Tracking of signal and its derivatives in Gaussian white noise

P.-L. Chow* and R. Khasminskii *,
Department of Mathematics, Wayne State University, Detroit,
Detroit MI, 48202 USA
R. Liptser †,
Electrical Engineering Systems,
Tel Aviv University, 69978 - Ramat Aviv, Tel Aviv, Israel

Abstract

For the observation model "signal + white Gaussian noise", an on-line tracking algorithm for signal and its derivatives is proposed. The tracking algorithm applies to a class of signals with derivative up to the k-th order. The asymptotic optimality in the minimax sense, with respect to small intensity of noise, is established.

Keywords: Gaussian White Noise; On-line tracking algorithm; Kernel estimator, Ito's equations; Riccati equation.

1 Introduction and Main Result

Let $S(t), t \geq 0$ be a smooth signal, the information of which is given by the observation of a random process $X_t, t \geq 0$ with

$$\begin{aligned} dX_t &= S(t)dt + \varepsilon dW_t \\ X_0 &= 0, \end{aligned} \tag{1.1}$$

***Partially supported by the ONR Grant N00014-95-1-0793**
†Partially supported by the ONR Grant N00014-96-1-0413

where $W_t, t \geq 0$ is a Wiener process, and where ε is a small parameter.

Here we choose a filtering type of estimates for $S(t)$ and its derivatives $S^{(j)}(t) = \frac{d^j S(t)}{dt^j}$, $j = 1, ...k-1$ based on observations of $X_s, s \leq t$. It is clear that for getting meaningful quality of filtering it is necessary to limit ourselves by consideration of some restricted class of signals. Here, we choose a class denoted by $\Sigma^{k,\alpha}(\ell)$ characterized by:

(i): the values of $|S(0)|$, $|S^{(j)}(0)|, j = 1, ..., k-1$ are bounded, i.e. there exists a constant c such that

$$|S(0)| + \sum_{j=1}^{k-1} |S^{(j)}(0)| \leq c;$$

(ii): $S^{(k-1)}(t)$ is a Hölder continuous function so that

$$|S^{(k-1)}(t') - S^{(k-1)}(t)| \leq \ell |t' - t|^{\alpha},\ 0 < \alpha \leq 1.$$

It should be noted that smoothing type estimates for the probability density function are well known from Rosenblatt [1] and Parzen [2] and for the model considered from [3], [4]. Namely, estimates $\widehat{S}(t)$, $\widehat{S}^{(j)}(t)$, $j = 1, ..., k-1$, corresponding to the signal and derivatives respectively, are defined as:

$$\begin{aligned} \widehat{S}(t) &= \frac{1}{\delta} \int_{-\infty}^{\infty} K\Big(\frac{t-s}{\delta}\Big) dX_s \\ \widehat{S}^{(j)}(t) &= \frac{1}{\delta^{j+1}} \int_{-\infty}^{\infty} K^{(j)}\Big(\frac{t-s}{\delta}\Big) dX_s,\ j = 1, ..., k-1, \end{aligned} \tag{1.2}$$

where $K(u)$ is compactly supported smooth kernel ($K^{(j)}(u) = \frac{d^j K(u)}{du^j}$) and δ is a positive small parameter. In [4], for suitable choice of $K(u)$ and δ, depending on ε ($\delta = \delta_\varepsilon$), the mean square errors for these estimates provide the following rates of convergence with respect to $\varepsilon \to 0$: with $\beta = k - 1 + \alpha$

$$\begin{aligned} E\Big(S(t) - \widehat{S}(t)\Big)^2 &\asymp \varepsilon^{4\beta/(2\beta+1)} \\ E\Big(\frac{d^j S(t)}{dt^j} - \widehat{S}^{(j)}(t)\Big)^2 &\asymp \varepsilon^{4(\beta-j)/(2\beta+1)},\ j = 1, ..., k-1. \end{aligned} \tag{1.3}$$

Moreover, these rates are optimal in the 'minimax' sense (for more details see Theorem 2.1 in [4]).

In contrast to smoothing estimators, we shall establish the so called filtering type estimators. Of course, for that purpose the estimators given in (1.2) with $K(u) = 0$ for $u < 0$ also are valid and optimal in the minimax sense. In contrast to the (1.2) type of filtering estimators, we choose some on-line filtering ones which are inspired by Kalman's idea. The proposed filter has a simple recursive structure for any k while kernels in (1.2) become more and more delicate with increasing k.

The form of the proposed tracking on-line filter is given below:

$$\begin{aligned} d\widehat{S}(t) &= \widehat{S}^{(1)}(t)dt + \frac{q_1}{\varepsilon^{2/(2\beta+1)}}(dX_t - \widehat{S}(t)dt), \\ d\widehat{S}^{(j)}(t) &= \widehat{S}^{(j+1)}(t)dt + \frac{q_{j+1}}{\varepsilon^{2(j+1)/(2\beta+1)}}(dX_t - \widehat{S}(t)dt), j = 1, ..., k-2, \\ d\widehat{S}^{(k-1)}(t) &= \frac{q_k}{\varepsilon^{2k/(2\beta+1)}}(dX_t - \widehat{S}(t)dt), \end{aligned} \tag{1.4}$$

subject to initial conditions $\widehat{S}(0) = S_\circ$, $\widehat{S}^{(j)}(0) = S_\circ^{(j)}$, $j = 1, ..., k-1$, where the vector

$$\mathcal{Q} = \begin{pmatrix} q_1 \\ . \\ . \\ . \\ q_k \end{pmatrix} \tag{1.5}$$

belongs to a set of vectors, say, $\mathcal{B}$ characterized by the following property: for every vector $\widetilde{\mathcal{Q}}$ from $\mathcal{B}$, the polynomial

$$p^k(x) = x^k + \widetilde{q}_1 x^{k-1} + ... + \widetilde{q}_{k-1}x + \widetilde{q}_k \tag{1.6}$$

has roots with negative real parts.

Theorem 1.1. *For every vector $(q_1, ..., q_k)$ from $\mathcal{B}$ and every signal $S(t)$ from $\Sigma^{k,\alpha}(\ell)$, the tracking filter (1.4) has the property: there exist constants $L(\mathcal{Q})$ and $C^\circ(\mathcal{Q})$, depending only on $\mathcal{Q}$, and a boundary layer $[0, \Delta^\varepsilon]$ with*

$$\Delta^\varepsilon = L(\mathcal{Q})\varepsilon^{2/(2\beta+1)}\log(1/\varepsilon)$$

such that for $t > \Delta^\varepsilon$

$$E\Big(\Big[\frac{\widehat{S}(t) - S(t)}{\varepsilon^{2\beta/(2\beta+1)}}\Big]^2 + \sum_{j=1}^{k-1}\Big[\frac{\widehat{S}^{(j)}(t) - S^{(j)}(t)}{\varepsilon^{2(\beta-j)/(2\beta+1)}}\Big]^2\Big) \le C^\circ(\mathcal{Q}) + o(\varepsilon). \tag{1.7}$$

As was mentioned above the rates $\varepsilon^{2\beta/(2\beta+1)}$ and $\varepsilon^{2(\beta-j)/(2\beta+1)}$, $j=1,...,k-1$ can not be improved uniformly in the class $\Sigma^{k,\alpha}(\ell)$. On the other hand, the estimators given in (1.3) with compactly supported kernel, say, $K(u)=0, |u|>A$ has the boundary layer $[0, A\varepsilon^{2/(2\beta+1)}]$. For estimators given by differential equations (1.4), the boundary layer is inevitably a little wider. This reflects the fact that the estimators given by differential equations (1.4) obey the kernel type representation with not compactly supported kernel.

The paper is organized as follows. Section 2 contains a heuristic explanation for the x of the structure of tracking filter. The proof of Theorem 1.1 is given in Section 3. In Section 4, it is shown that in the case $\alpha=1$ there exists $\mathcal{Q}^*$ from the set $\mathcal{B}$ such that $C^\circ(\mathcal{Q}^*)=\inf_{\mathcal{Q}\in\mathcal{B}} C^\circ(\mathcal{Q})$. A minimal boundary layer is established in Appendix (Section 5).

2 Heuristic Choice of Tracking Filter

The aim of this section is to show how to choose an appropriate Kalman filter which, being applied to the observed signal X_t, guarantees the same mean square rates (1.3) as the kernel estimator (1.2) does. Under assumptions made on the class $\Sigma^{k,\alpha}(\ell)$ of signals, each $S(t)$ from this class can be described as

$$\begin{aligned}\dot{S}(t) &= S^{(1)}(t)\\ \dot{S}^{(j)}(t) &= S^{(j+1)}(t),\ j=1,...,k-2.\end{aligned} \tag{2.1}$$

Taking into account (2.1), introduce an auxiliary filtering model

$$\begin{aligned}dS(t) &= S^{(1)}(t)dt\\ dS^{(j)}(t) &= S^{(j+1)}(t)dt,\ j=1,...,k-2\\ dS^{(k-1)}(t) &= \sigma_\varepsilon dW'_t\\ dX_t &= S(t)dt+\varepsilon dW_t,\end{aligned}$$

in which $S(t)$ possesses the same "dynamics" as the real signal does, but the derivative $S^{(k-1)}(t)$ is simulated by a Wiener process with a specially chosen diffusion parameter σ_ε depending on ε. Assuming that in this model the Wiener processes W'_t and W_t are independent, we arrive at the filtering equations

$$d\widehat{S}(t) = \widehat{S}^{(1)}(t)dt+\frac{P_{1,1}(t)}{\varepsilon^2}\Big(dX_t-\widehat{S}(t)dt\Big),$$

$$
\begin{aligned}
d\widehat{S}^{(j)}(t) &= \widehat{S}^{(j+1)}(t)dt + \frac{P_{1,j+1}(t)}{\varepsilon^2}\Big(dX_t - \widehat{S}(t)dt\Big),\ j=1,...,k-2,\\
d\widehat{S}^{(k-1)}(t) &= \frac{P_{1,k}(t)}{\varepsilon^2}\Big(dX_t - \widehat{S}(t)dt\Big),
\end{aligned} \tag{2.2}
$$

where $P_{i,j}(t),\ i,j=1,...,k$ are elements of matrix P_t, being a solution of the Ricatti equation

$$
\dot{P}(t) = aP_t + P_t a^T + \sigma_\varepsilon^2 bb^T - \frac{P_t A^T A P_t}{\varepsilon^2},
$$

and where the matrices a and A of sizes $k\times k$ and $1\times k$ are the following

$$
a = \begin{pmatrix}
0 & 1 & 0 & . & . & . & 0 & 0\\
0 & 0 & 1 & . & . & . & 0 & 0\\
. & . & . & . & . & . & . & .\\
. & . & . & . & . & . & . & .\\
. & . & . & . & . & . & . & .\\
0 & 0 & 0 & . & . & . & 0 & 1\\
0 & 0 & 0 & . & . & . & 0 & 0
\end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} 1 & 0 & . & . & . & . & . & . & 0 \end{pmatrix} \tag{2.3}
$$

while the matrix b of size $k\times 1$ is given by

$$
b = \begin{pmatrix} 0\\ .\\ .\\ .\\ 0\\ 1 \end{pmatrix}. \tag{2.4}
$$

It can be shown that there exists the limit matrix $P(\sigma_\varepsilon,\varepsilon) = \lim_{t\to\infty} P_t$ (its elements are denoted by $P_{i,j}(\sigma_\varepsilon,\varepsilon)$) which is positive definite and satisfies the algebraic equation

$$
aP(\sigma_\varepsilon,\varepsilon) + P(\sigma_\varepsilon,\varepsilon)a^T + \sigma_\varepsilon^2 bb^T - \frac{P(\sigma_\varepsilon,\varepsilon)A^T A P(\sigma_\varepsilon,\varepsilon)}{\varepsilon^2} = 0. \tag{2.5}
$$

Since

$$
\begin{aligned}
P_{1,1}(\sigma_\varepsilon,\varepsilon) &= \lim_{t\to\infty} E(S(t)-\widehat{S}(t))^2\\
P_{j,j}(\sigma_\varepsilon,\varepsilon) &= \lim_{t\to\infty} E(S^{(j-1)}(t)-\widehat{S}^{(j-1)}(t))^2,\ j=2,...,k,
\end{aligned}
$$

a proper choice of parameter σ_ε is based on the rates from (1.3). Below, we show that there exists such a σ_ε^* for which

$$P_{j,j}(\sigma_\varepsilon^*, \varepsilon) \asymp \varepsilon^{4(\beta+1-j)/(2\beta+1)}, \ j = 1, ..., k. \tag{2.6}$$

In fact, it is known from [6] that the elements of the matrix $P(\sigma_\varepsilon, \varepsilon)$ are defined as:

$$P_{i,j}(\sigma_\varepsilon, \varepsilon) = q_{i,j}\varepsilon^2\Big(\frac{\sigma_\varepsilon}{\varepsilon}\Big)^{\frac{i+j-1}{k}}, \tag{2.7}$$

where the parameters $q_{i,j}, i, j = 1, ..., k$, independent of ε, are the elements of a positive definite symmetric matrix $P(1,1)$, being the unique solution of the quadratic algebraic equation (compare with (2.5))

$$aP(1,1) + P(1,1)a^T + bb^T - P(1,1)A^T AP(1,1) = 0. \tag{2.8}$$

A straightforward calculation shows that (2.6) holds with $\sigma_\varepsilon^* = \varepsilon^{[1+2(\beta-k)]/(2\beta+1)}$. We find also, using q_j to designate $q_{1,j}$,

$$\frac{P_{1,j}(\sigma_\varepsilon^*, \varepsilon)}{\varepsilon^2} = \frac{q_j}{\varepsilon^{2j/(2\beta+1)}}, \ j = 1, ..., k. \tag{2.9}$$

Thus one can define a tracking filter by using (2.2) with $\frac{P_{1,j}(t)}{\varepsilon^2}$ replaced by $\frac{P_{1,j}(\sigma_\varepsilon^*,\varepsilon)}{\varepsilon^2}$:

$$\begin{aligned} d\widehat{S}(t) &= \widehat{S}^{(1)}(t)dt + \frac{q_1}{\varepsilon^{2/(2\beta+1)}}\Big(dX_t - \widehat{S}(t)dt\Big), \\ d\widehat{S}^{(j)}(t) &= \widehat{S}^{(j+1)}(t)dt + \frac{q_{j+1}}{\varepsilon^{2(j+1)/(2\beta+1)}}\Big(dX_t - \widehat{S}(t)dt\Big), \ j = 1, ..., k-2, \\ d\widehat{S}^{(k-1)}(t) &= \frac{q_k}{\varepsilon^{2k/(2\beta+1)}}\Big(dX_t - \widehat{S}(t)dt\Big) \end{aligned} \tag{2.10}$$

subject to the initial conditions $\widehat{S}(0) = S_\circ$, $\widehat{S}^{(j)}(0) = S_\circ^j$, $j = 1, ..., k-1$, which reflect a priori information on $S(0)$, $S^{(j)}(0)$, $j = 1, ..., k-1$. The parameters $q_1, ..., q_k$ involved in the filtering equation (2.10), being the elements of matrix Q, can be taken by any way that ensures the exponential stability for the system of differential equations

$$\begin{aligned} \dot{y}(t) &= y^{(1)}(t) - \frac{q_1}{\varepsilon^{2/(2\beta+1)}}y(t), \\ \dot{y}^{(j)}(t) &= y^{(j+1)}(t) - \frac{q_{j+1}}{\varepsilon^{2(j+1)/(2\beta+1)}}y(t), \ j = 1, ..., k-2, \\ \dot{y}^{(k-1)}(t) &= -\frac{q_k}{\varepsilon^{2k/(2\beta+1)}}y(t). \end{aligned} \tag{2.11}$$

3 Proof of Theorem 1.1.

Let the vector $\mathcal{Q}$ and the matrices a and A be defined in (1.5) and (2.3) respectively. Put

$$A^{\mathcal{Q}} = a - \mathcal{Q}A. \tag{3.1}$$

Lemma 3.1. *There exists a vector $\mathcal{Q}$ such that all the eigen-values of matrix $A^{\mathcal{Q}}$ have the a priori fixed negative real parts.*

Proof: The proof is implied by

$$\det\Big(Ix - A^{\mathcal{Q}}\Big) = x^k + q_1 x^{k-1} + ... + q_{k-1}x + q_k$$

and by the fact that $q_1, ..., q_k$ can be chosen such that the polynomial $p^k(x) = x^k + q_1 x^{k-1} + ... + q_{k-1}x + q_k$ has roots with the a priori fixed negative real parts.□

Let $\widehat{S}(t)$ and $\widehat{S}^{(j)}(t)$, $j = 1, ..., k-1$ be the estimates defined by the tracking filter (1.4). Consider the estimation errors given by the residuals

$$\begin{aligned} \Delta(t) &= \widehat{S}(t) - S(t), \\ \Delta^j(t) &= \widehat{S}^{(j)}(t) - S^{(j)}(t), j = 1, ..., k-1. \end{aligned} \tag{3.2}$$

Taking now (1.4) and (2.1) into account , we find for the residuals the following equations:

$$\begin{aligned} \Delta(t) &= \Delta(0) + \int_0^t [\Delta^1(s) - \frac{q_1}{\varepsilon^{2/(2\beta+1)}}\Delta(s)]ds + \frac{q_1\varepsilon}{\varepsilon^{2/(2\beta+1)}}W_t, \\ \Delta^j(t) &= \Delta^j(0) + \int_0^t [\Delta^{j+1}(s) - \frac{q_{j+1}}{\varepsilon^{2(j+1)/(2\beta+1)}}\Delta(t)]ds + \frac{q_{j+1}\varepsilon}{\varepsilon^{2(j+1)/(2\beta+1)}}W_t, \\ &\quad j = 1, ..., k-2, \\ \Delta^{(k-1)}(t) &= \Delta^{(k-1)}(0) - \int_0^t \frac{q_k}{\varepsilon^{2k/(2\beta+1)}}\Delta(s)ds + \frac{q_k\varepsilon}{\varepsilon^{2k/(2\beta+1)}}W_t \\ &\quad -\Big[S^{(k-1)}(t) - S^{(k-1)}(0)\Big], \end{aligned} \tag{3.3}$$

where $\Delta(0) = S_\circ - S(0)$, $\Delta^j(0) = S^j_\circ - S^{(j)}(0)$, $j = 1, ..., k-1$. For further convenience, introduce the rescaled residuals

$$\begin{aligned}\delta(t) &= \frac{\Delta\big(t\varepsilon^{2/(2\beta+1)}\big)}{\varepsilon^{2\beta/(2\beta+1)}},\\ \delta^j(t) &= \frac{\Delta^j\big(t\varepsilon^{2/(2\beta+1)}\big)}{\varepsilon^{2(\beta-j)/(2\beta+1)}},\ j=1,...,k-1.\end{aligned}\tag{3.4}$$

for which equations (3.3) are transformed into

$$\begin{aligned}\delta(t) &= \delta(0)+\int_0^t[\delta^1(s)-q_1\delta(s)]ds+q_1W_t^\varepsilon,\\ \delta^j(t) &= \delta^j(0)+\int_0^t[\delta^{j+1}(s)-q_{j+1}\delta(s)]ds+q_{j+1}W_t^\varepsilon,\\ &\quad j=1,....,k-2,\\ \delta^{k-1}(t) &= \delta^{k-1}(0)-\int_0^t q_k\delta(s)ds+q_kW_t^\varepsilon\\ &\quad -\frac{S^{(k-1)}\big(t\varepsilon^{2/(2\beta+1)}\big)-S^{(k-1)}(0)}{\varepsilon^{2\alpha/(2\beta+1)}}\end{aligned}\tag{3.5}$$

subject to the initial conditions $\delta(0)=\frac{S_\circ-S(0)}{\varepsilon^{2\beta/(2\beta+1)}}$, $\delta^j(0)=\frac{[S_\circ^j-S^{(j)}(0)]}{\varepsilon^{2(\beta-j)/(2\beta+1)}}$, $j=1,...,k-1$, where $W_t^\varepsilon=\varepsilon^{-1/(2\beta+1)}W_{t\varepsilon^{2/(2\beta+1)}}$ is a standard Wiener process in the sense that $E(W_t^\varepsilon)^2\equiv t$.

Put $D(t)=\begin{pmatrix}\delta(t)\\ \delta^1(t)\\ .\\ .\\ .\\ \delta^{k-1}(t)\end{pmatrix}$, $L^\varepsilon(t)=-\frac{S^{(k-1)}(t\varepsilon^{2/(2\beta+1)})-S^{(k-1)}(0)}{\varepsilon^{2\alpha/(2\beta+1)}}b$, where the matrix b is defined in (2.4), and rewrite (3.5) in a vector-matrix form

$$D(t)=D(0)+L^\varepsilon(t)+\int_0^t A^{\mathcal{Q}}D(s)ds+\mathcal{Q}W_t^\varepsilon.\tag{3.6}$$

We decompose $D(t)=D^0(t)+\big(ED(t)-D^0(t)\big)+\big(D(t)-ED(t)\big)$, where $D^0(t)$, $M(t)=ED(t)-D^0(t)$ and $V(t)=D(t)-ED(t)$ are defined by the integral equations

$$\begin{aligned}D^0(t) &= D(0)+\int_0^t A^{\mathcal{Q}}D^0(s)ds,\\ M(t) &= L^\varepsilon(t)+\int_0^t A^{\mathcal{Q}}M(s)ds,\\ V(t) &= \int_0^t A^{\mathcal{Q}}V(s)ds+\mathcal{Q}W_t^\varepsilon,\end{aligned}$$

where $D(0) = \begin{pmatrix} \frac{S_\circ - S(0)}{\varepsilon^{2k/(2\beta+1)}} \\ \frac{S_\circ^1 - S^{(1)}(0)}{\varepsilon^{2(k-1)/(2\beta+1)}} \\ \cdot \\ \cdot \\ \cdot \\ \frac{S_\circ^{k-1} - S^{(k-1)}(0)}{\varepsilon^{2/(2\beta+1)}} \end{pmatrix}$. Estimate now from above the values of $||D^0(t)||$, $||M(t)||$, and $||EV(t)V^T(t)||$. To find these upper bounds, we use the explicit form of solutions corresponding to the above-mentioned integral equations. Evidently, $D^0(t) = \exp\left(tA^{\mathcal{Q}}\right)D^0(0)$, $V(t) = \int_0^t \exp\left((t-s)A^{\mathcal{Q}}\right)\mathcal{Q}dW_s^\varepsilon$, and, since $L^\varepsilon(t)$ might not be an absolutely continuous function, $M(t) = L^\varepsilon(t) + \int_0^t \exp\left(sA^{\mathcal{Q}}\right)A^{\mathcal{Q}}L^\varepsilon(t-s)ds$. Also we emphasis two properties of the matrix $A^{\mathcal{Q}} = A^{\mathcal{Q}}_{k\times k}$ which, by Lemma 3.1, has negative real parts for all its eigenvalues ($-\lambda^\circ$ designates the maximal real part): there exists a positive constant c_2 such that

$$|| \exp\left\{tA^{\mathcal{Q}}\right\}||^2 \le c_2(1 + t^{2(k-1)})e^{-2\lambda^\circ t}. \tag{3.7}$$

and ($I = I_{k\times k}$ is the identity matrix)

$$|| \int_0^t \exp\left\{(t-s)A^{\mathcal{Q}}\right\}A^{\mathcal{Q}}ds + I||^2 \le c_2(1 + t^{2(k-1)})e^{-2\lambda^\circ t}. \tag{3.8}$$

It is clear that $||D^0(t)||^2 \le ||D^0(0)||^2||e^{A^{\mathcal{Q}}t}||^2$. Accordingly to the description of the class $\Sigma^{k,\alpha}(\ell)$ (see(i)) the initial conditions $S_\circ$, $S_\circ^{(j)}$, $j = 1, ..., k-1$ have to be chosen such that $|S_\circ| + \sum_{j=1}^{k-1}|S_\circ^{(j)}| \le c$. Then, due to the same condition (i), there exist constant c_1 such that $||D^0(0)||^2 \le c_1\varepsilon^{-4\beta/(2\beta+1)}$. Therefore, we get (hereafter we use a generic constant c)

$$||D^0(t)||^2 \le c\varepsilon^{-4\beta/(2\beta+1)}(1 + t^{2(k-1)})e^{-2\lambda^\circ t}. \tag{3.9}$$

An upper bound for $||EV(t)V^T(t)||$ is implied by the explicit formula for $V(t)$ and a property of Itô's stochastic integral:

$$\begin{aligned} trEV(t)V^T(t) &= \int_0^t tr\ \exp\left(sA^{\mathcal{Q}}\right)\mathcal{Q}\mathcal{Q}^T\exp\left(sA^{\mathcal{Q}}\right)^T ds \\ &\le c||\mathcal{Q}||^2\int_0^\infty (1 + s^{2(k-1)})e^{-2\lambda^\circ s}ds. \end{aligned} \tag{3.10}$$

To get an upper bound for $||M(t)M^T(t)||$, we use an obvious representation

$$
\begin{aligned}
M(t) &= \int_0^t \exp\big(sA^{\mathcal{Q}}\big) A^{\mathcal{Q}}[L^\varepsilon(t-s) - L^\varepsilon(t)]ds \\
&\quad + L^\varepsilon(t)\Big[I + \int_0^t \exp\big((t-s)A^{\mathcal{Q}}\big) A^{\mathcal{Q}} ds\Big]
\end{aligned}
$$

and the following estimates: $||L^\varepsilon(t-s) - L^\varepsilon(t)|| \le \ell s^\alpha$ (these estimates follow from (i) and (ii)). Hence we find an upper bound

$$
\begin{aligned}
||M(t)|| &\le \ell ||A^{\mathcal{Q}}|| \int_0^\infty s^\alpha \sqrt{c(1+s^{2(k-1)})} e^{-\lambda^\circ s} ds \\
&\quad + \ell t^\alpha \sqrt{c(1+t^{2(k-1)})} e^{-\lambda^\circ t}
\end{aligned}
\tag{3.11}
$$

Putting

$$
\begin{aligned}
\varphi(\varepsilon, t) &= c\varepsilon^{-4k/(2\beta+1)}(1+t^{2(k-1)}) + c\ell^2 t^{2\alpha}(1+t^{2(k-1)}), \\
p_1(\mathcal{Q}) &= c||\mathcal{Q}||^2 \int_0^\infty (1+s^{2(k-1)}) e^{-2\lambda^\circ s} ds, \\
p_2(\mathcal{Q}) &= \ell ||A^{\mathcal{Q}}|| \int_0^\infty s^\alpha \sqrt{c(1+s^{2(k-1)})} e^{-\lambda^\circ s} ds,
\end{aligned}
$$

and combining all the above-mentioned upper bounds, one can conclude that with $C^\circ(\mathcal{Q}) = p_1(\mathcal{Q}) + 2p_2^2(\mathcal{Q})$

$$
\begin{aligned}
& E\Big(\Big[\frac{\widehat{S}(t) - S(t)}{\varepsilon^{2\beta/(2\beta+1)}}\Big]^2 + \sum_{j=1}^{k-1} \Big[\frac{\widehat{S}^j(t) - S^{(j)}(t)}{\varepsilon^{2(\beta-j)/(2\beta+1)}}\Big]^2\Big)(\mathcal{Q}) \\
\le\ & 3\varphi(\varepsilon, t/\varepsilon^{2/(2\beta+1)}) e^{-2\lambda^\circ [t/\varepsilon^{2/(2\beta+1)}]} + C^\circ(\mathcal{Q}).
\end{aligned}
$$

Hence, the desired conclusion holds.□

4 Choice of Relevant Parameters in Class $\Sigma^{k,1}(\ell)$.

For fixed ε and $t > \Delta^\varepsilon$, in Theorem 1.1 an upper bound $C^\circ(\mathcal{Q})$ for the mean-square error of tracking for any of signals from the class $\Sigma^{k,\alpha}(\ell)$ and its derivatives is proposed. In this section, restricting ourselves by considerations only the class $\Sigma^{k,1}(\ell)$, we minimize, by choosing the vector $\mathcal{Q}$, an appropriate type of upper bound in (1.7). Note that in the case of the signal $S(t)$ belongs to $\Sigma^{k,1}(\ell)$ its $k-1$-the derivative

$S^{(k-1)}(t)$ is Lipschitz continuous, i.e. there exists a measurable function $u(t)$ such that $dS^{(k-1)}(t) = u(t)dt$ and $|u(t)| \le \ell$. Also note that in the present case, $\beta = k$. These facts allow, for any admissible vector $\mathcal{Q}$, to find other type of an upper bound in (1.7). Namely, (3.5) and (3.6) can be rewritten in the differential form

$$\begin{aligned} d\delta(t) &= [\delta^1(t) - q_1\delta(t)]dt + q_1 dW_t^\varepsilon, \\ d\delta^j(t) &= [\delta^{j+1}(t) - q_{j+1}\delta(t)]dt + q_{j+1}dW_t^\varepsilon, \\ & \qquad j = 1, ..., k-2, \\ d\delta^{(k-1)}(t) &= -u(t\varepsilon^{2/(2k+1)})dt - q_k\delta(t)dt + q_k dW_t^\varepsilon, \end{aligned}$$

and

$$dD(t) = -u(t\varepsilon^{2/(2k+1)})bdt + A^{\mathcal{Q}}D(t)dt + \mathcal{Q}dW_t^\varepsilon.$$

From the proof of Theorem 1.1, it follows that an upper bound in (1.7) can be defined as $\sup_{t\ge 0}\ trU_t$, where $U_t = ED(t)D^T(t)$ provided that $D(0) = 0$. The Itô formula, applying to $D(t)D^T(t)$, gives

$$\begin{aligned} dD(t)(D)^T(t) &= -u(t\varepsilon^{2/(2k+1)})[b(D)^T(t) + D(t)b^T]dt \\ &\quad + [A^{\mathcal{Q}}D(t)(D)^T(t) + D(t)(D)^T(t)(A^{\mathcal{Q}})^T + \mathcal{Q}\mathcal{Q}^T]dt \\ &\quad + [Q(D)^T(t) + D(t)\mathcal{Q}^T]dW_t^\varepsilon \end{aligned}$$

which in turn provides a representation

$$\frac{dU_t}{dt} = -u(t\varepsilon^{2/(2k+1)})[bE(D)^T(t) + ED(t)b^T] + [A^{\mathcal{Q}}U_t + U_t(A^{\mathcal{Q}})^T + \mathcal{Q}\mathcal{Q}^T].$$

Estimate now from above the symmetric matrix $-u(t\varepsilon^{2/(2k+1)})[bE(D)^T(t) + ED(t)b^T]$ $(= G(t))$ by $\alpha(t)I$ (I and $\alpha(t)$ are the identity matrix and a positive function respectively) in the sense that $\alpha(t)I - G(t)$ is a nonnegative definite matrix. To this end, denoting by $\delta_{1,j}(t), j = 1, ..., k$, elements of the vector $D(t)$ and taking a column vector Y with elements $y_1, ..., y_k$, we obtain

$$\begin{aligned} Y^T G(t) Y &= -u(t\varepsilon^{2/(2k+1)})Y^T[bE(D)^T(t) + ED(t)b^T]Y \\ &= -2u(t\varepsilon^{2/(2k+1)})\sum_{j=1}^k y_j y_k E\delta_{1,j}(t) \end{aligned}$$

$$
\begin{aligned}
&\le 2\ell |y_k| \sum_{j=1}^{k} |y_j||E\delta_{1,j}(t)| \\
&\le 2\ell |y_k| \sqrt{\sum_{j=1}^{k} y_j^2} \sqrt{\sum_{j=1}^{k} (E\delta_{1,j}(t))^2} \\
&\le 2\ell ||Y||^2 \sqrt{tr\ U_t},
\end{aligned}
$$

that is taking $\alpha(t) = 2\ell\sqrt{tr\ U_t}$ we get a nonnegative definite matrix $\Psi(t) = 2\ell\sqrt{tr\ U_t} I$ $-G(t)$ and since $\frac{dU_t}{dt} = 2\ell\sqrt{tr\ U_t} I + A^{\mathcal{Q}} U_t + U_t (A^{\mathcal{Q}})^T + \mathcal{Q}\mathcal{Q}^T - \Psi(t)$, the matrix U_t is defined as:

$$
\begin{aligned}
U_t &= \int_0^t \exp\left(sA^{\mathcal{Q}}\right) \mathcal{Q}\mathcal{Q}^T \exp\left(s(A^{\mathcal{Q}})^T\right) ds \\
&\quad + 2\ell \int_0^t \sqrt{tr\ U_{t-s}} \exp\left(sA^{\mathcal{Q}}\right) \exp\left(s(A^{\mathcal{Q}})^T\right) ds \\
&\quad - \int_0^t \exp\left(sA^{\mathcal{Q}}\right) \Psi(t-s) \exp\left(s(A^{\mathcal{Q}})^T\right) ds.
\end{aligned}
$$

This representation provides for $\gamma(t) = \sup_{s\le t} tr\ U_s$ the following inequality $\gamma(t) \le 2\ell\sqrt{\gamma(t)} p_1 + p_2$, where

$$
\begin{aligned}
p_1 &= \int_0^\infty tr\ \exp\left(sA^{\mathcal{Q}}\right) \exp\left(s(A^{\mathcal{Q}})^T\right) ds \\
p_2 &= \int_0^\infty tr\ \exp\left(sA^{\mathcal{Q}}\right) \mathcal{Q}\mathcal{Q}^T \exp\left(s(A^{\mathcal{Q}})^T\right) ds
\end{aligned}
\tag{4.1}
$$

Thus, the value

$$
C^0(\mathcal{Q}) = \left(\ell p_1 + \sqrt{\ell^2 p_1^2 + p_2}\right)^2 \tag{4.2}
$$

can be taken as an upper bound in(1.7) for any vector $\mathcal{Q}$ from the set $\mathcal{B}$. Below we show that there exists a vector $\mathcal{Q}^*$ from $\mathcal{B}$ such that

$$
C^0(\mathcal{Q}) \ge C^0(\mathcal{Q}^*). \tag{4.3}
$$

We define it in the following way. Let P be positive definite matrix of size $k \times k$ being the solution of the matrix equation

$$
2\ell\sqrt{tr\ P} I + aP + Pa^T - PA^T AP = 0, \tag{4.4}
$$

where matrices a and A are defined in (2.3).

Theorem 4.1 *For every $\mathcal{Q}$ from $\mathcal{B}$, the vector $\mathcal{Q}^* = PA^T$ obeys (4.3).*

Proof: The proof of the theorem consists of a few steps.

1. Let the vector $\mathcal{Q}$ be fixed. Introduce a positive definite matrix U of size $k \times k$ such that $tr\ U = C^\circ(\mathcal{Q})$. To this end, taking into account that the eigen-values of the matrix $A^{\mathcal{Q}}$ defined in (3.1) have negative real parts, consider the Lyapunov equation

$$2\ell\sqrt{c}I + A^{\mathcal{Q}}U + U(A^{\mathcal{Q}})^T + \mathcal{Q}\mathcal{Q}^T = 0 \tag{4.5}$$

the solution of which is the matrix

$$\begin{aligned} U(c) &= 2\ell\sqrt{c}\int_0^\infty \exp\Big(tA^{\mathcal{Q}}\Big)\exp\Big(t(A^{\mathcal{Q}})^T\Big)dt \\ &\quad + \int_0^\infty \exp\Big(tA^{\mathcal{Q}}\Big)\mathcal{Q}\mathcal{Q}^T\exp\Big(t(A^{\mathcal{Q}})^T\Big)dt. \end{aligned} \tag{4.6}$$

The matrix $U(c)$ is increasing in $c \nearrow$ in the sense that $c_1 \le c_2$ implies $U(c_1) \le U(c_2)$ (i.e. $U(c_2) - U(c_1)$ is nonnegative definite matrix). Choose the parameter c, such that $c = trU(c)$. This equation is equivalent to the quadratic algebraic equation with respect to $\sqrt{c}$: $\Big(\sqrt{c}\Big)^2 - 2\ell\sqrt{c}p_1 - p_2 = 0$ (positive constants p_1 and p_2 are defined in (4.1)) which has the unique positive solution $\sqrt{c} = \sqrt{C^\circ(\mathcal{Q})}$, that is the matrix $U = U(C^\circ(\mathcal{Q}))$ is the unique solution in the class of positive definite matrices having the same tr as a solution of equation (4.5) with $\sqrt{c}$ replaced by $\sqrt{tr\ U}$ or, by virtue of (3.1), the solution of the equation below

$$2\ell\sqrt{tr\ U}I + aU + Ua^T + \Big[\mathcal{Q}\mathcal{Q}^T - \mathcal{Q}AU - UA^T\mathcal{Q}^T\Big] = 0. \tag{4.7}$$

Observe the following property of the matrix form in bracket

$$\begin{aligned} \mathcal{Q}\mathcal{Q}^T - \mathcal{Q}AU - UA^T\mathcal{Q}^T &= \Big(\mathcal{Q} - UA^T\Big)\Big(\mathcal{Q} - UA^T\Big)^T - UA^TAU \\ &\ge -UA^TAU \end{aligned}$$

which implies (compared with (4.4))

$$2\ell\sqrt{tr\ U}I + aU + Ua^T - UA^TAU \le 0. \tag{4.8}$$

The last inequality is the key to finding 'optimal' vector $\mathcal{Q}^*$ by eliminating $\Big(\mathcal{Q} - UA^T\Big)\Big(\mathcal{Q} - UA^T\Big)^T$ and therefore converting (4.8) into (4.4).

2. Consider an auxiliary filtering problem for the unobservable and observable signals ξ_t and η_t, being a vector of size k and a scalar respectively, defined by Itô's equation

$$
\begin{aligned}
d\xi_t &= a\xi_t dt + \sqrt{2\ell\sqrt{c}}I dw'_t, \\
\xi_0 &= 0, \\
d\eta_t &= A\xi_t dt + dw''_t, \\
\eta_0 &= 0,
\end{aligned}
$$

where a and A are matrices defined in (2.3), and where w''_t is the Wiener process independent of the vector Wiener process w'_t with independent components. Consider the Kalman type filter for ξ_t given observations η_s, $s \le t$ with the gain which is a vector $\mathcal{Q}$ from the set $\mathcal{B}$:

$$
\begin{aligned}
d\widehat{\xi}_t &= a\widehat{\xi}_t dt + \mathcal{Q}(d\eta_t - A\widehat{\xi}_t dt), \\
\widehat{\xi}_0 &= 0.
\end{aligned}
$$

One can derive the following matrix differential equation for the mean square error $U_t = E(\xi_t - \widehat{\xi}_t)(\xi_t - \widehat{\xi}_t)^T$

$$
\begin{aligned}
\frac{dU_t}{dt} &= aU_t + U_t a^T + 2\ell\sqrt{c}I + \Big[\mathcal{Q}\mathcal{Q}^T - \mathcal{Q}AU_t - U_t A^T \mathcal{Q}^T\Big], \\
U_0 &= 0.
\end{aligned}
\tag{4.9}
$$

On the other hand, the Kalman filter for the model considered is well known

$$
\begin{aligned}
d\overline{\xi}_t &= a\overline{\xi}_t dt + P_t A^T(d\eta_t - A\overline{\xi}_t dt), \\
\overline{\xi}_0 &= 0,
\end{aligned}
\tag{4.10}
$$

where P_t is defined by Ricatti's equation

$$
\begin{aligned}
\frac{dP_t}{dt} &= aP_t + P_t a^T + 2\ell\sqrt{c}I - P_t A^T A P_t, \\
P_0 &= 0.
\end{aligned}
$$

Since the vector $\mathcal{Q}$ belongs to the set $\mathcal{B}$, the stability of the matrix differential equations (4.9) is guaranteed. The last together with the formula for the matrix $A^{\mathcal{Q}}$ (see (3.1)) imply $\lim_{t\to\infty} U_t (= U(c))$ being the solution of Lyapunov's equation (4.5).

Further, taking into account that the pair of matrices a, A obeys the following property $\begin{pmatrix} A \\ Aa \\ . \\ . \\ . \\ Aa^{k-1} \end{pmatrix} = I$, i.e. the rank of the above-mentioned block matrix is equal k, and the matrix $2\ell\sqrt{c}I$ is non singular, one can conclude (see e.g. Theorem 16.2 from [5]) that also the limit $\lim_{t\to\infty} P_t (= P(c))$ exists, where $P(c)$ is the positive definite matrix being the solution of the algebraic equation

$$aP(c) + P(c)a^T + 2\ell\sqrt{c}I - P(c)A^T AP(c) = 0. \tag{4.11}$$

Since for every $t \geq 0$ the matrix P_t defines the minimal mean square filtering error we have $U_t \geq P_t, t \geq 0$ and so the same property is inherited by their limits $U(c)$ and $P(c)$: $U(c) \geq P(c)$ and consequently for any $c > 0$, $tr\ U(c) \geq tr\ P(c)$. Now, evaluate from below $tr\ P(c)$. Denote by $\lambda_j(c), j = 1, .., k$ the eigen-values of the matrix $P(c)$. Let ϕ_j be the right eigen vector of the matrix $P(c)$ corresponding to $\lambda_j(c)$. Without the loss of generality one can assume that $||\phi_j|| = 1$. Then we get the following bounds $|\phi_j^T a\phi_j| \leq ||a||$ and $\phi_j A^T A\phi_j^T \leq 1$. Multiplying both sides of (4.11) from the left by ϕ^T and from the right by ϕ and taking into account the above-mentioned bounds, we arrive at the inequality $-2||a||\lambda_j(c) + 2\ell\sqrt{c} - \lambda_j^2(c) \leq 0$ which implies the same lower bound for all $\lambda_j(c)$: $\lambda_j(c) \geq -||a|| + \sqrt{||a||^2 + 2\ell\sqrt{c}}$. Therefore

$$tr\ P(c) = \sum_{j=1}^{k} \lambda_j(c) \geq k\Big(-||a|| + \sqrt{||a||^2 + 2\ell\sqrt{c}}\Big)\ (:= g(c)).$$

Thus, we arrive at the upper and lower bounds:

$$tr\ U(c)(\equiv 2\ell\sqrt{c}p_1 + p_2) \geq tr\ P(c) \geq g(c)$$

3. Letting $\inf\{\emptyset\} = \infty$, define three constants

$$\begin{aligned} c^+ &= \inf\{c > 0 : 2\ell\sqrt{c}p_1 + p_2 = c\}, \\ c* &= \inf\{c > 0 : tr\ P(c) = c\}, \\ c^- &= \inf\{c > 0 : g(c) = c\}. \end{aligned}$$

Evidently constants c^+ and c^- are finite and since $c^- \leq c^* \leq c^+$ also $0 < c^* < \infty$. Consequently c^* is the solution of the equation $c^* = tr\ P(c^*)$, or in other words $P(c^*)$ is the solution of (4.4). Now, taking into account the inequality $c^* \leq c^+$ and noticing that $c^+ = C^\circ(\mathcal{Q})$, we have

$$c^* = tr\ P(c^*) \leq tr\ U(c^*) \leq tr\ U(C^\circ(\mathcal{Q})) = C^\circ(\mathcal{Q}),$$

that is c^* is the lower bound for $C^\circ(\mathcal{Q})$.

Thus it remains to show only that this lower bound is attainable at the vector $\mathcal{Q}^* = PA^T$, i.e. $c^* = C^\circ(\mathcal{Q}^*)$. Note first that PA^T is the gain for the Kalman filter (4.10) for stationary regime that is, the differential equations (2.11) are stable provided that the vector PA^T has the parameters $q_1, ..., q_k$ as its components. Thereby PA^T belongs to the set $\mathcal{B}$. Consider now equation (4.7) with $\mathcal{Q}$ replaced by $\mathcal{Q}^*$. It is clear that

$$\begin{aligned} 0 &= 2\ell\sqrt{tr\ U^*}I + aU^* + U^*a^T + \Big[\mathcal{Q}^*(\mathcal{Q}^*)^T - \mathcal{Q}^*AU^* - U^*A^T(\mathcal{Q}^*)^T\Big] \\ &= 2\ell\sqrt{tr\ U^*}I + aU^* + U^*a^T + \Big((P-U^*)A^T\Big)\Big((P-U^*)A^T\Big)^T - U^*A^TAU^* \end{aligned}$$

and, due to the uniqueness of solution for equation (4.7) in the class of matrices with the same trace, we obtain $U^* = P$.

Hence, $c^* = tr\ U^* = C^\circ(\mathcal{Q}^*)$.□

5 Appendix

Denote by $\Delta_\varepsilon = \gamma(\varepsilon)\varepsilon^{2/(2\beta+1)}$, where $\gamma(\varepsilon)$ is a function decreasing to 0, as $\varepsilon \to 0$. Consider a boundary layer $[0, \Delta_\varepsilon]$. Let $\pi^\varepsilon = \{S_\varepsilon^{(0)}(t), S_\varepsilon^{(j)}(t), j = 1, ..., k-1\}$ be an estimator for the signal $S(t)$ and its derivatives $S^{(j)}(t), j = 1, ..., k-1$ based on the observation of $X_s, s \leq t$. Denote by Π_ε a class of all such estimators.

Lemma 5.1. *For every $j = 0, 1, ..., k-1$, $0 < t < \Delta_\varepsilon$, and $\delta > 0$*

$$\liminf_{\varepsilon\to 0} \inf_{S_\varepsilon^{(j)}\in\Pi_\varepsilon} \sup_{S\in\Sigma^{k,\alpha}(\ell)} P\Big((\gamma(\varepsilon))^{(\beta-j)/2\beta}\Big|\frac{S_\varepsilon^{(j)}(t) - S^{(j)}(t)}{\varepsilon^{2(\beta-j)/(2\beta+1)}}\Big| > \delta\Big) > 0.$$

Proof. The proof uses the approach from [3], [4]. Take as a signal the function

$$S_\varepsilon(s,\theta) = \theta\varepsilon^{2\beta/(2\beta+1)}(\gamma(\varepsilon))^{-1/2}\varphi((s-t)\varepsilon^{-2/(2\beta+1)}(\gamma(\varepsilon))^{1/(2\beta)}), \tag{5.1}$$

where $\varphi : R \to R$ is arbitrary sufficiently smooth and compactly supported function such that $\varphi(0) \neq 0$ and also its derivatives obey the same property: $\varphi^{(j)}(0) \neq 0, j = 1, ..., k-1$, and where θ is an unknown parameter taking values in the compact set $\Theta = [-a, a]$. It can be easily verified that, under the above-mentioned assumptions, the function $S_\varepsilon(s, \theta)$ belongs to the class $\Sigma^{k,\alpha}(\ell)$. Taking into account that in this case the observation process is defined as (compared with (1.1))

$$\begin{aligned} dX^\varepsilon_{t'} &= S_\varepsilon(t', \theta)dt' + \varepsilon dW_{t'}, \\ X^\varepsilon_0 &= 0 \end{aligned} \tag{5.2}$$

one can find the Fisher information corresponding to time interval Δ_ε:

$$\begin{aligned} I_\varepsilon &= \frac{1}{\varepsilon^2}\int_0^{\Delta_\varepsilon} \Big[\frac{\partial S_\varepsilon(t', \theta)}{\partial \theta}\Big]^2 dt' \\ &= \varphi^2(0) + o(1). \end{aligned}$$

Note now that

$$\begin{aligned} S_\varepsilon(t, \theta) &= \theta\varepsilon^{2\beta/(2\beta+1)}(\gamma(\varepsilon))^{-1/2}\varphi(0) \\ S^{(j)}_\varepsilon(t, \theta) &= \theta\varepsilon^{2(\beta-j)/(2\beta+1)}(\gamma(\varepsilon))^{-(\beta-j)/(2\beta)}\varphi^{(j)}(0), \ j = 1, ..., k-1. \end{aligned}$$

Therefore estimating of the parameter θ by the observation $X^\varepsilon_{t'}, t' \leq t$ is equivalent to estimating of the values

$$\begin{aligned} &S_\varepsilon(t, \theta)\varepsilon^{-2\beta/(2\beta+1)}(\gamma(\varepsilon))^{1/2} \\ &S^{(j)}_\varepsilon(t, \theta)\varepsilon^{-2(\beta-j)/(2\beta+1)}(\gamma(\varepsilon))^{-(j-\beta)/(2\beta)}(j = 1, ..., k-1). \end{aligned}$$

A quality of estimation for the parameter θ for a wide class of loss function can be characterized by the Le Cam-Hajek minimax lower bound (see e.g. Remark 2.12.3 in [4]). So, applying this lower bound to the loss function $\mathcal{W}(x) = \mathbf{I}_{\{|x|>\delta\}}$ and taking into account $\{S_\varepsilon(t', \theta), \theta \in \Theta\} \subset \Sigma^{k,\alpha}(\ell)$ we verify the statement of the lemma. $\square$

Thus, it follows from the lemma that the band width of a boundary layer for the estimator $S^{(j)}_\varepsilon(t)$ can not be less, uniformly in the class $\Sigma^{k,\alpha}(\ell)$, than $c\varepsilon^{2/(2\beta+1)}$. In fact, restricting ourselves by the quadratic loss function, we have by Chebyshev's inequality that for arbitrary decreasing function $\gamma(\varepsilon)$ and for $0 < t < \Delta_\varepsilon$

$$\begin{aligned} &\liminf_{\varepsilon\to 0} \inf_{S^{(j)}_\varepsilon \in \Pi_\varepsilon} \sup_{S\in\Sigma^{k,\alpha}(\ell)} E\Big((\gamma(\varepsilon))^{(\beta-j)/(2\beta)}\frac{S^{(j)}_\varepsilon(t) - S^{(j)}(t)}{\varepsilon^{2(\beta-j)/(2\beta+1)}}\Big)^2 \\ \geq\ &\delta^2 \liminf_{\varepsilon\to 0} \inf_{S^{(j)}_\varepsilon \in \Pi_\varepsilon} \sup_{S\in\Sigma^{k,\alpha}(\ell)} P\Big((\gamma(\varepsilon))^{(\beta-j)/2\beta}\Big|\frac{S^{(j)}_\varepsilon(t) - S^{(j)}(t)}{\varepsilon^{2(\beta-j)/(2\beta+1)}}\Big| > \delta\Big) > 0. \end{aligned}$$

Acknowledgements : The authors gratefully acknowledge the careful review by the anonymous referee. Due to his comments the paper could be significantly improved.

References

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

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

[3] I. Ibragimov and R. Khasminskii *Statistical estimation: Asymptotic theory.* Springer Verlag. 1981

[4] I. Ibragimov and R. Khasminskii "Estimation of signal, its derivatives, and point of maximum for Gaussian distributions". *Theory of Probab. Appl.*, **15**, No 4, 1980, pp. 703–720.

[5] R.Sh. Liptser and A.N. Shiryayev *Statistics of Random Processes, II.* Springer Verlag. 1978

[6] M.G. Kalachev "One method of multiple differentiation applied to a signal in automatic regulation systems". *Automation and Remote Control*, No 6, 1970, pp. 890–896.