ON-LINE ESTIMATION OF A SMOOTH REGRESSION FUNCTION

KHASMINSKII, R. AND LIPTSER, R.

Abstract. The tracking (recursive) type estimator for one-dimensional regression estimation problem with equidistant design is proposed. It is proven that, out of inevitable initial layer, this estimator has the optimal rate of convergence of quadratic risk to zero if the sample size goes to infinity.

1. Introduction.

Let (t, X) be a pair of random variables, $t \in [0,1], X \in \mathbb{R}^1$, and $f(t) = E(X|t)$ be a regression function. There are two the most popular setting for estimation of $f(t)$. For the first setting (*random design*), statistician estimates $f(t)$ on the base of a sample $(t_1, X_1), ..., (t_n, X_n)$, where $(X_1, t_1), \dots, (X_n, t_n)$ are independent copies of (X, t), while for the second one (*equidistant design*) statistician uses has m independent "measurements" $X_{i1}, ...X_{im}$ for each $t_{in} = i/[n^\gamma],\ i = 1, 2, ..., [n^\gamma];\ m = n/[n^\gamma], 0 < \gamma \le 1$ and in additional

$$P(X_{ij} \in A|t = t_{in}) = P(X \in A|t = t_{in}).$$

For *nonparametric* regression estimation the function f is assumed to be belong to a class of functions Σ which cannot be specified by a finite number of parameters. In this paper, following to [9], [10], [4] and [5], we fix the class

$$\Sigma(\beta, L) = f \begin{cases} \text{obeys } k \text{ derivatives, } f^{(0)}, f^{(1)}, \dots, f^{(k)}; \\ |f^{(k)}(t_2) - f^{(k)}(t_1)| \le L|t_2 - t_1|^\alpha,\ \forall\ t_1, t_2 \text{ and } \alpha \in (0,1]; \\ \beta = k + \alpha. \end{cases}$$

It is well known from the above-mentioned citations that for both designs there are estimators $\widehat{f_n^{(j)}}(t)$, $j = 0, 1, \dots, k$ such that for a wide class of loss functions $\mathcal{L}(*)$

$$\sup_{f \in \Sigma(\beta, L)} E\mathcal{L}\Big(n^{\frac{\beta - j}{2\beta+1}} \|\widehat{f_n^{(j)}} - f^{(j)}\|_{L_p}\Big) < C,\ j = 0, 1, ..., k \tag{1.1}$$

(C is positive constant) and does not exists an estimator with the better rate of convergence to zero for estimation risks in $n \to \infty$ uniformly in $\sum(\beta, L)$. It is also known that the same rate in n holds for the estimation risk $E\big(\widehat{f_n^{(j)}}(t) - f^{(j)}(t)\big)^2$ under fixed value t and this rate cannot be exceeded uniformly on any nonempty open set from $(0,1)$. In this paper, we intend to analyse an on-line estimation problem of a regression function f. To clarify our method, we mention related on-line estimation problem for a signal $f(t)$ observed in additive Gaussian white

Date: 16/06/2001.
1991 *Mathematics Subject Classification.* 62G05; Secondary 62M99.
Key words and phrases. On-line tracking estimator, Equidistant design.
Partially supported by NSF Grant DMS 9971608.

noise of a small intensity (see, [2]) when as observation is an Itô process (with respect to Wiener process W_t)

$$X_t^\varepsilon = \int_0^t f(s)ds + \varepsilon W_t,\ t \in [0,1], \tag{1.2}$$

where ε is a small parameter. For $f \in \Sigma(\beta, L)$, the following on-line filter for tracking of $f(t) = f^{(0)}(t)$ and $f^{(j)}(t), j = 1, \dots, k$ was proposed in [2]:

$$\begin{aligned} d\widehat{f^{(j)}}(t) &= \widehat{f^{(j+1)}}(t)dt + \frac{q_j}{\varepsilon^{2(j+1)/(2\beta+1)}}\big(dX_t^\varepsilon - \widehat{f^{(0)}}(t)dt\big), \\ j &= 0,1,\dots,k-1 \\ d\widehat{f^{(k)}}(t) &= \frac{q_k}{\varepsilon^{2(k+1)/(2\beta+1)}}\big(dX_t^\varepsilon - \widehat{f^{(0)}}(t)dt\big), \end{aligned} \tag{1.3}$$

where $q_0, \dots, q_k$ are chosen so that all roots of the polynomial

$$p^k(u) = u^{k+1} + q_0 u^k + q_1 u^{k-1} + \dots + q_{k-1}u + q_k \tag{1.4}$$

have negative real parts. Out of an initial (boundary) layer $[0, \Delta^\varepsilon]$, a performance of that filter is characterized by the optimal in the minimax sense rates in $\varepsilon \to 0$ of the mean square errors (compare (1.1)):

$$\frac{E\big(f^{(j)}(t) - \widehat{f^{(j)}}(t)\big)^2}{\varepsilon^{4(\beta-j)/(2\beta+1)}} < C,\ j = 0,1,\dots,k. \tag{1.5}$$

In this paper, we consider rather natural (from application point of view) model for observation: under $t_{in} = \frac{i}{n},\ \ i = 0, \dots, n$

$$X_{in} = f(t_{in}) + \sigma(t_{in})\xi_{in}, \tag{1.6}$$

where $(\xi_{in})_{i \le n}$ is a sequence of i.i.d. random variables with $E\xi_{in} \equiv 0,\ \ E\xi_{in}^2 = 1$ and $\sigma^2(t_{in}) < C$. It is known from [1] and [3] that the regression estimation via the observations (1.6) is asymptotically equivalent in the Le Cam sense, [6] to the estimation f for the model (1.2) (in [1] under Gaussian distribution of the noise and in [3] under arbitrary one). In view of [1] and [3] it is natural to assume that an appropriate discrete time tracking estimator can be created on the analogy of (1.3) (hereafter for brevity t_{in} is replaced by t_i):

$$\begin{aligned} \widehat{f_n^{(j)}}(t_i) &= \widehat{f_n^{(j)}}(t_{i-1}) + \frac{1}{n}\widehat{f_n^{(j+1)}}(t_{i-1}) + \frac{q_j}{n^{\frac{(2\beta-j)}{2\beta+1}}}\big(X_i - \widehat{f_n^{(0)}}(t_{i-1})\big) \\ j &= 0,1,\dots,k-1 \\ \widehat{f_n^{(k)}}(t_i) &= \widehat{f_n^{(k)}}(t_{i-1}) + \frac{q_k}{n^{\frac{(2\beta-k)}{2\beta+1}}}\big(X_i - \widehat{f_n^{(0)}}(t_{i-1})\big) \end{aligned} \tag{1.7}$$

subject to some initial conditions $\widehat{f_n^{(0)}}(0), \widehat{f_n^{(1)}}(0), \dots, \widehat{f_n^{(k)}}(0)$. The estimator given in (1.7) is of on-line type. Its implementation is simpler than for Čentsov's projection estimator (see, e.g. [5]) or kernel type ones [8], [7]. An important and useful for many applications property of this estimator is its recurrent structure when the estimator for $f(t_{i+1})$ receives a small correction, with respect to the estimator for $f(t_i)$, based on new arrived observation.

The main goal of this paper is to show that, out of the boundary layer of a volume $Cn^{-\frac{1}{2\beta+1}}\log n$, the estimator given in (1.7) possesses the best possible rate in $n\to\infty$ of convergence to zero for the mean square error.

The paper is organized as follows. In Section 2 the main results are formulated. Section 3 contains some preliminaries. Upper bounds for the normalized bias and variance are given in Sections 4 and 5. The final part of the proof for the main result is given in Section 6.

2. Properties of estimator (1.7).

2.1. **Generic constant.** All results of this paper have possess an asymptotical character with respect to a large parameter n. As a result of that all positive constants independent of n are denoted by a generic letter C.

2.2. **Formulation of main results.**

Theorem 2.1. *Let $q_0,\dots,q_k$ are chosen such that all roots of the polynomial given in (1.4) are different and have negative real parts. Let the observation model is defined in (1.6), $f\in\Sigma(\beta,L)$ and $\sigma^2(t)<C$. Then the estimator (1.7) with arbitrary bounded initial conditions $\widehat{f_n^{(0)}}(0),\widehat{f_n^{(1)}}(0),\dots,\widehat{f_n^{(k)}}(0)$ possesses the property: for $t_\ell>Cn^{-\frac{1}{2\beta+1}}\log n$*

$$\sup_{f\in\sum(\beta,L)}\sum_{j=0}^{k}E\big(f^{(j)}(t_\ell)-\widehat{f_n^{(j)}}(t_\ell)\big)^2 n^{\frac{2(\beta-j)}{2\beta+1}}\le C. \tag{2.1}$$

Remark 1. As it was noticed in the Introduction, the rate of convergence of risks to zero for $n\to\infty$ in (2.1) is unimprovable. From the other hand, the boundary layer of order $n^{-\frac{1}{2\beta+1}}$, where the optimal rate fails, is inevitable. More precisely, the statement, completely analogous to Lemma 5.1, [2], can be proven.

Proposition 2.1. *Let X_{in} be defined in (1.6), $f\in\Sigma(\beta,L)$ and $t_{in}\in\Delta_n$, where*

$$\Delta_n=[0,\psi(n)n^{-\frac{1}{2\beta+1}}],\ \psi(n)\to 0,\ n\to\infty.$$

Then for any estimator $\{\widetilde{f_n^{(0)}}(t_{in}),\widetilde{f_n^{(1)}}(t_{in}),\dots,\widetilde{f_n^{(k)}}(t_{in})\}$ and any $\delta>0$

$$\liminf_{n\to\infty}\sup_{f\in\Sigma(\beta,L)}P\Big(|\widetilde{f_n^{(j)}}(t_{in})-f^{(j)}(t_{in})|n^{\frac{\beta-j}{2\beta+1}}>\delta(\psi(n))^{-\frac{\beta-j}{2\beta}}\}>0.$$

Remark 2. Estimator given in (1.7) can be applied for evaluating of $f(t)$ and its derivatives for all t out of the boundary layer. For instance, one can take $\widehat{f_n^{(j)}}(t)=\widehat{f_n^{(j)}}(t_\ell)$ for $t_\ell\le t<t_{\ell+1}$. Making use a smoothness of f, it is readily to examine that the above-mentioned estimators have the best rate of convergence in $n\to\infty$ for all t out of the boundary layer, so that an additional interpolation or even the Taylor polynomial approximation do not improve them. It is not the case for the extrapolation problem as well, if only observations for $t_{in}\le t_{\ell n}=t_\ell$ are available while the value of $f(t_\ell+h),h>0$ is required to be estimated. The upper bound (2.1) allows to establish that the estimator $\widetilde{f_n^{(0)}}(t_\ell+h)=\sum_{j=0}^{k}\frac{h^j}{j!}\widehat{f_n^{(j)}}(t_\ell)$

possesses the rate of convergence $\big(\max\{h, n^{-\frac{1}{2\beta+1}}\}\big)^\beta$ in a sense

$$E\left(\frac{\widetilde{f}_n(t_\ell+h)-f(t_\ell+h)}{(\max\{h,n^{-\frac{1}{2\beta+1}}\})^\beta}\right)^2 < C$$

and that rate is unimprovable too.

3. Notations and Preliminaries.

Hereafter, $\|x\|$ is the Euclidian norm of vector x; $\|A\|$ is a norm of matrix A. For notational convenience we rewrite (1.7) in a vector-matrix form

$$\widehat{F}^n(t_i) = \widehat{F}^n(t_{i-1}) + \frac{1}{n}a\widehat{F}^n(t_{i-1}) + \mathfrak{q}_n\big(X_i - \widehat{f_n^{(0)}}(t_{i-1})\big) \tag{3.1}$$

subject to fixed $\widehat{F}^n(0)$, where (all matrices and vectors are of sizes $(k+1)\times(k+1)$ and $k+1$ respectively)

$$\widehat{F}^n(t_i) = \begin{pmatrix} \widehat{f_n^{(0)}}(t_i) \\ \widehat{f_n^{(1)}}(t_i) \\ \vdots \\ \widehat{f_n^{(k)}}(t_i) \end{pmatrix} \quad a = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \mathfrak{q}_\mathfrak{n} = \begin{pmatrix} q_0 n^{-\frac{2\beta}{2\beta+1}} \\ q_1 n^{-\frac{2\beta-1}{2\beta+1}} \\ \vdots \\ q_{k-1} n^{-\frac{2\beta-(k-1)}{2\beta+1}} \\ q_k n^{-\frac{2\beta-k}{2\beta+1}} \end{pmatrix}.$$

The following matrix and vectors also are used in the sequel

$$Q_n = \begin{pmatrix} q_0 n^{-\frac{2\beta}{2\beta+1}} & 0 & 0 & 0 & \dots & 0 \\ q_1 n^{-\frac{2\beta-1}{2\beta+1}} & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ q_{k-1} n^{-\frac{2\beta-(k-1)}{2\beta+1}} & 0 & 0 & 0 & \dots & 0 \\ q_k n^{-\frac{2\beta-k}{2\beta+1}} & 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \mathfrak{q} = \begin{pmatrix} q_0 \\ q_1 \\ \vdots \\ q_k \end{pmatrix}, \quad F(t_i) = \begin{pmatrix} f^{(0)}(t_i) \\ f^{(1)}(t_i) \\ \vdots \\ f^{(k)}(t_i) \end{pmatrix}.$$

4. Upper bound for normalized bias.

4.1. **Preliminaries.** Set $M^n(t_i) = E\widehat{F}^n(t_i)$. Taking the expectation from the both sides of (3.1) we find (here I is the unit matrix of the size $(k+1)\times(k+1)$):

$$\begin{aligned} M^n(t_i) &= \Big(I + \frac{1}{n}a\Big)M^n(t_{i-1}) + Q_n\big(F(t_i) - M^n(t_{i-1})\big) && (4.1) \\ &= B_n M^n(t_{i-1}) + Q_n F(t_i) && (4.2) \end{aligned}$$

subject to $M^n(0) = \widehat{F}^n(0)$, where $B_n = \Big(I + \frac{1}{n}a - Q_n\Big)$, that is

$$B_n = \begin{pmatrix} 1 - q_0 n^{-\frac{2\beta}{2\beta+1}} & \frac{1}{n} & 0 & 0 & \dots & 0 & 0 \\ -q_1 n^{-\frac{2\beta-1}{2\beta+1}} & 1 & \frac{1}{n} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -q_{k-1} n^{-\frac{2\beta-(k-1)}{2\beta+1}} & 0 & 0 & 0 & \dots & 1 & \frac{1}{n} \\ -q_k n^{-\frac{2\beta-k}{2\beta+1}} & 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

The next identity implied by (4.1)

$$\begin{aligned} M^n(t_i)-F(t_i) &= M^n(t_{i-1})-F(t_{i-1})+\Big(F(t_{i-1})-F(t_i)+\frac{1}{n}aF(t_{i-1})\Big)\\ &+\frac{1}{n}a\Big(M^n(t_{i-1})-F(t_{i-1})\Big)\\ &+Q_n\Big(F(t_i)-F(t_{i-1})+F(t_{i-1})-M^n(t_{i-1})\Big), \end{aligned}$$

allows to derive a recursion for $\triangle_i=M^n(t_i)-F(t_i)$:

$$\begin{aligned} \triangle_i &= \Big(I-Q_n+\frac{1}{n}a\Big)\triangle_{i-1}\\ &+\Big(F(t_{i-1})-F(t_i)+\frac{1}{n}aF(t_{i-1})\Big)+Q_n\Big(F(t_i)-F(t_{i-1})\Big). \end{aligned} \tag{4.3}$$

Introduce now the matrix of the size $(k+1)\times(k+1)$

$$C_n=\begin{pmatrix} n^{\frac{\beta}{2\beta+1}} & 0 & 0 & 0 & \dots & 0 & 0\\ 0 & n^{\frac{\beta-1}{2\beta+1}} & 0 & 0 & \dots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & 0 & \dots & n^{\frac{\beta-(k-1)}{2\beta+1}} & 0\\ 0 & 0 & 0 & 0 & \dots & 0 & n^{\frac{\beta-k}{2\beta+1}} \end{pmatrix} \tag{4.4}$$

and the vector of normalized biases

$$\delta_i=C_n\triangle_i. \tag{4.5}$$

It is readily to verify that

$$C_nQ_n=n^{-\frac{\beta}{2\beta+1}}\begin{pmatrix} q_0 & 0 & 0 & 0 & \dots & 0 & 0\\ q_1 & 0 & 0 & 0 & \dots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ q_{k-1} & 0 & 0 & 0 & \dots & 0 & 0\\ q_k & 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}:=n^{-\frac{\beta}{2\beta+1}}Q,$$

$$QC_n^{-1}=n^{-\frac{\beta}{2\beta+1}}Q,\qquad C_na=n^{\frac{1}{2\beta+1}}aC_n. \tag{4.6}$$

For the notational convenience, let us introduce the matrix

$$D_n=I-n^{-\frac{2\beta}{2\beta+1}}Q+n^{-\frac{2\beta}{2\beta+1}}a \tag{4.7}$$

and vector $b=\begin{pmatrix}0\\ \vdots\\ 0\\ 1\end{pmatrix}$ of size $k+1$. Then, taking into account (4.3) and (4.6), we find the recursion

$$\begin{aligned} \delta_i &= D_n\delta_{i-1}+C_n\Big(F(t_{i-1})-F(t_i)+\frac{1}{n}aF(t_{i-1})\Big)\\ &+ n^{-\frac{\beta}{2\beta+1}}Q\Big(F(t_i)-F(t_{i-1})\Big) \end{aligned} \tag{4.8}$$

subject to $\delta_0 = C_n\big(\widehat{F}(0) - F(0)\big)$. To simplify a further analysis, a decomposition for δ_i is used: $\delta_i \equiv \sum\limits_{p=1}^{4} \delta_i^p$ with

$$
\begin{aligned}
\delta_i^1 &= D_n\delta_{i-1}^1, \\
\delta_i^2 &= D_n\delta_{i-1}^2 + n^{-\frac{\beta}{2\beta+1}} Q\Big(F(t_i) - F(t_{i-1})\Big) \\
\delta_i^3 &= D_n\delta_{i-1}^3 + C_n\Big(F(t_{i-1}) - F(t_i) + \frac{1}{n} aF(t_{i-1}) - b\big(f^{(k)}(t_{i-1}) - f^{(k)}(t_i)\big)\Big) \\
\delta_i^4 &= D_n\delta_{i-1}^4 + C_n b\big(f^{(k)}(t_{i-1}) - f^{(k)}(t_i)\big)
\end{aligned}
$$

and

$$\delta_0^1 = C_n\big(\widehat{F}(0) - F(0)\big), \ \ \delta^2(0) = 0, \ \ \delta^3(0) = 0, \ \ \delta^4(0) = 0. \tag{4.9}$$

Hereafter,

$$\gamma = \frac{2\beta}{2\beta + 1}, \tag{4.10}$$

and

$$c_\circ = -\frac{1}{2} \max_{j=1,\dots,k+1} \Re(u_j), \tag{4.11}$$

where $u_j = \Re u_j + \sqrt{-1}\Im u_j$, $j = 1, \dots, k+1$ are the roots of polynomial (1.4).

Proposition 4.1. *Let $q_0, \dots, q_k$ (components of the vector q) are chosen such that all roots of the polynomial given (1.4) are different and have negative real parts.*

Then,

1. *with $c_\circ$ (> 0) from (4.11) and n large enough*

$$\|D_n^i\| \le Ce^{-\frac{c_\circ}{n^\gamma} i}, \ \ i = 1, 2, \dots;$$

2. *norms of vectors $P(t_i)$'s generated by the recursion*

$$P(t_i) = D_n P(t_{i-1}) + C_n(i) \tag{4.12}$$

subject to $P(0) = 0$ with $\|C_n(i)\| \le \frac{r}{n^\gamma}$ are bounded by a constant independent of n, $\|P(t_i)\| < C$.

Proof. 1. Since (here $\mathsf{det}(*)$ is the determinant of matrix)

$$
\mathsf{det}(\lambda I - D_n) = \begin{vmatrix}
\lambda - 1 + q_0 n^{-\gamma} & -n^{-\gamma} & 0 & 0 & \dots & 0 & 0 \\
q_1 n^{-\gamma} & \lambda - 1 & -n^{-\gamma} & 0 & \dots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
q_{k-1} n^{-\gamma} & 0 & 0 & 0 & \dots & \lambda - 1 & -n^{-\gamma} \\
q_k n^{-\gamma} & 0 & 0 & 0 & \dots & 0 & \lambda - 1
\end{vmatrix},
$$

with $\lambda - 1 = un^{-\gamma}$ the equation $\mathsf{det}(\lambda I - D_n) = 0$ is transformed into

$$
\mathsf{det}(uI + Q - a) = \begin{vmatrix}
u + q_0 & -1 & 0 & 0 & \dots & 0 & 0 \\
q_1 & u & -1 & 0 & \dots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
q_{k-1} & 0 & 0 & 0 & \dots & u & -1 \\
q_k & 0 & 0 & 0 & \dots & 0 & u
\end{vmatrix} = 0 \tag{4.13}
$$

what is nothing but $p^k(u) = 0$, where $p^k(u)$ is the polynomial defined in (1.4). Recall that all roots of $p^k(u)$ have strictly negative real parts $\Re u_j$. Since $\lambda_n(j) = 1 + \frac{u_j}{n^\gamma}$, $j = 1, \ldots, k+1$, we have $\Re u_j \le -2c_\circ$. Hence, for n large enough,

$$\begin{aligned} |\lambda_n(j)| &= \sqrt{\Big(1 + \frac{\Re u_j}{n^\gamma}\Big)^2 + \frac{(\Im u_j)^2}{n^{2\gamma}}} \le \sqrt{\Big(1 - \frac{2c_\circ}{n^\gamma}\Big)^2 + \frac{C}{n^{2\gamma}}} \\ &\le \Big(1 - \frac{2c_\circ}{n^\gamma}\Big)\big(1 + o(n^{-2\gamma})\big) \le \exp\Big(-\frac{c_\circ}{n^\gamma}\Big). \end{aligned}$$

Further, whereas eigenvalues of D_n are different, the Jordan form $J(D_n)$ of D_n is diagonal matrix

$$J(D_n) = \mathsf{diag}\Big(1 + \frac{u_0}{n^\gamma}, 1 + \frac{u_1}{n^\gamma}, \ldots, 1 + \frac{u_k}{n^\gamma}\Big)$$

and there is a nonsingular matrix T so that $J(D_n) = T^{-1} D_n T$. Consequently,

$$D_n = T J(D_n) T^{-1} \quad \text{and} \quad D_n^i = T J^i(D_n) T^{-1}, \ i \ge 2.$$

Particularly, for $c_\circ$ from (4.11) we have

$$\begin{aligned} \|D_n^i\| &\le \|T\|\|T^{-1}\| \Big\|\mathsf{diag}\Big(|1 + \frac{u_0}{n^\gamma}|^i, |1 + \frac{u_1}{n^\gamma}|^i, \ldots, |1 + \frac{u_k}{n^\gamma}|^i\Big)\Big\| \\ &\le C e^{-\frac{c_\circ}{n^\gamma} i}. \end{aligned} \tag{4.14}$$

2. Since $P(t_\ell) = \sum_{i=1}^{\ell} D_n^{\ell - i} C_n(i)$, we find

$$\|P(t_\ell)\| \le \frac{r}{n^\gamma} \sum_{i=1}^{n} \|D_n^i\| \le \frac{C}{n^\gamma} \sum_{i=0}^{\infty} e^{-\frac{c_\circ}{n^\gamma} i} = \frac{C}{n^\gamma\big(1 - e^{-\frac{c_\circ}{n^\gamma}}\big)} < C.$$

□

4.2. **Boundary layer and estimation of $\|\delta_\ell^1\|$.**

Lemma 4.1. $\sup_{\ell > C n^\gamma \log n} \|\delta_\ell^1\| \le C.$

Proof. Since $\delta_\ell^1 = D_n^\ell \delta_0^1$, we have $\|\delta_\ell^1\| \le \|D_n^\ell\| \|\delta_0^1\| \le C e^{-\frac{c_\circ}{n^\gamma}\ell} \|\delta_0^1\|$ (see Proposition 4.1). On the other hand, whereas by virtue of (4.4) and (4.9) $\|\delta_0^1\| \le C n^{\gamma/2}$, the desired conclusion holds. □

4.3. **Boundedness of $\|\delta_\ell^2\| \div \|\delta_\ell^4\|$.**

Lemma 4.2. *For any $\ell \le n$, $\|\delta_\ell^2\| < C$.*

Proof. Due to Proposition 4.1 it suffices to show that

$$\Big\| n^{-\frac{\beta}{2\beta+1}} Q\Big(F(t_i) - F(t_{i-1})\Big\| \le \frac{r}{n^\gamma}. \tag{4.15}$$

To this end, we note that all components of the vector $Q\big(F(t_i) - F(t_{i-1})\big)$ are in a proportion to $f^{(0)}(t_i) - f^{(0)}(t_{i-1})$ the absolute value of which is bounded by rn^{-1}. Hence the left hand side of (4.15) is bounded above by $\frac{r}{n^{1+\gamma/2}}$ and, whereas $\gamma < 1$, we have $\frac{r}{n^{1+\gamma/2}} \le \frac{r}{n^\gamma}$. □

Lemma 4.3. *For any $\ell \le n$, $\|\delta_\ell^3\| < C$.*

Proof. Making use the Proposition 4.1, it suffices to show that

$$\Big\|C_n\Big(F(t_{i-1})-F(t_i)+\frac{1}{n}aF(t_{i-1})-b\big(f^{(k)}(t_{i-1})-f^{(k)}(t_i)\big)\Big)\Big\|\le\frac{r}{n^\gamma}. \tag{4.16}$$

The first k components of

$$\mathfrak{F}_i=C_n\Big(F(t_{i-1})-F(t_i)+\frac{1}{n}aF(t_{i-1})-b\big(f^{(k)}(t_{i-1})-f^{(k)}(t_i)\big)\Big)$$

coincide with the first k components of $\mathfrak{F}_i'=C_n\Big(F(t_{i-1})-F(t_i)+\frac{1}{n}aF(t_{i-1})\Big)$ while the last component of $\mathfrak{F}_i$ is zero. At the same time absolute values of the first k components of $\mathfrak{F}_i'$ are bounded by $Cn^{\frac{\beta}{2\beta+1}-2}$ while the last one by $n^{\frac{\beta-k}{2\beta+1}-(1+\alpha)}$ (recall that $\beta=k+\alpha$). Hence the left hand side of (4.16) is bounded from above by

$$r\Big(\frac{1}{n^{2-\frac{\beta}{2\beta+1}}}+\frac{1}{n^{(1+\alpha)-\frac{\alpha}{2\beta+1}}}\Big)\le r\Big(\frac{1}{n^{\frac{3\beta+2}{2\beta+1}}}+\frac{1}{n^{\frac{2\beta(1+\alpha)+1}{2\beta+1}}}\Big)\le 2\frac{r}{n^\gamma}.$$

□

Lemma 4.4. *For any* $Cn^\gamma\log n<\ell\le n$, $\|\delta^4_\ell\|<C$.

Proof. Due to $C_nb=n^{\frac{\beta-k}{2\beta+1}}b$, it holds $\delta^4_i=D_n\delta^4_{i-1}+n^{\frac{\beta-k}{2\beta+1}}b\big(f^{(k)}(t_{i-1})-f^{(k)}(t_i)\big)$, so that $\delta^4_\ell=n^{\frac{\beta-k}{2\beta+1}}\sum_{i=1}^{\ell}D_n^{\ell-i}b\big(f^{(k)}(t_{i-1})-f^{(k)}(t_i)\big)$. We use obvious identities

$$\begin{aligned}\sum_{i=1}^{\ell}D_n^{\ell-i}bf^{(k)}(t_{i-1}) &:= D_n^{\ell-1}bf^{(k)}(0)+\sum_{i=1}^{\ell-1}D_n^{\ell-i-1}bf^{(k)}(t_i)\\ \sum_{i=1}^{\ell}D_n^{\ell-i}bf^{(k)}(t_i) &:= bf^{(k)}(t_\ell)+\sum_{i=1}^{\ell-1}D_n^{\ell-i-1}D_nbf^{(k)}(t_i)\\ I &:= D_n^{\ell-1}+\sum_{i=1}^{\ell-1}D_n^{\ell-i-1}\big(I-D_n\big)\end{aligned}$$

which allow us to find that

$$bf^{(k)}(t_\ell)=D_n^{\ell-1}bf^{(k)}(t_\ell)+\sum_{i=1}^{\ell-1}D_n^{\ell-i-1}\big(I-D_n\big)bf^{(k)}(t_\ell)$$

and

$$\delta^4_\ell=n^{\frac{\beta-k}{2\beta+1}}\Bigg(D_n^{\ell-1}b\Big(f^{(k)}(0)-f^{(k)}(t_\ell)\Big)+\sum_{i=1}^{\ell-1}D_n^{\ell-i-1}\big(I-D_n\big)b\Big(f^{(k)}(t_i)-f^{(k)}(t_\ell)\Big)\Bigg).$$

The latter provides

$$\|\delta^4_\ell\| \;\le\; Cn^{\frac{\beta-k}{2\beta+1}}\Bigg(\|D_n^{\ell-1}\|\Big(\frac{\ell}{n}\Big)^\alpha+\sum_{i=1}^{\ell-1}\|D_n^{\ell-i-1}\|\|I-D_n\|\Big(\frac{\ell-i}{n}\Big)^\alpha\Bigg).$$

Then, taking into account $\|D_n^i\| \le Ce^{-\frac{c_\circ}{n^\gamma}i}$, (see Proposition 4.1), $\|I-D_n\| \le \frac{C}{n^\gamma}$ and $\frac{n^{\frac{\beta-k}{2\beta+1}}n^{\alpha\gamma}}{n^\alpha} \equiv 1$ (recall $\beta-k=\alpha$), we find

$$\begin{aligned}\|\delta_\ell^4\| &\le Cn^{\frac{\beta-k}{2\beta+1}}\Bigg(e^{-\frac{c_\circ}{n^\gamma}(\ell-1)}+\frac{1}{n^\gamma}\sum_{i=1}^{\ell-1}e^{-\frac{c_\circ}{n^\gamma}(\ell-i-1)}\Big(\frac{\ell-i}{n}\Big)^\alpha\Bigg)\\ &\le C\Bigg(n^{\frac{\beta-k}{2\beta+1}}e^{-\frac{c_\circ}{n^\gamma}(\ell-1)}+\frac{e^{\frac{c_\circ}{n^\gamma}}}{n^\gamma}\sum_{i=1}^{\infty}e^{-c_\circ\frac{i}{n^\gamma}}\Big(\frac{i}{n^\gamma}\Big)^\alpha\Bigg).\end{aligned}$$

So, the assertion of the lemma follows, since

$$\sum_{i=1}^{\infty}e^{-c_\circ\frac{i}{n^\gamma}}\Big(\frac{i}{n^\gamma}\Big)^\alpha\frac{1}{n^\gamma}<C. \tag{4.17}$$

Indeed, for any K

$$\sum_{i=1}^{[Kn^\gamma]}e^{-c_\circ\frac{i}{n^\gamma}}\Big(\frac{i}{n^\gamma}\Big)^\alpha\frac{1}{n^\gamma}\le K^\alpha\sum_{i=1}^{[Kn^\gamma]}e^{-c_\circ\frac{i}{n^\gamma}}\frac{1}{n^\gamma}\le K^\alpha\int_0^\infty e^{-c_\circ x}dx<\infty$$

and at the same time, whereas $e^{-c_\circ\frac{i}{n^\gamma}}(\frac{i}{n^\gamma})^\alpha$ decreases in $i\ge K$ for K large enough,

$$\frac{1}{n^\gamma}\sum_{i\ge Kn^\gamma}^{\infty}e^{-c_\circ\frac{i}{n^\gamma}}\Big(\frac{i}{n^\gamma}\Big)^\alpha\le\int_K^\infty x^\alpha e^{-c_\circ x}dx<\infty.$$

□

5. Upper bound for variance.

Set

$$\Gamma(t_i)=EC_n\big(\widehat{F}^n(t_i)-M^n(t_i)\big)\big(\widehat{F}^n(t_i)-M^n(t_i)\big)^*C_n^*,$$

where C_n is defined in (4.4) and * is the transposition symbol. For notation convenience denote $V(t_i)=C_n\big(\widehat{F}^n(t_i)-M^n(t_i)\big)$. The recursion given in (3.1) and the definition of D_n (see, (4.2)) provide

$$V(t_i)=D_nV(t_{i-1})+n^{-\frac{\beta}{2\beta+1}}\mathfrak{q}\sigma(t_{in})\xi_{in}.$$

Hence

$$\begin{aligned}V(t_i)V^*(t_i) &= D_nV(t_{i-1})V^*(t_{i-1})D_n^*+n^{-\frac{2\beta}{2\beta+1}}\mathfrak{q}\mathfrak{q}_n^*\sigma^2(t_{in})\xi_{in}^2\\ &\quad+n^{-\frac{2\beta}{2\beta+1}}\big(D_nV(t_{i-1})\mathfrak{q}^*+\mathfrak{q}V^*(t_{i-1})D_n^*\big)\sigma(t_{in})\xi_{in}.\end{aligned}$$

Now, taking the expectation from both sides of that equality we find a recursion for $\Gamma(t_i)\equiv EV(t_i)V^*(t_i)$:

$$\Gamma(t_i)=D_n\Gamma(t_{i-1})D_n^*+n^{-\frac{2\beta}{2\beta+1}}\mathfrak{q}\mathfrak{q}^*\sigma_{in}^2$$

subject to an obvious initial condition $\Gamma(0)=0$. Hence

$$\Gamma(t_\ell)=\sigma_{in}^2n^{-\gamma}\sum_{i=1}^{\ell}D_n^{\ell-i}\mathfrak{q}\mathfrak{q}^*\big(D_n^{\ell-i}\big)^*$$

and, whereas σ_{in}^2's are bounded, we have

$$\begin{aligned}\big\|\Gamma(t_\ell)\big\| &\le Cn^{-\gamma}\sum_{i=1}^{\ell}\|D_n^{\ell-i}\|^2 = Cn^{-\gamma}\sum_{i=0}^{\ell-1}\|D_n^i\|^2 \\ &\le Cn^{-\gamma}\sum_{i=1}^{\infty}e^{-\frac{2c_\circ}{n^\gamma}i} = \frac{C}{n^\gamma\big(1-e^{-2c_\circ n^{-\gamma}}\big)} < C.\end{aligned}$$

Thus, for any $\ell \le n$

$$\sum_{j=0}^{k} E\big(f^{(j)}(t_\ell)-\widehat{f_n^{(j)}}(t_\ell)\big)^2 n^{\frac{2(\beta-j)}{2\beta+1}} < C. \tag{5.1}$$

□

6. Proof of Theorem 2.1.

Lemmas 4.1-4.4 provide

$$\sum_{j=0}^{k}\big(f^{(j)}(t_\ell)-E\widehat{f}^{(j,n)}(t_\ell)\big)^2 n^{\frac{2(\beta-j)}{2\beta+1}} < C, \text{ under } \ell > Cn^\gamma \log n. \tag{6.1}$$

The desired statement is implied by (5.1) and (6.1). □

References

[1] Brown, L. and Low, M. (1996) Asymptotic equivalence of nonparametric regression and white noise, *Ann. Statist.*, **24**, pp. 2384–2398.

[2] Chow, P.-L., Khasminskii, R. and Liptser, R.Sh. (1997) Tracking of signal and its derivatives in Gaussian white noise, *Stochastic processes and their application.* **69**, 2 pp. 259-273.

[3] Grama, I. and Nussbaum, M. (1998) Asymptotic equivalence for nonparametric generalized linear models, *Probab. Theory Relat. Fields*, **111** pp. 167–214.

[4] Ibragimov, I. and Khasminskii, R. (1980) On nonparametric estimation of regression,*Soviet Math.Dokl.*,**21**, pp. 810–814.

[5] Ibragimov, I. and Khasminskii, R. *Statistical estimation: Asymptotic theory.* Springer Verlag, 1981 (Russian ed.1979).

[6] Le Cam, L. *Asymptotic Methods in Statistical Decision Theory*, Springer, 1986, NY.

[7] Parzen, E. (1962) On estimation of a probability density function and mode. *Ann. Math. Statist.*, **33**, No 3, pp. 1065–1073.

[8] Rosenblatt, M. (1956) Remarks on some nonparametric estimates of a density function". *Ann. Math. Statist.*, **27**, No 3, pp. 832–837.

[9] Stone, C. (1980) Optimal rates of convergence for nonparametric estimators, *Ann. Statist.*, **8**, pp.1348–1360.

[10] Stone, C. (1982) Optimal global rates of convergence for nonparametric regression, *Ann. Statist.*, **10**, pp.1040–1053.

Department of Mathematics, Wayne State University, Detroit,, Detroit MI, 48202 USA

E-mail address: `rafail@math.wayne.edu`

Electrical Engineering Systems, Tel Aviv University, 69978 - Ramat Aviv, Tel Aviv, Israel

E-mail address: `liptser@eng.tau.ac.il`