(x) - \overrightarrow{s}_n(x)) \geq \epsilon/6 \text{ for infinitely many } n \right) = 0
\]

by Step 10, proving Step 12.

**Step 13** For any \( \epsilon > 0 \),

\[
\sup_{x \in \mathcal{H}, \overrightarrow{s}_n \in S_n} (\overrightarrow{s}_n(x) - f(x)) \leq \epsilon
\]

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 \( F p = y - x_0 \), where \( p = (p^1, \ldots, p^d)^T \) and \( F = (F_{ij} : 0 \leq i, j \leq d) \) is a square \( d \times d \) matrix in which the \( i \)-th column is \( x_i - x_0 \) for \( 0 \leq i \leq d \). Note that \( F \) is invertible for sufficiently small \( r > 0 \) because we have

\[
|F_{ii}| |F_{jj}| > \left( \sum_{k=1}^n F_{ik} \right) \left( \sum_{k=1}^n F_{jk} \right)
\]

for all \( i \neq j \) and \( 1 \leq i, j \leq d \) with sufficiently small \( r \), 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 \( G = (G_{ij} : 0 \leq i, j \leq d) \) be a square \( d \times d \) matrix in which the first column is \( x_1 \) and the \( i \)-th column is \( e_i \) for \( 2 \leq i \leq d \). Observe that \( q^1, \ldots, q^d \) defined by

\[
q^1 = y^1 / x^1_1 \\
q^i = y^i - y^1 x^i_1 / x^1_1
\]

for \( 2 \leq i \leq d \),...
for \(2 \leq i \leq d\) satisfy \(G q = y\), where \(q = (q^1, \ldots, q^d)\). Note also that \(1/(8d) \leq q^i \leq 2y^i\) and \(1/(8d) \leq q^i \leq y^i\) for \(2 \leq i \leq d\) for \(\tau\) sufficiently small.

Set \(\|F\| = \sup_{|y| = 1} \|F y\|\). 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 \(\tau\) sufficiently small. Also, \(\|F^{-1}\| \leq 1/\max_{1 \leq i \leq d} \|F_{i,i}\| \leq 4\) for \(\tau\) sufficiently small. So,

\[
\|p - q\| = \|F^{-1}(y - x_0) - G^{-1}y\|
\]

\[
\leq \|(F^{-1} - G^{-1})y\| + \|F^{-1}x_0\|
\]

\[
\leq \|\gamma\| \|F^{-1}\| + \|\|F^{-1}\||\|\tau\|
\]

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

A.2 Proof of Lemma 3

First, we show that

\[
\inf_{x \in A_1 \cap S_n} \overline{\gamma}_n(x) \geq \tilde{\gamma}
\]

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

\[
\inf_{x \in B_1 \cap S_n} \overline{\gamma}_n(x) \geq \tilde{\gamma}
\]

a.s. for \(n\) sufficiently large.

By Step 3, there exists a positive constant \(\gamma\) such that

\[
\sup_{\overline{x} \in S_n} \inf_{x \in A_1} \left| \overline{\gamma}_n(x) - f_\gamma(x) \right| \leq \gamma
\]

a.s. for all \(0 \leq i \leq d + 1\) and \(n\) sufficiently large.

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

\[
\sup_{\overline{x} \in S_n} \inf_{x \in A_1} \left| \overline{\gamma}_n(x) \right| \leq M + \gamma
\]

(8)

a.s. for all \(0 \leq i \leq d + 1\) and \(n\) sufficiently large.

Set \(\tilde{\gamma} = -32d(M + \gamma + 1)\). For any \(\overline{x}_n \in S_n\), if \(\overline{\gamma}_n(x_1) \leq \tilde{\gamma}\) for some \(x_1 \in A_1\) and \(\overline{\gamma}_n(x_i) \leq (M + \gamma + 1)\) for some \(x_i \in A_1\) (\(i = 0, 2, \ldots, d\)), 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 \(p^1 \geq 1/(16d)\). So, we have

\[
\overline{\gamma}_n(y) = \overline{\gamma}_n(p^0x_0 + \cdots + p^dx_d)
\]

\[
\leq p^0\overline{\gamma}_n(x_0) + p^1\overline{\gamma}_n(x_1) + \cdots + p^d\overline{\gamma}_n(x_d) \quad \text{because \(\overline{\gamma}_n\) is convex}
\]

\[
\leq \tilde{\gamma}/(16d) + (M + \gamma + 1) \quad \text{because \(p^1 \geq 1/(16d)\) and \(\overline{\gamma}_n(x_i) \leq \tilde{\gamma} \leq 0\)
\]

\[
= -(M + \gamma + 1).
\]

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

\[
\inf_{x \in K_i} \overline{\gamma}_n(x) \geq M + \gamma + 1
\]

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

\[
\sup_{x \in A_{d+1}} \overline{\gamma}_n(x) \leq -(M + \gamma + 1).
\]
Thus,

\[ P \left( \inf_{x \in \mathcal{S}, \|x\| \leq \gamma} \bar{\|x\|n} \leq \bar{\gamma} \text{ for infinitely many } n \right) \]

\[ \leq \sum_{i=0}^{n-1} \sum_{j=0}^{d} P \left( \sup_{x \in \mathcal{S}, \|x\| \leq \gamma} \inf_{x \in \mathcal{S}, \|x\| \leq \gamma} \bar{\|x\|n} \geq M + \gamma + 1 \text{ for infinitely many } n \right) \]

\[ + P \left( \inf_{x \in \mathcal{S}, \|x\| \leq \gamma} \bar{\|x\|n} \leq -(M + \gamma + 1) \text{ for infinitely many } n \right) \]

\[ \leq \sum_{i=0}^{n-1} \sum_{j=0}^{d} P \left( \sup_{x \in \mathcal{S}, \|x\| \leq \gamma} \|x\| \geq M + \gamma + 1 \text{ for infinitely many } n \right) \]

\[ + P \left( \sup_{x \in \mathcal{S}, \|x\| \leq \gamma} \|x\| \geq M + \gamma + 1 \text{ for infinitely many } n \right) \]

\[ = 0 \]

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 \( y \in \mathcal{H} \) and for any \( x_i \in C_i (1 \leq i \leq d) \), where \( C_i = \{ x \in [0, 1]^d : \|x - e_i\| \leq \tau(c) \} \), there exist nonnegative real numbers \( p^1, \ldots, p^d \) such that \( p^1 x_1 + \cdots + p^d x_d = y \) and that \( p^i \leq 1 \) for \( 1 \leq i \leq d \).

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 \( 0 \leq p^i \leq 1 \) for \( 1 \leq i \leq d \), where \( H = (H_{ij} : 1 \leq i, j \leq d) \) is a square \( d \times d \) matrix in which the \( i \)th column is \( x_i \) for \( 1 \leq i \leq d \). Set \( \|H\|_{\infty} = \max_{1 \leq i \leq d} \sum_{j=1}^{d} |H_{ij}| \) and note that \( \|H - I_d\|_{\infty} \leq \tau(c) \), where \( I_d \) is the \( d \times d \) identity matrix. Hence, for \( \tau(c) < 1/2 \), \( H \) is invertible and we have

\[ \|H^{-1}\|_{\infty} = \|(I_d + H - I_d)^{-1}\|_{\infty} \leq (1 - \|H - I_d\|_{\infty})^{-1} \leq 2. \]

Therefore,

\[ \|p - y\|_{\infty} = \|H^{-1}y - y\|_{\infty} \leq \|H^{-1} - I_d\|_{\infty} \|y\|_{\infty} \leq \|H^{-1} - I_d\|_{\infty}. \]

Since mapping a \( d \times d \) matrix to its inverse matrix is continuous with respect to \( \| \cdot \|_{\infty} \) in a neighborhood of \( H \) and \( \|H - I_d\|_{\infty} \leq \tau(c) \), there exists a positive number \( \tau(c) \) that guarantees \( \|H^{-1} - I_d\|_{\infty} \leq c/2 \). So, for such \( \tau(c) \), \( \|p - y\|_{\infty} \leq c/2 \).

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 \( 1 \leq i \leq d, r > 0 \), and \( \bar{\|x\|n} \in \mathcal{S}_n \),

\[ \frac{1}{n} \sum_{j=1}^{n} I \left( X_j \in C_i, \bar{\|x\|n} \geq r \right) \]

\[ \geq \frac{1}{n} \sum_{j=1}^{n} I \left( X_j \in C_i \right) - \frac{1}{n} \sum_{j=1}^{n} I \left( X_j \in C_i, \bar{\|x\|n} > r \right). \]

However, Markov inequality and Step 2 imply that

\[ \sup_{\bar{\|x\|n} \in \mathcal{S}_n} \frac{1}{n} \sum_{j=1}^{n} I \left( X_j \in C_i, \bar{\|x\|n} > r \right) \leq \sup_{\bar{\|x\|n} \in \mathcal{S}_n} r^{-1} \frac{1}{n} \sum_{j=1}^{n} \bar{\|x\|n} \leq \beta/r \]

a.s. for \( n \) sufficiently large. Choose \( r_0 \) so large that \( \beta/r_0 \leq \bar{\gamma} \triangleq \min[\|\bar{\|x\|n} \| : 1 \leq i \leq d]/2 \), then

\[ \inf_{\bar{\|x\|n} \in \mathcal{S}_n} \frac{1}{n} \sum_{j=1}^{n} I \left( X_j \in C_i, \bar{\|x\|n} \geq r_0 \right) \geq \bar{\gamma} \]

a.s. for \( n \) sufficiently large.

For each such \( n \), there exists \( X_{R(i)} \in C_i \) with \( 1 \leq I(i) \leq n \) and \( \bar{\|x\|n}(X_{R(i)}) \leq r_0 \). For each \( y \in [c, 1-c]^d \) and \( X_{R(i)} \in C_i \) for \( 1 \leq i \leq d \), there exist \( p^1, \ldots, p^d \) such that

\[ y = p^1 X_{R(1)} + \cdots + p^d X_{R(d)} \]
and that $0 \leq \theta^j \leq 1$ for $1 \leq i \leq d$. So, the convexity of $\overline{g}_n$ yields

$$\overline{g}_n (y) \leq \theta^1 \overline{g}_n (X_{i(1)}) + \cdots + \theta^d \overline{g}_n (X_{i(d)}) \leq \delta_0,$$

proving that

$$\sup_{x \in [c-1,c]} \overline{g}_n (x) \leq \delta_0$$

as.s. for $n$ sufficiently large.

**A.4 Proof of Lemma 5**

Let

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

We will prove that $P(C) = 1$ by showing that $P(A \cap B \cap C^c) = 0$, where $A$ and $B$ will be defined subsequently. (It will also be shown later that $P(A) = P(B) = 1$.)

Let

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

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

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

$$E(I(X_1 \notin H_\nu)) \to 0$$

as $\nu \to 0$ because $I(X_1 \notin H_\nu) \downarrow 0$ a.s. as $\nu \downarrow 0$. So, take $\nu_0$ small enough so that

$$E(I(X_1 \notin H_{\nu_0})) \leq \epsilon/((24(M + |\tilde{y}|))$$

and note that

$$\frac{1}{n} \sum_{i=1}^{n} I(X_i \notin H_{\nu_0}) \leq \epsilon/((12(M + |\tilde{y}|))$$

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

$$\sup_{x \in S} \frac{1}{n} \sum_{i=1}^{n} (f_n(X_i) - \overline{g}_n(X_i) - \epsilon/2)^+ I(X_i \notin H_{\nu_0}) \leq \epsilon/12$$

as.s. for $n$ sufficiently large.

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

$$\frac{1}{n} \sum_{i=1}^{n} (f_n(X_i) - h_j(X_i) - \epsilon/2)^+ (1/2 - I(\varepsilon_i \leq 0)) I(X_i \in H_{\nu_0}) \to 0$$

as $n \to \infty$ because the $X_i$s and the $\varepsilon_i$s are independent and the $\varepsilon_i$s have zero median. So,

$$\max_{1 \leq j \leq m} \left| \frac{1}{n} \sum_{i=1}^{n} (f_n(X_i) - h_j(X_i) - \epsilon/2)^+ (1/2 - I(\varepsilon_i \leq 0)) I(X_i \in H_{\nu_0}) \right| \leq \epsilon/24$$

as.s. for $n$ sufficiently large.

We let $B$ be the set

$$\left\{ \sup_{x \in S} \frac{1}{n} \sum_{i=1}^{n} (f_n(X_i) - \overline{g}_n(X_i) - \epsilon/2)^+ I(X_i \notin H_{\nu_0}) \leq \epsilon/12 \text{ for } n \text{ sufficiently large},
$$

$$\max_{1 \leq j \leq m} \left| \frac{1}{n} \sum_{i=1}^{n} (f_n(X_i) - h_j(X_i) - \epsilon/2)^+ (1/2 - I(\varepsilon_i \leq 0)) I(X_i \in H_{\nu_0}) \right| \leq \epsilon/24$$

for $n$ sufficiently large}.
then by (9) and (10) we have \( P(B) = 1 \).

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 \( \omega \), there exists \( \tilde{g}_n \in S_n \) such that

\[
\inf_{x \in B(n) \cap S_n} (f_n(x) - \tilde{g}_n(x)) > \epsilon
\]

for infinitely many \( n \).

Define \( k_n : [0, 1]^d \to \mathbb{R} \) by \( k_n = \max (f_n(x) - \epsilon/2, \tilde{g}_n(x)) \) 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,

\[
0 \leq \varphi_n(k_n) - \varphi(\tilde{g}_n)
= \frac{1}{n} \sum_{i=1}^{n} (Y_i - k_n(X_i)) - \frac{1}{n} \sum_{i=1}^{n} (Y_i - \tilde{g}_n(X_i))
= \frac{1}{n} \sum_{X \in P_n} |Y_i - k_n(X)| - \frac{1}{n} \sum_{X \in P_n} |Y_i - \tilde{g}_n(X)|,
\]

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

We denote

\[
Q_{i,n} = \{ X_i \in P_n \} \cap \{ \varepsilon_i + \epsilon/2 < - (f_n(X_i) - \tilde{g}_n(X_i) - \epsilon/2) \}
\]

\[
R_{i,n} = \{ X_i \in P_n \} \cap \{ - (f_n(X_i) - \tilde{g}_n(X_i) - \epsilon/2) \leq \varepsilon_i + \epsilon/2 < 0 \}
\]

\[
S_{i,n} = \{ X_i \in P_n \} \cap \{ 0 \leq \varepsilon_i + \epsilon/2 \}
\]

for \( 1 \leq i \leq n \) and observe that

\[
0 \leq \varphi_n(k_n) - \varphi(\tilde{g}_n)
= \frac{1}{n} \sum_{X \in P_n} |Y_i - f_n(X_i) - \epsilon/2| - \frac{1}{n} \sum_{X \in P_n} |Y_i - \tilde{g}_n(X_i)|
= \frac{1}{n} \sum_{X \in P_n} |\varepsilon_i + \epsilon/2| - \frac{1}{n} \sum_{X \in P_n} |\varepsilon_i + \epsilon/2 + (f_n(X_i) - \tilde{g}_n(X_i) - \epsilon/2)|
= \frac{1}{n} \sum_{i=1}^{n} (f_n(X_i) - \tilde{g}_n(X_i) - \epsilon/2) I(Q_{i,n})
- \frac{1}{n} \sum_{i=1}^{n} (2\varepsilon_i + \epsilon + f_n(X_i) - \tilde{g}_n(X_i) - \epsilon/2) I(R_{i,n})
- \frac{1}{n} \sum_{i=1}^{n} (f_n(X_i) - \tilde{g}_n(X_i) - \epsilon/2) I(S_{i,n})
= -\frac{1}{n} \sum_{i=1}^{n} (f_n(X_i) - \tilde{g}_n(X_i) - \epsilon/2) (1 - 2I(Q_{i,n}))
- \frac{2}{n} \sum_{i=1}^{n} (\varepsilon_i + \epsilon/2) I(R_{i,n})
= -\frac{2}{n} \sum_{i=1}^{n} (f_n(X_i) - \tilde{g}_n(X_i) - \epsilon/2) (1/2 - I(Q_{i,n}))
- \frac{2}{n} \sum_{i=1}^{n} (\varepsilon_i + \epsilon/2) I(R_{i,n})
= -\frac{2}{n} \sum_{i=1}^{n} (f_n(X_i) - \tilde{g}_n(X_i) - \epsilon/2) (I(\varepsilon_i) - 0) - I(Q_{i,n})
\]

(12)
\[-\frac{2}{n} \sum_{i=1}^{n} (f_i(X_i) - \tilde{g}_n(X_i) - \epsilon/2)^+ (1/2 - I(\varepsilon_i \leq 0)) - \frac{2}{n} \sum_{i=1}^{n} (\epsilon_i + \epsilon/2) I(R_{i,n}) \]

\[= -\frac{2}{n} \sum_{i=1}^{n} (f_i(X_i) - \tilde{g}_n(X_i) - \epsilon/2)^+ (I(R_{i,n}) + I(-\epsilon/2 \leq \varepsilon_i \leq 0)) \]

\[= -\frac{2}{n} \sum_{i=1}^{n} (f_i(X_i) - \tilde{g}_n(X_i) - \epsilon/2)^+ (1/2 - I(\varepsilon_i \leq 0)) - \frac{2}{n} \sum_{i=1}^{n} (\epsilon_i + \epsilon/2) I(R_{i,n}) \]

\[\leq -\frac{2}{n} \sum_{i=1}^{n} (f_i(X_i) - \tilde{g}_n(X_i) - \epsilon/2)^+ (1/2 - I(\varepsilon_i \leq 0)) \]

\[-\frac{2}{n} \sum_{i=1}^{n} (f_i(X_i) - \tilde{g}_n(X_i) - \epsilon/2)^+ (1/2 - I(\varepsilon_i \leq 0)) = \text{I + II, say.} \quad (13)\]

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

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

\[1 \leq -\frac{1}{n} \sum_{i=1}^{n} \epsilon I(X_i \in B(z, \delta), -\epsilon/2 \leq \varepsilon_i \leq 0) \leq -\epsilon \eta/2 \quad (14)\]

for infinitely many \(n\).

On the other hand, for each \(1 \leq j \leq m\),

\[\text{II} = -(2/n) \sum_{i=1}^{n} (f_i(X_i) - \tilde{g}_n(X_i) - \epsilon/2)^+ (1/2 - I(\varepsilon_i \leq 0)) \]

\[= -(2/n) \sum_{i=1}^{n} (f_i(X_i) - \tilde{g}_n(X_i) - \epsilon/2)^+ (1/2 - I(\varepsilon_i \leq 0)) I(X_i \notin H_{i,n}) \]

\[= -(2/n) \sum_{i=1}^{n} (f_i(X_i) - \tilde{g}_n(X_i) - \epsilon/2)^+ (1/2 - I(\varepsilon_i \leq 0)) I(X_i \in H_{i,n}) \]

\[\leq -(2/n) \sum_{i=1}^{n} (f_i(X_i) - \tilde{g}_n(X_i) - \epsilon/2)^+ (1/2 - I(\varepsilon_i \leq 0)) I(X_i \notin H_{i,n}) \]

\[= -(2/n) \sum_{i=1}^{n} (f_i(X_i) - \tilde{g}_n(X_i) - \epsilon/2)^+ (1/2 - I(\varepsilon_i \leq 0)) I(X_i \in H_{i,n}) \]

\[+ (2/n) \sum_{i=1}^{n} |h_i(X_i) - \tilde{g}_n(X_i)| 1/2 - I(\varepsilon_i \leq 0) I(X_i \in H_{i,n}) \]

because \(- (a + b)^+ c \leq -a^+ c + |b| \) for \(a, b, c \in \mathbb{R}\)

\[\leq (2/n) \sum_{i=1}^{n} (f_i(X_i) - \tilde{g}_n(X_i) - \epsilon/2)^+ 1/2 - I(\varepsilon_i \leq 0) I(X_i \notin H_{i,n}) \]

\[+ 2 \max_{1 \leq j \leq m} \left| \frac{1}{n} \sum_{X_i \in H_{i,n}} \left( f_i(X_i) - h_j(X_i) - \epsilon/2 \right)^+ 1/2 - I(\varepsilon_i \leq 0) \right| \]

\[+ (2/n) \sum_{X_i \notin H_{i,n}} \sup_{x \in H_{i,n}} |h_i(x) - \tilde{g}_n(x)| 1/2 - I(\varepsilon_i \leq 0) I(X_i \in H_{i,n}). \quad (15)\]
Since (15) holds for any $j \in \{1, \ldots, m\}$,

$$II \leq \left( \frac{2}{n} \right) \sum_{i=1}^{n} \left( f_i(X_i) - \bar{g}_n(X_i) - \epsilon/2 \right)^+ \left| 1/2 - I(\epsilon_i \leq 0) \right| I(X_i \notin \mathcal{H}_n)$$

$$+ 2 \max_{1 \leq j \leq m} \left[ \frac{1}{n} \sum_{X_i \in \mathcal{H}_n} \left( f_i(X_i) - h_j(X_i) - \epsilon/2 \right)^+ \left( 1/2 - I(\epsilon_i \leq 0) \right) \right]$$

$$+ \epsilon \eta/12 \leq \epsilon \eta/12 + \epsilon \eta/12 + \epsilon \eta/12 \text{ because } \omega \in B$$

$$= \epsilon \eta/4 \quad (16)$$

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

$$C = \{ \sup_{\tilde{g}_n \in S_n, B(\tilde{g}_n, \delta)} (\tilde{g}_n(x) - f(x)) \leq \epsilon \text{ for } n \text{ sufficiently large} \}$$

We will prove that $\mathbb{P}(C) = 1$ by showing that $\mathbb{P}(A \cap B \cap C^c) = 0$, where $A$ and $B$ will be defined subsequently. (It will also be shown later that $\mathbb{P}(A) = \mathbb{P}(B) = 1$.)

Let

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

where $\eta \overset{\Delta}{=} \mathbb{P}(X_1 \in B(z, \delta), 0 < \epsilon_1 < \epsilon/2)$. 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

$$\mathbb{P} \left( \frac{1}{n} \sum_{i=1}^{n} (1/2 - I(\epsilon_i > 0)) \geq \frac{1}{n} \sum_{i=1}^{n} (I(\epsilon_i \leq 0) - 1/2) \geq -\eta/16 \right)$$

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

$$\inf_{\tilde{g}_n \in S_n} \frac{1}{n} \sum_{i=1}^{n} (f_i(X_i) - \bar{g}_n(X_i) + \epsilon)^+ (1/2 - I(\epsilon_i \leq 0)) \geq -\epsilon \eta/16$$

a.s. for $n$ sufficiently large.

So, if we let

$$B = \left\{ \frac{1}{n} \sum_{i=1}^{n} (1/2 - I(\epsilon_i > 0)) \geq -\eta/16 \text{ for } n \text{ sufficiently large} \right\},$$

$$\inf_{\tilde{g}_n \in S_n} \frac{1}{n} \sum_{i=1}^{n} (f_i(X_i) - \bar{g}_n(X_i) + \epsilon)^+ (1/2 - I(\epsilon_i < 0)) \geq -\epsilon \eta/16$$

for $n$ sufficiently large,

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 $\omega$, there exists $\tilde{g}_n \in S_n$ such that

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

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

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

or equivalently,

$$0 \leq \varphi_n(k_n) - \varphi_n(\hat{g}_n)$$

$$= \frac{1}{n} \sum_{i=1}^{n} |Y_i - k_n(X_i)| - \frac{1}{n} \sum_{i=1}^{n} |Y_i - \hat{g}_n(X_i)|$$

$$= \frac{1}{n} \sum_{X_i \in P_n} |Y_i - \hat{g}_n(X_i) + \epsilon| - \frac{1}{n} \sum_{X_i \in P_n} |Y_i - \hat{g}_n(X_i)|$$

$$+ \frac{1}{n} \sum_{X_i \in P_n} |\epsilon_i| - \frac{1}{n} \sum_{X_i \in P_n} |\epsilon_i + f_\ast(X_i) - \hat{g}_n(X_i)|$$

$$= I + II + III + IV, \text{ say,}$$

(18)

where $P_n = \{x \in [0, 1]^d : \hat{g}_n(x) - \epsilon \geq f_\ast(x)\}$.

We denote

$$Q_{i,n} = \{X_i \in P_n\} \cap \{|\epsilon_i| \geq (f_\ast(X_i) - \hat{g}_n(X_i))\}$$

$$R_{i,n} = \{X_i \in P_n\} \cap \{-(f_\ast(X_i) - \hat{g}_n(X_i)) \leq \epsilon_i \leq -(f_\ast(X_i) - \hat{g}_n(X_i))\}$$

$$S_{i,n} = \{X_i \in P_n\} \cap \{|\epsilon_i| < -(f_\ast(X_i) - \hat{g}_n(X_i)) + \epsilon\}$$

and observe that

$$I + II = \frac{1}{n} \sum_{X_i \in P_n} |Y_i - \hat{g}_n(X_i)| + \frac{1}{n} \sum_{X_i \in P_n} |Y_i - \hat{g}_n(X_i)|$$

$$= \frac{1}{n} \sum_{X_i \in P_n} |f_\ast(X_i) - \hat{g}_n(X_i) + \epsilon_i| + \frac{1}{n} \sum_{X_i \in P_n} |f_\ast(X_i) - \hat{g}_n(X_i) + \epsilon_i|$$

$$= \frac{1}{n} \sum_{X_i \in P_n} \epsilon I(Q_{i,n}) + \frac{1}{n} \sum_{i=1}^{n} (2f_\ast(X_i) - 2\hat{g}_n(X_i) + 2\epsilon_i + \epsilon) I(R_{i,n}) - \frac{1}{n} \sum_{i=1}^{n} \epsilon I(S_{i,n})$$

$$= -\frac{1}{n} \sum_{X_i \in P_n} \epsilon (1 - 2I(S_{i,n}^\ast)) + \frac{2}{n} \sum_{i=1}^{n} (f_\ast(X_i) - \hat{g}_n(X_i) + \epsilon_i) I(R_{i,n})$$

$$= -\frac{2}{n} \sum_{X_i \in P_n} \epsilon (1/2 - I(S_{i,n}^\ast)) + \frac{2}{n} \sum_{i=1}^{n} (f_\ast(X_i) - \hat{g}_n(X_i) + \epsilon_i) I(R_{i,n})$$

$$= -\frac{2}{n} \sum_{X_i \in P_n} \epsilon (I(\epsilon_i > 0) - I(S_{i,n}^\ast)) - \frac{2}{n} \sum_{i=1}^{n} \epsilon (1/2 - I(\epsilon_i > 0))$$

$$+ \frac{2}{n} \sum_{i=1}^{n} (f_\ast(X_i) - \hat{g}_n(X_i) + \epsilon_i) I(R_{i,n})$$

$$= -\frac{2}{n} \sum_{X_i \in P_n} \epsilon I(0 < \epsilon_i < -(f_\ast(X_i) - \hat{g}_n(X_i) + \epsilon)) - \frac{2}{n} \sum_{X_i \in P_n} \epsilon (1/2 - I(\epsilon_i > 0))$$

$$+ \frac{2}{n} \sum_{X_i \in P_n} (f_\ast(X_i) - \hat{g}_n(X_i) + \epsilon_i) I(R_{i,n})$$

$$\leq -\frac{2}{n} \sum_{X_i \in P_n} \epsilon I(0 < \epsilon_i < -(f_\ast(X_i) - \hat{g}_n(X_i) + \epsilon)) - \frac{2}{n} \sum_{X_i \in P_n} \epsilon (1/2 - I(\epsilon_i > 0))$$

$$+ \frac{2}{n} \sum_{X_i \in P_n} (f_\ast(X_i) - \hat{g}_n(X_i) + \epsilon_i)$$

$$\cdot I(-(f_\ast(X_i) - \hat{g}_n(X_i) + \epsilon) \leq \epsilon_i < -(f_\ast(X_i) - \hat{g}_n(X_i) + \epsilon)) (19)$$
For case 1), we consider the three cases

\[
\epsilon < -(f_i(X_i) - \tilde{g}_n(X_i)) - \frac{2}{n} \sum_{X \in S_n} \epsilon (1/2 - I(\epsilon_i > 0))
\]

\[
- \frac{1}{n} \sum_{X \in S_n} \epsilon (-(f_i(X_i) - \tilde{g}_n(X_i)) + \epsilon) \leq \epsilon_i < -(f_i(X_i) - \tilde{g}_n(X_i)) + \epsilon/2)
\]

\[
- \frac{1}{n} \sum_{X \in S_n} \epsilon (0 < \epsilon_i < -(f_i(X_i) - \tilde{g}_n(X_i)) + \epsilon/2)
\]

\[
- \frac{2}{n} \sum_{X \in S_n} \epsilon (1/2 - I(\epsilon_i > 0))
\]

On the other hand, to handle III + IV, we consider the cases when 1) \(0 \leq f_i(X_i) - \tilde{g}_n(X_i)\) and 2) \(-\epsilon < f_i(X_i) - \tilde{g}_n(X_i) < 0\).

For case 1), we consider the three cases \(\epsilon_i \leq -(f_i(X_i) - \tilde{g}_n(X_i))\), \((f_i(X_i) - \tilde{g}_n(X_i)) < \epsilon_i \leq 0\), and \(0 < \epsilon_i\). For case 2), we consider the three cases \(\epsilon_i \leq 0\), \(0 < \epsilon_i \leq -(f_i(X_i) - \tilde{g}_n(X_i))\), and \(-(f_i(X_i) - \tilde{g}_n(X_i)) < \epsilon_i\). This yields the following relation:

\[
\text{III + IV}
\]

\[
= \frac{1}{n} \sum_{X \in S_n} |\epsilon_i| - \frac{1}{n} \sum_{X \in S_n} |\epsilon_i + f_i(X_i) - \tilde{g}_n(X_i)|
\]

\[
= -\frac{2}{n} \sum_{X \in S_n} (f_i(X_i) - \tilde{g}_n(X_i) + \epsilon_i)
\]

\[
\cdot I(0 \leq f_i(X_i) - \tilde{g}_n(X_i), -(f_i(X_i) - \tilde{g}_n(X_i)) < \epsilon_i \leq 0)
\]

\[
- \frac{2}{n} \sum_{X \in S_n} (f_i(X_i) - \tilde{g}_n(X_i))
\]

\[
\cdot \frac{1}{2} I(0 \leq f_i(X_i) - \tilde{g}_n(X_i)) - I(0 \leq f_i(X_i) - \tilde{g}_n(X_i), \epsilon_i \leq 0))
\]

\[
+ \frac{2}{n} \sum_{X \in S_n} (f_i(X_i) - \tilde{g}_n(X_i) + \epsilon_i)
\]

\[
\cdot I(-\epsilon < f_i(X_i) - \tilde{g}_n(X_i) < 0, 0 < \epsilon_i \leq -(f_i(X_i) - \tilde{g}_n(X_i)))
\]

\[
+ \frac{2}{n} \sum_{X \in S_n} (f_i(X_i) - \tilde{g}_n(X_i))
\]

\[
\cdot ((1/2) I(\epsilon_i > 0) \leq f_i(X_i) - \tilde{g}_n(X_i)) - I(-\epsilon < f_i(X_i) - \tilde{g}_n(X_i) < 0, \epsilon_i > 0)
\]

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

\[
\text{III + IV}
\]

\[
\leq -\frac{2}{n} \sum_{X \in S_n} (f_i(X_i) - \tilde{g}_n(X_i))
\]

\[
\cdot ((1/2) I(0 \leq f_i(X_i) - \tilde{g}_n(X_i)) - I(0 \leq f_i(X_i) - \tilde{g}_n(X_i), \epsilon_i \leq 0))
\]

\[
+ \frac{2}{n} \sum_{X \in S_n} (f_i(X_i) - \tilde{g}_n(X_i))
\]

\[
\cdot ((1/2) I(\epsilon_i > 0) \leq f_i(X_i) - \tilde{g}_n(X_i)) - I(-\epsilon < f_i(X_i) - \tilde{g}_n(X_i) < 0, \epsilon_i > 0))
\]

\[
= -\frac{2}{n} \sum_{X \in S_n} (f_i(X_i) - \tilde{g}_n(X_i) + \epsilon)
\]

\[
\cdot ((1/2) I(0 \leq f_i(X_i) - \tilde{g}_n(X_i)) - I(0 \leq f_i(X_i) - \tilde{g}_n(X_i), \epsilon_i \leq 0))
\]

\[
+ \frac{2}{n} \sum_{X \in S_n} \epsilon
\]
From (20) and (21), we obtain

\begin{align*}
\cdot (1/2) & I(0 \leq f_i(X_i) - \tilde{g}_n(X_i)) - I(0 \leq f_i(X_i) - \tilde{g}_n(X_i), \epsilon_i \leq 0)) \\
+ & \frac{2}{n} \sum_{X_i \in P_n} (f_i(X_i) - \tilde{g}_n(X_i) + \epsilon) \\
\cdot ((1/2) & I(-\epsilon < f_i(X_i) - \tilde{g}_n(X_i) < 0) - I(-\epsilon < f_i(X_i) - \tilde{g}_n(X_i) < 0, \epsilon_i > 0)) \\
- & \frac{2}{n} \sum_{X_i \in P_n} \epsilon \\
\cdot ((1/2) & I(-\epsilon < f_i(X_i) - \tilde{g}_n(X_i) < 0) - I(-\epsilon < f_i(X_i) - \tilde{g}_n(X_i) < 0, \epsilon_i > 0)) \\
= & \frac{2}{n} \sum_{X_i \in P_n} (f_i(X_i) - \tilde{g}_n(X_i) + \epsilon) \\
\cdot ((1/2) & I(0 \leq f_i(X_i) - \tilde{g}_n(X_i)) - I(0 \leq f_i(X_i) - \tilde{g}_n(X_i), \epsilon_i \leq 0)) \\
- & \frac{2}{n} \sum_{X_i \in P_n} \epsilon \\
\cdot ((1/2) & I(-\epsilon < f_i(X_i) - \tilde{g}_n(X_i) < 0) - I(-\epsilon < f_i(X_i) - \tilde{g}_n(X_i) < 0, \epsilon_i > 0)).
\end{align*}

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

III + IV

\begin{align*}
\leq & \frac{2}{n} \sum_{X_i \in P_n} (f_i(X_i) - \tilde{g}_n(X_i) + \epsilon) \\
\cdot ((1/2) & I(-\epsilon < f_i(X_i) - \tilde{g}_n(X_i)) - I(-\epsilon < f_i(X_i) - \tilde{g}_n(X_i), \epsilon_i \leq 0)) \\
- & \frac{2}{n} \sum_{X_i \in P_n} \epsilon \\
\cdot ((1/2) & I(-\epsilon < f_i(X_i) - \tilde{g}_n(X_i)) - I(-\epsilon < f_i(X_i) - \tilde{g}_n(X_i), \epsilon_i > 0)) \\
= & -\frac{2}{n} \sum_{i=1}^{n} (f_i(X_i) - \tilde{g}_n(X_i) + \epsilon)^+ (1/2 - I(\epsilon_i \leq 0)) - \frac{2}{n} \sum_{X_i \in P_n} \epsilon (1/2 - I(\epsilon_i > 0)).
\end{align*}

From (20) and (21), we obtain

I + II + III + IV

\begin{align*}
\leq & -\frac{1}{n} \sum_{X_i \in P_n} \epsilon I(0 < \epsilon_i < \epsilon/2) - \frac{2}{n} \sum_{X_i \in P_n} \epsilon (1/2 - I(\epsilon_i > 0)) \\
- & \frac{2}{n} \sum_{i=1}^{n} (f_i(X_i) - \tilde{g}_n(X_i) + \epsilon)^+ (1/2 - I(\epsilon_i \leq 0)) \\
- & \frac{2}{n} \sum_{X_i \in P_n} \epsilon (1/2 - I(\epsilon_i > 0)) \\
= & -\frac{1}{n} \sum_{i=1}^{n} \epsilon I(X_i \in P_n, 0 < \epsilon_i < \epsilon/2) - \frac{2}{n} \sum_{i=1}^{n} \epsilon (1/2 - I(\epsilon_i > 0)) \\
- & \frac{2}{n} \sum_{i=1}^{n} (f_i(X_i) - \tilde{g}_n(X_i) + \epsilon)^+ (1/2 - I(\epsilon_i \leq 0)).
\end{align*}
By (17),
\[
-\frac{1}{n} \sum_{i=1}^{n} \epsilon I(X_i \in P_n, 0 < \epsilon_i < \epsilon/2) \leq -\frac{1}{n} \sum_{i=1}^{n} \epsilon I(X_i \in B(\epsilon, \delta), 0 < \epsilon_i < \epsilon/2)
\]
for infinitely many \(n\) and because \(\omega \in A\),
\[
-\frac{1}{n} \sum_{i=1}^{n} \epsilon I(X_i \in P_n, 0 < \epsilon_i < \epsilon/2) \leq -\epsilon \eta /2 \tag{23}
\]
for infinitely many \(n\). However, because \(\omega \in B\),
\[
-\frac{2}{n} \sum_{i=1}^{n} \epsilon (1/2 - I(\epsilon_i > 0)) - \frac{2}{n} \sum_{i=1}^{n} (f_i(X_i) - \tilde{g}_n(X_i) + \epsilon)^+ (1/2 - I(\epsilon_i \leq 0)) \leq \epsilon \eta /4 \tag{24}
\]
for \(n\) sufficiently large. Combination of (18), (22), (23), and (24) gives \(0 \leq \varphi(k_\omega) - \varphi(\tilde{g}_n) = I + \Pi + III + IV \leq -\epsilon \eta /4\) 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.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 \(f_i\) is differentiable on \([c, 1 - c]^d\). Take \(c_0 < c\) and let
\[
A = \left\{ \sup_{x \in [c_0, 1 - c]^d, \varphi \in \mathcal{S}_n} |\tilde{g}_n(x) - f_i(x)| \to 0 \text{ as } n \to \infty \right\},
\]
then \(\mathbb{P}(A) = 1\) by Theorem 1. We will show that \(\mathbb{P}(B) = 1\), where
\[
B = \left\{ \sup_{x \in [c_0, 1 - c]^d, \varphi \in \partial \mathcal{S}_n} ||x - \nabla f_i(x)|| \to 0 \text{ as } n \to \infty \right\},
\]
by proving that \(A \cap B^c = 0\). Suppose, on the contrary, that \(\omega \in A \cap B^c\) exists. For such an \(\omega\), 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
\[
||\xi_n - \nabla f_i(x_n)|| > \epsilon
\]
for infinitely many \(n\). Furthermore, there exists an index \(i \in \{1, \ldots, d\}\) such that
\[
|e_i^T \xi_n - e_i^T \nabla f_i(x_n)| > \epsilon /d \tag{25}
\]
for infinitely many \(n\), where \(e_i\) is the \(i\)th unit vector. Equation (25) implies that either
\[
e_i^T \xi_n > e_i^T \nabla f_i(x_n) + \epsilon /d \tag{26}
\]
or
\[
e_i^T \xi_n < e_i^T \nabla f_i(x_n) - \epsilon /d \tag{27}
\]
holds. We first consider the case where (26) holds. Since \([c, 1 - c]^d\) is compact, there exists a subsequence \((x_{n_k} : 1 \leq k)\) 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
\[
e_i^T \xi_n \leq (\tilde{g}_n(x_0 + \lambda e_i) - \tilde{g}_n(x_0)) / \lambda \tag{28}
\]
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 \(f_i(x_0 + \lambda e_i)\) and \(\tilde{g}_n(x_0)\) tends to \(f_i(x_0)\) as \(n \to \infty\). By Theorem 25.5 on p. 246 of [7], \(\nabla f_i\) is continuous on \([c, 1 - c]^d\), and hence, \(\nabla f_i(x_n)\) tends to \(\nabla f_i(x_0)\). Therefore,
\[
e_i^T \nabla f_i(x_0) + \epsilon /d = \lim_{n \to \infty} e_i^T \nabla f_i(x_n) + \epsilon /d \leq \lim_{n \to \infty} e_i^T \xi_n \text{ by (26)} \leq \lim_{n \to \infty} (\tilde{g}_n(x_n + \lambda e_i) - \tilde{g}_n(x_n)) / \lambda \text{ by (28)} = (f_i(x_0 + \lambda e_i) - f_i(x_0)) / \lambda \tag{29}
\]
This is supposed to hold for every sufficiently small $\lambda > 0$. However

$$e_i^T \nabla f_*(x_0) = \lim_{\lambda \to 0} \frac{f_*(x_0 + \lambda e_i) - f_*(x_0)}{\lambda},$$

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

References


Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).