# On the Hermite spline conjecture and its connection to $k$-monotone densities

**Fadoua Balabdaoui**

*Centre de Recherche en Mathématiques de la Décision*
*Université Paris-Dauphine, Paris, France*
*e-mail:* fadoua@ceremade.dauphine.fr

**Simon Foucart***

*Department of Mathematics, Drexel University*
*Philadelphia, PA 19104*
*e-mail:* foucart@math.drexel.edu

**and**

**Jon A. Wellner**†

*Department of Statistics, University of Washington*
*Seattle, WA 98195-4322*
*e-mail:* jaw@stat.washington.edu

**Abstract:**
The $k$-monotone classes of densities defined on $(0, \infty)$ have been known in the mathematical literature but were for the first time considered from a statistical point of view by Balabdaoui and Wellner (2007) and Balabdaoui and Wellner (2010). In these works, the authors generalized the results established for monotone ($k = 1$) and convex ($k = 2$) densities by giving a characterization of the Maximum Likelihood and Least Square estimators (MLE and LSE) and deriving minimax bounds for rates of convergence. For $k \geq 3$, the pointwise asymptotic behavior of the MLE and LSE studied by Balabdaoui and Wellner (2007) would show that the MLE and LSE attain the minimax lower bounds in a local pointwise sense. However, the theory assumes that a certain conjecture about the approximation error of a Hermite spline holds true. The main goal of the present note is to show why such a conjecture cannot be true. We also suggest how to bypass the conjecture and rebuilt the key proofs in the limit theory of the estimators.

**Keywords and phrases:** conjecture, asymptotic distribution, Hermite spline, $k$-monotone.

## 1. Introduction

For an integer $k \geq 1$, a density $g_0$ defined on $(0, \infty)$ is said to be $k$-monotone if it is nonincreasing when $k = 1$, and if $(-1)^j g_0^{(j)}$ is nonincreasing and convex for all $j \in \{0, \dots, k-2\}$ when $k \geq 2$. Considering the problem of estimating a

*Supported by NSF Grant DMS-1120622
†Supported in part by NSF Grants DMS-0804587 and DMS-1104832, by NI-AID grant 2R01 AI291968-04, and by the Alexander von Humboldt Foundation

density in one of these classes presents several interesting features. As shown in Balabdaoui and Wellner (2007) and Balabdaoui and Wellner (2010), both the MLE and LSE of a $k$-monotone density exist. These estimators generalize the Grenander estimator of a nonincreasing density ($k = 1$) and the MLE and LSE of a nonincreasing and convex density ($k = 2$) studied by Groeneboom et al. (2001).

While it is known that the Grenander estimator in the case $k = 1$ converges pointwise at the rate $n^{1/3}$ and that the MLE of a convex nonincreasing density converges pointwise at the rate $n^{2/5}$, the rate of convergence $n^{k/(2k+1)}$ for the MLE (or LSE) of a $k$-monotone density in the general case $k \geq 3$ studied in Balabdaoui and Wellner (2007) depends on a key conjecture which has not yet been verified. In fact we show here that the spline conjectures made in Balabdaoui and Wellner (2007) fail to hold. On the other hand, Gao and Wellner (2009) obtained a result concerning the global rate of convergence of the MLE of a $k$-monotone density for a general $k \geq 3$: they showed that the rate of convergence of the MLE with respect to the Hellinger metric is indeed $n^{k/(2k+1)}$.

The limit case $k = \infty$ corresponds to the intersection of all $k$-monotone classes, that is the class of completely monotone densities on $(0, \infty)$. The latter turns out to be equal to the the class of mixtures of Exponentials, a consequence of Bernstein–Widder theorem, see e.g. Cheney and Light (2009). The nonparametric MLE of a mixture of Exponentials was considered by Jewell (1982) who showed its consistency and developed an EM algorithm to compute the estimator. So far, there are no results available on the limit distribution of the completely monotone MLE. As noted in Balabdaoui and Wellner (2007), one natural approach seems to study the behavior of the MLE in the $k$-monotone class as $k \to \infty$ and $n \to \infty$. Such an approach requires evidently a deep understanding of the asymptotic behavior of the $k$-monotone MLE and of the distance between its knots.

For an arbitrary $k \geq 1$, the work of Balabdaoui and Wellner (2007) aims to give a general approach to derive the limit distribution of the MLE and LSE at a fixed point $x_0 > 0$. More precisely, their work can be seen as an extension of the approach used by Groeneboom et al. (2001) in convex estimation. Let $g_0$ denote the true $k$-monotone density. At $x_0$, and modulo the spline conjecture, it is shown that

$$\begin{pmatrix} n^{\frac{k}{2k+1}}(\bar{g}_n(x_0) - g_0(x_0)) \\ n^{\frac{k-1}{2k+1}}(\bar{g}_n^{(1)}(x_0) - g_0^{(1)}(x_0)) \\ \vdots \\ n^{\frac{1}{2k+1}}(\bar{g}_n^{(k-1)}(x_0) - g_0^{(k-1)}(x_0)) \end{pmatrix} \to_d \begin{pmatrix} c_0(x_0) H_k^{(k)}(0) \\ c_1(x_0) H_k^{(k+1)}(0) \\ \vdots \\ c_{k-1}(x_0) H_k^{(2k-1)}(0) \end{pmatrix}, \qquad (1)$$

where $\bar{g}_n$ is either the MLE or LSE, and

$$c_j(x_0) = \left\{ g_0(x_0)^{k-j} \left( \frac{(-1)^k g_0^{(k)}(x_0)}{k!} \right)^{2j+1} \right\}^{\frac{1}{2k+1}},$$

for $j = 0, \ldots, k-1$. Note that the constants $c_j(x_0)$, $j = 0, \ldots, k-1$, appear also in the asymptotic minimax lower bound for $L_1$ risk (see Balabdaoui and Wellner (2010)). Let

$$Y_k(t) = \begin{cases} \int_0^t \frac{(t-s)^{k-1}}{(k-1)!} dW(s) + \frac{(-1)^k k!}{(2k)!} t^{2k}, & t \geq 0 \\ \int_t^0 \frac{(t-s)^{k-1}}{(k-1)!} dW(s) + \frac{(-1)^k k!}{(2k)!} t^{2k}, & t < 0 \end{cases}$$

where $W$ is a two-sided Brownian motion on $\mathbb{R}$. The process $H_k$ appearing in the limit (1) is characterized by the following conditions:

(i) The process $H_k$ stays everywhere above the process $Y_k$:

$$H_k(t) \geq Y_k(t), \qquad t \in \mathbb{R}.$$

(ii) $(-1)^k H_k$ is $2k$-convex, i.e., $(-1)^k H_k^{(2k-2)}$ exists and is convex.

(iii) The process $H_k$ satisfies

$$\int_{-\infty}^{\infty} (H_k(t) - Y_k(t))\, dH_k^{(2k-1)}(t) = 0.$$

(iv) If $k$ is even, $\lim_{|t|\to\infty}(H_k^{(2j)}(t) - Y_k^{(2j)}(t)) = 0$ for $j = 0, \ldots, (k-2)/2$; if $k$ is odd, $\lim_{t\to\infty}(H_k(t) - Y_k(t)) = 0$ and $\lim_{|t|\to\infty}(H_k^{(2j+1)}(t) - Y_k^{(2j+1)}(t)) = 0$ for $j = 0, \ldots, (k-3)/2$.

Because there is so far no device equivalent to the switching relationship device used in the monotone problem (see e.g. Groeneboom (1985) and also Balabdaoui et al. (2011)), the proof by Groeneboom et al. (2001) of the limit of the convex estimators is more complex and built in several steps. One of the most crucial pieces of this proof is that the stochastic order $n^{-1/5}$ for the distance between two knot points of the estimators in a small neighborhood of $x_0$. This result holds true under the assumption that the true convex density $g_0$ is twice continuously differentiable in a neighborhood of $x_0$ such that $g_0''(x_0) > 0$.

In the monotone problem, one can also show that the distance between the jump points of the Grenander estimator is stochastically bounded above by $n^{-1/3}$ provided that $g_0$ is continuously differentiable in a neighborhood of $x_0$ such that $g_0'(x_0) < 0$. These working assumptions can be naturally put in the following general form: the true $k$-monotone density is $k$-times continuously differentiable in a neighborhood of $x_0$ such that $(-1)^k g_0^{(k)}(x_0) > 0$. Thus, it seems natural that $n^{-1/(2k+1)}$ gives the general stochastic order for all integers $k \geq 1$. As noted in Balabdaoui and Wellner (2007), Mammen and van de Geer (1997) have, in the context of fitting a regression curve via splines, already conjectured that $n^{-1/(2k+1)}$ is the order of the distance between the knot points of their regression spline under the assumption that the true regression curve satisfies our same working assumptions.

In the extension of the argument of Groeneboom et al. (2001) to an arbitrary $k$, we have found that there is a need to show that an envelope of a certain

VC-class is bounded. In the next section, we describe this fact more precisely, and give the connection to our two spline conjectures made in Balabdaoui and Wellner (2007). In Section 3, we show that these conjectures are false for $k = 3$. The argument can be generalized to $k \geq 4$ but the calculations rapidly become cumbersome. In Section 4, we give a number of suggestions for building an alternative proof for the limit theory of the $k$-monotone estimators.

## 2. Connection to splines and the conjectures

We begin with some notation. For integers $m \geq 0$ and $p \geq 1$, let us denote by $\mathcal{S}_m(a_1, a_2, ..., a_p)$ the space of splines on $[a, b]$ of degree $m$ and internal knots $a_1 < \cdots < a_p$. The points $a$ and $b$ can be seen as external knots and will be denoted by $a_0$ and $a_{p+1}$, respectively. Let $f$ be a differentiable function on $[a, b]$ (differentiable on $(a, b)$ and to the right and left of $a$ and $b$, respectively). If $m = 2k - 1$, $p = 2k - 4$, and $a < a_1 < \cdots < a_{2k-4} < b$, we know that there exists a unique (Hermite) spline $H_k \in \mathcal{S}_{2k-1}(a_1, a_2, ..., a_{2k-4})$ satisfying

$$H_k(a_j) = f(a_j) \quad \text{and} \quad H_k'(a_j) = f'(a_j), \quad \text{for } j = 0, \ldots, 2k-3.$$

Note that for $k = 2$ the Hermite spline reduces to the cubic polynomial interpolating $f$ at $a$ and $b$. We denote by $\mathcal{H}_k$ the spline interpolation operator which assigns to $f$ its spline interpolant $H_k$.

Let $\tilde{g}_n$ be the LSE of the true $k$-monotone density $g_0$ based on $n$ i.i.d. random variables $X_1, \cdots, X_n$. It was shown by Balabdaoui and Wellner (2010) that $\tilde{g}_n$ exists, is unique, and is a spline of degree $k-1$. Let $\tilde{H}_n$ denote its $k$-fold integral, that is

$$\tilde{H}_n(x) = \frac{1}{(k-1)!} \int_0^x (x-t)^{k-1} \tilde{g}_n(t) dt.$$

The function $\tilde{H}_n$ is important due to its direct involvement in the characterization of the estimator $\tilde{g}_n$. More precisely, if we consider the $(k-1)$-fold integral of the empirical distribution $\mathbb{G}_n$

$$\mathbb{Y}_n(x) = \frac{1}{(k-1)!} \int_0^x (x-t)^{k-1} d\mathbb{G}_n(t),$$

then the spline $\tilde{g}_n$ of degree $k-1$ is the LSE if and only if the following (Fenchel) conditions hold

$$\begin{aligned} \tilde{H}_n(x) &\geq \mathbb{Y}_n(x), \quad \text{for all } x \geq 0, \\ \tilde{H}_n(x) &= \mathbb{Y}_n(x), \quad \text{if } x \text{ is knot of } \tilde{g}_n. \end{aligned} \tag{2}$$

The greater focus put on the LSE is explained by the fact that the characterization in (2) is much simpler to study, especially when the empirical processes involved are localized (see Balabdaoui and Wellner (2007)). However, it was shown by Balabdaoui and Wellner (2007) that understanding the asymptotics of the LSE is enough as one can use strong consistency of the MLE to linearize its characterization and put it in a more familiar form.

One of the key points in the study of the asymptotics is to note that the characterization of the LSE implies $\tilde{H}_n(\tau) = \mathbb{Y}_n(\tau)$ and $\tilde{H}'_n(\tau) = \mathbb{Y}'_n(\tau)$ for a knot $\tau$ of $\tilde{g}_n$. Furthermore, given $2k-2$ knots $\tau_0 < \cdots < \tau_{2k-3}$, $\tilde{g}_n$ is uniquely determined on $[\tau_0, \tau_{2k-3}]$ by the interpolation equalities $\tilde{H}_n^{(i)}(\tau_j) = \mathbb{Y}_n^{(i)}(\tau_j)$, $i = 0, 1$, $j = 0, \ldots, 2k-3$. In other words, $\tilde{H}_n$ is a Hermite spline interpolant of $\mathbb{Y}_n$, i.e.,

$$\tilde{H}_n(x) = \mathcal{H}_k[\mathbb{Y}_n](x) \qquad \text{for } x \in [\tau_0, \tau_{2k-3}].$$

Note that in any small neighborhood of the estimation point $x_0$, strong consistency of the $(k-1)$-st derivative of $\tilde{g}_n$ combined with the assumption that $g_0^{(k)}(x_0) \neq 0$ guarantee that the number of knots in that neighborhood tends to $\infty$ almost surely as $n \to \infty$. Hence, finding at least $2k-2$ knots is possible with probability one. At this stage, we know that $\tau_{2k-3} - \tau_0 \to 0$ almost surely as $n \to \infty$, and our goal is to show that this convergence occurs with a rate equal to $n^{-1/(2k+1)}$. In the next section, we describe briefly the key argument in the proof of Balabdaoui and Wellner (2007) and recall the two related spline conjectures.

### *2.1. The spline conjectures*

Take an arbitrary point $\bar{\tau} \in [\tau_0, \tau_{2k-3}]$ such that $\bar{\tau} \notin \{\tau_0, \cdots, \tau_{2k-3}\}$. By the inequality in (2), we have that

$$\mathcal{H}_k[\mathbb{Y}_n](\bar{\tau}) \geq \mathbb{Y}_n(\bar{\tau}).$$

If $Y$ denotes the population counterpart of $\mathbb{Y}_n$, i.e., the $(k-1)$-fold integral of $g_0$

$$Y(x) = \frac{1}{(k-1)!} \int_0^x (x-t)^{k-1} g_0(t)dt,$$

then the latter inequality can be rewritten in the more useful form

$$[\mathcal{H}_k Y - Y](\bar{\tau}) \geq \mathcal{H}_k[Y - \mathbb{Y}_n](\bar{\tau}) - [Y - \mathbb{Y}_n](\bar{\tau}). \tag{3}$$

Both sides of the inequality can be recognized as the Hermite interpolation errors corresponding to the interpolated functions $Y$ and $Y - \mathbb{Y}_n$. While $Y$ is $(2k)$-times differentiable on $[\tau_0, \tau_{2k-3}]$ under our working assumptions, the function $Y - \mathbb{Y}_n$ is only $(k-2)$-times continuously differentiable since $\mathbb{Y}_n$ is the $(k-1)$-st fold integral of the (piecewise constant) empirical distribution function $\mathbb{G}_n$.

Taylor expansions of $Y$ and $\mathbb{Y}_n - Y$ up to the orders $2k$ and $k-1$, respectively, give yet another form for (3). On $[\tau_0, \tau_{2k-3}]$, consider the functions

$$f_0(x) = \frac{x^{2k}}{(2k)!}, \qquad b_u(x) = \frac{(x-u)_+^{k-1}}{(k-1)!}, \quad u \in (\tau_0, \tau_{2k-3}),$$

$$r(x) = \frac{1}{(2k-1)!} \int_{\bar{\tau}}^{\tau_{2k-3}} (x-t)_+^{2k-1} (g_0^{(k)}(t) - g_0^{(k)}(\bar{\tau}))dt.$$

Let $e_k = f_0 - \mathcal{H}_k f_0$ be the error associated with Hermite interpolation of $f_0$. Then, (3) is equivalent to

$$g_0^{(k)}(\bar\tau) e_k(\bar\tau) \le \mathbb{E}_n + \mathbb{R}_n$$

where, with $G_0$ denoting the c.d.f. of $g_0$,

$$\mathbb{E}_n = \int_{\tau_0}^{\tau_{2k-3}} \mathcal{H}_k[b_u](\bar\tau) d(\mathbb{G}_n(u) - G_0(u)) \qquad \text{and} \qquad \mathbb{R}_n = \mathcal{H}_k[r](\bar\tau).$$

Recalling that $(-1)^k g_0^{(k)}(x_0) > 0$, so that $(-1)^k g_0^{(k)}$ is positive on a neighborhood $[x_0 - \delta, x_0 + \delta]$ for some $\delta > 0$, (3) can also be rewritten as

$$(-1)^k g_0^{(k)}(\bar\tau)(-1)^k e_k(\bar\tau) \le \mathbb{E}_n + \mathbb{R}_n.$$

The term $\mathbb{E}_n$ is an empirical process indexed by the class of functions $h$ such that

$$h(u) = h_{s,s_0,\ldots,s_{2k-3}}(u) = \mathcal{H}_k[b_u](s) 1_{[s_0,s_{2k-3}]}(u),$$

for some $s_0 < \cdots < s_{2k-3}$ in $[x_0 - \delta, x_0 + \delta]$ and $s \in (s_0, s_{2k-3})$. Here $\mathcal{H}_k[f]$ denotes the Hermite spline interpolating $f$ at $s_j, j = 0, \cdots, 2k-3$. The second term $\mathbb{R}_n$ is equal to the interpolation error corresponding to the $(2k)$-times differentiable function $r$. The main goals are: (a) find upper stochastic bounds for $\mathbb{E}_n$ and $\mathbb{R}_n$; (b) find a lower bound for $(-1)^k e_k(\bar\tau)$ as a function of a power of the distance $\tau_{2k-3} - \tau_0$.

In the absence of any knowledge about the location and distribution of the random knots $\tau_0, \ldots, \tau_{2k-3}$, it seems naturally desirable to get of rid of any dependency on these points. This motivates the assumption that the interpolation error is uniformly bounded independently of the knots. Thus, the following conjectures were formulated in Balabdaoui and Wellner (2007) to tackle (a).

**Conjecture 1** *Let $a = 0$, $b = 1$, and $b_t(x) = (x-t)_+^{k-1}/(k-1)!$ for $t \in (0,1)$. There exists a constant $d_k > 0$ such that*

$$\sup_{t \in (0,1)} \sup_{0 < y_1 < \cdots < y_{2k-4} < 1} \|b_t - \mathcal{H}_k b_t\|_\infty \le d_k. \tag{4}$$

**Conjecture 2** *Let $a = 0$ and $b = 1$. Then there exists a constant $c_k > 0$ such that, for any $f \in C^{(2k)}[0,1]$,*

$$\sup_{0 < y_1 < \cdots < y_{2k-4} < 1} \|f - \mathcal{H}_k f\|_\infty \le c_k \|f^{(2k)}\|_\infty. \tag{5}$$

Note that Conjecture 1 cannot hold if Conjecture 2 does not: indeed, in view of the Taylor expansion

$$f(x) = \sum_{j=0}^{2k-1} f^{(j)}(0) \frac{x^j}{j!} + \int_0^1 f^{(2k)}(t) \frac{(x-t)_+^{2k-1}}{(2k-1)!} dt$$

and of the fact that polynomials of degree $\leq 2k-1$ are preserved by $\mathcal{H}_k$, we observe that (4) implies (5) with $c_k = d_{2k}$.

Let us fix $s_0$ and $R > 0$ such that $[s_0, s_0 + R] \subset [x_0 - \delta, x_0 + \delta]$. Conjecture 1 implies that the class

$$\mathcal{F}_{s_0,R} = \{h_{s,s_0,\ldots,s_{2k-3}} : [s_0, s_{2k-3}] \subset [s_0, s_0 + R] \subset [x_0 - \delta, x_0 + \delta]\}$$

admits a finite envelope, e.g.

$$F_{s_0,R}(x) = a_k R^{k-1} 1_{[s_0,s_0+R]}(x)$$

where $a_k > 0$ is a constant depending only on $k$ (through $d_k$). Together with the fact that the class $\mathcal{F}_{s_0,R}$ is a VC-subgraph, this gives one of the most crucial results that helps establishing the stochastic order of the gap: the "right" stochastic bound

$$\mathbb{E}_n = O_p\big(n^{-2k/(2k+1)}\big) + o_p\big((\tau_{2k-3} - \tau_0)^{2k}\big). \tag{6}$$

On the other hand, the term $\mathbb{R}_n$ could be bounded using Conjecture 2. Since $\mathbb{R}_n$ is $(2k)$-times continuously differentiable on a neighborhood of $x_0$, Conjecture 2 yields

$$\mathbb{R}_n = o_p\big((\tau_{2k-3} - \tau_0)^{2k}\big). \tag{7}$$

It follows that

$$\sup_{\bar{\tau} \in [\tau_{j_0}, \tau_{j_0+1}]} (-1)^k e_k(\bar{\tau}) \leq O_p\big(n^{-2k/(2k+1)}\big) + o_p\big((\tau_{2k-3} - \tau_0)^{2k}\big),$$

where $[\tau_{j_0}, \tau_{j_0+1}]$ is a largest knot interval among $[\tau_j, \tau_{j+1}], j = 0, \ldots, 2k-4$.

At this stage of the argument, the stochastic order of the gap can be shown to be $n^{-1/(2k+1)}$ if there exists $M > 0$ such that

$$\sup_{\bar{\tau} \in [\tau_{j_0}, \tau_{j_0+1}]} (-1)^k e_k(\bar{\tau}) > M(\tau_{2k-3} - \tau_0)^{2k}.$$

This can be shown using some known results on monosplines and Chebyshev polynomials (see Balabdaoui and Wellner (2005, 2006, 2007)).

Conjecturing boundedness of the Hermite spline interpolant was a crucial assumption to obtain the right stochastic bound for the empirical process $\mathbb{E}_n$ and the remainder term $\mathbb{R}_n$. However, this boundedness served only as a sufficient condition. In the next section, we show that Conjecture 2 (hence Conjecture 1) is in fact answered negatively.

## 3. Unboundedness of the Hermite interpolation error

We now prove that the statement of Conjecture 2 (and even a weaker statement where $c_k$ would be allowed to depend on $f$) is violated for the function $f = S_*$ defined by

$$S_*(t) = S_*(t; \tau_1, \ldots, \tau_{2k-4}) = \frac{1}{(2k)!}\left(t^{2k} + 2\sum_{i=1}^{2k-4}(-1)^i (t - \tau_i)_+^{2k}\right). \tag{8}$$

This choice is dictated by the fact (not necessary here, so not proven) that, for $0 = \tau_0 < \tau_1 < \cdots < \tau_{2k-4} < \tau_{2k-3} = 1$ and for $t \in [0,1]$,

$$\sup_{f \in W_\infty^{2k},\, \|f^{(2k)}\|_\infty \le 1} \Big| [\mathcal{H}_k f](t) - f(t) \Big| = \Big| [\mathcal{H}_k S^*](t) - S^*(t) \Big|. \tag{9}$$

Setting $\mathcal{E}_k := \mathcal{H}_k(S_*) - S_*$, the Landau–Kolmogorov inequality (see e.g. Kolmogorov (1939) or Schoenberg (1973)) guarantees the existence of a constant $D_k > 0$ depending only on $k$ such that

$$\|\mathcal{E}_k^{(2k-1)}\|_\infty \le D_k \; \|\mathcal{E}_k\|_\infty^{\frac{1}{2k}} \; \|\mathcal{E}_k^{(2k)}\|_\infty^{\frac{2k-1}{2k}}.$$

Since $\mathcal{E}_k^{(2k)} = -S_*^{(2k)}$ alternates between $+1$ and $-1$, so that $\|\mathcal{E}_k^{(2k)}\|_\infty = 1$, it follows that if Conjecture 2 was true, then $\|\mathcal{E}_k^{(2k-1)}\|_\infty$ would be bounded independently of the knots. Studying the latter turns out to be easier than studying $\|\mathcal{E}_k\|_\infty$ itself, as $\mathcal{E}_k^{(2k-1)}$ is a piecewise linear function (not necessarily continuous at the knots) whose slope alternates between $+1$ and $-1$. Let us note that $\mathcal{E}_k$ belongs to the space

$$\Omega_k(\tau_1, \cdots, \tau_{2k-4}) = \Big\{ \gamma S_*(t) + s(t),\ \gamma \in \mathbb{R},\ s \in \mathcal{S}_{2k-1}(\tau_1, \cdots, \tau_{2k-4}) \Big\},$$

which is a $(4k-3)$-dimensional weak Chebyshev space (see e.g. Lemma 1 in Bojanov and Naidenov (2002)). Let us also note that $\mathcal{E}_k$ has double zeros occurring at the knots $\tau_0, \tau_1, \cdots, \tau_{2k-4}, \tau_{2k-3}$. Since $4k-4$ is the maximal number of zeros for a nonzero function in a weak Chebyshev space of dimension $4k-3$, there exists a constant $C \in \mathbb{R}$ such that, for all $t \in [0,1]$, $\mathcal{E}_k(t)$ equals

$$C \begin{vmatrix} B_1(0) & B_1'(0) & \cdots & \cdots & B_1(1) & B_1'(1) & B_1(t) \\ B_2(0) & B_2'(0) & \cdots & \cdots & B_2(1) & B_2'(1) & B_2(t) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ B_{4k-3}(0) & B_{4k-3}'(0) & \cdots & \cdots & B_{4k-3}(1) & B_{4k-3}'(1) & B_{4k-3}(t) \end{vmatrix},$$

where $(B_1, \ldots, B_{4k-3})$ is any basis for $\Omega_k(\tau_1, \cdots, \tau_{2k-4})$. The value of $C$ is determined by $\mathcal{E}_k^{(2k)}(t) = -1$ for $0 \le t < \tau_1$. Our objective is now to prove the unboundedness of $\|\mathcal{E}_k^{(2k-1)}\|_\infty$, which we do in the particular case $k = 3$. We consider the basis for $\Omega(\tau_1, \tau_2)$ (which has dimension 9) given by

$$\big(1, t, t^2, t^2(t-\tau_1), t^2(t-\tau_1)^2, t^2(t-\tau_1)^2(t-\tau_2), (t-\tau_1)_+^5, (t-\tau_2)_+^5, S_*(t)\big).$$

The determinantal expression of $\mathcal{E}_3(t)$ can be explicitly written as

$$\mathcal{E}_3(t) = C \left|\begin{array}{cc|ccccc|c} 1 & 0 & \mathrm{x} & \cdots & \cdots & \cdots & \mathrm{x} & 1 \\ 0 & 1 & \mathrm{x} & \cdots & \cdots & \cdots & \mathrm{x} & t \\ \hline 0 & 0 & & & & & & t^2 \\ 0 & 0 & & & & & & t^2(t-\tau_1) \\ \vdots & \vdots & & & & & & t^2(t-\tau_1)^2 \\ \vdots & \vdots & & & D & & & t^2(t-\tau_1)^2(t-\tau_2) \\ \vdots & \vdots & & & & & & (t-\tau_1)_+^5 \\ 0 & 0 & & & & & & (t-\tau_2)_+^5 \\ 0 & 0 & & & & & & S_*(t) \end{array}\right|, \tag{10}$$

where $D$ is the $7 \times 6$ matrix

$$\begin{bmatrix} \tau_1^2 & 2\tau_1 & \tau_2^2 & 2\tau_2 & 1 & 2 \\ 0 & \tau_1^2 & \tau_2^2(\tau_2-\tau_1) & \tau_2(3\tau_2-2\tau_1) & 1-\tau_1 & 3-2\tau_1 \\ 0 & 0 & \tau_2^2(\tau_2-\tau_1)^2 & p(\tau_1,\tau_2) & (1-\tau_1)^2 & 2(1-\tau_1)(2-\tau_1) \\ 0 & 0 & 0 & \tau_2^2(\tau_2-\tau_1)^2 & (1-\tau_1)^2(1-\tau_2) & q(\tau_1,\tau_2) \\ 0 & 0 & (\tau_2-\tau_1)^5 & 5(\tau_2-\tau_1)^4 & (1-\tau_1)^5 & 5(1-\tau_1)^4 \\ 0 & 0 & 0 & 0 & (1-\tau_2)^5 & 5(1-\tau_2)^4 \\ \frac{\tau_1^6}{6!} & \frac{\tau_1^5}{5!} & \frac{\tau_2^6-2(\tau_2-\tau_1)^6}{6!} & \frac{\tau_2^5-2(\tau_2-\tau_1)^5}{5!} & \frac{1-2(1-\tau_1)^6+2(1-\tau_2)^6}{6!} & \frac{1-2(1-\tau_1)^5+2(1-\tau_2)^5}{5!} \end{bmatrix},$$

$p(\tau_1,\tau_2) = 2\tau_2(\tau_2-\tau_1)(2\tau_2-\tau_1)$, $q(\tau_1,\tau_2) = (1-\tau_1)(2(1-\tau_2)(2-\tau_1)+1-\tau_1)$. Taking the 5th derivative in (10) and expanding along the last columns yields, for $0 \le t < \tau_1$,

$$\mathcal{E}_3^{(5)}(t) = C(-5!\delta_1 + \delta_2 t),$$

where $\delta_1$ and $\delta_2$ are the determinants of the submatrices of $D$ obtained by removing the fourth row and the last row, respectively. From $\mathcal{E}_3^{(6)}(t) = -1$ for $0 \le t < \tau_1$, we derive $C = -1/\delta_2$, and in turn

$$\mathcal{E}_3^{(5)}(0) = 120\,\frac{\delta_1}{\delta_2}.$$

In the case $\tau_2 = 2\tau_1$, an explicit calculation (facilitated by a computer algebra software) reveals that

$$\delta_1 = \frac{1}{360}\tau_1^{12}(1-2\tau_1)^6(4-32\tau_1+189\tau_1^2-312\tau_1^3+159\tau_1^4),$$
$$\delta_2 = 4\tau_1^{13}(1-2\tau_1)^6(1-\tau_1)(7-5\tau_1).$$

Thus, as $\tau_1 \to 0$, we have

$$\mathcal{E}_3^{(5)}(0) \sim 120\frac{4\tau_1^{12}/360}{28\tau_1^{13}} = \frac{1}{21\tau_1} \to +\infty.$$

This shows that Conjecture 2 does not hold. □

## 4. Alternative arguments

Although the results of Section 3 show that the methods of proof used in Balabdaoui and Wellner (2007) (which are heavily based on the methods used in Kim and Pollard (1990) and Groeneboom et al. (2001)) do not suffice for proving the desired rate results as stated there, we continue to believe that the rate will be $n^{-1/(2k+1)}$ for the "gap conjecture" of Balabdaoui and Wellner (2007), and $n^{k/(2k+1)}$ for the MLE of the $k$-monotone density $f_0$. Here we sketch several possible routes toward proofs of these conjectured results.

### *4.1. Option A: lower bound for the gaps*

Note that the arguments in the preceding section showing unboundedness of the envelope of the interpolation error relied on taking $\tau_2 = 2\tau_1$ so that $\tau_2 - \tau_1 = \tau_1 \to 0$ where the $\tau$'s are regarded as parameters or variables indexing the entire class of interpolation errors for a scaling of the problem with

$$0 \equiv \tau_0 < \tau_1 < \cdots < \tau_{2k-4} < \tau_{2k-3} \equiv 1.$$

Thus a "coalesence" of the knots leads to failure of the conjectures made in Balabdaoui and Wellner (2007).

On the other hand, on the original time scale for the (random!) knots $\tau_0 < \tau_1 < \cdots < \tau_{2k-3}$ we want to show that $\tau_{2k-3} - \tau_0 = O_p(n^{-1/(2k+1)})$. It seems likely that these random knots for the LSE actually do not "coalesce", but stay bounded away from each other asymptotically (at the rate $n^{-1/(2k+1)}$), and hence we expect to have

$$\max_{1 \le j \le 2k+3} \frac{1}{(\tau_j - \tau_{j-1})} = O_p(n^{1/(2k+1)}), \tag{11}$$

or, equivalently

$$\max_{1 \le j \le 2k+3} \frac{1}{n^{1/(2k+1)}(\tau_j - \tau_{j-1})} = O_p(1). \tag{12}$$

If we could show that (12) holds, then the classes of functions involved in the interpolation errors could be restricted to classes involving separated knots and the conjectures may be more plausible for these restricted classes.

### *4.2. Option B: alternative inequalities*

While the methods of proof used in Balabdaoui and Wellner (2007) (and Kim and Pollard (1990), Groeneboom et al. (2001)) are based on empirical process inequalities which rely on the small or scaling properties of envelopes (see e.g. Lemma A.1, page 2560 of Balabdaoui and Wellner (2007) or Lemma 4.1 of Kim and Pollard (1990)), as opposed to smallness of the individual functions in the class relative to an envelope as in Lemmas 3.4.2 and 3.4.3 of van der Vaart and Wellner (1996) or van der Vaart and Wellner (2011). On the other hand the proofs of the (global) rate of convergence of Hellinger distance from the MLE $\widehat{f}_n$ to $f_0$ established in Gao and Wellner (2009) rely on inequalities for suprema of empirical processes based on uniform or bracketing entropy for function classes in which the $L_2-$norms of individual functions are small relative to envelope functions (which may possibly be unbounded) (see e.g. Lemmas 3.4.2 and 3.4.3 and Theorem 3.4.4 of van der Vaart and Wellner (1996) for bracketing entropy type bounds, and see van der Vaart and Wellner (2011) for classes with well behaved uniform entropy bounds; the results of Giné and Koltchinskii (2006) might also be helpful in connection with the latter classes).

Thus there is some possibility that alternative inequalities for suprema of the empirical processes involved may be needed in establishing the desired rate results when $k \geq 3$.

### *4.3. Option C: alternative inequalities involving "weak parameters"*

While the inequalities discussed in option B above require application of empirical process inequalities involving "strong parameters" such as the expected values of envelope functions, there remains some possibility for the development of new inequalities based on "weak parameters"; see e.g. the discussion on page 51 of Massart (2007) and the material on page 209 of Boucheron et al. (2013). This option is the most speculative of the three.

## References

Balabdaoui, F., Jankowski, H., Pavlides, M., Seregin, A. and Wellner, J. (2011). On the Grenander estimator at zero. *Statist. Sinica* **21** 873–899.

Balabdaoui, F. and Wellner, J. (2005). Conjecture of error boundedness in a new Hermite interpolation problem via splines of odd-degree. Tech. Rep. 480, University of Washington, Department of Statistics.

Balabdaoui, F. and Wellner, J. (2006). Estimation of a $k$-monotone density: limit distribution theory and the spline connection, with complete proofs. Tech. Rep. 503, University of Washington, Department of Statistics.

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.

Balabdaoui, F. and Wellner, J. A. (2010). Estimation of a $k$-monotone density: characterizations, consistency and minimax lower bounds. *Stat. Neerl.* **64** 45–70.

Bojanov, B. and Naidenov, N. (2002). Exact Markov-type inequalities for oscillating perfect splines. *Constr. Approx.* **18** 37–59.

Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence.* Oxford University Press.

Cheney, W. and Light, W. (2009). *A course in approximation theory*, vol. 101 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI. Reprint of the 2000 original.

Gao, F. and Wellner, J. A. (2009). On the rate of convergence of the maximum likelihood estimator of a $k$-monotone density. *Sci. China Ser. A* **52** 1525–1538.

Giné, E. and Koltchinskii, V. (2006). Concentration inequalities and asymptotic results for ratio type empirical processes. *Ann. Probab.* **34** 1143–1216.

Groeneboom, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*. Wadsworth Statist./Probab. Ser., Wadsworth, Belmont, CA.

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

Jewell, N. P. (1982). Mixtures of exponential distributions. *Ann. Statist.* **10** 479–484.

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

Kolmogorov, A. N. (1939). On inequalities between the upper bounds of successive derivatives of an arbitrary function on an infinite interval. In *Amer. Math. Soc. Transl., Vols. 1,2, pp. 233-343*, vol. 30. Amer. Math. Soc., Providence, RI, 1962, 3–16.

Mammen, E. and van de Geer, S. (1997). Locally adaptive regression splines. *Ann. Statist.* **25** 387–413.

Massart, P. (2007). *Concentration inequalities and model selection*, vol. 1896 of *Lecture Notes in Mathematics*. Springer, Berlin. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

Schoenberg, I. J. (1973). The elementary cases of Landau's problem of inequalities between derivatives. *Amer. Math. Monthly* **80** 121–158.

van der Vaart, A. and Wellner, J. A. (2011). A local maximal inequality under uniform entropy. *Electron. J. Stat.* **5** 192–203.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics, Springer-Verlag, New York. With applications to statistics.