Proceedings of the World Congress on Engineering 2011 Vol I
WCE 2011, July 6 - 8, 2011, London, U.K.

Regularized Data-Based Nonparametric Filtration of Stochastic Signals

Alexander V. Dobrovidov, and Gennady M. Koshkin

***Abstract*—The data-based filtration method is proposed on the basis of the recent results for bandwidth selection by using smoothed cross-validation procedure. The optimal regularization procedure was developed to obtain the stable nonparametric estimator of filtration. Simulation has shown a high quality of the proposed filtration estimators as compared with the optimal Kalman filter.**

***Index Terms*—Bandwidth selection, kernel estimates, nonparametric filtration, regularization**

I. Introduction

MORE then twenty years ago in [1] there was proposed the filtration method of a stochastic signal with an unknown distribution from mixture with the noise. It was assumed that the noise distribution in the observation model is known due to the principal opportunity to observe the noise without a signal (for instance, in hydroacoustics), and one can restore the noise distribution from the noise observations. Inverse situation – the signal observation without a noise – is unreal one. So, in this situation the estimation of the signal distribution is impossible, and therefore the signal distribution is assumed to be unknown.

Let $(S_n, X_n)_{n\geq 1}$, $S_n \in \mathbb{R}^m$, $X_n \in \mathbb{R}^l$ be partly observable random sequence, where S_n and X_n are unobservable and observable components of this sequence. The problem is to estimate the vector S_n or the known function $Q(S_n)$ from the observations $x_1^n = (x_1,...,x_n)^{\mathrm{T}}$ of $(X_n)_{n\geq 1}$. The optimal mean square estimate of $Q(S_n)$ is the conditional mean $\hat{Q}(S_n) = \mathsf{E}(Q(S_n) \mid x_1^n)$.

The principal result of the theory of filtration is to obtain the optimal filtering equation for $\hat{Q}(S_n)$ not depending on unknown distribution of a signal S_n. This is possible for a class of observation models when the observation conditional density under the fixed signal $S_n = s_n$ belongs to the following *conditionally-exponent* family of densities [4]:

$$f(x_n \mid s_n, x_{n-L}^{n-1}) = \tilde{C}(s_n; x_{n-L}^{n-1}) V(x_n; x_{n-L}^{n-1}) \exp\{T^{\mathrm{T}}(x_{n-L}^n) Q(S_n)\}, \quad (1)$$

where $T = (T_1,...,T_m)^{\mathrm{T}}$; $Q = (Q^{[1]},...,Q^{[m]})^{\mathrm{T}}$; $V(\cdot)$, $Q^{[j]}(\cdot)$ and $T_j(\cdot)$, $(j = 1,...,m)$ are the given Borelean functions, and $\tilde{C}(\cdot)$ is the normalizing factor, $1 \leq L \leq n-1$.

In general case, under condition (1), the equation for the optimal estimate $\hat{Q}(S_n)$ takes the form

$$\Lambda^{\mathrm{T}}(x_{n-L}^n) \hat{Q}(s_n) = \nabla_{x_n} \ln \frac{f(x_n \mid x_1^{n-1})}{V(x_{n-L}^n)}, \quad (2)$$

where Λ is the matrix of size $m \times l$ with the elements $\lambda_{ij} = \partial T_i(x_{n-L}^n)/\partial x_n^{[j]}$ $(i = 1,...,m,\ \ j = 1,...,l)$, ∇ denotes the gradient operator, and $f(x_n \mid x_1^{n-1})$ is the conditional density depending on observable variables. We use the same notation $f(\cdot)$ for the densities of different variables without anxiety of ambiguity because the exact function form is not important now. Note that equation (2) is independent on a priori characteristics of an unobservable signal (S_n), and it is not recurrent. As the conditional density $f(x_n \mid x_1^{n-1})$ in equation (2) is unknown, we will estimate $f(x_n \mid x_1^{n-1})$ from the dependent observations x_1^n by using methods of nonparametric statistics.

In Section 2, the main idea of equation derivation for the optimal mean square estimate $\hat{Q}(S_n)$ is illustrated by the example of the Gaussian conditional density of observation. The nonparametric counterpart of the optimal equation is also derived. Methods of bandwidth selection for kernel density and derivative estimates are stated in Section 3. The regularization problems of unstable nonparametric estimates are considered in Section 4. Section 5 presents the simulation results to compare the optimal Kalman filtration with the nonparametric filtration.

II. Basic Models and Optimal Equation

Compare the proposed algorithm with the Kalman filter, which is optimal when all the statistical information about signals and noises is available. Note that for the Kalman filter the state and observation equations should be linear. As an example we consider a scalar autoregressive process

$$S_{n+1} = aS_n + b\xi_n, \quad n \geq 1, \quad (3)$$

where ξ_n is the Gaussian noise. The observation process is described by the additive model

$$X_n = AS_n + \sigma\eta_n, \quad n \geq 1, \quad (4)$$

where A and σ are the known constants and η_n is the standard Gaussian noise.

The optimal Kalman filter can be obtained from these equations. It will be used in Section 5, devoted to simulation

Manuscript received March 5, 2011. This work was supported by the Russian Foundation for Basic Research, project no. 09-08-00595-a and Program no. 29 of the RAS Presidium.

A. V. Dobrovidov is with Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia (phone: (8-495)-334-79-59; e-mail: dobrovid@ipu.ru).

G. M. Koshkin is with Tomsk State University and with Department of Informatization Problem, Tomsk Scientific Center SB RAS, Tomsk, Russia (phone: 8-(382-2)-529-828; fax: 8-(382-2)-529-895; e-mail: kgm@mail.tsu.ru).

ISBN: 978-988-18210-6-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

results, and therefore the constants a and b in (3) are not specified now.

In the case under consideration, state equation (3) is unknown and we have only observation equation (4). Note that the assignment of equation (4) exactly corresponds to the assignment of the conditionally-Gaussian density

$$f(x_n \mid s_n) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(x_n - As_n)^2}{2\sigma^2} \right\},\ x_n \in \mathbb{R}^1,\ s_n \in \mathbb{R}^1 \qquad (5)$$

belonging to the family (1).

If $w(s_n \mid x_1^n)$ is a posteriori density of an unobservable signal S_n, then $\int w(s_n \mid x_1^n) ds_n = 1$. So, using the total probability formula, we have

$$f^{-1}(x_1^n)\int w(s_n, x_1^n) ds_n = f^{-1}(x_1^n)\int w(s_n, x_1^{n-1}) f(x_n \mid s_n, x_{n-L}^{n-1}) ds_n$$
$$= f^{-1}(x_1^n)\int w(s_n \mid x_1^{n-1}) f(x_1^{n-1}) f(x_n \mid s_n, x_{n-L}^{n-1}) ds_n$$
$$= f^{-1}(x_n \mid x_1^{n-1})\int w(s_n \mid x_1^{n-1}) f(x_n \mid s_n, x_{n-L}^{n-1}) ds_n = 1. \qquad (6)$$

Transpose the multiplier $f(x_n \mid x_1^{n-1})$ in the right side of the last equation (6), differentiate it with respect to x_n and obtain

$$\frac{d}{dx_n} f(x_n \mid x_1^{n-1}) = \int \frac{d}{dx_n} f(x_n \mid s_n, x_{n-L}^{n-1}) w(s_n \mid x_1^{n-1}) ds_n, \qquad (7)$$

where, according to equation (4), the conditional density $f(x_n \mid s_n, x_{n-L}^{n-1}) = f(x_n \mid s_n)$, and its derivative is

$$\frac{d}{dx_n} f(x_n \mid s_n) = \sigma^{-2}(As_n - x_n) f(x_n \mid s_n). \qquad (8)$$

Substitution of (8) in (7) provides

$$\frac{d}{dx_n} f(x_n \mid x_1^{n-1}) = \int \sigma^{-2}(As_n - x_n) f(x_n \mid s_n) w(s_n \mid x_1^{n-1}) ds_n$$
$$= f^{-1}(x_1^{n-1}) \int \sigma^{-2}(As_n - x_n) w(s_n, x_1^n) ds_n$$
$$= f(x_n \mid x_1^{n-1}) \sigma^{-2} \left(A \int s_n w(s_n \mid x_1^n) ds_n - x_n \right), \qquad (9)$$

and the exact equation for the optimal filtering estimate $\hat{S}_n = \int s_n w(s_n \mid x_1^n) ds_n$ can be written as

$$\frac{A}{\sigma^2} \hat{S}_n = \frac{d}{dx_n} \ln f(x_n \mid x_1^{n-1}) + \frac{x_n}{\sigma^2}. \qquad (10)$$

Equation (10) contains the logarithmic derivative of the conditional density of observations and does not contain any characteristics of an unobservable signal (S_n). If the probability distribution of a signal (S_n) is unknown, then the distribution of an observable signal (X_n) is unknown too. Therefore, equation (10) can not be used directly. However, relying on the strong stationarity of the sequence (X_n), the logarithmic derivative of the density in (10) may be estimated from observations x_1^n. As the logarithmic density derivative has the form

$$\frac{d}{dx_n} \ln f(x_n \mid x_1^{n-1}) = \frac{\frac{d}{dx_n} f(x_1^n)}{f(x_1^n)}, \qquad (11)$$

then according to a plug-in method it is necessary to estimate the derivative and density separately. For large n, i.e., for a long realization of the sequence (X_n), a dimension of the multivariate density in (10) is very high. Therefore, taking into account a strong mixing condition of the sequence (X_n), accepted in this approach, one can replace (with a small error) the conditional density $f(x_n \mid x_1^{n-1})$ by the truncated conditional density $f(x_n \mid x_{n-\tau}^{n-1})$, $1 \le \tau \le n-1$, where the number τ is called a *degree of dependence* and represents an order of connectivity of the Markov process approximating the non-Markovian process (X_n). Then equality (11) takes the form

$$\frac{d}{dx_n} \ln f(x_n \mid x_{n-\tau}^{n-1}) = \frac{\frac{d}{dx_n} f(x_{n-\tau}^n)}{f(x_{n-\tau}^n)} \doteq \psi(x_{n-\tau}^n). \qquad (12)$$

The denominator in (12) is a $(\tau+1)$-dimensional marginal density. The nonparametric density estimate of a small dimension can be obtained by using the single series realization x_1^n, which is divided into the overlapping fragments $x_k^{k+\tau}$, $1 \le k \le n-\tau-1$ of the length $\tau+1$. All the realizations contain $N = n-\tau-1$ fragments. The last fragment $x_{n-\tau}^n$ is used as the argument of the function $\hat{f}_h(\cdot)$ in formula (13). The nonparametric kernel estimates of a density and its derivative in (12) have the forms

$$\hat{f}_h(x_{n-\tau}^n) = n^{-1} h^{-(\tau+1)} \sum_{i=1}^{n-\tau-1} \prod_{j=1}^{\tau+1} K\left(\frac{x_{n-j+1} - x_{n-j-i+1}}{h} \right), \qquad (13)$$

$$\hat{f}_{h_1}^{(1)}(x_{n-\tau}^n) = n^{-1} h_1^{-(\tau+2)} \sum_{i=1}^{n-\tau-1} K'\left(\frac{x_{n-j+1} - x_{n-j-i+1}}{h_1} \right)$$
$$\cdot \prod_{j=1}^{\tau+1} K\left(\frac{x_{n-j+1} - x_{n-j-i+1}}{h_1} \right), \qquad (14)$$

where K' denote the partial derivatives with respect to x_n. So, the nonparametric estimate of the logarithmic density derivative $\psi(x_{n-\tau}^n)$ can be written as

$$\hat{\psi}_n(x_{n-\tau}^n) = \frac{\hat{f}_{h_1}^{(1)}(x_{n-\tau}^n)}{\hat{f}_h(x_{n-\tau}^n)}. \qquad (15)$$

To calculate (15) it needs to select bandwidths h and h_1 in (13) and (14).

III. Bandwidth Selection for Densities and Their Derivatives

For the time being, several data-based selection methods of bandwidths are known of which the methods of cross-validation CV [2, 3], smoothed cross-validation SCV [4], and plug-in [5] seem to be the basic ones as the most clear and rapidly converging procedures. In [6] the method SCV, proposed in [7] for density estimation, was extended to the kernel estimates of density derivatives. The SCV method generates data-based bandwidth estimates with a higher rate of convergence and substantially smaller scatter than in the CV method.

Take a measure of distance between $f^{(r)}(\cdot)$ and its estimator $\hat{f}_h^{(r)}(\cdot)$ as the mean integrated square error $(MISE)$

$$MISE_r(h) = \mathsf{E}\int \left(\hat{f}_h^{(r)}(x) - f^{(r)}(x) \right)^2 dx,$$

$r = 0,\ 1,\quad f^{(0)}(x) = f(x).$

This criterion depends on the bandwidth h and it would be natural to select such an h, which will minimize the $MISE_r(h)$. Using the aforementioned SCV method and Gaussian kernels $K(\cdot)$ in (13) provides [9]

ISBN: 978-988-18210-6-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$SCV(h) = \frac{1}{2\sqrt{\pi} nh} + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \left(\varphi_{\sqrt{2h^2+2g^2}} - 2\varphi_{\sqrt{h^2+2g^2}} + \varphi_{\sqrt{2}g} \right)(x_i - x_j), \quad (16)$$

where $\varphi_t(u)$ is a Gaussian density with zero mean and standard deviation t, and a new constant g is responsible for the data presmoothing. Select g by minimization of the mean square error (MSE) of the bandwidth estimate $\hat{h}(g)$, which minimizes (16):

$$\hat{g} = \left(\frac{15}{16\sqrt{\pi}\nu_6} \right)^{1/7} n^{-1/7}, \quad (17)$$

where

$$\nu_k = \int f^{(k)}(x) f(x) dx, \quad k = 0,1,...,8. \quad (18)$$

Analogous technique provides an estimate for the $MISE_1$ of the derivative in a more complicated form [6]

$$SCV_1(h_1) = \frac{1}{4\sqrt{\pi} n h_1^3} + \frac{1}{n}\left(\frac{1}{4\sqrt{\pi} g^3} - \frac{2}{\sqrt{2\pi}(h_1^2 + 2g^2)^{3/2}} \right) + \frac{1}{n} \frac{(n-1)/n}{\sqrt{2\pi}(2h_1^2 + 2g^2)^{3/2}}$$
$$+ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{2g^2 - (x_i - x_j)^2}{(2g^2)^2} \varphi_{(\sqrt{2g})}(x_i - x_j)$$
$$- 2\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{h_1^2 + 2g^2 - (x_i - x_j)^2}{(h_1^2 + 2g^2)^2} \varphi_{(h_1^2+2g^2)^{1/2}}(x_i - x_j)$$
$$+ \frac{n-1}{n} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{2h_1^2 + 2g^2 - (x_i - x_j)^2}{(2h_1^2 + 2g^2)^2} \varphi_{(2h_1^2+2g^2)^{1/2}}(x_i - x_j), \quad (19)$$

where g, minimizing the MSE of $\hat{h}_1(g)$, is as follows

$$\hat{g}_1 = \left(\frac{105}{32\sqrt{\pi}\nu_8} \right)^{1/9} n^{-1/9}. \quad (20)$$

Formulae (17) and (20) contain the parameters ν_6 and ν_8, which depend upon an unknown density $f(x)$ and its derivatives. They are also can be estimated using the cross-validation method for a density and the *rule of thumb* for a higher derivative.

According to the law of large numbers, integral (18) is approximated by the sum

$$\frac{1}{n} \sum_{i=1}^{n} f_{h,i}^{(k)}(X_i),$$

where $f_{h,i}^{(k)}(X_i)$ can be estimated by the CV method:

$$\hat{f}_{h,i}^{(k)}(X_i) = \frac{1}{n-1} \sum_{j \neq i}^{n} K_h^{(k)}(X_i - X_j). \quad (21)$$

Such estimates, unlike to (13), are estimates of the second level, where a less precision is admissible. For the Gaussian kernels $K(x) = \varphi_1(x)$, where $\varphi_1(x)$ is the standard normal density, the derivatives in (21) are calculated by making use of the well known formula

$$\varphi_1^{(k)}(x) = (-1)^k H_k(x) \varphi_1(x), \quad (22)$$

where $H_k(x)$ is the Hermitian polynomial, which may be found by the recurrent formula

$$H_{k+1}(x) = xH_k(x) - kH_{k-1}(x), \quad H_0(x) = 1, \quad k = 1,2,...$$

At last, the bandwidth h on the second level is found roughly from the observations by the *rule of thumb*:

$$\tilde{h} = 1{,}06\, \hat{\sigma}\, n^{-1/5},$$

where $\hat{\sigma}$ is the sample standard deviation, estimated from x_1^n.

As a result, we obtain the following data-based expressions:

$$\nu_6 = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{1}{\tilde{h}^6} \left(\frac{b_{ij}^6}{\tilde{h}^6} - 15\frac{b_{ij}^4}{\tilde{h}^4} + 45\frac{b_{ij}^2}{\tilde{h}^2} - 15 \right) \varphi_{\tilde{h}}(b_{ij}),$$

$$\nu_8 = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{1}{\tilde{h}^8} \left(\frac{b_{ij}^8}{\tilde{h}^8} - 28\frac{b_{ij}^6}{\tilde{h}^6} + 210\frac{b_{ij}^4}{\tilde{h}^4} - 420\frac{b_{ij}^2}{\tilde{h}^2} + 105 \right) \varphi_{\tilde{h}}(b_{ij}),$$

where $b_{ij} = (X_i - X_j)$.

IV. Regularized Estimate

Logarithmic density derivative estimate (15) is the special case of the plug-in estimate of a composite function $G(t_n(x))$, where $x \in \mathbb{R}^{\tau+1}$, $t_n : \mathbb{R}^{\tau+1} \to \mathbb{R}^m$, $G : \mathbb{R}^m \to \mathbb{R}^1$. In our case $m = 2$, $t_n = (t_{1n}, t_{2n})^{\mathrm{T}}$, $t_{1n} = \hat{f}_h(x_{n-\tau}^n)$, $t_{2n} = \hat{f}_{h_1}^{(1)}(x_{n-\tau}^n)$, $G(t_n) = t_{2n}/t_{1n}$. If the statistic t_n converges to a function t in the mean square sense as $n \to \infty$, then under some regularity conditions $G(t_n) \to G(t)$ in the same sense too.

Write the main regularity conditions:

1) the existence and boundedness of several derivatives of $G(t_n)$;
2) the sequence $\{|G(t_n)|\}$ is dominated by the number sequence $\{C_0 d_n^{\gamma}\}$, where C_0 is a constant, $d_n \to \infty$ as $n \to \infty$, and $0 \le \gamma < \infty$.

These conditions provide the mean square convergence of $G(t_n)$ to $G(t)$ [8].

If the mean Euclidean distance $\mathrm{E}\|t_n - t\| < \varepsilon$, $\varepsilon > 0$, then for a small ε the following equality holds:

$$G(t_n) - G(t) = \nabla G(\vartheta_n)(t_n - t), \quad \vartheta_n \in (t_n, t),$$

where ∇ is the gradient with respect to t. From here according to [8]

$$\left| \mathrm{E}(G(t_n) - G(t))^2 - \mathrm{E}(\nabla G(\vartheta_n)(t_n - t))^2 \right| = O(d_n^{-3/2}), \quad (23)$$

i.e., the mean square closeness of the composite functions $G(t_n)$ and $G(t)$ is replaced by the mean square closeness of more simple statistics t_n and t.

There are a number of cases when conditions 1) and 2) do not hold. For example, the function $G(t) = 1/t$ does not satisfy both the conditions, and the estimator $G(t_n) = 1/t_n$ becomes unstable because of its possible unboundedness. For the one-dimensional Gaussian density $f(x)$, we have $G = -x$. This function is unbounded on $\mathbb{R}^1$. As proposition (23) is valid only for bounded functions G, we apply here some procedure of regularization, called the *piecewise smooth approximation* [8]. In the special case the procedure coincides with the Tychonoff regularization method. Using this procedure, we obtain the following stable approximation of G:

$$\Phi(G(t), \delta_n) = \tilde{\Phi}(t, \delta_n) = \frac{G(t)}{1 + \delta_n |G(t)|^4},$$

ISBN: 978-988-18210-6-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

where $\delta_n > 0$ is a regularization parameter. As it is proved in [8], $\tilde{\Phi}(t_n, \delta_n)$ satisfies both the above mentioned conditions, and therefore is dominated by the power function of n. Moreover, $\tilde{\Phi}(t_n, \delta_n)$ converges to $G(t)$ in the mean square sense, i.e., as $\mathrm{E}\|t_n - t\| \to 0$ and $\delta_n \to 0$, then

$$\lim_{n\to\infty} \mathrm{E}(\tilde{\Phi}(t_n, \delta_n) - G(t))^2 = 0. \qquad (24)$$

The statistic $\hat{\psi}_n(x_{n-\tau}^n)$ in (15) is unstable when its denominator is close to zero. So, we use the stable estimate

$$\breve{\psi}_n(x_{n-\tau}^n) = \frac{\hat{\psi}_n(x_{n-\tau}^n)}{1 + \delta_n \,|\, \hat{\psi}_n(x_{n-\tau}^n)\,|^4}, \qquad (25)$$

where the regularization parameter δ_n has to be found. One can obtain an optimal parameter, which minimizes the mean square deviation of $\breve{\psi}_n(x_{n-\tau}^n)$ from $\psi(x_{n-\tau}^n)$ at each point $x_{n-\tau}^n$. But this approach is not so good for practice because a minimization procedure has to be repeated for each signal processing. We propose to make an optimization procedure only once before signal processing using the criterion of the $MISE$ for estimating the logarithmic density derivative with a weight function $\omega(\cdot)$, i.e.,

$$MISE(\delta_n) = \int u^2\left(\breve{\psi}_n(x_{n-\tau}^n)\right)\omega(x_{n-\tau}^n)dx_{n-\tau}^n,$$

$$u^2\left(\breve{\psi}_n(x_{n-\tau}^n)\right) \doteq \mathrm{E}(\breve{\psi}_n(x_{n-\tau}^n) - \psi(x_{n-\tau}^n))^2. \qquad (26)$$

To exist the criterion, we should select the weight function as $\omega(\cdot) = f^2(\cdot)$.

Calculating of the expectation of the ratio in (26) is laborious. According to (24), for the mean square convergence of the regularized estimate $\breve{\psi}_n(x_{n-\tau}^n)$ to the logarithmic density derivative $\psi(x_{n-\tau}^n)$ it is necessary that $\delta_n \to 0$ as $n \to \infty$. Therefore, under the assumption of a small δ_n we expand (25) with respect to the parameter δ_n and approximately obtain

$$\breve{\psi}_n(x_{n-\tau}^n) \approx \hat{\psi}_n(x_{n-\tau}^n) - \delta_n \hat{\psi}_n^5(x_{n-\tau}^n). \qquad (27)$$

Substituting (27) into the $MISE$ (26) and using Theorem 2 from [8], we receive

$$\int u^2\left(\breve{\psi}_n(x_{n-\tau}^n)\right) f(x_{n-\tau}^n)dx_{n-\tau}^n \approx \int H_1^2 u^2\left(\hat{f}'(x_{n-\tau}^n)\right) f(x_{n-\tau}^n)dx_{n-\tau}^n$$
$$+ 2\int H_1 H_2 \operatorname{cov}\left(\hat{f}'(x_{n-\tau}^n), \hat{f}(x_{n-\tau}^n)\right) f(x_{n-\tau}^n)dx_{n-\tau}^n$$
$$+ \int H_2^2 u^2\left(\hat{f}(x_{n-\tau}^n)\right) f(x_{n-\tau}^n)dx_{n-\tau}^n, \qquad (28)$$

where $H_1 = \dfrac{1 - 5\delta\psi^4}{f}$, $H_2 = \dfrac{-\psi + 5\delta\psi^5}{f}$.

Now, minimizing (26) with respect to δ, we find

$$\delta_{opt} = \frac{\int u^2\left(\hat{f}'\right)f(\cdot) - 2\int \psi \operatorname{cov}\left(\hat{f}', \hat{f}\right)f(\cdot) + \int \psi^2 u^2\left(\hat{f}\right)f(\cdot)}{5\int \psi^4 u^2\left(\hat{f}'\right)f(\cdot) - 10\int \psi^5 \operatorname{cov}\left(\hat{f}', \hat{f}\right)f(\cdot) + 5\int \psi^6 u^2\left(\hat{f}\right)f(\cdot)}. \qquad (29)$$

The integrals in the numerator and denominator of δ_{opt} depend on unknown densities. Therefore, they will be estimated from the observations.

The main parts of $u^2(\cdot)$ and $\operatorname{cov}(\cdot,\cdot)$ equal as $n \to \infty$

$$u^2\left(\hat{f}'\right) \approx \frac{f}{nh_n^3}\int \left(K^{(1)}(u)\right)^2 du + \frac{h_n^4}{4}\left(f^{(3)}\right)^2\left(\int u^2 K(u)du\right)^2,$$

$$\operatorname{cov}\left(\hat{f}', \hat{f}\right) \approx \frac{f}{nh_n^2}\int K^{(1)}(u)K(u)du + \frac{h_n^4}{4} f^{(3)} f^{(2)}\left(\int u^2 K(u)du\right)^2,$$

$$u^2\left(\hat{f}\right) \approx \frac{f}{nh_n}\int K^2(u)du + \frac{h_n^4}{4}\left(f^{(2)}\right)^2\left(\int u^2 K(u)du\right)^2.$$

Substituting these formulae into (29), we find δ_{opt}, in which it is necessary to estimate the following integrals:

$$J_k = \int \left(f^{(k)}(u)\right)^q f(u)du, \quad \nu = 0,\ldots,4, \quad q = 1,2,\ldots$$

It can be done by the CV method, described above.

V. Simulation Results

First, we generate a sequence of dependent observations using state equation (3) for S_n and observation equation (4) for X_n. The equation for the Kalman filter is well known and is not given here.

When the state equation is unknown, we use the nonparametric counterpart of optimal equation (10), which, taking into account expression (15), can be written as

$$\tilde{S}_n = \frac{B^2}{A}\hat{\psi}_n(x_{n-\tau}^n) + \frac{x_n}{A},$$

where

$$\hat{\psi}_n(x_{n-\tau}^n) = \frac{h_{1n}^{-(\tau+3)} \sum_{i=1}^{n-\tau-1} (x_{n-i} - x_n) \prod_{j=1}^{\tau} \exp\left(-\dfrac{(b_{ij})^2}{2h_{1n}^2}\right)}{h_n^{-(\tau+1)} \sum_{i=1}^{n-\tau-1} \prod_{j=1}^{\tau+1} \exp\left(-\dfrac{(b_{ij})^2}{2h_n^2}\right)}, \qquad (30)$$

$b_{ij} = x_{n-j+1} - x_{n-j-i+1}$. The plug-in nonparametric estimate $\hat{\psi}_n(x_{n-\tau}^n)$ is constructed from the realization of an observed sequence (X_n). Unfortunately, the plug-in estimate is unstable when the denominator of (30) is close to zero. In this case, the estimate may have spikes, which can be seen in Fig.1.

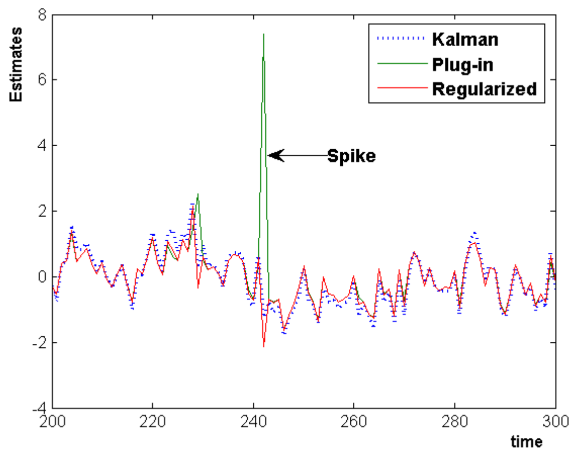

Fig. 1. Comparison of the nonparametric and optimal Kalman filtration when there are spikes.

These spikes sharply impaire the performance of the plug-in the nonparametric estimate (see Table 1). To eliminate the spikes, we use the regularized estimates, introduced in (25). This leads to the following regularized nonparametric equation:

$$\breve{S}_n = \frac{B^2}{A}\breve{\psi}_n(x_{n-\tau}^n) + \frac{x_n}{A}.$$

ISBN: 978-988-18210-6-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Comparison of nonparametric estimates $\tilde{S}_n$ and $\breve{S}_n$ with the optimal Kalman estimate $\hat{S}_n$ is carried out by calculating the relative error ε in percentage

$$\varepsilon = \frac{u_{non} - u_{kal}}{u_{kal}} 100,$$

where

$$u_{non} = (\tilde{u}_{non} \text{ or } \breve{u}_{non}), \quad \tilde{u}_{non} = \sqrt{\frac{1}{n}\sum_k (S_k - \tilde{S}_k)^2},$$

$$\breve{u}_{non} = \sqrt{\frac{1}{n}\sum_k (S_k - \breve{S}_k)^2}, \quad u_{kal} = \sqrt{\frac{1}{n}\sum_k (S_k - \hat{S}_k)^2}.$$

The nonparametric filtering estimates $\tilde{S}_n$, $\breve{S}_n$ and optimal Kalman estimate $\hat{S}_n$ are given in Fig. 1 and 2.

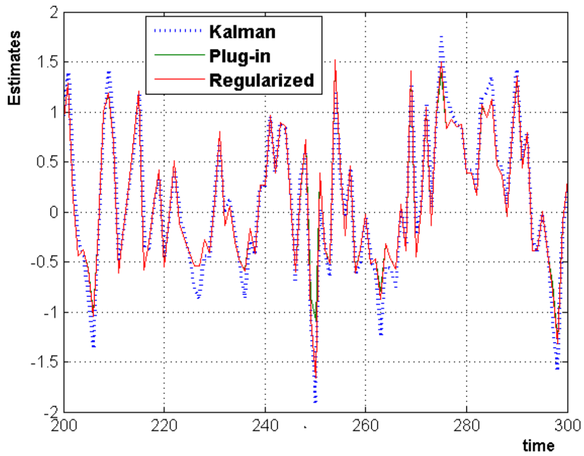

Fig. 2. Comparison of the nonparametric and optimal Kalman filtration without spikes.

One can see that the discrepancy ε between both the estimates is very small without spikes. But when spikes are present, the advantage of the regularization procedure becomes obvious. The distances between the nonparametric estimates $\tilde{S}_n$, $\breve{S}_n$ and optimal Kalman estimate $\hat{S}_n$ in the ε -units are given in Table 1.

TABLE 1
MEASURE OF CLOSENESS OF THE ESTIMATES $\tilde{S}_n$ AND $\breve{S}_n$ TO THE KALMAN ESTIMATE $\hat{S}_n$

Plug-in $\tilde{\varepsilon}$	Regularized $\breve{\varepsilon}$	Spikes
83.13%	1.42%	yes
1.13%	1.31%	no

It should be noted that the quality of the nonparametric filtering estimates depends strongly on the bandwidth and regularization parameters. So, the problem of theirs optimal selection is an important part of the signal processing.

VI. CONCLUSION

The new results in nonparametric bandwidth selection [2, 5] and regularization methods allow to synthesize the data-based algorithms of the nonparametric signal filtration. Such algorithms are based on the optimal filtering equation for partly observable stochastic sequences (not only Gaussian). This equation does not include the probability characteristics of an unobservable component of the sequence.

For the strong stationary sequences the nonparametric counterpart of the optimal equation was constructed in the theory of nonparametric signal processing. This approach was developed when the state equation and the probability distribution of an unobservable signal are unknown, and the stochastic observation equation is known completely. The estimation equation includes the kernel estimator of the logarithmic density derivative, which depends on bandwidths of density estimates and its derivatives.

The data-based filtration method is suggested, using the recent results of [6] and [7] for bandwidth selection by the SCV method. The optimal regularization procedure was developed to obtain the formula of the stable non-parametric algorithm of filtration.

Simulation, carried out to compare the behavior of the nonparametric filtration algorithms with the optimal Kalman filter, has showed a high quality of the proposed procedures.

REFERENCES

[1] A.V. Dobrovidov, "Nonparametric methods of nonlinear filtering of stationary random sequences", *Automat. and Remote Control*, vol.44, no. 6, pp. 757-768, Jun. 1983.

[2] A. Bowman, "An alternative method of cross-validation for the smoothing of density estimates", *Biometrica*, vol. 71, pp. 353-360, 1984.

[3] M. Rudemo, "Empirical Choice of Histograms and Kernel Density Estimators," *Scand. J. Stat*,. vol. 9, pp. 65-78, 1982.

[4] P. Hall, J. Marron, and B. Park, "Smoothed Cross-validation", *Prob. Theory Related Fields*, vol. 90, pp. 1-20, 1992.

[5] B. Park and J. Marron, "Comparison of data-driven bandwidth selectors", *J. Amer. Statist. Assoc.*, vol. 85, pp. 66-72, 1990.

[6] A. V. Dobrovidov and I. M. Rudko, "Bandwidth selection in nonparametric estimator of density derivative by smoothed cross-validation method", *Automat. and Remote Control*, vol. 71, no. 2, pp. 45-57, Feb. 2010.

[7] T. Duong and M. L. Hazelton, "Cross-validation bandwidth matrices for multivariate kernel density estimation", *Scand. J. Statist*, vol. 32, pp. 485-506, 2005.

[8] G. M. Koshkin, "Deviation moments of the substitution estimate and its piecewise smooth approximations", *Siberian Math. J.*, vol. 40, no. 3, pp. 515-527, Jun. 1999.

ISBN: 978-988-18210-6-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)