ON-LINE TRACKING OF A SMOOTH REGRESSION FUNCTION

GOLDENTAYER, L. AND LIPTSER, R.

Abstract. We construct an on-line estimator with equidistant design for tracking a smooth function from Stone-Ibragimov-Khasminskii class. This estimator has the optimal convergence rate of risk to zero in sample size. The procedure for setting coefficients of the estimator is controlled by a single parameter and has a simple numerical solution. The off-line version of this estimator allows to eliminate a boundary layer. Simulation results are given.

1. Introduction.

In this paper, we consider a tracking problem for smooth function \( f = f(t), \, 0 \leq t \leq T \), under observation

\[ X_{in} = f(t_{in}) + \sigma \xi_i, \tag{1.1} \]

for \( t_{in} = \frac{i}{n}, \quad i = 0, \ldots, n \) (\( n \) is large), where \( (\xi_i) \) is a sequence of i.i.d. random variables with \( E\xi_i = 0, \quad E\xi_i^2 = 1 \), and \( \sigma^2 \) is a positive constant. Without additional assumptions on the function \( f \) it is difficult to create an estimator even for large \( n \). The filtering approach, see Bar-Shalom and Li [1], proposes an estimator in the form of Kalman filter corresponding to a stochastic model for \( f \), e.g. \( f \) is differentiable \( k \) times, and \( k \)-th derivatives of \( f \) is simulated by a white noise with a certain intensity. Since \( f \) is deterministic function, the non-trivial part of such approach is a choice of filter parameters and asymptotic analysis of estimation risk in \( n \to \infty \). On the other hand, nonparametric statistic approach to the regression estimation of a function \( f \) assumes that \( f \) belong to some limited class. We take \( f \) from the class \( \Sigma(\beta, L) \) (introduced by Stone, [10], [11] and Ibragimov and Khasminskii, [3], [4]) of \( k \) times continuously differentiable functions with Hölder continuous last derivative (here \( f^{(0)} = f \), \( L \) and \( \alpha \) are the same for any function from the class):

\[ \Sigma(\beta, L) = \left\{ f : \begin{array}{c} \text{obeys } k \text{ derivatives, } f^{(0)}, f^{(1)}, \ldots, f^{(k)}; \\ \left| f^{(k)}(t_2) - f^{(k)}(t_1) \right| \leq L|t_2 - t_1|^{\alpha}, \forall t_1, t_2, \, \alpha \in (0, 1]; \end{array} \right\}. \]

It is known [10], [11], [3], [4] that there are kernel type estimators \( \hat{f}_n^{(j)}(t) \) of \( f^{(j)}(t), \quad j = 0, 1, \ldots, k \) such that for a wide class of loss functions \( L(\ast) \) and...
Remark 1. The left side boundary layer $c(q)n^{-\frac{1}{2\beta+1}} \log n$ is due to on-line limitations of the above tracking system. One can readily suggest an off-line modification with the same recursion in the backward time subject to some boundary conditions independent of observation $X_i$’s. This modification possesses the right side boundary layer $[T - c(q)n^{-\frac{1}{2\beta+1}} \log n, T]$ and accuracy (1.3) on $[0, T - c(q)n^{-\frac{1}{2\beta+1}} \log n]$. So, some combination of the forward and
backward time tracking algorithms allows (1.3) accuracy on \([0, T]\). For instance, a combination of forward time tracking on the interval \([\frac{T}{2}, T]\) and backward time tracking on \([0, \frac{T}{2}]\) can be used.

In this paper, we deal with the estimator given in (1.3) and restrict ourselves by considering \(f\) from the class \(\Sigma(L, k + 1)\), i.e. the class of \(k\)-times differentiable functions \(f\) with Lipschitz continuous \(f^{(k)}(t)\).

A suitable choice of filtering gain \(q\) should satisfy multiple requirements regarding the cost function \(C(q)\) and parameter \(c(q)\), involved in the description of boundary layer. Moreover, a correct choice of \(q\) should guarantee that the roots of characteristic polynomial \(p^k(u, q)\) are different and have negative real parts. These requirements might contradict each other. To avoid contradictions, we use the fact that estimator (1.3) has a structure of Kalman filter. We build a Kalman filter according to Bar-Shalom and Li [1], so that \(f^{(k)}(t)\) is generated by a white noise with intensity \(\gamma\). For each \(\gamma\), we choose the Kalman gain \(q(\gamma)\) and use it for minimization of \(C(q(\gamma))\) in \(\gamma\). So, the minimization problem of the cost function is controlled by single parameter and allows to establish a reasonable relationship between \(C(q)\) and \(c(q)\).

Moreover, this type of minimization automatically guarantees negative real parts of the roots for characteristic polynomial \(p^k(u, q)\). For \(k \leq 4\) the roots of \(p^k(u, q)\) are different and numerical verification of the same fact for \(k > 4\) is available.

2. The filter gain choice

2.1. Preliminaries. Henceforth \(\beta = k + 1\). For notational convenience we describe our problem in matrix notation. Introduce the following matrices:

\[
\begin{align*}
\mathbf{a} &= \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & 0 & \ldots & 0 \end{pmatrix} \in \mathbb{R}^{(k+1) \times (k+1)}, \\
\mathbf{A} &= \begin{pmatrix} 1 & 0 & \ldots & 0 \end{pmatrix} \in \mathbb{R}^{1 \times (k+1)}, \\
\mathbf{b} &= \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^{(k+1) \times 1}.
\end{align*}
\]

Notice that by Lemma 3.1 in [2] the roots of \(p^k(u, q)\) and eigenvalues of \((\mathbf{a} - q\mathbf{A})\) coincide. In accordance with this remark, while eigenvalues of \((\mathbf{a} - q\mathbf{A})\) have negative real parts we may describe the cost function \(C(q)\) in terms of the bias \(\tilde{M}(q)\) and variance \(P(q)\) for tracking errors (see, (3.11)) (hereafter * is the transposition symbol):

\[
C(q) = \text{trace} \left( P(q) + \tilde{M}(q)\tilde{M}^*(q) \right)
\]

where \(\tilde{M}(q) = L(\mathbf{a} - q\mathbf{A})^{-1}\mathbf{b}\) and the matrix \(P(q)\) solves the Lyapunov equation \((\mathbf{a} - q\mathbf{A})P(q) + P(q)(\mathbf{a} - q\mathbf{A})^* + \sigma^2 qq^* = 0\). In Section 3, we select...
the filter gain $q(\gamma)$ from one-parameter family
\[
\Gamma = \left\{ \gamma \geq \gamma_\varepsilon > 0 : q(\gamma) = \frac{Q(\gamma)A^*}{\sigma^2} \right\}
\]
where $Q(\gamma)$ is a positive definite matrix given by the algebraic Riccati equation
\[
aQ(\gamma) + Q(\gamma)a^* + \gamma^2 b^* b - \frac{Q(\gamma)A^* A Q(\gamma)}{\sigma^2} = 0.
\]
Finally we choose
\[
\gamma^\circ = \arg\min_{\gamma \in \Gamma} C(q(\gamma))
\]
and the filter gain $q(\gamma^\circ)$. A relevant choice of $\gamma_\varepsilon$ allows to have an acceptable
value of the constant $c(q(\gamma^\circ))$.

2.2. Explicit formulae. In Section 5, we show that the cost function is
expressed as:
\[
C(q) = \sigma^2 \left( \int_0^\infty q^* e^{(a-qA)^* t} e^{(a-qA) t} q dt + \left( \frac{L}{\sigma} \right)^2 \left[ \left( \frac{1}{q_k} \right)^2 + \sum_{j=0}^{k-1} \left( \frac{q_j}{q_k} \right)^2 \right] \right).
\]
Furthermore, we give the explicit structure of $q(\gamma)$ as a function of the
control parameter $(\gamma/\sigma)^{1/(k+1)}$. Namely
\[
q_0(\gamma) = U_{00} \left( \frac{\gamma}{\sigma} \right)^{1/k+1},
q_1(\gamma) = U_{01} \left( \frac{\gamma}{\sigma} \right)^{2/k+1},
\]
\[\vdots\]
\[
q_k(\gamma) = U_{0k} \left( \frac{\gamma}{\sigma} \right),
\]
where $U_{ij}, i, j = 0, 1, \ldots, k$ are entries of the matrix $U$ being solution of the
algebraic Riccati equation
\[
aU + Ua^* + b^* b - U A^* A U = 0.
\]

2.3. Example 1. Here, we consider the tracking problem for Lipschitz continuous function $f$. Since $k = 0$, we have $a = 0$, $A = 1$, $b = 1$, $1 - U_{00} = 0$ and so, $q_0 = \frac{\gamma}{\sigma}$. Therefore,
\[
C(q(\gamma)) = \frac{\sigma \gamma}{2} + \frac{L^2 \sigma^2}{\gamma^2}.
\]
With $\gamma_\varepsilon < (2L)^{2/3} \sigma^{1/3}$, we have $\gamma^\circ = (2L)^{2/3} \sigma^{1/3}$ and $q(\gamma^\circ) = \left( \frac{2L}{\sigma} \right)^{2/3}$.
The following estimator is constructed (here $\hat{f}(t_i) := \tilde{f}^{(0)}_{n}(t_i)$)
\[
\hat{f}(t_i) = \hat{f}(t_{i-1}) + \left( \frac{2L}{n \sigma} \right)^{2/3} (X_i - \hat{f}(t_{i-1})).
\]

Remark 2. Notice that the direct minimization of $C(q) = P(q) + \tilde{M}^2(q)$
with respect to $q_0$ provides the optimal $q_0 = q(\gamma^\circ)$. For $k \geq 1$, this coincidence
is not guaranteed. Under the direct minimization of the cost function
$C(q)$ with respect to $q$ the eigenvalues of $(a - q^* A)$ might have nonnegative
real parts.
2.4. Example 2. Let us consider a numerical solution for $k = 2$, i.e. $f$ is twice differentiable function. Its second derivative is Lipschitz continuous with constant $L = 100$. For $\sigma = 0.25$, we find (see Figure 2) $\gamma^\circ = 24.533$ and $\gamma^\circ / \sigma = 98.132$. According to Table 1, $U_{00} = 2$, $U_{01} = 2$, $U_{02} = 1$. Hence $q_0 = 9.225$, $q_1 = 42.550$, and $q_2 = 98.132$. So, the following estimator is constructed

$$\hat{f}(0)(t_i) = \hat{f}(0)(t_{i-1}) + 1/n \hat{f}(1)(t_{i-1}) + \frac{9.225}{n^{6/7}} (X_i - \hat{f}(0)(t_{i-1}))$$

$$\hat{f}(1)(t_i) = \hat{f}(1)(t_{i-1}) + 1/n \hat{f}(2)(t_{i-1}) + \frac{42.550}{n^{5/7}} (X_i - \hat{f}(0)(t_{i-1}))$$

$$\hat{f}(2)(t_i) = \hat{f}(2)(t_{i-1}) + \frac{98.132}{n^{4/7}} (X_i - \hat{f}(0)(t_{i-1})).$$

The combination of forward and backward tracking practically allows to eliminate the boundary layer (see Figure 2).

3. Error analysis

In this section, we present a vector of normalized tracking errors and derive the expression for $C(q)$. 
3.1. **Notation.** For notational convenience, set

\[ F(t_i) = \begin{pmatrix} f^{(0)}(t_i) \\ f^{(1)}(t_i) \\ \vdots \\ f^{(k)}(t_i) \end{pmatrix}, \quad \widehat{F}^n(t_i) = \begin{pmatrix} \widehat{f}^{(0)}_n(t_i) \\ \widehat{f}^{(1)}_n(t_i) \\ \vdots \\ \widehat{f}^{(k)}_n(t_i) \end{pmatrix} \]

and introduce a diagonal matrix \( C_n \) and vector \( q_n \):

\[ C_n = \text{diag}(n^{\frac{\beta}{2\beta+1}}, n^{\frac{\beta-1}{2\beta+1}}, \ldots, n^{\frac{2-k}{2\beta+1}}) \quad \text{and} \quad q_n = \begin{pmatrix} q_0 n^{-\frac{2\beta}{2\beta+1}} \\ q_1 n^{-\frac{2\beta-1}{2\beta+1}} \\ \vdots \\ q_k n^{-\frac{2\beta-k}{2\beta+1}} \end{pmatrix}. \]

We use a vector-matrix form of estimator (1.3)

\[ \widehat{F}^n(t_i) = \widehat{F}^n(t_{i-1}) + \frac{1}{n} a \widehat{F}^n(t_{i-1}) + q_n (X_i - A \widehat{F}(t_{i-1})) \]  

(3.1)

and obvious identity

\[ F(t_i) \equiv F(t_{i-1}) + \frac{1}{n} a F(t_{i-1}) + b (f^{(k)}(t_i) - f^{(k)}(t_{i-1})). \]  

(3.2)

3.2. **Normalized errors.** Denote \( \delta_i = \widehat{F}(t_i) - F(t_i) \) and introduce normalized error \( \Delta_i = C_n \delta_i \). Recursions (3.1) and (3.2) provide

\[ \delta_i = \delta_{i-1} + \frac{1}{n} a \delta_{i-1} + q_n \sigma \xi_i - q_n A \delta_{i-1} + (f(t_i) - f(t_{i-1})) q_n - b (f^{(k)}(t_i) - f^{(k)}(t_{i-1})). \]

Multiplying both sides of this equation from the left by \( C_n \) we find

\[ \Delta_i = \Delta_{i-1} + \frac{1}{n} C_n a \delta_{i-1} + C_n q_n \sigma \xi_i - C_n q_n A \delta_{i-1} + (f(t_i) - f(t_{i-1})) C_n q_n - C_n b (f^{(k)}(t_i) - f^{(k)}(t_{i-1})). \]  

(3.3)

A special structure (see [6]) of the objects involved in (3.3)

\[ C_n a = n^{\frac{1}{2\beta+1}} a C_n \]

\[ C_n q_n = n^{-\frac{\beta}{2\beta+1}} q \]

\[ \frac{1}{n} C_n a \delta_{i-1} = n^{-\frac{2\beta}{2\beta+1}} a \Delta_{i-1} \]

\[ C_n q_n A \delta_{i-1} = n^{-\frac{2\beta}{2\beta+1}} q A \Delta_{i-1} \]

\[ C_n b = n^{\frac{1}{2\beta+1}} b \]

allows to simplify (3.3) significantly:

\[ \Delta_i = \Delta_{i-1} + n^{-\frac{2\beta}{2\beta+1}} (a - q A) \Delta_{i-1} + n^{-\frac{\beta}{2\beta+1}} q \sigma \xi_i \]

\[ + \ n^{-\frac{1}{2\beta+1}} q (f(t_i) - f(t_{i-1})) - n^{-\frac{1}{2\beta+1}} b (f^{(k)}(t_i) - f^{(k)}(t_{i-1})). \]
With $D_n = I + n^{-\frac{2\beta}{2\beta+1}}(a - qA)$ we rewrite the recursion for $\Delta_i$’s to

$$
\Delta_i = D_n \Delta_{i-1} + n^{-\frac{\beta}{2\beta+1}} q \sigma_i
$$

and

$$
n^{-\frac{1}{2\beta+1}} q(f(t_i) - f(t_{i-1})) - n^{-\frac{1}{2\beta+1}} b(f^{(k)}(t_i) - f^{(k)}(t_{i-1})).
$$

In Proposition 4.1 in [3], it is shown that for $n$ large enough the magnitudes of all eigenvalues of $D_n$ are strictly less than 1. We shall use this property for asymptotic analysis and continuous time approximation.

### 3.3. Normalized bias and variance

Denote the bias and variance of the normalized error $\Delta_i$:

$$
M^n_i = E \Delta_i \quad \text{and} \quad P^n_i = E(\Delta_i - M^n_i)(\Delta_i - M^n_i)^*.
$$

Taking the expectation from both side of (3.4) we find

$$
M^n_i = D_n M^n_{i-1} + n^{-\frac{1}{2\beta+1}} q \sigma_i
$$

From (3.4) and (3.5), we get $(\Delta_i - M^n_i) = D_n(\Delta_{i-1} - M^n_{i-1}) + n^{-\frac{\beta}{2\beta+1}} q \sigma_i$, so that $P^n_i = E(\Delta_i - M^n_i)(\Delta_i - M^n_i)^*$ is defined by the recursion

$$
P^n_i = D_n P^n_{i-1} + n^{-\frac{2\beta}{2\beta+1}} q \sigma^n.
$$

Since $|f^{(k)}(t_i) - f^{(k)}(t_{i-1})| \leq \frac{L}{n}$, it is natural to choose $f$ with

$$
f^{(k)}(t_i) - f^{(k)}(t_{i-1}) \equiv \frac{L}{n} \quad \text{or} \quad -\frac{L}{n}
$$

and substitute sup $f \in \mathcal{F}(\beta, L)$ $M^n_i(M^n_i)^*$ by $\tilde{M}_n(M^n_i)^*$. Henceforth, $\tilde{M}_n$ is defined by recursion (3.3) with such $f$, i.e.

$$
\tilde{M}_n = \tilde{M}_{n-1} + n^{-\frac{2\beta}{2\beta+1}} (a - qA) \tilde{M}_{n-1} + n^{-\frac{2\beta}{2\beta+1}} q \sigma^n - n^{-\frac{2\beta}{2\beta+1}} L b,
$$

where $\sigma^n = \frac{L q}{(k+1)n^2}$. Thus, trace $(P^n_i + \tilde{M}_n(M^n_i)^*)$ determines the normalized mean square tracking error

$$
C(q) = \lim_{n \to \infty} \text{trace} \left( P^n_i + \tilde{M}_n(M^n_i)^* \right), \quad t_i \geq c(q)n^{-\frac{1}{2\beta+1}} \log n.
$$

### 3.4. Continuous time approximation

To find $\lim_{n \to \infty}$ in (3.6), we give a continuous time approximation of $(M^n_i, P^n_i)$. To this end, let us introduce the time stretching

$$
(t_i - t_{i-1}) = n^{-1} \quad \Rightarrow \quad (s_i - s_{i-1}) = n^{-\frac{2\beta}{2\beta+1}},
$$

with $t_0 = s_0 = 0$. The boundary layer $[0, c(q)n^{-\frac{1}{2\beta+1}} \log n]$ is transformed to $[0, c(q) \log n]$ and the interval $[0, T]$ to $[0, T n^{\frac{2\beta}{2\beta+1}}]$. Let us define $\tilde{M}_s = \tilde{M}_n$ and $P_s = P^n_i$, $i = 0, 1, \ldots$, and for $s \in [s_{i-1}, s_i)$

$$
\tilde{M}_s = \tilde{M}_{s_{i-1}} + \int_{s_{i-1}}^{s} (a - qA) \tilde{M}_{s'_{i-1}} ds' + (\sigma^n - bL)(s - s_{i-1})
$$

$$
P_s = P^n_{s_{i-1}} + \int_{s_{i-1}}^{s} (a - qA) P^n_{s_{i-1}} + P^n_{s_{i-1}} (a - qA)^* + n^{-\frac{2\beta}{2\beta+1}} (a - qA) P^n_{s_{i-1}} (a - qA)^* + \sigma^n q \sigma^n ds'.
$$

(3.7)
We shall consider these recursions for $s_i$ from $(c(q) \log n, n^{\frac{1}{x+n} T}]$, where $(\tilde{M}_n^q, P_n^q)$ have entries bounded in $n$ (see, [3]). Taking into account that recursions (3.4) are homogeneous in $s$, let us replace $\{s, s_i\}$ by $\{u = s - c(q) \log n$ and $u_i = s_i - c(q) \log n\}. Then, the entries of $(\tilde{M}_u^n, P_u^n)$ are bounded in $n$ for $0 \leq u \leq n^{\frac{1}{x+n} T} - c(q) \log n$. Therefore without loss of generality we may consider (3.7) with initial conditions bounded in $n$:

$$C(q) = \lim_{u \to \infty} \lim_{n \to \infty} \text{trace} \left( P_u^n + \tilde{M}_u^n (\tilde{M}_u^n)^* \right).$$

To determine the right hand side of (3.8), we apply the Arzela-Ascoli theorem. For any $T > 0$, the functions $(M_u^n, P_u^n)_{0 \leq u \leq T}$ are uniformly bounded and equicontinuous. So, by the Arzela-Ascoli theorem, any converging subsequence $(\tilde{M}_u^n, P_u^n)$ obeys the limit $(\tilde{M}_u^n, P_u^n)$ in the local uniform topology:

$$\lim_{n' \to \infty} \sum_{N=1}^{\infty} \frac{1}{2 N} \min \left( 1, \sup_{0 \leq u \leq N} \left( \| \tilde{M}_u^n - \tilde{M}_u^n' \| + \| P_u^n - P_u^n' \| \right) \right) = 0,$$

where

$$\tilde{M}_u^n = \tilde{M}_0 + \int_0^u \left( (a - qA) \tilde{M}_v^n + bL \right) dv$$

$$P_u^n = P_0 + \int_0^u \left( (a - qA) P_v^n + P_v^n (a - qA)^* + \sigma^2 q q^* \right) dv.$$

Since the eigenvalues of $a - qA$ have negative real parts, the limits $\tilde{M}(q) := \lim_{u \to \infty} \tilde{M}_u^n$ and $P(q) := \lim_{u \to \infty} P_u^n$ exist and are defined as:

$$\tilde{M}(q) = -L (a - qA)^{-1} b$$

$$a - qA) P(q) + P(q) (a - qA)^* + \sigma^2 q q^* = 0,$$

that is $\tilde{M}(q)$ and $P(q)$ are independent of $\{n'\}$ and so

$$C(q) = \text{trace} \left( P(q) + \tilde{M}(q) \tilde{M}^*(q) \right).$$

4. Minimization of the cost function in one parameter class

4.1. Motivation. For large values of $k$ a direct minimization of $C(q)$ from (3.11) would be a difficult problem. Moreover,

$$q^o = \arg \min_q C(q)$$

could not a priori guarantee negative real parts of eigenvalues for $(a - q^o A)$.

To avoid implementation of a conditional minimization procedure, we propose to choose $q$ from some limited class given below.

4.2. Adaptation to Kalman filter design. Our estimator has a structure of Kalman filter in the discrete time. We assume that $F(0)$ is a random vector, $F(0) = EF(0)$, and $f^{(k)}(t_i)$ is generated by stochastic recursion

$$f^{(k)}(t_i) = f^{(k)}(t_{i-1}) + n^{\frac{1}{x+n} - 1} \gamma n_i,$$

where $(n_i)$ is a white noise, independent of $(\xi_i)$, with $E n_i = 0$, $E \eta_i = 1$ and $\gamma$ is an arbitrary nonzero parameter. For the observation model

$$X_i = f^{(0)}(t_{i-1}) + \sigma \xi_i$$

...
we apply the estimator given in (3.1). The resulting errors \( \delta_i = \tilde{F}(t_i) - F(t_i), \) \( i = 1, \ldots, \) are defined by a recursion

\[
\delta_i = \delta_{i-1} + \frac{1}{n} a \delta_{i-1} + q \sigma \xi_i - q_n A \delta_{i-1} - n^{-\frac{\delta+1}{2\delta+1}} \beta \gamma \eta_i.
\]

Then, for \( \Delta_i = C_n \delta_i, \) \( i \geq 1 \) we obtain

\[
\Delta_i = \Delta_{i-1} + n^{-\frac{2\beta+1}{2\beta+1}} (a - q A) \Delta_{i-1} + n^{-\frac{2\beta}{2\beta+1}} q \sigma \xi_i - n^{-\frac{2\beta}{2\beta+1}} \beta \gamma \eta_i \tag{4.2}
\]

and supply \( \Delta_0 = \delta_0. \) Denote \( Q^n_i = E \Delta_i \Delta^*_i. \) From (4.2) it follows

\[
Q^n_i = \left( I + n^{-\frac{2\beta}{2\beta+1}} (a - q A) \right) Q^n_{i-1} \left( I + n^{-\frac{2\beta}{2\beta+1}} (a - q A) \right)^* + n^{-\frac{2\beta}{2\beta+1}} \sigma^2 q q^* + n^{-\frac{2\beta}{2\beta+1}} \gamma^2 b b^* + D_n Q^n_{i-1} D_n^* + n^{-\frac{2\beta}{2\beta+1}} \sigma^2 q q^* + n^{-\frac{2\beta}{2\beta+1}} \gamma^2 b b^*.
\]

Similar to \( (\tilde{M}^n_i, \tilde{P}^n_i) \) (see previous section), let us introduce \( Q^n_s: \)

\[
Q^n_s = Q^n_{s_{i-1}} + \int_{s_{i-1}}^{s} \left( (a - q A) Q^n_{s_{i-1}} + Q^n_{s_{i-1}} (a - q A)^* + n^{-\frac{2\beta}{2\beta+1}} (a - q A) Q^n_{s_{i-1}} (a - q A)^* + \sigma^2 q q^* + \gamma^2 b b^* \right) ds'.
\]

Applying the Arzela-Ascoli theorem technique it can be readily shown that \( Q^n_s \) converges in the local uniform topology to \( Q_s, \) where

\[
Q_s = Q_0 + \int_0^s \left( (a - q A) Q_s + Q_s (a - q A)^* + \sigma^2 q q^* + \gamma^2 b b^* \right) ds,
\]

and \( \lim_{s \to \infty} Q_s := Q \) with \( Q \) being the unique solution of Lyapunov equation

\[
(a - q A) Q + Q (a - q A)^* + \sigma^2 q q^* + \gamma^2 b b^* = 0. \tag{4.3}
\]

The matrix \( Q \) is a function of arguments \( q \) and \( \gamma: Q = Q(q, \gamma). \) We choose \( q = q(\gamma) \) so that for any \( \gamma \)

\[
Q(q, \gamma) \geq Q(q(\gamma), \gamma) := Q(\gamma) > 0. \tag{4.4}
\]

Due to the Kalman filtering theory, the lower bound (4.4) holds true for

\[
q(\gamma) = \frac{Q(\gamma) A^*}{\sigma^2} \tag{4.5}
\]

with \( Q(\gamma) \) being solution of the algebraic Riccati equation

\[
a Q(\gamma) + Q(\gamma) a^* + \gamma^2 b b^* - \frac{Q(\gamma) A^* A Q(\gamma)}{\sigma^2} = 0. \tag{4.6}
\]

It is well known (see e.g. Theorem 16.2 in [7]) that (4.6) possesses a unique positive-definite solution provided that block-matrices

\[
G_1 = \begin{pmatrix}
A \\
A a \\
\vdots \\
A a^k
\end{pmatrix} \quad \text{and} \quad G_2 = (b b^* \ a b b^* \ \ldots \ a^k b b^*)
\]

have full ranks \( r = k + 1. \) Notice that \( G_1 \) is a unite matrix and the rank of \( G_2 \) is \( k + 1. \) Consequently, the eigenvalues of the matrix \( (a - q(\gamma) A) \) with \( q(\gamma) \) defined in (4.3) have negative real parts (see, Lemma 16.11 in [7]).
4.3. Minimization of the cost function. The one parameter family

$$\Gamma = \{ \gamma \geq \gamma_\varepsilon > 0 : q(\gamma) = \frac{Q(\gamma)A^*}{\sigma^2} \}$$

permits a simple numeric implementation and guarantees filtering stability mentioned above. In this class we use a constrain parameter \( \gamma_\varepsilon \) to compensate unacceptably large boundary layer when the minimization procedure yields small values of \( \gamma^0 = \arg\min_{\gamma > 0} C(q(\gamma)) \), with \( C(q(\gamma)) \) given in (3.11). The magnitude of \( \gamma_\varepsilon \) is dictated by \( k, \sigma \), initial conditions and boundary layer specifications. The minimization in our class provides

$$\gamma^0 = \arg\min_{\gamma \geq \gamma_\varepsilon} C(q(\gamma)) \quad \text{and} \quad q(\gamma^0) = \frac{Q(\gamma^0)A^*}{\sigma^2}. \quad (4.7)$$

5. Explicit minimization procedure

5.1. Filter gain. In this section we describe a structure of \( q(\gamma^0) \). Recall that \( q(\gamma) = \frac{Q(\gamma)A^*}{\sigma^2} \) and \( Q(\gamma) \) solves the Riccati equation (4.6) for any fixed \( \sigma \). For notational convenience replace \( Q(\gamma) \) by \( Q(\gamma, \sigma) \) and set \( U = Q(1, 1) \). Clearly, \( U \) solves the algebraic Riccati equation

$$aU + Ua^* + bb^* - UA^*AU = 0.$$

Kalachev [5] shows that

$$Q_{ij}(\gamma, \sigma) = U_{ij}\sigma^2\left(\frac{\gamma}{\sigma}\right)^{i+j+1}, \quad i, j = 0, 1, \ldots, k,$$

where \( Q_{ij}(\gamma, \sigma) \) and \( U_{ij} \) are entries of \( Q(\gamma, \sigma) \) and \( U \) respectively. Hence,

$$q_0(\gamma) = U_{00}\left(\frac{\gamma}{\sigma}\right)^{1/k+1} \quad q_1(\gamma) = U_{01}\left(\frac{\gamma}{\sigma}\right)^{2/k+1} \quad \cdots \quad q_k(\gamma) = U_{0k}\left(\frac{\gamma}{\sigma}\right).$$

For \( k \leq 4 \), these values are given in the table below.

| \( k \) | \( U_{00} \) | \( U_{01} \) | \( U_{02} \) | \( U_{03} \) | \( U_{04} \) |
|--------|--------|--------|--------|--------|--------|
| 0      | 1      | NA     |        |        |        |
| 1      | \sqrt{2} | 1      | NA     | NA     | NA     |
| 2      | 2      | 2      | 1      | NA     | NA     |
| 3      | \sqrt{4 + \sqrt{8}} | 2 + \sqrt{2} | \sqrt{4 + \sqrt{8}} | 1 | NA |
| 4      | 1 + \sqrt{5} | 3 + \sqrt{5} | 3 + \sqrt{5} | 1 + \sqrt{5} | 1 |

(Table 1)

The complex structure of \( C(q(\gamma)) \) does not provide an insight of \( \gamma \) and \( L \) connection. Numerical simulations show that for a wide range of values \( \log(\gamma^0) \) is almost proportional to \( \log(L) \) (see also Figure 3). This remark enables to construct a simple interpolation tables for the values of \( \gamma^0 \) and \( C(q(\gamma)) \) with respect to the parameter \( L \).
5.2. Eigenvalues of \((a - q(\gamma)A)\). Although the eigenvalues of \(q(\gamma)\) have negative real parts, we may not formally guarantee that they are different. So, for \(k \leq 4\) we give the eigenvalues:

\[
\begin{align*}
    k &= 0 : -\left(\frac{\gamma}{\sigma}\right) \\
    k &= 1 : -\left(\frac{\gamma}{\sigma}\right)^{1/2} \left(\frac{1}{\sqrt{2}} \pm i \frac{1}{\sqrt{2}}\right) \\
    k &= 2 : -\left(\frac{\gamma}{\sigma}\right)^{1/3} \left(1; \frac{1}{2} \pm i \frac{\sqrt{3}}{2}\right) \\
    k &= 3 : -\left(\frac{\gamma}{\sigma}\right)^{1/4} \left(0.924 \pm i0.383; 0.383 \pm i0.924\right) \\
    k &= 4 : -\left(\frac{\gamma}{\sigma}\right)^{1/5} \left(1; 0.809 \pm i0.588; 0.309 \pm i0.951\right).
\end{align*}
\]

For \(k > 4\), the fact that the polynomial has different roots should be verified.

Notice that \((\gamma/\sigma)^{1/(k+1)}\) is a natural control parameter defining the size of boundary layer and should be limited from below by \((\gamma_\varepsilon/\sigma)^{1/(k+1)}\) with appropriate \(\gamma_\varepsilon\).
5.3. **Cost function.** The index form of \((a - qA)\tilde{M}(q) = -Lb\) i.e.

\[
\begin{pmatrix}
-q_0 & 1 & 0 & 0 & \cdots & 0 \\
-q_1 & 0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
-q_{k-1} & 0 & 0 & 0 & \cdots & 1 \\
-q_k & 0 & 0 & 0 & \cdots & 0 \\
\end{pmatrix}
\begin{pmatrix}
\tilde{M}_0(q) \\
\tilde{M}_1(q) \\
\vdots \\
\tilde{M}_{k-1}(q) \\
\tilde{M}_k(q)
\end{pmatrix}
= -L
\begin{pmatrix}
0 \\
0 \\
\vdots \\
0 \\
1
\end{pmatrix}
\]

allows to find the solution

\[
\tilde{M}_0(q) = L \frac{1}{q_k}, \quad \tilde{M}_1(q) = L \frac{q_0}{q_k}, \ldots, \quad \tilde{M}_k(q) = L \frac{q_{k-1}}{q_k}.
\]

Thus, trace \(\tilde{M}(q)\tilde{M}^*(q)\) = \(L^2 \left( \frac{1}{q_k} \right)^2 + \sum_{j=0}^{k-1} \left( \frac{q_j}{q_k} \right)^2 \). From (B.9) it follows

\[
P(q) = \sigma^2 \int_0^{\infty} q e^{(a-qA)\tau} e^{(a-qA)\tau} q dt.
\]

As a result, the final expression for the cost function is

\[
C(q) = \sigma^2 \left( \int_0^{\infty} q e^{(a-qA)\tau} e^{(a-qA)\tau} q dt + \left( \frac{L}{\sigma} \right)^2 \left[ \frac{1}{q_k} \right]^2 + \sum_{j=0}^{k-1} \left( \frac{q_j}{q_k} \right)^2 \right).
\]

### 6. CONCLUSION REMARK

In this paper, we use the fact that a class of Kalman filters, being adapted to a nonparametric statistic setting, provides the optimal rate of convergence in sample size \((n \to \infty)\). We show how to evaluate a normalized risk function for large sample size and minimize that value in some subclass of Kalman filters with constant filter gain. The Kalman type estimator, as any online estimator, has inevitable boundary layer. We suggest to reduce the boundary layer by interpolation procedure and limitation from below for filtering gain.

### REFERENCES

1. Bar-Shalom, Yaakov and Li, Xiao-Rong Estimation and tracking. Principles, techniques, and software. Artech House, Inc., Boston, MA, 1993.
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. Ibragimov, I. and Khasminskii, R. (1980) On nonparametric estimation of regression, Soviet Math. Dokl. 21, pp. 810–814.
4. Ibragimov, I. and Khasminskii, R. Statistical estimation: Asymptotic theory. Springer Verlag, 1981 (Russian ed.1979).
5. 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.
6. Khasminskii, R and Liptser, R. (2001) On-line estimation of a smooth regression function
7. R.Sh. Liptser and A.N. Shiryaev Statistics of Random Processes, II. Springer Verlag, 2000
8. Parzen, E. (1962) On estimation of a probability density function and mode. Ann. Math. Statist., 33, No 3, pp. 1065–1073.
9. Rosenblatt, M. (1956) Remarks on some nonparametric estimates of a density function. Ann. Math. Statist., 27, No 3, pp. 832–837.
[10] Stone, C. (1980) Optimal rates of convergence for nonparametric estimators, *Ann. Statist.*, 8, pp.1348–1360.
[11] Stone, C. (1982) Optimal global rates of convergence for nonparametric regression, *Ann. Statist.*, 10, pp.1040–1053.

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

*E-mail address*: goldlev@eng.tau.ac.il

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

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