On convex least squares estimation when the truth is linear

Yining Chen and Jon A. Wellner

London School of Economics and Political Science and University of Washington
e-mail: y.chen101@lse.ac.uk; jaw@stat.washington.edu

Abstract: We prove that the convex least squares estimator (LSE) attains a $n^{-1/2}$ pointwise rate of convergence in any region where the truth is linear. In addition, the asymptotic distribution can be characterized by a modified invelope process. Analogous results hold when one uses the derivative of the convex LSE to perform derivative estimation. These asymptotic results facilitate a new consistent testing procedure on the linearity against a convex alternative. Moreover, we show that the convex LSE adapts to the optimal rate at the boundary points of the region where the truth is linear, up to a log-log factor. These conclusions are valid in the context of both density estimation and regression function estimation.

MSC 2010 subject classifications: Primary 62E20, 62G07, 62G08, 62G10, 62G20; secondary 60G15.
Keywords and phrases: Adaptive estimation, convexity, density estimation, least squares, regression function estimation, shape constraint.

Contents

1. Introduction

Shape-constrained estimation has received much attention recently. The attraction is the prospect of obtaining automatic nonparametric estimators with no smoothing parameters to choose. Convexity is among the popular shape constraints that are of both mathematical and practical interest. Groeneboom, Jongbloed and Wellner (2001b) show that under the convexity constraint, the least squares estimator (LSE) can be used to estimate both a density and a regression function. For density estimation, they showed that the LSE $\hat{f}_n$ of the true convex density f_0 converges pointwise at a $n^{-2/5}$ rate under certain assumptions. The corresponding asymptotic distribution can be characterized via a so-called "invelope" function investigated by Groeneboom, Jongbloed and Wellner (2001a). In the regression setting, similar results hold for the LSE $\hat{r}_n$ of the true regression function r_0.

However, in the development of their pointwise asymptotic theory, it is required that f_0 (or r_0) has positive second derivative in a neighborhood of the point to be estimated. This assumption excludes certain convex functions that may be of practical value. Two further scenarios of interest are given below:

Scenario 1. At the point x_0, the k-th derivative $f_0^{(k)}(x_0) = 0$ (or $r_0^{(k)}(x_0) = 0$) for $k = 2, 3, \ldots, 2s - 1$ and $f_0^{(2s)}(x_0) > 0$ (or $r_0^{(2s)}(x_0) > 0$), where s is an integer greater than one;

Scenario 2. There exists some region $[a, b]$ on which f_0 (or r_0) is linear.

Scenario 1 can be handled using techniques developed in Balabdaoui, Rufibach and Wellner (2009). The aim of this manuscript is to provide theory in the setting of Scenario 2.

We prove that for estimation of a convex density when Scenario 2 holds, at any fixed point $x_0 \in (a, b)$, the LSE $\hat{f}_n(x_0)$ converges pointwise to $f_0(x_0)$ at a $n^{-1/2}$ rate. Its (left or right) derivative $\hat{f}_n'(x_0)$ converges to $f_0'(x_0)$ at the same rate. The corresponding asymptotic distributions are characterized using a modified invelope process. More generally, we show that the processes $\left\{\sqrt{n}(\hat{f}_n(x) - f_0(x)) : x \in [a+\delta, b-\delta]\right\}$ and $\left\{\sqrt{n}(\hat{f}_n'(x) - f_0'(x)) : x \in [a+\delta, b-\delta]\right\}$ both converge weakly for any $\delta > 0$. We remark that unlike the case of Groeneboom, Jongbloed and Wellner (2001b), these processes are not local, because intuitively speaking, the linear behavior of the true function is non-local, which requires linearity on an interval with positive length. In addition, there does not exist a "universal" distribution of $\hat{f}_n$ on (a, b), i.e. the pointwise limit distributions at different points are in general different. In addition, we show that the derived asymptotic processes can be used in a more practical setting where one would like to perform tests on the linearity against a convex alternative. Moreover, we study the adaptation of the LSE $\hat{f}_n$ at the boundary points of the linear region (e.g. a and b). Note that the difficulty level of estimating $f_0(a)$ and $f_0(b)$ depends on the behavior of f_0 outside $[a, b]$. Nevertheless, we show that $\hat{f}_n(a)$ (or $\hat{f}_n(b)$) converges to $f_0(a)$ (or $f_0(b)$) at the minimax optimal rate up to a negligible factor of $\sqrt{\log \log n}$. Last but not least, we establish the analogous rate and asymptotic distribution results for the LSE $\hat{r}_n$ in the regression setting.

Our study yields a better understanding of the adaptation of the LSE in terms of pointwise convergence under the convexity constraint. It is also one of the first attempts to quantify the behavior of the convex LSE at non-smooth points. When the truth is linear, the minimax optimal $n^{-1/2}$ pointwise rate is indeed achieved by the LSE on (a, b). The optimal rate at the boundary points a and b is also achievable by the LSE up to a log-log factor. These results can also be applied to the case where the true function consists of multiple linear components. Furthermore, our results can be viewed as an intermediate stage for the development of theory under misspecification. Note that linearity is regarded as the boundary case of convexity: if a function is non-convex, then its projection to the class of convex functions will have linear components. We conjecture that the LSE in these misspecified regions converges at an $n^{-1/2}$ rate, with the asymptotic distribution characterized by a more restricted version of the invelope process. More broadly, we expect that this type of behavior will be seen in situations of other shape restrictions, such as log-concavity for $d = 1$ (Balabdaoui, Rufibach and Wellner, 2009) and k-monotonicity (Balabdaoui and Wellner, 2007).

The LSE of a convex density function was first studied by Groeneboom, Jongbloed and Wellner (2001b), where its consistency and some asymptotic distribution theory were provided. On the other hand, the idea of using the LSE for convex regression function estimation dates back to Hildreth (1954). Its consistency was proved by Hanson and Pledger (1976), with some rate results given in Mammen (1991). In this manuscript, for the sake of mathematical convenience, we shall focus on the non-discrete version discussed by Balabdaoui and Rufibach (2008). See Groeneboom, Jongbloed and Wellner (2008) for the computational aspects of all the above-mentioned LSEs.

There are studies similar to ours regarding other shape restrictions. See Remark 2.2 of Groeneboom (1985) and Carolan and Dykstra (1999) in the context of decreasing density function estimation (a.k.a. Grenander estimator) when the truth is flat and Balabdaoui (2014) with regard to discrete log-concave distribution estimation when the true distribution is geometric. In addition, Groeneboom and Pyke (1983) studied the Grenander estimator's global behavior in the L^2 norm under the uniform distribution and gave a connection to statistical problems involving combination of p-values and two-sample rank statistics, while Cator (2011) studied the adaptivity of the LSE in monotone regression (or density) function estimation. For estimation under misspecification of various shape constraints, we point readers to Jankowski (2014), Cule and Samworth (2010), Dümbgen, Samworth and Schuhmacher (2011), Chen and Samworth (2013), and Balabdaoui *et.al.* (2013). More recent developments on global rates of the shape-constrained methods can be found in Guntuboyina and Sen (2015), Doss and Wellner (2016), and Kim and Samworth (2014). See also Meyer (2013) and Chen and Samworth (2016) where an additive structure is imposed in shape-constrained estimation in the multidimensional setting.

The rest of the paper is organized as follows: in Section 2, we study the behavior of the LSE for density estimation. In particular, for notational convenience, we first focus on a special case where the true density function f_0 is taken to be

triangular. The convergence rate and asymptotic distribution are given in Section 2.1.1 and Section 2.1.2. Section 2.1.3 demonstrates the practical use of our asymptotic results, where a new consistent testing procedure on the linearity against a convex alternative is proposed. More general cases are handled later in Section 2.2 based on the ideas illustrated in Section 2.1. Section 2.3 discusses the adaptation of the LSE at the boundary points. Analogous results with regard to regression function estimation are presented in Section 3. Some proofs, mainly on the existence and uniqueness of a limit process and the adaptation of the LSE, are deferred to the appendices.

2. Estimation of a density function

Given n independent and identically distributed (IID) observations from a density function $f_0 : [0, \infty) \to [0, \infty)$. Let F_0 be the corresponding distribution function (DF). In this section, we denote the convex cone of all non-negative continuous convex and integrable functions on $[0, \infty)$ by $\mathcal{K}$. The LSE of f_0 is given by

$$\hat{f}_n = \operatorname*{argmin}_{g \in \mathcal{K}} \Big(\frac{1}{2} \int_0^\infty g(t)^2 dt - \int_0^\infty g(t) d\mathbb{F}_n(t) \Big),$$

where $\mathbb{F}_n$ is the empirical distribution function of the observations. Furthermore, we denote the DF of $\hat{f}_n$ by $\hat{F}_n$.

Throughout the manuscript, without specifying otherwise, the derivative of a convex function can be interpreted as either its left derivative or its right derivative.

2.1. A special case

To motivate the discussion, as well as for notational convenience, we take $f_0(t) = 2(1 - t)\mathbf{1}_{[0,1]}(t)$ in Section 2.1.1 and Section 2.1.2. Extensions to other convex density functions are presented in Section 2.2.

Consistency of $\hat{f}_n$ over $(0, \infty)$ in this setting can be found in Groeneboom, Jongbloed and Wellner (2001b). Balabdaoui (2007) has shown how $\hat{f}_n$ can be used to provide consistent estimators of $f_0(0)$ and $f_0'(0)$.

Here we concentrate on the LSE's rate of convergence and asymptotic distribution.

2.1.1. Rate of convergence

The following proposition shows that the convergence rate at any interior point where f_0 is linear is $n^{-1/2}$, thus achieving the optimal rate given in Example 1 of Cai and Low (2015).

Theorem 2.1 (Pointwise rate of convergence). *Suppose $f_0(t) = 2(1-t)\mathbf{1}_{[0,1]}(t)$. Then for any fixed $x_0 \in (0, 1)$,*

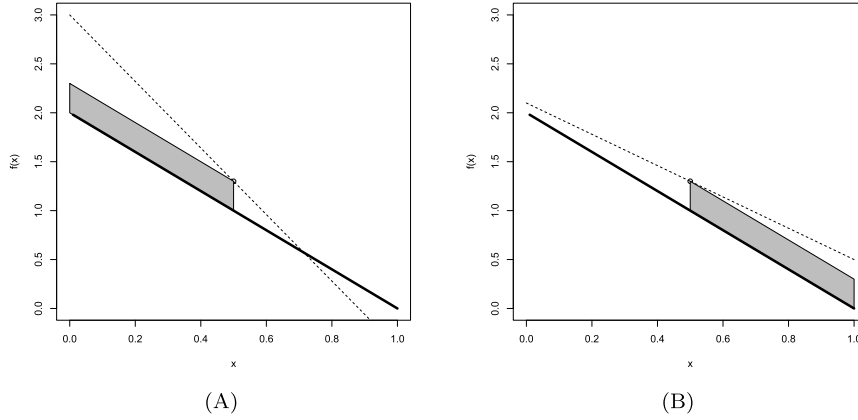$$|\hat{f}_n(x_0) - f_0(x_0)| = O_p(n^{-1/2}).$$

(A) (B)

FIG 1. *The case of $c > 0$: (A) illustrates the situation of $k \leq -2$, where $\hat{F}_n(x_0) - F_0(x_0) \geq cx_0$; (B) illustrates the situation of $k > -2$, where $\hat{F}_n(x_0) - F_0(x_0) < -c(1-x_0)$. In both scenarios, the dashed curve represents the supporting hyperplane of the LSE, while the thick solid line represents the true density f_0.*

PROOF OF THEOREM 2.1
A key ingredient of this proof is the version of Marshall's lemma in this setting (Dümbgen, Rufibach and Wellner, 2007, Theorem 1), which states that

$$\|\hat{F}_n - F_0\|_\infty \leq 2\|\mathbb{F}_n - F_0\|_\infty, \tag{2.1}$$

where $\|\cdot\|_\infty$ is the uniform norm.

Let $c = \hat{f}_n(x_0) - f_0(x_0)$. Two cases are considered here: (a) $c > 0$ and (b) $c < 0$.

In the first case, because $\hat{f}_n$ is convex, one can find a supporting hyperplane of $\hat{f}_n$ passing through $(x_0, f_0(x_0) + c)$ such that

$$\hat{f}_n(y) \geq k(y - x_0) + f_0(x_0) + c, \quad \text{for any } y \in [0, \infty),$$

where k is a negative slope. If $k \leq -2$, then $\hat{F}_n(x_0) - F_0(x_0) \geq cx_0 > 0$. Otherwise, $\hat{F}_n(x_0) - F_0(x_0) < -c(1 - x_0) < 0$. Figure 1 explains the above inequalities graphically. Consequently, $|\hat{F}_n(x_0) - F_0(x_0)| \geq c \min(x_0, 1 - x_0)$.

In the second case, if we can find $0 \leq t < x_0$ such that $\hat{f}_n$ and f_0 intersect at the point $(t, 2 - 2t)$, then

$$\hat{F}_n(t) - F_0(t) \geq \frac{1}{2}t\frac{|c|t}{x_0 - t} \geq 0;$$

$$\left(\hat{F}_n(x_0) - \hat{F}_n(t)\right) - \left(F_0(x_0) - F_0(t)\right) \leq -\frac{1}{2}|c|(x_0 - t) < 0.$$

Otherwise, by taking $t = 0$, we can still verify that the above two inequalities hold true. Figure 2 illustrates these findings graphically.
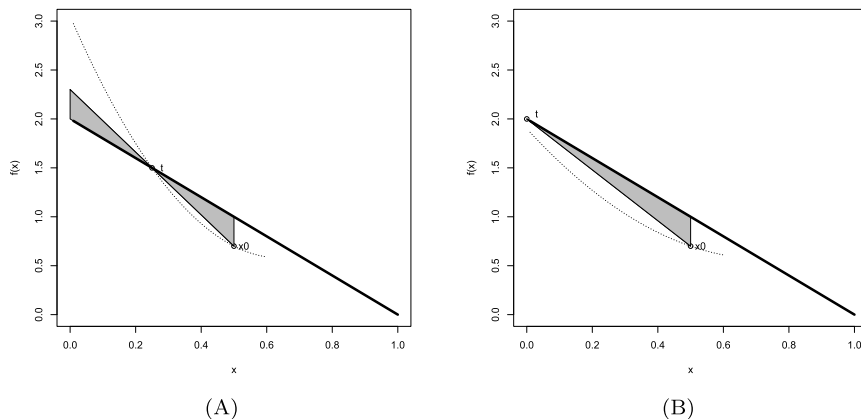
FIG 2. *The case of $c < 0$: (A) illustrates the inequalities when $\hat{f}_n$ and f_0 has an intersection on $[0, x_0)$; the situation where $\hat{f}_n$ and f_0 has no intersection on $[0, x_0)$ is represented in (B). In both scenarios, the dotted curve represents the LSE and the thick solid line represents the true density f_0.*

Therefore,

$$2\|\hat{F}_n - F_0\|_\infty \geq \max\left\{ |\hat{F}_n(t) - F_0(t)|, \left| \left(\hat{F}_n(x_0) - F_0(x_0)\right) - \left(\hat{F}_n(t) - F_0(t)\right)\right| \right\}$$
$$\geq \frac{|c|}{2} \max\left(\frac{t^2}{x_0 - t}, x_0 - t \right) \geq \frac{|c|x_0}{4},$$

where the last inequality uses the fact that for any $t \in [0, x_0)$, $t^2/(x_0 - t)$ is an increasing function of t while $x_0 - t$ is a decreasing function of t.

By (2.1), we have that

$$O_p(n^{-1/2}) = 4\|\mathbb{F}_n - F_0\|_\infty \geq 2\|\hat{F}_n - F_0\|_\infty \geq \min\left(2x_0, 2 - 2x_0, \frac{x_0}{4} \right) |c|.$$

It then follows that $c = O_p(n^{-1/2})$, as desired. $\square$

Corollary 2.2 (Uniform rate of convergence). *For any $0 < \delta \leq 1/2$,*

$$\sup_{x \in [\delta, 1-\delta]} |\hat{f}_n(x) - f_0(x)| = O_p(n^{-1/2}).$$

PROOF OF COROLLARY 2.2
A closer look at the proof of Theorem 2.1 reveals that we have

$$O_p(n^{-1/2}) = 4\|\mathbb{F}_n - F_0\|_\infty \geq \min\left(2x, 2 - 2x, \frac{x}{4} \right) |\hat{f}_n(x) - f_0(x)|$$

simultaneously for every $x \in [\delta, 1 - \delta]$. Therefore,

$$O_p(n^{-1/2}) = \inf_{x \in [\delta, 1-\delta]} \min\left(2x, 2 - 2x, \frac{x}{4} \right) \sup_{x \in [\delta, 1-\delta]} |\hat{f}_n(x) - f_0(x)|.$$

Consequently, $\sup_{x \in [\delta, 1-\delta]} |\hat{f}_n(x) - f_0(x)|$ is $O_p(n^{-1/2})$. $\square$

Let $\hat{f}_n^-$ and $\hat{f}_n^+$ denote respectively the left and right derivatives of $\hat{f}_n$. The same convergence rate also applies to these derivative estimators.

Corollary 2.3 (Uniform rate of convergence: derivatives)**.** *For any* $0 < \delta \leq 1/2$,

$$\sup_{x \in [\delta, 1-\delta]} \max\left(|\hat{f}_n^-(x) - f_0'(x)|, |\hat{f}_n^+(x) - f_0'(x)| \right) = O_p(n^{-1/2}).$$

PROOF OF COROLLARY 2.3
By the convexity of $\hat{f}_n$,

$$\sup_{x \in [\delta, 1-\delta]} \max\left(|\hat{f}_n^-(x) - f_0'(x)|, |\hat{f}_n^+(x) - f_0'(x)| \right)$$

$$\leq \max\left(|\hat{f}_n^-(\delta) - f_0'(\delta)|, |\hat{f}^+(1-\delta) - f_0'(1-\delta)| \right)$$

$$\leq \frac{2}{\delta} \max\left(\left|\hat{f}_n(\delta) - \hat{f}_n(\delta/2) - f_0(\delta) + f_0(\delta/2)\right|, \right.$$

$$\left. \left|\hat{f}_n(1-\delta/2) - \hat{f}_n(1-\delta) - f_0(1-\delta/2) + f_0(1-\delta)\right| \right)$$

$$= O_p(n^{-1/2}),$$

where the final equation follows from Corollary 2.2. □

2.1.2. Asymptotic distribution

To study the asymptotic distribution of $\hat{f}_n$, we start by characterizing the limit distribution.

Theorem 2.4 (Characterization of the limit process)**.** *Let* $X(t) = \mathbb{U}(F_0(t))$ *and* $Y(t) = \int_0^t X(s)ds$ *for any* $t \geq 0$, *where* $\mathbb{U}$ *is a standard Brownian bridge process on* $[0,1]$. *Then almost surely (a.s.), there exists a uniquely defined random continuously differentiable function* H *on* $[0,1]$ *satisfying the following conditions:*

(1) $H(t) \geq Y(t)$ *for every* $t \in [0,1]$;
(2) H *has convex second derivative on* $(0,1)$;
(3) $H(0) = Y(0)$ *and* $H'(0) = X(0)$;
(4) $H(1) = Y(1)$ *and* $H'(1) = X(1)$;
(5) $\int_0^1 \left(H(t) - Y(t) \right) dH^{(3)}(t) = 0$.

The above claim also holds if we instead consider a standard Brownian motion $\tilde{X}(t)$ *on* $[0,1]$ *and replace* $(X, Y, H)^T$ *by* $(\tilde{X}, \tilde{Y}, \tilde{H})^T$, *where* $\tilde{Y}(t) = \int_0^t \tilde{X}(s)ds$, *and where* $\tilde{H}$ *is different from* H.

Figure 3 shows a typical realization of $X(t)$, $Y(t)$, $H'(t)$ and $H(t)$. A detailed construction of the above limit invelope process can be found in Appendix I. Note that our process is defined on a compact interval, so is technically different from the process presented in Groeneboom, Jongbloed and Wellner (2001a) (which is defined on the whole real line). As a result, extra conditions regarding
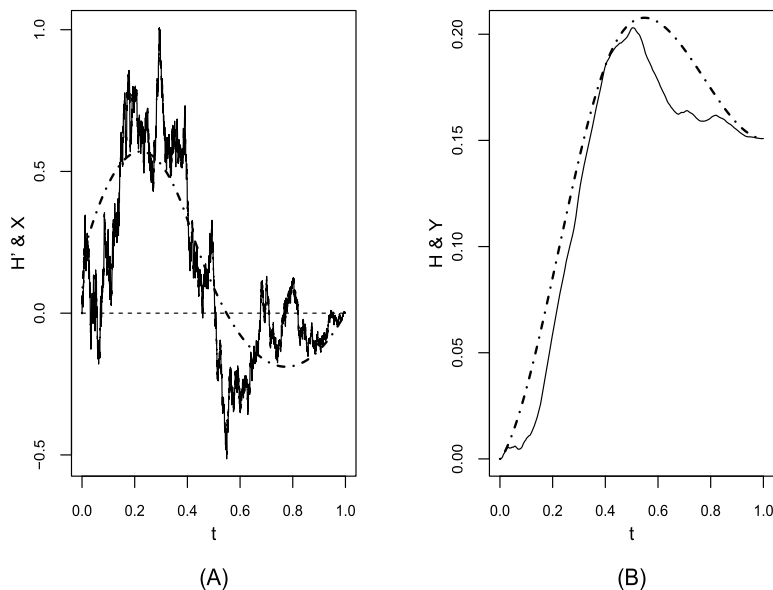
Fig 3. *A typical realization of $X(t)$, $Y(t)$, $H'(t)$ and $H(t)$ in Theorem 2.4 with $X(t) = \mathbb{U}(F_0(t))$ and F_0 the distribution function with triangular density $f_0(t) = 2(1-t)\mathbf{1}_{[0,1]}(t)$: in (A), $X(t)$ and $H'(t)$ are plotted in solid and dash-dotted curves respectively; in (B), the corresponding $Y(t)$ is plotted in solid curve, while the invelope process $H(t)$ is illustrated in dash-dotted curve.*

the behavior of H (and H') at the boundary points are imposed here to ensure its uniqueness.

Other characterization of the limit process is also possible. A slight variant is given below, which proof can also be found in Appendix I.

Corollary 2.5 (Different characterization). *Conditions (3) and (4) in the statement of Theorem 2.4 can be replaced respectively by*

(3') $\lim_{t\to 0^+} H^{(2)}(t) = \infty$, and
(4') $\lim_{t\to 1^-} H^{(2)}(t) = \infty$.

Now we are in the position to state our main result of this section.

Theorem 2.6 (Asymptotic distribution). *Suppose $f_0(t) = 2(1-t)\mathbf{1}_{[0,1]}(t)$. Then for any $\delta > 0$, the process*

$$\sqrt{n}\left(\begin{array}{c} \hat{f}_n(x) - f_0(x) \\ \hat{f}'_n(x) - f'_0(x) \end{array} \right) \Rightarrow \left(\begin{array}{c} H^{(2)}(x) \\ H^{(3)}(x) \end{array} \right) \ in \ \mathcal{C}[\delta, 1-\delta] \times \mathcal{D}[\delta, 1-\delta],$$

where $\mathcal{C}$ is the space of continuous functions equipped with the uniform norm, $\mathcal{D}$ is the Skorokhod space, and H is the invelope process defined in Theorem 2.4.

In particular, for any $x_0 \in (0,1)$,

$$\sqrt{n}\left(\hat{f}_n(x_0) - f_0(x_0), \hat{f}'_n(x_0) - f'_0(x_0) \right)^T \xrightarrow{d} \left(H^{(2)}(x_0), H^{(3)}(x_0) \right)^T. \qquad (2.2)$$

PROOF OF THEOREM 2.6
Before proceeding to the proof, we first define the following processes on $[0, \infty)$:

$$X_n(t) = n^{1/2}\Big(\mathbb{F}_n(t) - F_0(t)\Big);$$

$$Y_n(t) = n^{1/2}\int_0^t \Big(\mathbb{F}_n(s) - F_0(s)\Big)ds;$$

$$\hat{H}_n(t) = n^{1/2}\int_0^t \left\{\int_0^s \Big(\hat{f}_n(v) - f_0(v)\Big)dv\right\}ds. \tag{2.3}$$

Furthermore, define the set of "knots" of a convex function f on $(0,1)$ as

$$\mathcal{S}(f) = \left\{t \in (0,1) : f(t) < \frac{1}{2}\Big(f(t-\delta) + f(t+\delta)\Big), \ \forall \delta \in \Big(0, \min(1-t, t)\Big)\right\}. \tag{2.4}$$

We remark that the above definition of knots can be easily extended to convex functions with a different domain.

By Lemma 2.2 of Groeneboom, Jongbloed and Wellner (2001b), we have that $\hat{H}_n(t) \geq Y_n(t)$ for $t \in [0,1]$, with equality if $t \in \mathcal{S}(\hat{f}_n)$. In addition, $\int_0^1 (\hat{H}_n(t) - Y_n(t))d\hat{H}_n^{(3)}(t) = 0$.

Now define the space E_m of vector-valued functions ($m \geq 3$) as

$$E_m = \mathcal{C}[0,1] \times \mathcal{C}[0,1] \times \mathcal{C}[1/m, 1 - 1/m] \times \mathcal{D}[1/m, 1 - 1/m] \times \mathcal{C}[0,1] \times \mathcal{D}[0,1].$$

and endow E_m with the product topology induced by the uniform norm on $\mathcal{C}$ (i.e. the space of continuous functions) and Skorokhod metric on $\mathcal{D}$ (i.e. the Skorokhod space). Let E_m be supported by the stochastic process

$$Z_n = \left(\hat{H}_n, \hat{H}'_n, \hat{H}_n^{(2)}, \hat{H}_n^{(3)}, Y_n, X_n\right)^T.$$

Corollary 2.3 entails the tightness of $\hat{H}_n^{(3)}$ in $\mathcal{D}[1/m, 1 - 1/m]$, while Corollary 2.2 entails the tightness of $\hat{H}_n^{(2)}$ in $\mathcal{C}[1/m, 1 - 1/m]$. In addition, both $\hat{H}_n$ and $\hat{H}'_n$ are tight in $\mathcal{C}[0,1]$ (via an easy application of Marshall's lemma). Since X_n converges to a Brownian bridge, it is tight in $\mathcal{D}[0,1]$ as well. Finally, the same is also true for Y_n.

Note that E_m is separable. For any subsequence of Z_n, by Prohorov's theorem, we can construct a further subsequence Z_{n_j} such that $\{Z_{n_j}\}_j$ converges weakly in E_m to some

$$Z_0 = \left(H_0, H'_0, H_0^{(2)}, H_0^{(3)}, Y, X\right)^T.$$

In addition, it follows from a diagonalization argument that a further extraction can make $\{Z_{n_j}\}_j$ converge weakly to Z_0 in every E_m. Using Skorokhod's representation theorem, we can assume without loss of generality that for almost every ω in the sample space, $Z_{n_j}(\omega) \to Z_0(\omega)$. Moreover, since X has a continuous

path a.s., the convergence of X_{n_j} can be strengthened to $\|X_{n_j}(\omega) - X(\omega)\|_\infty \to 0$, where $\|\cdot\|_\infty$ is the uniform norm. In the rest of the proof, ω is suppressed for notational convenience, so depending on the context, Z_{n_j} (or Z_0) can either mean a random variable or a particular realization.

In the following, we shall verify that H_0 satisfies all the conditions listed in the statement of Theorem 2.4.

(1) The fulfillment of this condition follows from the fact that

$$\inf_{t \in [0,1]} \left(\hat{H}_n(t) - Y_n(t) \right) \geq 0 \text{ for all } n \in \mathbb{N}.$$

(2) Since f_0 is linear on $[0, 1]$, $\hat{H}_n^{(2)}$ is convex on $[0, 1]$ for every $n \in \mathbb{N}$. The pointwise limit of convex functions is still convex, so $H_0^{(2)}$ is convex on $[1/m, 1 - 1/m]$. The condition is then satisfied by letting $m \to \infty$.

(3) This condition always holds in view of our construction of $\hat{H}_n$ in (2.3).

(4) We consider two cases: (a) if $H_0^{(2)}(1 - 1/m) \to \infty$ as $m \to \infty$, then the conditions are satisfied in view of Corollary 2.5; (b) otherwise, it must be the case that $H_0^{(2)}(1^-)$ is bounded from above. Note that $H_0^{(2)}(1^-)$ is also bounded from below a.s., which can be proved by using Corollary 2.3 and the fact that $\hat{H}_n^{(2)}$ is convex. Denote by τ_{n_j} the knot of $\hat{f}_{n_j}$ closest to 1. Then by Lemma 2.2 of Groeneboom, Jongbloed and Wellner (2001b), $\hat{H}_{n_j}(\tau_{n_j}) = Y_{n_j}(\tau_{n_j})$ and $\hat{H}'_{n_j}(\tau_{n_j}) = X_{n_j}(\tau_{n_j})$. Since $t = 1$ is a knot of f_0, consistency of $\hat{f}_{n_j}$ allows us to see that $\lim_{j \to \infty} \tau_{n_j} = 1$. Because $H_0^{(2)}(1^-)$ is finite and both $Y(t)$ and $X(t)$ are sample continuous processes, taking $\tau_{n_j} \to 1^-$ yields $H_0(1) = Y(1)$ and $H'_0(1) = X(1)$. Note that this argument remains valid even if $\tau_{n_j} > 1$, because in this scenario, $\hat{H}_{n_j}^{(2)}$ is linear and bounded on $[2 - \tau_{n_j}, \tau_{n_j}]$.

(5) It follows from $\int_{1/m}^{1-1/m} \left(\hat{H}_{n_j}(t) - Y_{n_j}(t) \right) d\hat{H}_{n_j}^{(3)}(t) = 0$ that

$$\int_{1/m}^{1-1/m} \left(H_0(t) - Y(t) \right) dH_0^{(3)}(t) = 0.$$

Since this holds for any m, one necessarily has that

$$\int_0^1 \left(H_0(t) - Y(t) \right) dH_0^{(3)}(t) = 0.$$

Consequently, in view of Theorem 2.4, the limit Z_0 is the same for any subsequences of Z_n in E_m. Fix any $m > 1/\delta$. It follows that the full sequence $\{Z_n\}_n$ converges weakly in E_m and has the limit $(H, H', H^{(2)}, H^{(3)}, Y, X)^T$. This, together with the fact that $H^{(3)}$ is continuous at any fixed $x_0 \in (0, 1)$ with probability one (which can be proved using Conditions (1) and (5) of H), yields (2.2). □

It can be inferred from Corollary 2.5 and Theorem 2.6 that both $\hat{f}_n(0)$ and $\hat{f}_n(1)$ do not converge to the truth at a $n^{-1/2}$ rate. In fact, Balabdaoui (2007)

proved that $\hat{f}_n(0)$ is an inconsistent estimator of $f_0(0)$. Nevertheless, the following proposition shows that $\hat{f}_n(0)$ is at most $O_p(1)$. For the case of the maximum likelihood estimator of a k monotone density, we refer readers to Gao and Wellner (2009) for a similar result.

Proposition 2.7 (Behavior at zero). $\hat{f}_n(0) = O_p(1)$.

PROOF OF PROPOSITION 2.7
Let τ_n be the left-most point in $\mathcal{S}(\hat{f}_n)$. Since $\hat{f}_n(0)$ is finite for every n, the linearity of $\hat{f}_n$ on $[0, \tau_n]$ means that

$$\hat{f}_n(t) \geq (1 - t/\tau_n)\hat{f}_n(0)$$

for every $t \geq 0$. By Corollary 2.1 of Groeneboom, Jongbloed and Wellner (2001b),

$$\mathbb{F}_n(\tau_n) = \int_0^{\tau_n} \hat{f}_n(t)dt \geq \int_0^{\tau_n} (1 - t/\tau_n)\hat{f}_n(0)dt = \tau_n \hat{f}_n(0)/2.$$

It follows from Theorem 9.1.2 of Shorack and Wellner (1986) that

$$\hat{f}_n(0) \leq 2\mathbb{F}_n(\tau_n)/\tau_n \leq 2 \sup_{t>0} \frac{\mathbb{F}_n(t)}{F_0(t)} \frac{F_0(t)}{t} \leq O_p(1). \qquad \square$$

2.1.3. Testing against a general convex density

The asymptotic results established in Section 2.1.2 can be used to test the linearity of a density function against a general convex alternative. To illustrate the main idea, in this section, we focus on the problem of testing $\mathcal{H}_0 : f_0(t) = 2(1-t)\mathbf{1}_{[0,1]}(t)$ against

$$\mathcal{H}_1 : f_0 \in \mathcal{K} \text{ and } f_0(t) \neq 2(1-t)\mathbf{1}_{[0,1]}(t) \text{ for some } t \in (0, \infty).$$

The test we propose is free of tuning parameters. Since the triangular distribution is frequently used in practice, and is closely related to the uniform distribution (e.g. the minimum of two independent $U[0,1]$ is triangular), our test could be a valuable addition to practitioners' toolkit. For other tests on the linearity in the regression setting, we point readers to Meyer (2003) and Sen and Meyer (2013).

Our test is based on the statistic

$$T_n := \sqrt{n} \sup_{t \in [0,\infty)} \big\{ 2(1-t)\mathbf{1}_{[0,1]}(t) - \hat{f}_n(t) \big\}.$$

The behavior of T_n under $\mathcal{H}_0$ is established below.

Proposition 2.8 (Behavior of T_n under $\mathcal{H}_0$). *Write* $T := -\inf_{t \in [0,1]} H^{(2)}(t)$, *where H is the invelope process defined in Theorem 2.4. Then $T_n \xrightarrow{d} T$. Moreover, $P(T \geq 0) = 1$.*

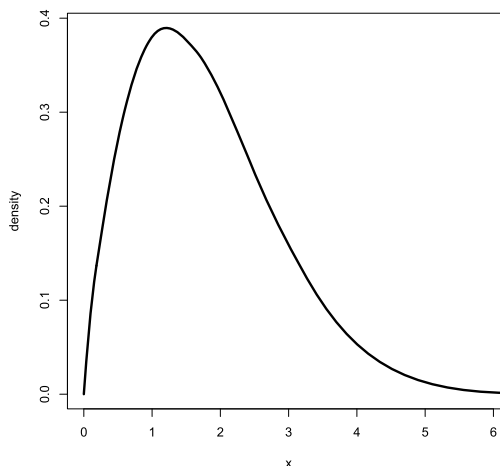| α	1%	2.5%	5%	10%	20%
t_α	4.76	4.21	3.75	3.23	2.63

FIG 4. *The estimated density function of T based on 10^5 simulations. Here we used the kernel density estimator with the boundary correction at zero.*

PROOF OF PROPOSITION 2.8
The first part follows directly from Theorem 2.6 and Corollary 2.5. For the second part, note that if $T < 0$, then $\inf_{t\in[0,1]} H^{(2)}(t) > 0$, which would imply $H'(1) - H'(0) > 0$. But this would violate the characterization of H (i.e. $H'(1) - H'(0) = X(1) - X(0) = 0$, since X is a Brownian bridge). Consequently, $P(T \geq 0) = 1$. □

The quantiles and the density function of T can be approximated numerically using Monte Carlo methods. They are given in Table 1 and Figure 4. Here we denote the upper α quantile of T by t_α. Our numerical evidence also suggests that the density function of T appears to be log-concave. For related work on log-concavity of Chernoff's distribution (i.e. a different random variable also associated with a Brownian motion), see Balabdaoui and Wellner (2014).

For a test of size $\alpha \in (0,1)$, we propose to reject $\mathcal{H}_0$ if $T_n > t_\alpha$. According to Proposition 2.8, this test is asymptotically of size α. In the following, we prove that our test is also consistent, i.e., its power goes to one as $n \to \infty$.

Theorem 2.9 (Consistency of the test). *Under $\mathcal{H}_1$, for any $\alpha > 0$, $P(T_n > t_\alpha) \to 1$ as $n \to \infty$.*

PROOF OF THEOREM 2.9
Suppose that $f_0(t) \neq 2(1-t)\mathbf{1}_{[0,1]}(t)$ for some $t \in (0,\infty)$. First, we show that there exists some $t^* \in (0,1)$ such that $f_0(t^*) < 2(1-t^*)$. Suppose the conclusion fails, then $f_0(t) \geq 2(1-t)\mathbf{1}_{[0,1]}(t)$ for all $t > 0$. Hence $\int_0^\infty f_0(t)dt \geq$

$\int_0^\infty 2(1-t)\mathbf{1}_{[0,1]}(t)dt = 1$ with strict inequality if $f_0(t) > 2(1-t)\mathbf{1}_{[0,1]}(t)$ for some $t > 0$. But then f_0 is not a density and we conclude that $f_0(t) = 2(1-t)\mathbf{1}_{[0,1]}(t)$ for every $t \in (0,\infty)$. This contradiction yields the conclusion.

Next, it follows from Theorem 3.1 of Groeneboom, Jongbloed and Wellner (2001b) that the LSE $\hat{f}_n$ is consistent at t^*, i.e., $|\hat{f}_n(t^*) - f_0(t^*)| \overset{a.s.}{\to} 0$. Therefore, almost surely,

$$T_n/\sqrt{n} > 2(1-t^*) - \hat{f}_n(t^*) = 2(1-t^*) - f_0(t^*) + f_0(t^*) - \hat{f}_n(t^*)$$
$$\to 2(1-t^*) - f_0(t^*) > 0.$$

It follows that $P(T_n > t_\alpha) \to 1$. □

2.2. More general settings

The aim of this subsection is to extend the conclusions presented in Section 2.1 to more general convex densities. We assume that f_0 is positive and linear on (a, b) for some $0 \le a < b$, where the open interval (a, b) is picked as the "largest" interval on which f_0 remains linear. More precisely, it means that there does not exist a bigger open interval (a', b') (i.e. $(a, b) \subset (a', b') \subseteq (0, \infty)$) on which f_0 is linear.

For the sake of notational convenience, we suppress the dependence of H_* on a, b and F_0 in the following two theorems. Their proofs are similar to those given in Section 2.1, so are omitted for brevity.

Theorem 2.10 (Characterization of the limit process). *Let* $X(t) = \mathbb{U}(F_0(t))$ *and* $Y(t) = \int_0^t X(s)ds$ *for any* $t > 0$. *Then a.s., there exists a uniquely defined random continuously differentiable function* H_* *on* $[a, b]$ *satisfying the following conditions:*

(1) $H_*(t) \ge Y(t)$ *for every* $t \in [a, b]$;
(2) H_* *has convex second derivative on* (a, b);
(3) $H_*(a) = Y(a)$ *and* $H'_*(a) = X(a)$;
(4) $H_*(b) = Y(b)$ *and* $H'_*(b) = X(b)$;
(5) $\int_a^b \left(H_*(t) - Y(t)\right) dH_*^{(3)}(t) = 0$.

Theorem 2.11 (Rate and asymptotic distribution). *For any* $0 < \delta \le (b-a)/2$,

$$\sup_{x \in [a+\delta, b-\delta]} \left(|\hat{f}_n(x) - f_0(x)|, \ |\hat{f}'_n(x) - f'_0(x)|\right) = O_p(n^{-1/2}).$$

Moreover,

$$\sqrt{n}\begin{pmatrix} \hat{f}_n(x) - f_0(x) \\ \hat{f}'_n(x) - f'_0(x) \end{pmatrix} \Rightarrow \begin{pmatrix} H_*^{(2)}(x) \\ H_*^{(3)}(x) \end{pmatrix} \ in \ \mathcal{C}[a + \delta, b - \delta] \times \mathcal{D}[a + \delta, b - \delta],$$

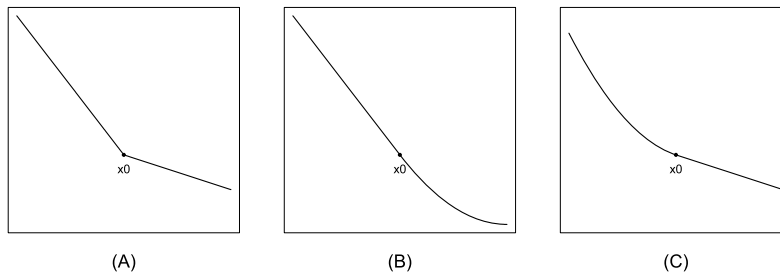*where* H_* *is the invelope process defined in Theorem 2.10.*

FIG 5. *Typical f_0 in different cases are illustrated, where we set $\alpha = 2$ in (B) and (C).*

2.3. Adaptation at the boundary points

In this subsection, we study the pointwise convergence rate of the convex LSE $\hat{f}_n$ at the boundary points of the region where f_0 is linear. Examples of such points include a and b given in Section 2.2.

To begin our discussion, we assume that $x_0 \in (0, \infty)$ is such a boundary point in the interior of the support (i.e. $f_0(x_0) > 0$). Here again f_0 is a convex (and decreasing) density function on $[0, \infty)$. Three cases are under investigation as below. These cases are illustrated in Figure 5.

(A) $f_0(t) = f_0(x_0) + K_1(t - x_0) + K_2(t - x_0)\mathbf{1}_{[x_0, \infty)}(t)$ for every t in a fixed (small) neighborhood of x_0, with $K_1 + K_2 < 0$ and $K_2 > 0$;

(B) $f_0(t) = f_0(x_0) + K_1(t - x_0) + K_2(t - x_0)^{\alpha}\mathbf{1}_{[x_0, \infty)}(t)$ for every t in a fixed (small) neighborhood of x_0, with $K_1 < 0$, $K_2 > 0$ and $\alpha > 1$;

(C) $f_0(t) = f_0(x_0) + K_1(t - x_0) + K_2(x_0 - t)^{\alpha}\mathbf{1}_{[0, x_0)}(t)$ for every t in a fixed (small) neighborhood of x_0, with $K_1 < 0$, $K_2 > 0$ and $\alpha > 1$.

Note that in all cases above, only the behavior of f_0 in a small neighborhood of x_0 is relevant. As pointed out in Example 2 of Cai and Low (2015), the minimax optimal convergence rate at x_0 is $n^{-1/3}$ in (A). Furthermore, in (B) and (C), Example 4 of Cai and Low (2015) suggests that the optimal rate at x_0 is $n^{-\alpha/(2\alpha+1)}$. In the following, we prove that the convex LSE automatically adapts to optimal rates, up to a factor of $\sqrt{\log \log n}$. Though almost negligible, the factor of $\sqrt{\log \log n}$ here indicates that there might be room for improvement. These results should be viewed as a first step for the investigation of the adaptation of the convex LSE. One could also compare our results with Cator (2011), where the adaptation of the LSE in the context of decreasing density function estimation was tackled.

Theorem 2.12 (Adaptation at the boundary points: I). *In the case of (A),*

$$|\hat{f}_n(x_0) - f_0(x_0)| = O_p\big(n^{-1/3}\sqrt{\log \log n}\big).$$

Moreover,

$$\min\big(\hat{f}_n(x_0) - f_0(x_0), 0\big) = O_p(n^{-1/2}).$$

PROOF OF THEOREM 2.12

Suppose that for some fixed $\delta > 0$, (A) holds for every $t \in [x_0 - 2\delta, x_0 + 2\delta]$. Our first aim is to show that

$$\inf_{t \in [x_0 - \delta, x_0 + \delta]} \min \left(\hat{f}_n(t) - f_0(t), 0 \right) = O_p(n^{-1/2}). \tag{2.5}$$

Suppose that $c := \hat{f}_n(t) - f_0(t) < 0$. If we can find $x_0 - \delta \leq t < x_0$ such that $\hat{f}_n$ and f_0 intersect at the point $(t, f_0(t))$, then

$$\left(\hat{F}_n(t) - \hat{F}_n(x_0 - \delta) \right) - \left(F_0(t) - F_0(x_0 - \delta) \right) \geq \frac{1}{2} \frac{|c|(t - x_0 + \delta)^2}{x_0 - t} \geq 0;$$

$$\left(\hat{F}_n(x_0) - \hat{F}_n(t) \right) - \left(F_0(x_0) - F_0(t) \right) \leq -\frac{1}{2} |c|(x_0 - t) < 0.$$

Otherwise, by taking $t = x_0 - \delta$, we can still verify the above two inequalities. Therefore,

$$\begin{aligned}
O_p(n^{-1/2}) = 2\|\hat{F}_n - F_0\|_\infty \\
\geq \max \Big\{ \left| \left(\hat{F}_n(t) - F_0(t) \right) - \left(\hat{F}_n(x_0 - \delta) - F_0(x_0 - \delta) \right) \right|, \\
\left| \left(\hat{F}_n(x_0) - F_0(x_0) \right) - \left(\hat{F}_n(t) - F_0(t) \right) \right| \Big\} \\
\geq \frac{|c|}{2} \max \left\{ \frac{(t - x_0 + \delta)^2}{x_0 - t}, x_0 - t \right\} \geq \frac{|c|\delta}{4},
\end{aligned}$$

where the first equation follows from Marshall's lemma (see (2.1)). Consequently, (2.5) holds true. As a remark, the above arguments are essentially the same as those illustrated in the proof of Theorem 2.1, where one only relies on the fact that f_0 is linear on $[x_0 - \delta, x_0]$ (i.e. "one-sided linearity").

In the rest of the proof, it suffices to only consider the situation of $\hat{f}_n(x_0) - f_0(x_0) > 0$. Recall that to handle this scenario, the proof of Theorem 2.1 makes use of the fact that the triangular density is linear on the whole $[0, 1]$ (i.e. "two-sided linearity"). This is no longer true here since f_0 is not linear on $[x_0 - \delta, x_0 + \delta]$. In the following, we deploy a different strategy to establish the rate.

Let $\tau_n^- = \max\{t \in \mathcal{S}(\hat{f}_n), t < x_0\}$ and $\tau_n^+ = \min\{t \in \mathcal{S}(\hat{f}_n), t \geq x_0\}$, where $\mathcal{S}(\cdot)$ is defined in (2.4). We consider three different cases separately.

(a) $\max \left(\tau_n^+ - x_0, x_0 - \tau_n^- \right) > \delta$. Then f_0 and $\hat{f}_n$ are linear on either $[x_0 - \delta, x_0]$ or $[x_0, x_0 + \delta]$. It follows from the line of reasoning as in the proof of Theorem 2.1 that $\max \left(\hat{f}_n(x_0) - f_0(x_0), 0 \right) = O_p(n^{-1/2})$.

(b) $\max \left(\tau_n^+ - x_0, x_0 - \tau_n^- \right) \leq \delta$ and $\tau_n^+ - \tau_n^- > n^{-1/3}$. Note that being in the set $\mathcal{S}(\hat{f}_n)$ implies that $\mathbb{F}_n(\tau_n^-) = \hat{F}_n(\tau_n^-)$ and $\mathbb{F}_n(\tau_n^+) = \hat{F}_n(\tau_n^+)$ for τ_n^- and τ_n^+. Since $\hat{f}_n$ is linear on $[\tau_n^-, \tau_n^+]$,

$$\left(\mathbb{F}_n(\tau_n^+) - F_0(\tau_n^+) \right) - \left(\mathbb{F}_n(\tau_n^-) - F_0(\tau_n^-) \right) \tag{2.6}$$

$$= \int_{\tau_n^-}^{\tau_n^+} \left(\hat{f}_n(t) - f_0(t) \right) dt$$

$$\geq \left(\hat{f}_n(x_0) - f_0(x_0) \right) (\tau_n^+ - \tau_n^-)/2$$
$$+ \min \left(\hat{f}_n(\tau_n^-) - f_0(\tau_n^-), \hat{f}_n(\tau_n^+) - f_0(\tau_n^+), 0 \right) (\tau_n^+ - \tau_n^-).$$

By the law of the iterated logarithm for local empirical processes (cf. Lemma 4.3.5 of Csörgő and Horvath (1993)), the term displayed in (2.6) is at most $\sqrt{\tau_n^+ - \tau_n^-} O_p(n^{-1/2}\sqrt{\log\log n})$. In view of (2.5), rearranging the terms in the above inequality yields

$$\hat{f}_n(x_0) - f_0(x_0) \leq (\tau_n^+ - \tau_n^-)^{-1/2} O_p(n^{-1/2}\sqrt{\log\log n}) + O_p(n^{-1/2})$$
$$\leq O_p(n^{-1/3}\sqrt{\log\log n}).$$

(c) $\tau_n^+ - \tau_n^- \leq n^{-1/3}$. Now we define $\tau_n^{++} = \max\{t \in \mathcal{S}(\hat{f}_n), t < x_0 + 2n^{-1/3}\}$ and $\tau_n^{+++} = \min\{t \in \mathcal{S}(\hat{f}_n), t \geq x_0 + 2n^{-1/3}\}$. Here the existence of τ_n^{++} is guaranteed by the condition that $\tau_n^+ - \tau_n^- \leq n^{-1/3}$. Furthermore, we note that in our definitions τ_n^+ and τ_n^{++} might not be distinct. Within this setting, three further scenarios are to be dealt with.

(c1) $\tau_n^{+++} - x_0 > \delta$. Since both f_0 and $\hat{f}_n$ are linear on $[\tau_n^{++}, \tau_n^{+++}]$, one can apply a strategy similar to that used in the proof of Theorem 2.1 to show that $|\hat{f}_n(\tau_n^{++}) - f_0(\tau_n^{++})| = O_p(n^{-1/2})$. Consequently,

$$\hat{f}_n(x_0) - f_0(x_0) = \hat{f}_n(\tau_n^{++}) - f_0(\tau_n^{++}) - \int_{x_0}^{\tau_n^{++}} \left(\hat{f}'_n(t) - f'_0(t) \right) dt$$
$$= O_p(n^{-1/3}),$$

where we have used the facts that $|\tau_n^{++} - x_0| < 2n^{-1/3}$ and

$$\sup_{t \in [x_0, x_0 + \delta]} |\hat{f}'_n(t) - f'_0(t)| = O_p(1).$$

Here the second fact can be derived by invoking

$$K_1 + o_p(1) \leq \inf_{t \in [x_0 - \delta, x_0 + \delta]} \hat{f}'_n(t) \leq \sup_{t \in [x_0 - \delta, x_0 + \delta]} \hat{f}'_n(t) \leq K_1 + K_2 + o_p(1),$$

which follows easily from consistency of $\hat{f}'_n$ in estimating f'_0 at the points $x_0 \pm \delta$.

(c2) $3n^{-1/3} < \tau_n^{+++} - x_0 \leq \delta$. Then $|\tau_n^{+++} - \tau_n^{++}| > n^{-1/3}$. Using essentially the same argument as in (b), we see that

$$\max \left(\hat{f}_n(\tau_n^{++}) - f_0(\tau_n^{++}), 0 \right) = O_p(n^{-1/3}\sqrt{\log\log n}),$$

and hence,

$$|\hat{f}_n(\tau_n^{++}) - f_0(\tau_n^{++})| = O_p(n^{-1/3}\sqrt{\log\log n}).$$

We then apply the argument presented in (c1) to derive

$$|\hat{f}_n(x_0) - f_0(x_0)| = O_p\big(n^{-1/3}\sqrt{\log\log n}\big).$$

(c3) $\tau_n^{+++} - x_0 \leq 3n^{-1/3}$. Then $n^{-1/3} \leq \tau_n^{+++} - \tau_n^{+} \leq 2n^{-1/3}$. By proceeding as in the proof of Lemma 4.3 of Groeneboom, Jongbloed and Wellner (2001b), we are able to verify that

$$\inf_{t\in[\tau_n^{+},\tau_n^{+++}]} |\hat{f}_n(t) - f_0(t)| = O_p(n^{-1/3}).$$

Here we invoked Lemma A.1 of Balabdaoui and Wellner (2007) with $k = 1$ and $d = 2$ to verify the above claim. See also Kim and Pollard (1990). Finally, we can argue as in (c1) to show that $\hat{f}_n(x_0) - f_0(x_0) = O_p(n^{-1/3})$.

The proof is complete by taking into account all the above cases. We remark that in our proof the worst case scenarios that drive the convergence rate are (b) and (c2). □

Theorem 2.13 (Adaptation at the boundary points: II). *In the case of (B) or (C),*
$$|\hat{f}_n(x_0) - f_0(x_0)| = O_p\big(n^{-\alpha/(2\alpha+1)}\sqrt{\log\log n}\big).$$

Moreover,
$$\min\big(\hat{f}_n(x_0) - f_0(x_0), 0\big) = O_p(n^{-1/2}).$$

The proof of Theorem 2.13 is of more technical nature and is deferred to Appendix II.

3. Estimation of a regression function

Changing notation slightly from the previous section, we now assume that we are given pairs $\{(X_{n,i}, Y_{n,i}) : i = 1, \ldots, n\}$ for $n = 1, 2, \ldots$ with

$$Y_{n,i} = r_0(X_{n,i}) + \epsilon_{n,i}, \quad \text{for } i = 1, \ldots, n,$$

where $r_0 : [0, 1] \to \mathbb{R}$ is a convex function, and where $\{\epsilon_{n,i} : i = 1, \ldots, n\}$ is a triangular array of IID random variables satisfying

(i) $\mathbb{E}\epsilon_{1,1} = 0$ and $\sigma_0^2 = \mathbb{E}\epsilon_{1,1}^2 < \infty$.

To simplify our analysis, the following fixed design is considered:

(ii) $X_{n,i} = i/(n+1)$ for $i = 1, \ldots, n$.

The LSE of r_0 proposed by Balabdaoui and Rufibach (2008) is

$$\hat{r}_n = \operatorname*{argmin}_{g\in\mathcal{K}} \left(\frac{1}{2}\int_0^1 g(t)^2 dt - \frac{1}{n}\sum_{i=1}^n g(X_{n,i})Y_{n,i}\right), \tag{3.1}$$

where, in this section, $\mathcal{K}$ denotes the set of all continuous convex functions on $[0, 1]$.

We note that the above estimator is slightly different from the more "classical" LSE that minimizes $\frac{1}{2n} \sum_{i=1}^{n} g(X_{n,i})^2 - \frac{1}{n} \sum_{i=1}^{n} g(X_{n,i}) Y_{n,i}$ over $\mathcal{K}$. Since its criterion function has a completely discrete nature, different techniques are needed to prove analogous results. We will not pursue this direction in the manuscript.

3.1. Basic properties

In this subsection, we list some basic properties of $\hat{r}_n$ given as (3.1). For any $t \in [0, 1]$, define

$$R_0(t) = \int_0^t r_0(s) ds;$$

$$\hat{R}_n(t) = \int_0^t \hat{r}_n(s) ds;$$

$$\mathbb{R}_n(t) = \frac{1}{n} \sum_{i=1}^{n} Y_{n,i} \mathbf{1}_{[X_{n,i}, \infty)}(t).$$

Proposition 3.1. *The LSE $\hat{r}_n$ exists and is unique.*

Proposition 3.2. *Let $\mathcal{S}(\hat{r}_n)$ denote the set of knots of $\hat{r}_n$. The following properties hold:*

(a) $\int_0^t \left(\hat{R}_n(s) - \mathbb{R}_n(s) \right) ds \begin{cases} = 0, & \text{for all } t \in \mathcal{S}(\hat{r}_n) \cup \{0, 1\}, \\ \geq 0, & \text{for all } t \in [0, 1]; \end{cases}$

(b) $\hat{R}_n(t) = \mathbb{R}_n(t)$ *for any* $t \in \mathcal{S}(\hat{r}_n) \cup \{0, 1\}$;

(c) $\int_0^1 \left\{ \int_0^t \left(\hat{R}_n(s) - \mathbb{R}_n(s) \right) ds \right\} d\hat{r}_n'(t) = 0.$

Proposition 3.3 (Marshall's lemma)**.**

$$\|\hat{R}_n - R_0\|_\infty \leq 2\|\mathbb{R}_n - R_0\|_\infty. \tag{3.2}$$

PROOF OF PROPOSITIONS 3.1, 3.2 AND 3.3
Note that we can rewrite $\hat{r}_n$ as

$$\operatorname*{argmin}_{g \in \mathcal{K}} \left(\frac{1}{2} \int_0^1 g(t)^2 dt - \int_0^1 g(t) d\mathbb{R}_n(t) \right).$$

From this perspective, it is easy to check that Proposition 3.1, Proposition 3.2 and Proposition 3.3 follow from, respectively, slight modifications of Lemma 2.1, Lemma 2.2 of Groeneboom, Jongbloed and Wellner (2001b) and Theorem 1 of Dümbgen, Rufibach and Wellner (2007). □

We remark that this version of Marshall's lemma in the regression setting serves as an important tool to establish consistency and the rate. In particular, in

the following, we show that (3.2) easily yields consistency of $\hat{R}_n$, and consistency of $\hat{r}_n$ follow from this together with convexity of $\hat{r}_n$.

Proposition 3.4 (Consistency). *Under Assumptions (i)–(ii), for any $0 < \delta \leq 1/2$,*

$$\sup_{x\in[\delta,1-\delta]} |\hat{r}_n(x) - r_0(x)| \overset{a.s.}{\to} 0, \quad as\ n \to \infty.$$

PROOF OF PROPOSITION 3.4

The proof replies on the following intermediate result: suppose there are differentiable functions $G_0, G_1, \ldots$, all are $[0,1] \to \mathbb{R}$ and have convex first derivative, then for any fixed $0 < \delta \leq 1/2$, $\|G_k - G_0\|_\infty \to 0$ implies $\sup_{t\in[\delta,1-\delta]} |G_k'(t) - G_0'(t)| \to 0$.

To see this, we first show that

$$\limsup_{k\to\infty} \sup_{t\in[\delta,1-\delta]} \{G_k'(t) - G_0'(t)\} \leq 0. \tag{3.3}$$

Note that for any $t \in [\delta, 1 - \delta]$ and any $0 < \epsilon \leq \delta$,

$$G_k'(t) - G_0'(t)$$
$$\leq \max\left(\frac{G_k(t+\epsilon) - G_k(t)}{\epsilon} - G_0'(t), \frac{G_k(t) - G_k(t-\epsilon)}{\epsilon} - G_0'(t) \right).$$

It then follows that

$$\limsup_{k\to\infty} \sup_{t\in[\delta,1-\delta]} \{G_k'(t) - G_0'(t)\}$$
$$\leq \sup_{t\in[\delta,1-\delta]} \left(\frac{G_0(t+\epsilon) - G_0(t)}{\epsilon} - G_0'(t), \frac{G_0(t) - G_0(t-\epsilon)}{\epsilon} - G_0'(t) \right).$$

Our claim (3.3) can be verified by letting $\epsilon \to 0$.

Secondly, we show that

$$\liminf_{k\to\infty} \inf_{t\in[0,1]} \{G_k'(t) - G_0'(t)\} \geq 0.$$

We prove this by contradiction. Suppose that

$$\liminf_{k\to\infty} \inf_{t\in[0,1]} \{G_k'(t) - G_0'(t)\} = -M$$

for some $M > 0$. By extracting subsequences if necessary, we can assume that $\inf_{t\in[0,1]}\{G_k'(t) - G_0'(t)\} \to -M$ as $k \to \infty$. In view of (3.3), it follows from the convexity of G_k' and G_0' that one can find an interval $\mathcal{I}_k$ of positive length δ (which can depend on M) such that $\inf_{t\in\mathcal{I}_k}\{G_k'(t) - G_0'(t)\} \leq -M/2$ for every $k > K$, where K is a sufficiently large integer. This implies that

$$\|G_k - G_0\|_\infty \geq \sup_{t\in\mathcal{I}_k} |G_k(t) - G_0(t)| \geq M\delta/4 > 0,$$

for every $k > K$, which contradicts the fact that $\|G_k - G_0\|_\infty \to 0$ as $k \to \infty$.

Combining these two parts together completes the proof of the intermediate result.

Since $\|\mathbb{R}_n - R_0\|_\infty \overset{a.s.}{\to} 0$ by empirical process theory, Proposition 3.3 entails that $\|\hat{R}_n - R_0\|_\infty \overset{a.s.}{\to} 0$. Consistency of $\hat{r}_n$ then follows straightforwardly from the above intermediate result. □

3.2. Rate of convergence and asymptotic distribution

In the following, we assume that r_0 that is linear on $(a, b) \subseteq (0, 1)$. Moreover, (a, b) is "largest" in the sense that one can not find a bigger open interval (a', b') on which r_0 remains linear.

Theorem 3.5 (Rate and asymptotic distribution). *Under Assumptions (i)–(ii), for any $0 < \delta \leq (b - a)/2$,*

$$\sup_{x \in [a+\delta, b-\delta]} \left(|\hat{r}_n(x) - r_0(x)|, \ |\hat{r}_n'(x) - r_0'(x)| \right) = O_p(n^{-1/2}).$$

Moreover,

$$\sqrt{n} \left(\begin{array}{c} \hat{r}_n(x) - r_0(x) \\ \hat{r}_n'(x) - r_0'(x) \end{array} \right) \Rightarrow \sigma_0 \left(\begin{array}{c} (b-a)^{1/2} \, \tilde{H}^{(2)}\left(\frac{x-a}{b-a}\right) \\ (b-a)^{-1/2} \, \tilde{H}^{(3)}\left(\frac{x-a}{b-a}\right) \end{array} \right)$$

in $\mathcal{C}[a + \delta, b - \delta] \times \mathcal{D}[a + \delta, b - \delta]$, where $\tilde{H}$ is the invelope process defined in the second part of Theorem 2.4.

The proof of Theorem 3.5 is very similar to what has already been shown in Section 2.1, so is omitted for the sake of brevity.

In presence of the linearity of r_0 on (a, b), the limit distribution of the process $\sqrt{n}(\hat{r}_n - r_0)$ on (a, b) does not depend on $r_0(a)$ or $r_0(b)$. In addition, the above theorem continues to hold if we weaken Assumption (ii) to:

(ii') $\sup_{t \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[X_{n,i}, \infty)}(t) - t \right| = o\left(n^{-1/2}\right)$.

Theoretical results in the random design are also possible, where for instance, we can assume that $\{X_{n,i}, i = 1, \ldots, n\}$ are IID uniform random variables on $[0, 1]$. In this case, Theorem 3.5 is still valid, while a process different from $\tilde{H}$ is required to characterize the limit distribution. This follows from the fact that in the random design $\sqrt{n}(\mathbb{R}_n - R_0)$ can converge to a Gaussian process that is not a Brownian motion.

4. Appendices: proofs

4.1. Appendix I: existence of the limit process

Recall that $f_0(t) = 2(1 - t)\mathbf{1}_{[0,1]}(t)$, $X(t) = \mathbb{U}(F_0(t))$, and $Y(t) = \int_0^t X(s)ds$ for $t \geq 0$. In this section, we show the existence and uniqueness of the invelope

process H. The case of $\tilde{H}$ can be handled using essentially the same arguments, so is omitted here.

Lemmas 4.1–4.8 are needed to prove Theorem 2.4.

Lemma 4.1. *Let the functional $\phi(g)$ be defined as*

$$\phi(g) = \frac{1}{2} \int_0^1 g^2(t)\,dt - \int_0^1 g(t)\,dX(t)$$

for functions in the set $\mathcal{G}_k = \{g : [0,1] \to \mathbb{R}, g \text{ is convex}, g(0) = g(1) = k\}$. Then with probability one, the problem of minimizing $\phi(g)$ over $\mathcal{G}_k$ has a unique solution.

PROOF OF LEMMA 4.1
We consider this optimization problem in the metric space L^2. First, we show that if it exists, the minimizer must be in the subset

$$\mathcal{G}_{k,M} = \{g : [0,1] \to \mathbb{R}, g \text{ is convex}, g(0) = g(1) = k, \inf_{[0,1]} g(t) \geq -M\}$$

for some $0 < M < \infty$. To verify this, we need the following result

$$\sup_{g \in \mathcal{G}_{1,1}} \left| \int_0^1 g(t)\,dX(t) \right| < \infty, \quad \text{a.s.} \tag{4.1}$$

Let $W(t)$ be a standard Brownian motion. We note that $\int_0^1 g(t)\,dX(t)$ has the same distribution as

$$\int_0^1 g\big(1 - \sqrt{1-t}\big)dW(t) - W(1) \int_0^1 g(t)f_0(t)dt.$$

Using the entropy bound of $\mathcal{G}_{1,1}$ in L^2 (Theorem 2.7.1 of Guntuboyina and Sen (2013)) and Dudley's theorem (cf. Theorem 2.6.1 of Dudley (1999)), we can establish that

$$\mathcal{H} = \big\{h : [0,1] \to \mathbb{R} \mid h(t) = g\big(1 - \sqrt{1-t}\big), g \in \mathcal{G}_{1,1}\big\}$$

is a GC-set. As $\int_0^1 g\big(1 - \sqrt{1-t}\big)dW(t)$ is an isonormal Gaussian process indexed by $\mathcal{H}$, we have that a.s. $\sup_{g \in \mathcal{G}_{1,1}} \big| \int_0^1 g\big(1 - \sqrt{1-t}\big)dW(t) \big| < \infty$. Furthermore, it is easy to check that $W(1) < \infty$ a.s. and $\sup_{g \in \mathcal{G}_{1,1}} \big| \int_0^1 g(t)f_0(t)dt \big| \leq 2$. So our claim of (4.1) holds.

Now for sufficiently large M (with $M > k$),

$$\sup_{g \in \mathcal{G}_{k,M}} \left| \int_0^1 g(t)\,dX(t) \right| \leq M \sup_{g \in \mathcal{G}_{1,1}} \left| \int_0^1 g(t)\,dX(t) \right|.$$

Thus, for any $g \in \mathcal{G}_k$ with $\inf_{[0,1]} g = -M$, $\int_0^1 g(t)dX(t)$ is at most $O(M)$. Furthermore, $\frac{1}{2} \int_0^1 g^2(t)\,dt$ is of $O(M^2)$. Since ϕ at the minimizer could at most

be as large as $\phi(0) = 0$, we conclude that it suffices to only consider functions in $\mathcal{G}_{k,M}$ for some sufficiently large M.

Note that the functional ϕ is continuous (cf. Dudley's theorem) and strictly convex. Moreover, for $g_1, g_2 \in \mathcal{G}_{k,M}$, if $\int_0^1 \big(g_1(t) - g_2(t)\big)^2 dt = 0$, then $g_1 = g_2$ on $[0, 1]$. Since $\mathcal{G}_{k,M}$ is compact in L^2, the existence and uniqueness follow from a standard convex analysis argument in a Hilbert space. $\qquad\square$

As a remark, it can be seen from the proof of Lemma 4.1 that for a given $\omega \in \Omega$ from the sample space (which determines the value of $X(t)$), if the function ϕ has a unique minimizer over $\mathcal{G}_1$ (which happens a.s.), it also admits a unique minimizer over $\mathcal{G}_k$ for any $k > 1$.

Lemma 4.2. *Almost surely, $Y(t)$ does not have parabolic tangents at either $t = 0$ or $t = 1$.*

PROOF OF LEMMA 4.2
First, consider the case of $t = 0$. Theorem 1 of Lachal (1997) says that

$$\limsup_{t \to 0^+} \frac{\left| \int_0^t W(s)ds \right|}{\sqrt{(2/3)t^3 \log\log(1/t)}} = 1, \quad a.s.,$$

where W is a standard Brownian motion. From this, it follows that

$$\limsup_{t \to 0^+} \frac{|Y(t)|}{\sqrt{(2/3)t^3 \log\log(1/t)}} = \limsup_{t \to 0^+} \frac{\left| \int_0^t \{W(2s - s^2) - (2s - s^2)W(1)\}ds \right|}{\sqrt{(2/3)t^3 \log\log(1/t)}}$$

$$= \limsup_{t \to 0^+} \frac{\left| \int_0^t \sqrt{2}W(s)ds \right|}{\sqrt{(2/3)t^3 \log\log(1/t)}} = \sqrt{2}, \quad a.s.$$

thanks to the scaling properties of Brownian motion ($W(at) \stackrel{d}{=} \sqrt{a}W(t)$). This implies that $Y(t)$ does not have a parabolic tangent at $t = 0$.

Second, consider the case of $t = 1$. Note that $\lim_{t \to 1^-} \frac{Y(1) - Y(t)}{1-t} = X(1) = 0$ and

$$\limsup_{t \to 1^-} \left| \frac{Y(1) - Y(t)}{(1-t)^2} \right| = \limsup_{t \to 1^-} \left| \frac{\int_t^1 \{W(1) - W(F_0(s))\}ds - W(1)(1-t)^3/3}{(1-t)^2} \right|$$

$$= \lim_{t \to 1^-} \left| \frac{\int_t^1 \{W(1) - W(F_0(t))\}ds}{(1-t)^2} \right|$$

$$\stackrel{a.s.}{=} \lim_{t \to 0^+} \left| \frac{\int_0^t W(s^2)ds}{t^2} \right|,$$

where $F_0(t) = 2t - t^2$. Therefore, to prove that $Y(t)$ does not have a parabolic tangent at $t = 1$, it suffices to show that

$$\limsup_{t \to 0^+} \left| \frac{\int_0^t W(s^2)ds}{t^2} \right| = \infty, \quad a.s.$$

Denote by $Z(t) = \int_0^t W(s^2)ds$. For any $0 < t_1 < t_2 < 1$, we argue that the random variable

$$Z(t_2) - Z(t_1) - (t_2 - t_1)W(t_1^2)$$

follows a distribution of $N\left(0, t_2^4/6 - t_1^2 t_2^2 + 4t_1^3 t_2/3 - t_1^4/2\right)$. This is because

$$
\begin{aligned}
& Z(t_2) - Z(t_1) - (t_2 - t_1)W(t_1^2) \\
&= \int_{t_1}^{t_2} \{W(s^2) - W(t_1^2)\}ds \\
&= \int_{t_1^2}^{t_2^2} \frac{W(s) - W(t_1^2)}{2\sqrt{s}}ds = \int_{t_1^2}^{t_2^2} \frac{1}{2\sqrt{s}} \int_{t_1^2}^s dW(u)ds \\
&= \int_{t_1^2}^{t_2^2} \int_u^{t_2^2} \frac{1}{2\sqrt{s}}dsdW(u) = \int_{t_1^2}^{t_2^2} (t_2 - \sqrt{u})dW(u),
\end{aligned}
$$

where we invoked the stochastic Fubini's theorem in the last line, and thus,

$$
\begin{aligned}
\mathrm{Var}\left\{ Z(t_2) - Z(t_1) - (t_2 - t_1)W(t_1^2) \right\} &= \int_{t_1^2}^{t_2^2} (t_2 - \sqrt{u})^2 du \\
&= t_2^4/6 - t_1^2 t_2^2 + 4t_1^3 t_2/3 - t_1^4/2.
\end{aligned}
$$

Now setting $t_1 = 1/2$ and $t_{i+1} = t_i^2$ for every $i \in \mathbb{N}$. It is easy to check that the collection of random variables

$$\left\{ Z(t_{i-1}) - Z(t_i) - (t_{i-1} - t_i)W(t_i^2) \right\}_{i=1}^{\infty}$$

is mutually independent, so

$$\limsup_{i \to \infty} \frac{Z(t_{i-1}) - Z(t_i) - (t_{i-1} - t_i)W(t_i^2)}{t_{i-1}^2} = \infty, \quad a.s., \tag{4.2}$$

where we made use of the fact that

$$\lim_{i \to \infty} \frac{\sqrt{t_{i-1}^4/6 - t_i^2 t_{i-1}^2 + 4t_i^3 t_{i-1}/3 - t_i^4/2}}{t_{i-1}^2} = 1/\sqrt{6}.$$

Assume that there exists some $K > 0$ such that $|Z(t)| \le Kt^2$ for all sufficiently small $t > 0$. But it follows from (4.2) that a.s. one can find a subsequence of $\mathbb{N}$ (denoted by $\{i_j\}_{j=1}^{\infty}$) satisfying

$$\frac{Z(t_{i_j-1}) - Z(t_{i_j}) - (t_{i_j-1} - t_{i_j})W(t_{i_j}^2)}{t_{i_j-1}^2} \ge K + 1, \quad \text{for } j = 1, 2, \ldots$$

Consequently,

$$\frac{Z(t_{i_j-1})}{t_{i_j-1}^2} \ge K + 1 + \frac{Z(t_{i_j})}{t_{i_j-1}^2} + \frac{W(t_{i_j}^2)(t_{i_j-1} - t_{i_j})}{t_{i_j-1}^2}$$

$$\geq K + 1 - K t_{i_j} - \frac{|W(t_{i_j}^2)|}{\sqrt{t_{i_j}}} \frac{\sqrt{t_{i_j}}(t_{i_j-1} - t_{i_j})}{t_{i_j}} \overset{a.s.}{\to} K + 1,$$

as $j \to \infty$. The last step is due to $t_{i_j} \to 0^+$ and $\limsup_{s\to0^+} |W(s)|/s^{1/4} = 0$ a.s. (which is a direct application of the law of the iterated logarithm). The proof is completed by contradiction. $\square$

Now denote by f_k the unique function which minimizes $\phi(g)$ over $\mathcal{G}_k$. Let H_k be the second order integral satisfying $H_k(0) = Y(0)$, $H_k(1) = Y(1)$ and $H_k^{(2)} = f_k$.

Lemma 4.3. *Almost surely, for every $k \in \mathbb{N}$, f_k and H_k has the following properties:*

   (i) $H_k(t) \geq Y(t)$ for every $t \in [0,1]$;

  (ii) $\int_0^1 \big(H_k(t) - Y(t)\big) df_k'(t) = 0$, where the derivative can be interpreted as either the left or the right derivative;

 (iii) $H_k(t) = Y(t)$ and $H_k'(t) = X(t)$ for any $t \in \mathcal{S}(f_k)$, where $\mathcal{S}$ is the set of knots;

 (iv) $\int_0^t f_k(s)ds \leq X(t) - X(0)$ and $\int_t^1 f_k(s)ds \leq X(1) - X(t)$ for any $t \in \mathcal{S}(f_k)$;

  (v) f_k is a continuous function on $[0,1]$.

PROOF OF LEMMA 4.3

To show (i), (ii), (iii) and (iv), one may refer to Lemma 2.2 and Corollary 2.1 of Groeneboom, Jongbloed and Wellner (2001a) and use a similar functional derivative argument.

For (v), we note that since f_k is convex, discontinuity can only happen at $t = 0$ or $t = 1$. In the following, we show that it is impossible at $t = 0$. Suppose that f_k is discontinuous at zero. Consider the class of functions $g_\delta(t) = \max(1 - t/\delta, 0)$. Then

$$\Big(f_k(t) + \epsilon g_\delta(t)\Big)\mathbf{1}_{(0,1)}(t) + k\mathbf{1}_{\{0,1\}}(t) \in \mathcal{G}_k$$

for every $\epsilon \in (0, k - \liminf_{t\to0^+} f_k(t)]$. By considering the functional derivative of $\phi(g)$ and using integration by parts, we obtain that for any $\delta > 0$,

$$k\delta/2 \geq \int_0^1 g_\delta(t) f_k(t) dt \geq \int_0^1 g_\delta(t) dX(t) = Y(\delta)/\delta,$$

which implies that $k\delta^2/2 \geq Y(\delta)$ for every $\delta > 0$. However, this contradicts Lemma 4.2, which says that a.s. Y does not have a parabolic tangent at $t = 0$. Consequently, f_k is continuous at $t = 0$. The same argument can also be applied to show the continuity of f_k at $t = 1$. $\square$

Lemma 4.4. *Fix any $t \in (0,1)$. Denote by τ_k^- the right-most knot of f_k on (0,t]. If such knot does not exist, then set $\tau_k^- = 0$. Similarly, denote by τ_k^+ the left-most knot of f_k on [t,1), and set $\tau_k^+ = 1$ if such knot does not exist. Then, for almost every $\omega \in \Omega$, there exists $K > 0$ such that for every $k \geq K$, $\tau_k^- \neq 0$ and $\tau_k^+ \neq 1$. Here we suppressed the dependence of K and f_k (as well as τ_k^- and τ_k^+) on ω (via X) in the notation.*

PROOF OF LEMMA 4.4

Recall that $X(t)$ is said to be sample-bounded on $[0,1]$ if $\sup_{[0,1]} |X(t)|$ is finite for almost all $\omega \in \Omega$. Since both $X(t)$ and $Y(t)$ are sample bounded. Without loss of generality, in the following, wee can fix $\omega \in \Omega$ and assume that $\sup_{[0,1]} |X(t)| < \infty$ and $\sup_{[0,1]} |Y(t)| < \infty$.

First, we show the existence of at least one knot on $(0,1)$. Note that the cubic polynomial P_k with $P_k(0) = Y(0) = 0$, $P_k(1) = Y(1)$, $P_k^{(2)}(0) = P_k^{(2)}(1) = k$ can be expressed as

$$P_k(s) = \frac{k}{2}s^2 + \left(Y(1) - \frac{k}{2}\right)s,$$

Therefore, take for instance $s = 0.5$ and consider the event $P_k(0.5) \geq Y(0.5)$. This event can be reexpressed as

$$\frac{1}{2}Y(1) - Y(0.5) \geq \frac{1}{8}k,$$

which will eventually become false as $k \to \infty$. This is due to the fact that $Y(t)$ is sample bounded. In view of (i) and (v) of Lemma 4.3, we conclude that f_k has at least one knot in the open interval $(0,1)$ for sufficiently large k.

Next, take k large enough so that f_k has one knot in $(0,1)$, which we denote by τ_k. By (iii) of Lemma 4.3, $H_k(\tau_k) = Y(\tau_k)$ and $H_k'(\tau_k) = X(\tau_k)$. Without loss of generality, we may assume that $\tau_k > t$. Now the cubic polynomial P_k with $P_k(0) = Y(0) = 0$, $P_k(\tau_k) = Y(\tau_k)$, $P_k^{(2)}(0) = k$ and $P_k'(\tau_k) = X(\tau_k)$ can be expressed as

$$P_k(s) = a_{k1}s^3 + a_{k2}s^2 + a_{k3}s,$$

where

$$a_{k1} = \frac{X(\tau_k)}{2\tau_k^2} - \frac{Y(\tau_k)}{2\tau_k^3} - \frac{k}{4\tau_k};$$
$$a_{k2} = \frac{k}{2};$$
$$a_{k3} = \frac{3Y(\tau_k)}{2\tau_k} - \frac{X(\tau_k)}{2} - \frac{k\tau_k}{4}.$$

By taking, say for example, $s = t/2$, it can then be verified that the event $P_k(t/2) \geq Y(t/2)$, which is equivalent to

$$\frac{X(\tau_k)t^3}{16\tau_k^2} - \frac{Y(\tau_k)t^3}{16\tau_k^3} + \frac{3Y(\tau_k)t}{4\tau_k} - \frac{X(\tau_k)t}{4} \geq \frac{kt}{32\tau_k}(t - 2\tau_k)^2,$$

will eventually stop happening as $k \to \infty$. This is due to the sample boundedness of both $X(t)$ and $Y(t)$. Consequently, $\tau_k^- \neq 0$ for sufficiently large k. Furthermore, using essentially the same argument, one can also show that $\tau_k^+ \neq 1$ for large k, which completes the proof of this lemma. $\square$

Lemma 4.5. *For almost every $\omega \in \Omega$ (which determines $X(t)$ and f_k), we can find an $M > 0$ such that*

$$\inf_{k \in \mathbb{N}} \inf_{[0,1]} f_k > -M.$$

PROOF OF LEMMA 4.5

Fix any $\delta \in (0, 1/2]$. In view of Lemma 4.4, we may assume that there exist knots τ_k^- and τ_k^+ on $(0, \delta]$ and $[1 - \delta, 1)$ respectively for sufficiently large k. If $\inf_{[0,1]} f_k \geq 0$ for every $k > K$, then we are done. Otherwise, we focus on those $\inf_{[0,1]} f_k < 0$ and find $0 < t_{k,1} < t_{k,2} < 1$ with $f_k(t_{k,1}) = f_k(t_{k,2}) = 0$. Note that the existence of $t_{k,1}$ and $t_{k,2}$ are guaranteed by Lemma 4.3 (v). In the following, we take $\delta = 1/12$ and consider two scenarios.

(a) $t_{k,2} - t_{k,1} < 2\delta$. Let $a_{k,1} \in \partial f_k(t_{k,1})$ and $a_{k,2} \in \partial f_k(t_{k,2})$, where ∂ is the subgradient operator. Then Lemma 4.3(v) implies that $a_{k,1} < 0$ and $a_{k,2} > 0$. Since both $a_{k,1}(s - t_{k,1})$ and $a_{k,2}(s - t_{k,2})$ can be regarded as supporting hyperplanes of f_k due to its convexity, it follows that $f_k(s) \geq \max\left(a_{k,1}(s - t_{k,1}), a_{k,2}(s - t_{k,2})\right)$. Set $C_k = -a_{k,1}a_{k,2}(t_{k,2} - t_{k,1})/(a_{k,2} - a_{k,1})$ (i.e. $-C_k$ is the value of the above hyperplanes at their intersection, which is negative), so $\inf_{[0,1]} f_k \geq -C_k$. Now

$$2 \sup_{t \in [0,1]} |X(t)|$$

$$\geq X(\tau_k^+) - X(\tau_k^-) = H_k'(\tau_k^+) - H_k'(\tau_k^-) = \int_{\tau_k^-}^{\tau_k^+} f_k(s)\, ds$$

$$\geq \int_{\tau_k^-}^{\tau_k^+} \max\left(a_{k,1}(s - t_{k,1}), a_{k,2}(s - t_{k,2})\right) ds$$

$$\geq \int_{\delta}^{1-\delta} \max\left(a_{k,1}(s - t_{k,1}), a_{k,2}(s - t_{k,2})\right) ds - 2\delta C_k$$

$$\geq \int_{\delta}^{1-\delta} \max\left(a_{k,1}(s - t_{k,1}), a_{k,2}(s - t_{k,2}), 0\right) ds$$

$$\qquad - \int_{\delta}^{1-\delta} \max\left(-a_{k,1}(s - t_{k,1}), -a_{k,2}(s - t_{k,2}), 0\right) ds - 2\delta C_k$$

$$\geq \inf_{u \in [0, 1-4\delta]} \left\{|a_{k,1}| u^2 + |a_{k,2}|(1 - 4\delta - u)^2\right\}/2 - C_k(t_{k,2} - t_{k-1})/2 - 2\delta C_k$$

$$\geq \frac{(1 - 4\delta)^2}{4\delta} C_k - \delta C_k - 2\delta C_k = \left(\frac{1}{4\delta} + \delta - 2\right) C_k \geq C_k.$$

Consequently, a.s., $\sup_{k \in \mathbb{N}} C_k < \infty$.

(b) $t_{k,2} - t_{k,1} \geq 2\delta$. Now consider $f_{k+} = \max(f_k, 0)$. It follows that

$$0 \geq \phi(f_k) - \phi(f_{k+})$$

$$= \frac{1}{2} \int_0^1 \min(0, f_k(t))^2\, dt - \int_0^1 \min(0, f_k(t))\, dX(t)$$

Let $M_k = -\inf_{[0,1]} f_k$. By the convexity of f_k, we see that the first term is no smaller than $2\delta M_k^2/3$. For the second term, we can again use Dudley's theorem and the fact that the following class

$$\mathcal{G}_{\text{unit}} = \left\{ g : [0,1] \to [-1,0] \middle| \exists\, 0 \leq x_1 < x_2 \leq 1 \right.$$
$$\left. \text{s.t. } g(s) \text{ is convex on } [x_1, x_2]; g(s) = 0, \forall s \in [0, x_1] \cup [x_2, 1] \right\}$$

has entropy of order $\eta^{-1/2}$ in L^2 to argue that

$$\sup_{g \in \mathcal{G}_{\text{unit}}} \left| \int_{[0,1]} g\, dX(t) \right| < \infty, \quad \text{a.s.}$$

Therefore, the second term is at most $O(M_k)$. Then we can use the argument in the proof of Lemma 4.1 to establish that a.s. $\limsup_{k \to \infty} M_k < \infty$. $\square$

Lemma 4.6. *For any fixed $t \in (0,1)$ and almost every $\omega \in \Omega$, $\sup_k |f_k(t)|$, $\sup_k |f_k^-(t)|$ and $\sup_k |f_k^+(t)|$ are bounded.*

PROOF OF LEMMA 4.6
Let $\Delta = \min(t, 1-t)/2$. In view of Lemma 4.4, for sufficiently large k, we can assume that f_k has at least one knot in $(0, t - \Delta]$, and one knot in $[t + \Delta, 1)$. Denote these two points by τ_k^- and τ_k^+ respectively. By the convexity of f_k, there exists some $c \in \mathbb{R}$ such that $f_k(s) \geq c(s - t) + f_k(t)$ for every $s \in [0,1]$. It follows that

$$2 \sup_{t \in [0,1]} |X(t)| \geq X(\tau_k^+) - X(\tau_k^-) = \int_{\tau_k^-}^{\tau_k^+} f_k(s)\, ds$$
$$\geq \int_{[\tau_k^-, t-\Delta] \cup [t+\Delta, \tau_k^+]} \left(\inf_{[0,1]} f_k \right) ds + \int_{t-\Delta}^{t+\Delta} \{c(s-t) + f_k(t)\}\, ds$$
$$\geq -M_k + 2\Delta f_k(t),$$

where $M_k = -\inf_{[0,1]} f_k$. By Lemma 4.5, we see that for almost every $\omega \in \Omega$, $\limsup_{k \to \infty} f_k(t)$ is bounded. Combining this with the lower bound we established previously entails the boundedness of $\{f_k(t)\}_k$.

Next, note that both $\{f_k(t-\Delta)\}_k$ and $\{f_k(t+\Delta)\}_k$ are bounded. The boundedness of $\{f_k^-(t)\}_k$ and $\{f_k^+(t)\}_k$ immediately follows from the convexity of f_k by utilizing

$$\partial f_k(t) \subseteq \left[\frac{f_k(t) - f_k(t-\Delta)}{\Delta}, \frac{f_k(t+\Delta) - f_k(t)}{\Delta} \right]. \qquad \square$$

Lemma 4.7. *For almost every $\omega \in \Omega$, we have that both $\left\{ \sup_{t \in [0,1]} |H_k(t)| \right\}_k$ and $\left\{ \sup_{t \in [0,1]} |H_k'(t)| \right\}_k$ are bounded.*

By Lemma 4.3(iv), for any $t \in [0, 1]$,

$$\min\left(0, \inf_{k \in \mathbb{N}} \inf_{[0,1]} f_k\right) \leq t \inf_{[0,1]} f_k \leq \int_0^t f_k(s)ds = \int_0^1 f_k(s)ds - \int_t^1 f_k(s)ds$$

$$\leq X(1) - X(0) - (1-t) \inf_{[0,1]} f_k \leq X(1) - X(0) - \min\left(0, \inf_{k \in \mathbb{N}} \inf_{[0,1]} f_k\right).$$

Consequently, Lemma 4.5 entails the boundedness of $\left\{\sup_{t \in [0,1]} \left| \int_0^t f_k(s)ds \right|\right\}_k$. Furthermore, Lemma 4.4 says that one can always find a knot $\tau_k^- \in (0, 1)$ with $X(\tau_k^-) = H_k'(\tau_k^-)$ for all sufficiently large k. Thus, the boundedness of $\left\{\sup_{t \in [0,1]} |H_k'(t)|\right\}_k$ follows from the fact that

$$\sup_{t \in [0,1]} |H_k'(t)| \leq \sup_{t \in [0,1]} \left| X(\tau_k^-) - \int_0^{\tau_k^-} f_k(s)ds + \int_0^t f_k(s)ds \right|$$

$$\leq \sup_{t \in [0,1]} |X(t)| + 2 \sup_{t \in [0,1]} \left| \int_0^t f_k(s)ds \right|.$$

Finally, one can derive the sample boundedness of $\left\{\sup_{t \in [0,1]} |H_k(t)|\right\}_k$ by using the equality $H_k(t) = Y(0) + \int_0^t H_k'(s)ds$. $\square$

Lemma 4.8. *For almost every $\omega \in \Omega$, both $\{H_k\}_k$ and $\{H_k'\}_k$ are uniformly equicontinuous on $[0, 1]$. In fact, they are uniformly Hölder continuous with exponent less than $1/4$.*

Here we only show that the family $\{H_k'\}_k$ is uniformly equicontinuous. Fix any $0 < \delta < 1$. For any $0 \leq t_1 \leq t_2 \leq 1$ with $t_2 - t_1 < \delta$, by the convexity of f_k,

$$H_k'(t_2) - H_k'(t_1) = \int_{t_1}^{t_2} f_k(s)ds \leq \max\left(\int_0^\delta f_k(s)ds, \int_{1-\delta}^1 f_k(s)ds\right).$$

In the following, we shall focus on $\int_0^\delta f_k(s)ds$. The term $\int_{1-\delta}^1 f_k(s)ds$ can be handled in exactly the same fashion. By Lemma 4.4, $\mathcal{S}(f_k)$ is non-empty for every $k > K$, where K is a sufficiently large positive integer. Furthermore, in view of Lemma 4.5, we can assume that $\inf_{k \in \mathbb{N}} \inf_{[0,1]} f_k \geq -M$ for some $M > 0$.

Three scenarios are discussed in the following. Here we fix $\alpha \in (0, 1/2)$.

(a) There exists at least one knot $\tau_k \in \mathcal{S}(f_k)$ with $\tau_k \in [\delta, \delta^\alpha]$. Then

$$\int_0^\delta f_k(s)ds \leq \int_0^{\tau_k} f_k(s)ds + M\delta^\alpha$$

$$\leq X(\tau_k) - X(0) + M\delta^\alpha \leq M'\delta^{\alpha^2} + M\delta^\alpha,$$

for some $M' > 0$ (which is the α-Hölder constant of this particular realization of $X(t)$). The last line follows from Lemma 4.3(iv) and the fact that $X(t)$ is α-Hölder-continuous, so

$$|X(\tau_k) - X(0)| \leq M'|\tau_k|^\alpha \leq M'(\delta^\alpha)^\alpha = M'\delta^{\alpha^2}.$$

(b) $\mathcal{S}(f_k) \cap (0, \delta^\alpha] = \emptyset$. Let $\tau_k \in \mathcal{S}(f_k)$ be the left-most knot in $[\delta^\alpha, 1)$. Then f_k is linear on $[0, \tau_k]$ and $f_k(\tau_k) \geq -M$. Note that

$$M'\tau_k{}^\alpha \geq X(\tau_k) - X(0) \geq \int_0^{\tau_k} f_k(s)ds = \frac{f_k(0) + f_k(\tau_k)}{2}\tau_k$$
$$\geq (f_k(0) - M)\tau_k/2,$$

so $f_k(0) \leq 2M'\tau_k^{\alpha-1} + M \leq 2M'\delta^{\alpha^2-\alpha} + M$. It now follows that

$$\int_0^\delta f_k(s)ds \leq (2M'\delta^{\alpha^2-\alpha} + M)\delta \leq (2M' + M)\delta^\alpha.$$

We remark that since $f_k(0) = k$, the above conclusion also implies that $\mathcal{S}(f_k) \cap (0, \delta^\alpha] \neq \emptyset$ for all sufficiently large k.

(c) $\mathcal{S}(f_k) \cap [\delta, \delta^\alpha] = \emptyset$, but $\mathcal{S}(f_k) \cap (0, \delta] \neq \emptyset$. Let τ_k^- be the right-most knot in $(0, \delta]$ and τ_k^+ be the left-most knot in $[\delta^\alpha, 1)$. As a convention, we set $\tau_k^+ = 1$ if such a knot does not exist. Note that f_k is linear on $[\tau_k^-, \tau_k^+]$, so we can use essentially the same argument as above to see that $\int_{\tau_k^-}^\delta f_k(s)ds \leq (2M' + M)\delta^\alpha$. Finally,

$$\int_0^\delta f_k(s)ds \leq X(\tau_k^-) - X(0) + \int_{\tau_k^-}^\delta f_k(s)ds$$
$$\leq M'\delta^\alpha + (2M' + M)\delta^\alpha = (3M' + M)\delta^\alpha.$$

To finish the proof, we shall apply Lemma 4.5 to verify that $H_k'(t_2) - H_k'(t_1) \geq -M\delta$.

In addition, combining results from all the scenarios listed above, we see that $\{H_k'\}_k$ is uniformly Hölder continuous with exponent α^2 ($< 1/4$). □

PROOF OF THEOREM 2.4

For every $m \in \mathbb{N}$ with $m \geq 3$, define the following norms

$$\|H\|_m = \sup_{t \in [0,1]} |H(t)| + \sup_{t \in [0,1]} |H'(t)| + \sup_{t \in [1/m, 1-1/m]} |H^{(2)}(t)|.$$

First, we show the existence of such a function for almost all $\omega \in \Omega$ by construction. Fix ω (thus we focus on a particular realization of $X(t)$ but suppress its dependence on ω in the notation). Let H_k be the function satisfying $H_k^{(2)} = f_k$, $H_k(0) = 0$ and $H_k(1) = Y(1)$. We claim that the sequence H_k admits a convergent subsequence in the topology induced by the norm $\|\cdot\|_m$.

By Lemma 4.6, we may assume that for $t = 1/m$ and $t = 1 - 1/m$, $\{f_k(t)\}_k$, $\{f_k^-(t)\}_k$ and $\{f_k^+(t)\}_k$ are bounded. By the convexity, $\{f_k\}_k$ have uniformly bounded derivatives on $[1/m, 1 - 1/m]$ so are uniformly bounded and equicontinuous. Therefore, the Arzelà–Ascoli theorem guarantees that the sequence f_k has a convergent subsequence f_{k_l} in the supremum metric on $[1/m, 1 - 1/m]$. Extracting further subsequences if necessary, one can get f_{k_l} converging in the topology induced by the L^∞ norm on $[1/m, 1 - 1/m]$ for $m = 3, 4, \ldots$.

Now by Lemma 4.7 and Lemma 4.8, we can assume that $\{H_k\}_k$ and $\{H_k'\}_k$ are bounded and equicontinuous on $[0, 1]$. By the Arzelà–Ascoli theorem again, we are able to extract further subsequences if necessary to make H_{k_l} converge in the topology induced by the norms $\|\cdot\|_m$ for $m = 3, 4, \ldots$. We denote the function that H_{k_l} converges to by H.

In the following, we show that H has the properties listed in the statement of the theorem.

(1) Because $H_k(t) \geq Y(t)$ for every $k \in \mathbb{N}$ and every $t \in [0, 1]$, $H(t) \geq Y(t)$ for $t \in [0, 1]$.

(2) $H^{(2)}$ is convex on $(0, 1)$ since every f_k is convex.

(3)–(4) Since $H_k(0) = Y(0)$ and $H_k(1) = Y(1)$ for all $k \in \mathbb{N}$, we have $H(0) = Y(0)$ and $H(1) = Y(1)$. Now let $\tau_k^- = \inf\{\tau_k : \tau \in \mathcal{S}(f_k)\}$ and $\tau_k^+ = \sup\{\tau_k : \tau \in \mathcal{S}(f_k)\}$. In light of Lemma 4.4, one can assume that $\lim_{k \to \infty} \tau_k^- = 0$ and $\lim_{k \to \infty} \tau_k^+ = 1$. The sample continuity of $X(t)$, together with the property of knots (see (iii) of Lemma 4.3), entails that $H_k'(\tau_k^-) \to X(0)$ and $H_k'(\tau_k^+) \to X(1)$ as $k \to \infty$. Finally, we use the uniform convergence and the continuity of H' to establish $H'(0) = X(0)$ and $H'(1) = X(1)$.

(5) It follows from Lemma 4.3(ii) that $\int_{[1/m, 1-1/m]} \big(H(t) - Y(t)\big)\, dH^{(3)}(t) = 0$. Now let $m \to \infty$ to see the required property.

This completes the proof of existence.

It remains to show the uniqueness of H. Suppose that there are H_1 and H_2 satisfying Conditions (1)–(5) listed in the statement of Theorem 2.4. For notational convenience, we write $h_1 = H_1^{(2)}$ and $h_2 = H_2^{(2)}$. Then,

$$\left\{ \frac{1}{2} \int_0^1 h_1^2(t)dt - \int_0^1 h_1(t)dX(t) \right\} - \left\{ \frac{1}{2} \int_0^1 h_2^2(t)dt - \int_0^1 h_2(t)dX(t) \right\}$$

$$= \frac{1}{2} \int_0^1 \Big(h_1(t) - h_2(t) + h_2(t) \Big)^2 dt - \int_0^1 \Big(h_1(t) - h_2(t) + h_2(t) \Big) dX(t)$$

$$\quad - \frac{1}{2} \int_0^1 h_2^2(t)dt + \int_0^1 h_2(t)dX(t)$$

$$= \frac{1}{2} \int_0^1 \big(h_1(t) - h_2(t) \big)^2 dt$$

$$\quad + \left\{ \int_0^1 \big(h_1(t) - h_2(t) \big) h_2(t)dt - \int_0^1 \big(h_1(t) - h_2(t) \big) dX(t) \right\}$$

$$\geq \frac{1}{2}\int_0^1 \big(h_1(t) - h_2(t)\big)^2 dt,$$

where we used Conditions (1)–(5) of H_2 to derive the last inequality. By swapping H_1 and H_2, we further obtain the following inequality

$$\left\{\frac{1}{2}\int_0^1 h_2^2(t)dt - \int_0^1 h_2(t)dX(t)\right\} - \left\{\frac{1}{2}\int_0^1 h_1^2(t)dt - \int_0^1 h_1(t)dX(t)\right\}$$
$$\geq \frac{1}{2}\int_0^1 \big(h_2(t) - h_1(t)\big)^2 dt.$$

Adding together the above two inequalities yields $0 \geq \int_0^1 \big(h_1(t) - h_2(t)\big)^2 dt$, which implies the uniqueness of $H^{(2)}$ on $(0,1)$. The uniqueness of H then follows from its third condition.

The proof for the second part (i.e. the existence and uniqueness of $\tilde{H}$) is similar and is therefore omitted. $\qquad\square$

Proof of Corollary 2.5

We can easily verify the existence of such a function by using the same construction in the proof of Theorem 2.4. In particular, if $Y(t)$ does not have parabolic tangents at both $t = 0$ and $t = 1$ (which happens a.s. according to Lemma 4.2), then $f_k(0) \to \infty$ and $f_k(1) \to \infty$ as $k \to \infty$.

On the other hand, if $H^{(2)}(0^+) \to \infty$, there must be a sequence of knots $\tau_1, \tau_2, \ldots$ of $H^{(2)}$ with $\lim_{j\to\infty} \tau_j = 0$. In views of Conditions (1), (2) and (5), one necessarily has $H(\tau_j) = Y(\tau_j)$ and $H'(\tau_j) = X(\tau_j)$ for every j. The fact that H, H', Y and X are all continuous entails that $H(0) = Y(0)$ and $H'(0) = X(0)$. Consequently, Condition (3') implies Condition (3). We now apply the same argument to $H^{(2)}(1^-)$ to conclude that Condition (4') implies Condition (4). Hence, in view of Theorem 2.4, H is unique. $\qquad\square$

4.2. Appendix II: pointwise adaptation for cases (B) and (C)

The following three lemmas are required to prove Theorem 2.13.

Lemma 4.9. *For any $\alpha > 1$, $\inf_{k\in[0,1]}\left\{4k^{\alpha+2} + (1+k)^{\alpha+1}(\alpha - 2k)\right\} > 0$.*

Proof of Lemma 4.9

It suffices to show that for any $k \in [0,1]$,

$$4k^{\alpha+2} + (1+k)^{\alpha+1}(\alpha - 2k) > 0.$$

First, it is easy to check that the above inequality holds true when $k = 0$. In the case of $k > 0$, we can restate the inequality to be proved as

$$4\Big(\frac{k}{1+k}\Big)^{\alpha+1} > 2 - \frac{\alpha}{k}. \tag{4.3}$$

Next, we define $m = k/(1+k) \in [0, 1/2]$, so that (4.3) can be rewritten as

$$4m^{\alpha+1} + \frac{\alpha}{m} > 2 + \alpha. \tag{4.4}$$

Note that

$$\inf_{m \in [0,1/2]} \left(4m^{\alpha+1} + \frac{\alpha}{m} \right) \geq \frac{\alpha(\alpha+2)}{(\alpha+1)} \left(\frac{4(\alpha+1)}{\alpha} \right)^{1/(\alpha+2)}$$

$$= (\alpha+2) \left\{ 4 \left(1 - \frac{1}{\alpha+1} \right)^{\alpha+1} \right\}^{1/(\alpha+2)}.$$

The inequality (4.4) now follows easily from that fact that $\left(1 - \frac{1}{\alpha+1} \right)^{\alpha+1} > 1/4$ for any $\alpha > 1$. $\square$

Lemma 4.10. Let $f_0(t) = f_0(x_0)K_1(t - x_0) + K_2(t - x_0)^\alpha \mathbf{1}_{[x_0,\infty)}(t)$ for every $t \in [x_0 - \delta, x_0 + \delta]$ for some $\delta \in (0, x_0)$ small, with $K_1 < 0$, $K_2 > 0$ and $\alpha > 1$. Then for any $x_0 - \delta \leq \tau^- \leq x_0 \leq \tau^+ \leq x_0 + \delta$,

$$\int_{\tau^-}^{\tau} \left\{ \frac{1}{2}(\tau^- + \tau) - t \right\} f_0(t) dt + \int_{\tau}^{\tau^+} \left\{ t - \frac{1}{2}(\tau + \tau^+) \right\} f_0(t) dt$$
$$\geq K_{f_0}(\tau^+ - x_0)^{\alpha+1}(\tau^+ - \tau^-),$$

where $\tau = (\tau^- + \tau^+)/2$, and where $K_{f_0} > 0$ is a constant that only depends on f_0.

PROOF OF LEMMA 4.10
First, it is easy to check that

$$\int_{\tau^-}^{\tau} \left\{ \frac{1}{2}(\tau^- + \tau) - t \right\} f_0(t) dt + \int_{\tau}^{\tau^+} \left\{ t - \frac{1}{2}(\tau + \tau^+) \right\} f_0(t) dt \tag{4.5}$$

$$= K_2 \int_{\tau^-}^{\tau} \left\{ \frac{1}{2}(\tau^- + \tau) - t \right\} (t - x_0)^\alpha \mathbf{1}_{[x_0,\infty)}(t) dt$$

$$+ K_2 \int_{\tau}^{\tau^+} \left\{ t - \frac{1}{2}(\tau + \tau^+) \right\} (t - x_0)^\alpha \mathbf{1}_{[x_0,\infty)}(t) dt.$$

If $\tau \leq x_0$, then (4.5) can be expressed as

$$K_2 \int_0^{\tau^+ - x_0} \left\{ t - \frac{1}{2}(\tau + \tau^+) + x_0 \right\} t^\alpha dt$$

$$= \frac{K_2(\tau^+ - x_0)^{\alpha+1}\{(\tau^+ - \tau)\alpha + 2(x_0 - \tau)\}}{2(\alpha+1)(\alpha+2)}$$

$$\geq \frac{K_2 \alpha}{4(\alpha+1)(\alpha+2)}(\tau^+ - x_0)^{\alpha+1}(\tau^+ - \tau^-).$$

On the other hand, if $\tau > x_0$, then after some elementary calculations, we can show that (4.5) is equal to

$$\frac{K_2[4(\tau - x_0)^{\alpha+2} + (\tau^+ - x_0)^{\alpha+1}\{\alpha(\tau^+ - \tau) - 2(\tau - x_0)\}]}{2(\alpha + 1)(\alpha + 2)}. \quad (4.6)$$

Denote by $k = (\tau - x_0)/(\tau^+ - \tau)$, so that (4.6) can be rewritten as

$$\frac{K_2}{2(\alpha + 1)(\alpha + 2)}\{4k^{\alpha+2} + (1 + k)^{\alpha+1}(\alpha - 2k)\}(\tau^+ - \tau)^{\alpha+2} \geq C_{f_0}(\tau^+ - \tau)^{\alpha+2},$$

where $C_{f_0} > 0$ is a constant that only depends on f_0, and where we applied Lemma 4.9 with the fact that $k \in [0, 1]$ to derive the above displayed equation. Consequently, by setting $K_{f_0} = C_{f_0}/2^{\alpha+2}$, it is straightforward to check that (4.6) is greater than or equal to $K_{f_0}(\tau^+ - x_0)^{\alpha+1}(\tau^+ - \tau^-)$. □

Lemma 4.11. *Let $\mathcal{F}$ be a collection of functions defined on $[x_0 - \delta, x_0 + \delta]$, with $\delta > 0$ small. Suppose that for a fixed $x \in [x_0 - \delta, x_0 + \delta]$ and every $0 < R \leq R_0$ such that $[x, x + R] \subseteq [x_0 - \delta, x_0 + \delta]$, the collection*

$$\mathcal{F}_{x,R} = \left\{f_{x,y} = f1_{[x,y]}, \quad f \in \mathcal{F}, \quad x \leq y \leq x + R\right\}$$

admits an envelope $F_{x,R}$ such that

$$\mathbb{E}F^2_{x,R}(X_1) \leq KR^{2d-1}$$

for some $d \geq 1/2$ fixed and $K > 0$ depending only on x_0 and δ, where $X_1 \sim F_0$. Moreover, suppose that

$$\sup_Q \int_0^1 \sqrt{\log N(\eta\|F_{x,R}\|_{Q,2}, \mathcal{F}_{x,R}, L_2(Q))}d\eta < \infty.$$

Then, for every $\epsilon > 0$ and $s_0 > 0$, there exist a random variable M_n of $O_p(1)$ such that

$$\sup_{s \geq s_0} \sup_{x \leq y \leq x+R_0} n^{\frac{s+d}{2s+1}} \max\left(\sup_{f_{x,y} \in \mathcal{F}_{x,y-x}} \left|\int f_{x,y}d(\mathbb{F}_n - F_0)\right| - \epsilon|y - x|^{s+d}, 0\right)$$
$$= M_n.$$

PROOF OF LEMMA 4.11
This lemma slightly generalizes Lemma A.1 of Balabdaoui and Wellner (2007). Its proof proceeds as in Balabdaoui and Wellner (2007) with minor modifications, so is omitted for brevity. We remark that here only the collection of functions defined on $[x, x + R]$ are considered. For functions on $[x - R, x]$, an analogous version of this lemma also holds true by symmetry. □

PROOF OF THEOREM 2.13
Here we only consider case (B). Case (C) can be handled similarly by symmetry.

Suppose that (B) holds true for every $t \in [x_0 - 2\delta, x_0 + 2\delta]$ for some fixed $\delta > 0$. Let $\tau_n^- = \max\{t \in \mathcal{S}(\hat{f}_n), t < x_0\}$ and $\tau_n^+ = \min\{t \in \mathcal{S}(\hat{f}_n), t \geq x_0\}$. Note that consistency of $\hat{f}_n$ implies that $\tau_n^+ - x_0 = o_p(1)$. Moreover, since f_0 is linear on $[x_0 - \delta, x_0]$, we can proceed as in the proofs of Theorem 2.1 and Corollary 2.3 to show that

$$\inf_{t \in [x_0-\delta, x_0]} \min \left(\hat{f}_n(t) - f_0(t), 0 \right) = O_p(n^{-1/2}) \tag{4.7}$$

and

$$\sup_{t \in [x_0-\delta, x_0]} |\hat{f}_n^-(t) - f_0'(t)| = O_p(n^{-1/2}),$$

where $\hat{f}_n^-$ is the left derivative of $\hat{f}_n$. Therefore, it suffices to only consider the behavior of $\max \left(\hat{f}_n(x_0) - f_0(x_0), 0 \right)$.

The proof can be divided into four parts.

(i) Suppose that $x_0 - \tau_n^- \geq n^{-1/(4\alpha+2)}$. Denote by $\tau_n^* = \max(\tau_n^-, x_0 - \delta)$. Then because both $\hat{f}_n$ and f_0 are linear on $[\tau_n^*, x_0]$,

$$\left(\hat{F}_n(x_0) - F_0(x_0) \right) - \left(\hat{F}_n(\tau_n^*) - F_0(\tau_n^*) \right)$$
$$= (x_0 - \tau_n^-)\left(\hat{f}_n(x_0) - f_0(x_0) + \hat{f}_n(\tau_n^*) - f_0(\tau_n^*) \right)/2.$$

Marshall's lemma entails that the left-hand side of the above equality is $O_p(n^{-1/2})$. Furthermore, it has been shown that

$$\min \left(\hat{f}_n(\tau_n^*) - f_0(\tau_n^*), 0 \right) = O_p(n^{-1/2}).$$

Rearranging the above equation yields

$$\max \left(\hat{f}_n(x_0) - f_0(x_0), 0 \right) = O_p(n^{-\alpha/(2\alpha+1)}).$$

(ii) Suppose that $\tau_n^+ - x_0 < n^{-1/(2\alpha+1)}$. First, we modify Lemma 4.2 of Groeneboom, Jongbloed and Wellner (2001b) to the following:

Let ξ_n be an arbitrary sequence of numbers converging to x_0. Define $\eta_n^- = \max\{t \in \mathcal{S}(\hat{f}_n) : t < \xi_n\}$ and $\eta_n^+ = \min\{t \in \mathcal{S}(\hat{f}_n) : t \geq \xi_n\}$. Then, $\max(\eta_n^+, \tau^+) - \max(\eta_n^-, \tau_n^+) = O_p(n^{-\alpha/(2\alpha+1)})$.

Here one can apply Lemma A.1 of Balabdaoui and Wellner (2007) with $k = \alpha$ and $d = 2$ to verify the above extension. By proceeding as in the proof of Lemma 4.3 of Groeneboom, Jongbloed and Wellner (2001b), one can verify that there exists some $\tau_n^{+++} > \tau_n^+$ such that $\tau_n^{+++} - x_0 = O_p(n^{-1/(2\alpha+1)})$ and

$$\inf_{t \in [\tau_n^+, \tau_n^{+++}]} |\hat{f}_n(t) - f_0(t)| = O_p(n^{-\alpha/(2\alpha+1)}).$$

Since $|\hat{f}_n^-(x_0) - f_0'(x_0)| = O_p(n^{-1/2})$, it follows that

$$\max(\hat{f}_n(x_0) - f_0(x_0), 0)$$

$$\leq \inf_{t\in[\tau_n^+,\tau_n^{+++}]} |\hat{f}_n(t) - f_0(t)| - \hat{f}_n^-(x_0)(\tau_n^{+++} - x_0) - f_0(\tau_n^{+++}) + f_0(x_0)$$

$$\leq O_p(n^{-\alpha/(2\alpha+1)}) - (\hat{f}_n^-(x_0) - f_0'(x_0))(\tau_n^{+++} - x_0) + K_2(\tau_n^{+++} - x_0)^\alpha$$

$$= O_p(n^{-\alpha/(2\alpha+1)}).$$

In fact, it is easy to see that the above conclusion still holds if we change our assumption from $\tau_n^+ - x_0 < n^{-1/(2\alpha+1)}$ to $\tau_n^+ - x_0 = O_p(n^{-1/(2\alpha+1)})$.

(iii) Suppose that $x_0 - \tau_n^- \leq \tau_n^+ - x_0$. In view of the discussion in (ii), it suffices to show that $\tau_n^+ - x_0 = O_p(n^{-1/(2\alpha+1)})$. Consider the behavior of $\int_0^t \hat{F}_n(s)ds$ (as a function of t) at the middle point $\tau_n = (\tau_n^- + \tau_n^+)/2$. Let

$$E_{1n} = \int_{\tau_n^-}^{\tau_n} \left\{\frac{1}{2}(\tau_n^- + \tau_n) - t\right\}d\big(F_0(t) - \mathbb{F}_n(t)\big)$$

$$+ \int_{\tau_n}^{\tau_n^+} \left\{t - \frac{1}{2}(\tau_n + \tau_n^+)\right\}d\big(F_0(t) - \mathbb{F}_n(t)\big);$$

$$E_{2n} = \int_{\tau_n^-}^{\tau_n} \left\{\frac{1}{2}(\tau_n^- + \tau_n) - t\right\}f_0(t)dt + \int_{\tau_n}^{\tau_n^+} \left\{t - \frac{1}{2}(\tau_n + \tau_n^+)\right\}f_0(t)dt.$$

Lemma 4.2 of Groeneboom, Jongbloed and Wellner (2001b) says that

$$E_{1n} \geq E_{2n}. \tag{4.8}$$

In view of Lemma A.1 of Balabdaoui and Wellner (2007) with $k = \alpha$ and $d = 2$,

$$E_{1n} \leq \epsilon(\tau_n^+ - \tau_n^-)^{\alpha+2} + O_p(n^{-(\alpha+2)/(2\alpha+1)})$$

$$\leq \epsilon\{2(\tau_n^+ - x_0)\}^{\alpha+2} + O_p(n^{-(\alpha+2)/(2\alpha+1)}) \tag{4.9}$$

for any $\epsilon > 0$. On the other hand, by Lemma 4.10, we have

$$E_{2n} \geq K_{f_0}(\tau_n^+ - x_0)^{\alpha+2} \tag{4.10}$$

for some constant $K_{f_0} > 0$ that only depends on the density function f_0. Combining (4.8), (4.9) and (4.10) yields $\tau_n^+ - x_0 = O_p(n^{-1/(2\alpha+1)})$.

(iv) Finally, suppose that $x_0 - \tau_n^- < n^{-1/(4\alpha+2)}$, $\tau_n^+ - x_0 \geq n^{-1/(2\alpha+1)}$ and $x_0 - \tau_n^- > \tau_n^+ - x_0$. Our first aim here is to show that

$$\frac{(\tau_n^+ - x_0)^{\alpha+1}}{\tau_n^+ - \tau_n^-} = O_p(n^{-\alpha/(2\alpha+1)}). \tag{4.11}$$

Let

$$s_n = (\alpha + 1)\frac{\log(\tau_n^+ - x_0)}{\log(\tau_n^+ - \tau_n^-)} - 1. \tag{4.12}$$

Since $\tau_n^+ \overset{a.s.}{\to} x_0$ and $x_0 - \tau_n^- < n^{-1/(4\alpha+2)}$, $0 < \tau_n^+ - x_0 < \tau_n^+ - \tau_n^- \leq 1$ for large n. This implies that $s_n \geq \alpha$ for sufficiently large n. It now follows from Lemma 4.11 (with $d = 2$ and $s_0 = \alpha$) that

$$n^{(s_n+2)/(2s_n+1)} \max\left(E_{1n} - \epsilon(\tau_n^+ - \tau_n^-)^{s_n+2}, 0\right) = O_p(1).$$

for any small $\epsilon > 0$. On the other hand, in view of Lemma 4.10 and identity (4.12), some elementary calculations yield

$$E_{2n} \geq K_{f_0}(\tau_n^+ - \tau_n^-)^{s_n+2}.$$

Plugging the above two equations into (4.8) entails that

$$\tau_n^+ - \tau_n^- \leq n^{-1/(2s_n+1)} O_p(1).$$

As a result,

$$\frac{(\tau_n^+ - x_0)^{\alpha+1}}{\tau_n^+ - \tau_n^-} = (\tau_n^+ - \tau_n^-)^{s_n} = n^{-s_n/(2s_n+1)} O_p(1) \leq O_p(n^{-\alpha/(2\alpha+1)}).$$

The rest of the proof is similar to (b) in the proof of Theorem 2.12. By the law of the iterated logarithm for local empirical processes,

$$\sqrt{\tau_n^+ - \tau_n^-} O_p\left(n^{-1/2}\sqrt{\log\log n}\right)$$

$$= \left(\mathbb{F}_n(\tau_n^+) - F_0(\tau_n^+)\right) - \left(\mathbb{F}_n(\tau_n^-) - F_0(\tau_n^-)\right)$$

$$= \int_{\tau_n^-}^{\tau_n^+} \left(\hat{f}_n(t) - f_0(t)\right) dt$$

$$= \left\{ \left(1 - \frac{\tau_n^+ - \tau_n^-}{2(x_0 - \tau_n^-)}\right)(\hat{f}_n(\tau_n^-) - f_0(\tau_n^-)) \right.$$

$$\left. + \frac{\tau_n^+ - \tau_n^-}{2(x_0 - \tau_n^-)}(\hat{f}_n(x_0) - f_0(x_0)) \right\}(\tau_n^+ - \tau_n^-) - \frac{K_2}{\alpha+1}(\tau_n^+ - x_0)^{\alpha+1},$$

where the last equality follows from the linearity of $\hat{f}_n(t)$ and $f_0(t) - K_2(t - x_0)^\alpha \mathbf{1}_{[x_0,\infty)}(t)$ on $[\tau_n^-, \tau_n^+]$. Since our assumptions in (iv) guarantee that $1 \leq (\tau_n^+ - \tau_n^-)/(x_0 - \tau_n^-) \leq 2$, rearranging the terms in the above displayed equation leads to

$$\max\left(\hat{f}_n(x_0) - f_0(x_0), 0\right)$$

$$\leq 2\left\{ (\tau_n^+ - \tau_n^-)^{-1/2} O_p\left(n^{-1/2}\sqrt{\log\log n}\right) \right.$$

$$\left. - \frac{1}{2}\min(\hat{f}_n(\tau_n^-) - f_0(\tau_n^-), 0) + \frac{K_2}{\alpha+1}\frac{(\tau_n^+ - x_0)^{\alpha+1}}{\tau_n^+ - \tau_n^-} \right\}. \quad (4.13)$$

Finally, as $\tau_n^+ - \tau_n^- \geq n^{-1/(2\alpha+1)}$, we can plug (4.7) and (4.11) into (4.13) to verify

$$\max\left(\hat{f}_n(x_0) - f_0(x_0), 0\right) \leq O_p\left(n^{-\alpha/(2\alpha+1)}\sqrt{\log\log n}\right). \qquad \square$$

Acknowledgments

References

Balabdaoui, F. (2007) Consistent estimation of a convex density at the origin. *Math. Methods Statist.*, **16**, 77–95. MR2335091

Balabdaoui, F. (2014) Global convergence of the log-concave MLE when the truth is geometric. *J. Nonparametric. Statist.*, **26**, 21–59. MR3174308

Balabdaoui, F., Jankowski, H., Rufibach, K. and Pavlides, M. (2013). Asymptotic distribution of the discrete log-concave MLE and related applications. *J. Roy. Statist. Soc. Ser. B*, **75**, 769–790. MR3091658

Balabdaoui, F. and Rufibach, K. (2008) A second Marshall inequality in convex estimation. *Statist. Probab. Lett.*, **78**, 118–126. MR2382064

Balabdaoui, F., Rufibach, K. and Wellner, J. A. (2009) Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.*, **37**, 1299–1331. MR2509075

Balabdaoui, F. and Wellner, J. A. (2007) Estimation of a k-monotone density: limit distribution theory and the spline connection. *Ann. Statist.*, **35**, 2536–2564. MR2382657

Balabdaoui, F. and Wellner, J. A. (2014) Chernoff's density is log-concave. *Bernoulli*, **20**, 231–244. MR3160580

Cai, T. T. and Low, M. (2015) A framework for estimation of convex functions. *Statist. Sinica.*, **25**, 423–456. MR3379081

Cator, E. (2011) Adaptivity and optimality of the monotone least-squares estimator. *Bernoulli*, **17**, 714-735. MR2787612

Carolan, C. and Dykstra, R. (1999) Asymptotic behavior of the Grenander estimator at density flat regions. *Canad. J. Statist.*, **27**, 557–566. MR1745821

Chen, Y. and Samworth, R. J. (2013) Smoothed log-concave maximum likelihood estimation with applications. *Statist. Sinica.*, **23**, 1373–1398. MR3114718

Chen, Y. and Samworth, R. J. (2016) Generalized additive and index models with shape constraints. *J. R. Stat. Soc. Ser. B.*, to appear. Available at http://arxiv.org/abs/1404.2957.

Csörgő, M. and Horvath, L. (1993) Weighted approximations in probability and statistics. John Wiley & Sons Inc., Chichester, England. MR1215046

Cule, M. and Samworth, R. (2010) Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron. J. Statist.*, **4**, 254–270. MR2645484

Doss, C. and Wellner, J. A. (2016) Global rates of convergence of the MLEs of log-concave and s-concave densities. *Ann. Statist.*, to appear. Available at http://arxiv.org/abs/1306.1438.

Dudley, R. M. (1999) Uniform central limit theorems. Cambridge University Press, Cambridge, U.K. MR1720712

Dümbgen, L., Rufibach, K. and Wellner, J. A. (2007) Marshall's lemma for convex density estimation. In *Asymptotics: particles, processes and inverse problems: festschrift for Piet Groeneboom* (E. Cator, G. Jongbloed, C. Kraaikamp, R. Lopuhaä and J.A. Wellner, eds.), 101–107. Institute of Mathematical Statistics, Ohio. MR2459933

Dümbgen, L., Samworth, R. and Schuhmacher, D. (2011) Approximation by log-concave distributions with applications to regression. *Ann. Statist.*, **39**, 702-730. MR2816336

Gao, F. and Wellner, J. A. (2009) On the rate of convergence of the maximum likelihood estimator of a k-monotone density. *Science in China Series A: Mathematics*, **52**, 1–14. MR2520591

Groeneboom, P. (1985) Estimating a monotone density. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer*, 539–555, Wadsworth, Belmont, CA. MR0822052

Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2001a) A canonical process for estimation of convex functions: the "invelope" of integrated Brownian motion $+t^4$. *Ann. Statist.*, **29**, 1620–1652. MR1891741

Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2001b) Estimation of a convex function: characterizations and asymptotic theory. *Ann. Statist.*, **29**, 1653–1698. MR1891742

Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2008) The support reduction algorithm for computing nonparametric function estimates in mixture models. *Scand. J. Statist.*, **35**, 385-399. MR2446726

Groeneboom, P. and Pyke, R. (1983) Asymptotic normality of statistics based on the convex minorants of empirical distribution functions. *Ann. Prob.*, **11**, 328–345. MR0690131

Guntuboyina, A. and Sen, B. (2013) Covering numbers for convex functions. *IEEE Trans. Inf. Theory*, **59**, 1957–1965. MR3043776

Guntuboyina, A. and Sen, B. (2015) Global risk bounds and adaptation in univariate convex regression. *Prob. Theory Related Fields*, **163**, 379–411. MR3405621

Hanson, D.L. and Pledger, G. (1976) Consistency in concave regression. *Ann. Statist.*, **4**, 1038–1050. MR0426273

Hildreth, C. (1954) Point estimators of ordinates of concave functions. *J. Amer. Statist. Assoc.*, **49**, 598–619. MR0065093

Lachal, A. (1997) Local asymptotic classes for the successive primitives of Brownian motion *Ann. Prob.*, **25**, 1712–1734. MR1487433

Jankowski, H. (2014) Convergence of linear functionals of the Grenander estimator under misspecification. *Ann. Statist.*, **42**, 625–653. MR3210981

Kim, J. and Pollard, D. (1990) Cube root asymptotics. *Ann. Statist.*, **18**, 191–219. MR1041391

Kim, A. K. H. and Samworth, R. J. (2014) Global rates of convergence in log-concave density estimation. Available at `http://arxiv.org/abs/1404.2298`.

Mammen, E. (1991) Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.*, **19**, 741–759. MR1105842

Meyer, M. C. (2003). A test for linear versus convex regression function using shape-restricted regression. *Biometrika*, **90**, 223–232. MR1966562

Meyer, M. C. (2013). Semi-parametric additive constrained regression. *J. Nonparametric Statist.*, **25**, 715–743. MR3174293

Rockafellar, R. T. (1997) *Convex analysis.* Princeton University Press, Princeton, NJ. MR1451876

Shorack, G. R. and Wellner, J. A. (1986) *Empirical processes with applications to statistics.* John Wiley & Sons Inc, New York, NY. MR0838963

Sen, B. and Meyer, M. C. (2013). Testing against a parametric regression function using ideas from shape restricted estimation. Available at `http://arxiv.org/abs/1311.6849`.