Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Regularized Nonparametric Interpolation of Unobservable Markov Sequences ⋆

A. Dobrovidov * & G. Koshkin **

* *Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia(e-mail: dobrovid@ipu.ru).*
** *Department of Applied Mathematics and Cybernetics, Tomsk State University, Tomsk, Russia (e-mail: kgm@mail.tsu.ru).*

Abstract: In this paper, the synthesis problem of interpolation algorithms for an unobservable stationary sequence in a partly observable (hidden) Markov process is considered. When the distributions of the compound Markov sequence are completely known, the problem solution can be found by applying the transformation equations for the posterior probability density of the unobservable sequence. This equations were firstly obtained for filtration and interpolation problems by Stratonovich (1966). Khazen (1978) managed to present the interpolation equation in the form of the product of filtration posterior probability densities in forward and backward time. The similar equation is also valid for dynamic observation models, but in this case the main equation is to be supplemented by another recursive equation connected with the dynamic properties of observations. Unfortunately, it is impossible to make use of this equations when the probability family for the unobservable sequence is unknown. However, for some conditional probability family of observations, the equation admits the representation which does not depend on statistical characteristics unknown *a priori*. The solution is based on the principles of the empirical Bayes approach and the theory of kernel non-parametric functional estimation. The solutions of the equation may be unstable in some points. Therefore the optimal regularization procedure was developed to obtain the stable nonparametric estimator of interpolation. Modeling showed a high quality of the proposed interpolation estimators as compared with the optimal backward interpolation.

Keywords: Interpolation, Markov sequence, unobservable component, nonparametric uncertainty, kernel estimates, regularization.

1. INTRODUCTION

This work is devoted to the problem of synthesis of interpolation algorithms for an unobservable stationary sequence $(S_n)_{n\geqslant 1}$ in conditional Markov scheme described by the compound Markov process $(S_n, X_n)_{n\geqslant 1}$. Classical solution for such problems consists in calculating the posterior density distribution of unobservable signals $(S_n)_{n\geqslant 1}$. The transformation equation for the posterior probability densities of unobservable sequences in filtration and interpolation problems were firstly obtained by Stratonovich (1966). Khazen (1978) was able to present the interpolation equation in the form of the normalizing product of filtration posterior probability densities in forward and backward time. This was done only for static observation models. The similar equation is also valid for dynamic observation models (see below Sections 2 and 3), but in this case the main equation is to be supplemented by another recursive equation connected with the dynamic properties of observations. Unfortunately, it is impossible to make use of this equations when the probability family for the unobservable sequence is unknown. However, for some conditional probability family of observations (determined the observation model), it is possible to transform these equations so as to eliminate dependence on statistic characteristics unknown *a priori*. The solution is found on the principles of the empirical Bayes approach and the theory of kernel non-parametric functional estimation (see Dobrovidov et al. (2012)). New results are mainly associated with an altered form of non-parametric estimation of the logarithmic derivative, its properties and the process of regularization of the estimation procedure to obtain stable estimators.

2. INTERPOLATION EQUATION FOR DYNAMIC OBSERVATION MODELS

Interpolation (smoothing) of partly observable (hidden) Markov random sequence $(S_n, X_n)_{n\geqslant 1}$, $S_n \in \mathcal{S} \subset \mathbb{R}^m$, $X_n \in \mathcal{X} \subset \mathbb{R}^l$, is the problem of constructing estimators of the unobservable vector S_k or a known one-to-one function $Q(S_k)$ by observations $x_1^n = (x_1, ..., x_n)^{\mathrm{T}}$ of the random sequence $(X_n)_{n\geqslant 1}$ for all $1 \leqslant k \leqslant n$. It is well known the optimal mean-square smoothing estimator of $Q(S_k), 1 \leqslant k \leqslant n$, is the conditional expectation

$$\mathsf{E}(Q(S_k)|x_1^n) = \int\limits_{\mathcal{S}} Q(s_k)\pi(s_k|x_1^n)ds_k, \tag{1}$$

⋆ Supported by the Russian Foundation for Basic Research, project no. 13-08-00744, and Tomsk State University Competitiveness Improvement Program.

Copyright © 2014 IFAC

where $\pi_k(s_k|x_1^n)$ is the posterior probability density of S_k given all observable realizations x_1^n, which is called the *interpolating* posterior density. There are some ways to calculate this conditional density. One way, examined below and referred to as the two-filter smoothing, is a recursive calculation of the *filtering* posterior probability density $w_k(s_k|x_1^k)$ in forward time and the *filtering* posterior density $\widetilde{w}_k(s_k|x_k^n)$ in backward time. This formula can be represented as follows Khazen (1978), Briers et al. (2003):

$$\pi_k(s_k|x_1^n) = \frac{f(x_1^k)f(x_k^n)}{f(x_1^n)} \cdot \frac{w_k(s_k|x_1^k)\widetilde{w}_k(s_k|x_k^n)}{f(s_k, x_k)}, \qquad (2)$$

where the first factor is a normalizing constant depending only on observations. Here and further the function f without index denotes any probability density that may differ from another one even in the same expression because its argument completely determines the object. For static and dynamic observation models, the algorithms for such computation are different. For the static models, described by the conditional density $f(x_k|s_k)$, the joint probability density $f(s_k, x_k)$ in the denominator of (2) can be represented in the form of the product $p(s_k)f(x_k|s_k)$, where $p(s_k)$ is the prior density. For dynamic observation models, described by the conditional density $f(x_k|x_{k-1}, s_k)$, such a product form can not be constructed because, in this case, the density $f(x_k|s_k)$ is unknown. Nevertheless, the joint probability density $f(s_k, x_k)$ can be represented via $f(x_k|x_{k-1}, s_k)$ by means of the recursive equation

$$f(s_k, x_k) = \int\limits_{\mathcal{S}_{k-1}} p(s_k|s_{k-1}) \int\limits_{\mathcal{X}_{k-1}} f(x_k|x_{k-1}, s_k) \times f(s_{k-1}, x_{k-1}) dx_{k-1} ds_{k-1} \qquad (3)$$

with the initial condition $f(s_1, x_1)$. It is given by the prior density of the composed Markov process $(S_n, X_n)_{n\geqslant 1}$. This equation is simply derived using the total probability formula. Consequently, pair (2), (3) is the system of equations for interpolating posterior probability density in the case of dynamic observation models.

The recursive computation in (2), (3) can be carried out if all the distributions of the composed Markov process $(S_n, X_n)_{n\geqslant 1}$ are known. To solve the interpolation problem with unknown distributions of the unobservable stationary Markov sequence $(S_n)_{n\geqslant 1}$, we use the empirical Bayes approach and the theory of non-parametric functional estimation by weakly dependent observations (see Dobrovidov et al. (2012)). According to the empirical Bayes approach, we have to construct the estimators that are explicitly independent of the probabilistic characteristics of the unobservable random variables. This can be done, for instance, by using the conditional densities of observations from the exponential density family (see Chentsov (1972)):

$$f(x_n|s_n) = \widetilde{C}(s_n)h(x_n)\exp\left\{T^{\mathrm{T}}(x_n)Q(s_n)\right\}, \qquad (4)$$

where $T = (T_1, \cdots, T_m)^{\mathrm{T}}; Q = (Q^{[1]}, \cdots, Q^{[m]})^{\mathrm{T}}; h(\cdot)$, $Q^{[j]}(\cdot)$, $T_j(\cdot)$, $j = \overline{1,m}$, are the given Borelean functions, and $\widetilde{C}(s_n)$ is the normalizing factor.

3. OPTIMAL INTERPOLATION ESTIMATOR EQUATION UNDER UNKNOWN PROBABILITY DISTRIBUTION OF UNOBSERVABLE SIGNAL

Problems of extracting a signal from the mixture with a noise have often to be solved when this signal is never observed, and therefore one can not get any statistics to restore its distribution. We propose to use the empirical Bayes approach to the interpolation problem and to find such a form of the estimation equation that does not depend explicitly on the unknown probability characteristics of the unobserved signal. Now there are no regular methods for constructing such equations for arbitrary observation models. So, we consider more narrow class of observation models, described by conditional densities from the exponential family (4). In this case, we construct an equation not for the signal posterior probability density (2), but directly for the optimal mean square estimator (1). The detailed derivation of this equation, based on the Markovian property of the sequence $(S_k, X_k)_{k\geqslant 1}$, have been fulfilled in Dobrovidov (2008). We present only some main steps of the derivation. First of all, it should be noted that the equation in the case of unknown signal distribution we construct only for the static observation models, described by conditional density $f(x_k|s_k)$. It was done by using only the equation (2). The equation synthesis for dynamic observation models with the additional recursive equation (3) is assumed to be implemented in the next time. Taking into account the Markovian property of the compound sequence $(S_k, X_k)_{k\geqslant 1}$, the equation (2) can be transformed to the expression

$$\pi_k(s_k|x_1^n) = \frac{\lambda_k(x_1^n \text{ without } x_k)}{f_k(x_k|x_1^n \text{ without } x_k)} f(x_k|s_k) \times w_k^{pr}(s_k|x_1^{k-1})\tilde{w}_k^{pr}(s_k|x_{k+1}^n)p^{-1}(s_k), \qquad (5)$$

where $w_k^{pr}(\cdot)$ and $\tilde{w}_k^{pr}(\cdot)$ are posterior *predicting* (one step ahead) densities in forward and backward time. Normalizing constants in this formula

$$\lambda_k = \lambda_k(x_1^n \text{ without } x_k) \triangleq \frac{f(x_1^{k-1})f(x_{k+1}^n)}{f(x_1^{k-1}, x_{k+1}^n)},$$
$$f_k(x_k|x_1^n \text{ without } x_k) \triangleq f(x_k|x_1^{k-1}, x_{k+1}^n)$$

depend only on statistical characteristics of observations (X_n). It should be noted that in equation (5) the conditional density $f(x_k|s_k)$ belongs to the exponential family (4) by assumption. With this in mind, we obtain a counterpart of equation (1) independent of the statistical characteristics of the unobservable process $(S_n)_{n\geqslant 1}$. For this, let us introduce

$$u_k(s_k) = w_k^{pr}(s_k|x_1^{k-1})\tilde{w}_k^{pr}(s_k|x_{k+1}^n)p^{-1}(s_k) \qquad (6)$$

and remark once more that u_k is independent of x_k. Then, the equation (5) can be written as

$$\pi_k(s_k|x_1^n) = \frac{\lambda_k(x_1^n \text{ without } x_k)}{f_k(x_k|x_1^n \text{ without } x_k)} f(x_k|s_k)u_k(s_k). \qquad (7)$$

Let us integrate this equation w.r.t. s_k and carry over the normalizing factor, which depends only on the observations, to the left-hand side:

$$\frac{f_k(x_k|x_1^n \text{ without } x_k)}{\lambda_k(x_1^n \text{ without } x_k)} = \int\limits_{\mathcal{S}_k} f(x_k|s_k)u_k(s_k)ds_k. \qquad (8)$$

Now, assuming the density $f(x_k|s_k)$ belongs to the exponential family (4), differentiate (8) w.r.t. x_k. The possibility of differentiating under the sign of integral is justified by the assumption of existence of the second prior moment $\mathsf{E}Q^T(S_k)Q(S_k)$, that is the natural restriction on a signal power. Differentiation w.r.t. x_k provides the equation

$$\frac{\bigtriangledown_{x_k} f_k(x_k|x_1^n \text{ without } x_k)}{\lambda_k(x_1^n \text{ without } x_k)} = \int\limits_{\mathcal{S}_k} \bigtriangledown_{x_k} f(x_k|s_k)u_k(s_k)ds_k. \tag{9}$$

For the exponential conditional density $f(x_k|s_k)$,

$$\bigtriangledown_{x_k} f(x_k|s_k) \\ = (\bigtriangledown_{x_k} \ln h(x_k) + \bigtriangledown_{x_k} T^{\mathrm{T}}(x_k)Q(s_k))f(x_k|s_k). \tag{10}$$

Substituting (10) to (9) and denoting by $Q(\widehat{s}_k)$ the integral $\int Q(s_k)\pi_k(s_k|x_1^n)ds_k$, we find the equation for the optimal mean square estimator $Q(\widehat{s}_k)$:

$$\mathcal{T}^{\mathrm{T}}(x_k)Q(\widehat{s}_k) = \bigtriangledown_{x_k} \ln \frac{f_k(x_k|x_1^n \text{ without } x_k)}{h(x_k)}, \tag{11}$$

where $\mathcal{T}$ is the Jacobi matrix with elements $\partial T_i/\partial x_k^{[j]}, i = \overline{1, m}, j = \overline{1, r}$. The equation (11) is a linear vector equation w.r.t. $Q(\hat{s}_k)$, but it can be solved only for a certain density $f_k(x_k|x_1^n$ without $x_k)$. If all the probability distributions are known, this density can be computed, and the result coincides with (1) and (2). But when $f_k(x_k|x_1^n$ without $x_k)$ can not be explicitly computed, we estimate it from the observations using the kernel non-parametric procedures.

4. NON-PARAMETRIC INTERPOLATION EQUATION

To solve the problem of interpolation on the basis of one realization x_1^n of a process $(X_k)_{1\leqslant k\leqslant n}, X_k \in \mathbb{R}^l$, we can use the asymptotically ε-optimal interpolating procedure from Dobrovidov et al. (2012), in which the truncated conditional density $\overline{f}(x_k|x_{k-\tau}^{k-1}, x_{k+1}^{k+\tau})$ is used instead of the conditional density $f(x_k|x_1^n$ without $x_k) \triangleq f(x_k|x_1^{k-1}, x_{k+1}^n)$, where the parameter τ is the order of the Markov process, which approximates the non-Markovian weak dependent process $(X_n)_{n\geqslant 1}$. The criteria and methods of estimation of τ, which were developed in Dobrovidov et al. (2012) with regard to filtration, can be extended to the interpolation problems. The truncated conditional density $\overline{f}(x_k|x_{k-\tau}^{k-1}, x_{k+1}^{k+\tau})$ can be written as a ratio

$$\overline{f}(x_k|x_{k-\tau}^{k-1}, x_{k+1}^{k+\tau}) = \frac{f(x_{k-\tau}^{k+\tau})}{f(x_{k-\tau}^{k-1}, x_{k+1}^{k+\tau})},$$

where the numerator is the marginal density of $(l \times (2\tau + 1))$-dimensional vector, and the denominator is the marginal density of $(l \times 2\tau)$-dimensional vector of observations. Substitute the multivariate non-parametric kernel estimators

$$f_N(x_1^n) = \frac{1}{Nh_N^{lr}} \sum_{i=1}^{N} \prod_{k=1}^{n} \prod_{j=1}^{l} K\left(\frac{(x_k^{[j]} - X_k^{[j]}(i))}{h_N}\right) \tag{12}$$

instead of truncated densities, and we get a non-parametric approximation of the equation (11)

$$\mathcal{T}^{\mathrm{T}}(x_k)Q({}^{\tau}\widehat{s}_{k,N}) = \frac{\bigtriangledown_{x_k} f_N(x_{k-\tau}^{k+\tau})}{f_N(x_{k-\tau}^{k+\tau})} - \frac{\bigtriangledown_{x_k} h(x_k)}{h(x_k)}. \tag{13}$$

So, to construct the interpolation estimator at the point k, one uses the data before and later of k by a distance not exceeding τ. The interpolation estimator in the equation (13) is consistent, but it depends on the logarithmic gradient of the conditional probability density, which can be unstable in some points x_k taking infinite values when the denominator is equal to zero. For a more strong convergence, one should construct a piecewise smooth approximation (see Dobrovidov et al. (2012)) providing the mean-square convergence under some additional regularity conditions.

5. NON-PARAMETRIC INTERPOLATION ESTIMATOR AND OPTIMAL ESTIMATORS

Since the nonparametric kernel estimation of multivariate densities of large dimension is very difficult for implementation, we illustrate the proposed method of estimation and its performance by the following example with univariate state and observation models ($m = l = 1$):

$$S_{n+1} = aS_n + b\xi_{n+1}, \quad b^2 = \sigma^2(1 - a^2); \tag{14}$$
$$X_n = AS_n + B\eta_n, \quad S_n,\, X_n \in \mathbb{R}. \tag{15}$$

Here, S_1, ξ_n, and η_n are the mutually independent random variables with the Gaussian distributions $\mathcal{N}\{0,\, \sigma^2\}$ for S_1, and $\mathcal{N}\{0,\, 1\}$ for ξ_n and η_n, $n \geqslant 1$. The coefficients a, b, A, and B are known, $|a| < 1$. With suitable initial conditions, such equations generate a strongly stationary sequence. For the model (14), (15), the conditional probability density of observations is Gaussian, and, therefore, the Kalman filter and the optimal forward and backward recursive linear interpolation equations, associated with it, can be obtained by using the results of Liptser and Shiryaev (1977).

The Kalman filter:

$$\widehat{S}_{k+1} = a\widehat{S}_k + \frac{Ab^2 + a^2A\gamma_k}{B^2 + A^2b^2 + A^2a^2\gamma_k}[x_{k+1} - Aa\widehat{S}_k],$$
$$\gamma_{k+1} = \frac{B^2(a^2\gamma_k + b^2)}{A^2(a^2\gamma_k + b^2) + B^2} \tag{16}$$

with the initial conditions

$$\widehat{S}_1 = \frac{A\sigma^2}{A^2\sigma^2 + B^2}x_1, \quad \gamma_1 = \frac{B^2\sigma^2}{A^2\sigma^2 + B^2},$$

where $\widehat{S}_k = \mathsf{E}[S_k|x_1^k]$, $\gamma_k = \mathsf{E}[(S_k - \widehat{S}_k)^2|x_1^k]$.

Forward interpolation:

$$D_k = A^2(a^2\gamma_k + \sigma^2(1 - a^2)) + B^2,$$
$$\widetilde{S}_k = \widehat{S}_k + Aa\gamma_k(X_{k+1} - Aa\widehat{S}_k)/D_k,$$
$$\widetilde{\gamma}_k = A^2\sigma^2(1 - a^2) + B^2)\gamma_k/D_k, \quad k = 2, ..., n, \tag{17}$$

where $\widehat{S}_k = \mathsf{E}[S_k|x_1^{k+1}]$, $\widetilde{\gamma}_k = \mathsf{E}[(S_k - \widetilde{S}_k)^2|x_1^{k+1}]$.

Backward interpolation:

$$\tilde{\tilde{S}}_k = \widetilde{S}_k + \tilde{\tilde{\gamma}}_k a\sigma^2(1 - a^2)(\tilde{\tilde{S}}_{k+1} - \widehat{S}_k))/d_k\gamma_{k+1},$$
$$\tilde{\tilde{\gamma}}_k = \widetilde{\gamma}_k + \frac{\widetilde{\gamma}_k^2(\sigma^2(1 - a^2))^2\tilde{\tilde{\gamma}}_{k+1}}{D_k^2\gamma_{k+1}}, \quad k = 2, ..., n-1, \tag{18}$$

where $\tilde{\tilde{S}}_k = \mathsf{E}[S_k|x_1^n]$, $\tilde{\tilde{\gamma}}_k = \mathsf{E}[(S_k - \tilde{\tilde{S}}_k)^2|x_1^n]$.

A non-parametric interpolation can be constructed by using only one observation equation (15) and the data sample of size n. In this univariate case, non-parametric interpolation equation (11) reduces to

$$\widehat{S}_k^\tau = \frac{B^2}{A}\frac{\frac{\partial}{\partial x_k}f(x_{k-\tau}^{k+\tau})}{f(x_{k-\tau}^{k+\tau})} + \frac{x_k}{A} = \frac{B^2}{A}\psi(x_{k-\tau}^{k+\tau}) + \frac{x_k}{A}, \quad (19)$$

$$\psi(x_{k-\tau}^{k+\tau}) = \frac{\frac{\partial}{\partial x_k}f(x_{k-\tau}^{k+\tau})}{f(x_{k-\tau}^{k+\tau})}. \quad (20)$$

So, we see that equation (19) does not involve the parameters a, b of state-space equation (18). The non-parametric kernel estimator for the denominator $f(x_{k-\tau}^{k+\tau})$ in (19) is defined by formula (12). The numerator in (19) contains a partial derivative of the multivariate density $f(x_{k-\tau}^{k+\tau})$ at the point x_k. Usually, the estimator of the density derivative is selected as the derivative of the density estimator. In this case, the optimal bandwidths for the density and its derivative do not match (see Dobrovidov and Rud'ko (2010)). Therefore, to calculate the data-driven bandwidths, one have to use different algorithms. Here, these parameters are estimated by cross-validation method.

6. REGULARIZED ESTIMATOR

The non-parametric estimator for the ratio (20) is the following expression:

$$\psi_N(x_{k-\tau}^{k+\tau}) = \frac{\frac{\partial}{\partial x_k}f_N(x_{k-\tau}^{k+\tau})}{f_N(x_{k-\tau}^{k+\tau})} \quad (21)$$

$$= \frac{h_N \sum\limits_{i=\tau+1}^{n-\tau}(x_i - x_k)\, e^{-\frac{(x_k - x_i)^2}{2{h_N}'^2}} \prod\limits_{\substack{j=-\tau \\ j\neq 0}}^{\tau} e^{-\frac{(x_{k+l} - x_{i+l})^2}{2{h_N}^2}}}{h_N'^2 \sum\limits_{i=\tau+1}^{n-\tau}\prod\limits_{l=-\tau}^{\tau} e^{-\frac{(x_{k+l} - x_{i+l})^2}{2{h_N}^2}}},$$

where $N = n - 2\tau$ and h_N and h_N' are the bandwidth parameters for a density and its derivative. Optimal values of them depend on unknown functions.

Estimator (21) is the special case of plug-in estimator of a composite function $G(t_N(x))$, where $x \in \mathbb{R}^s$, $t_N : \mathbb{R}^s \to \mathbb{R}^m$, $G : \mathbb{R}^m \to \mathbb{R}^1$. In our case $m = 2$, $t_N = (t_{1N}, t_{2N})^{\mathrm{T}}$, $t_{1N} = f_N(x_{k-\tau}^{k+\tau})$, $t_{2N} = \frac{\partial}{\partial x_k}f_N(x_{k-\tau}^{k+\tau})$, $G(t_N) = t_{2N}/t_{1N}$. 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 also. Write the main regularity conditions for convergence: 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 $0 \le \gamma < \infty$. So, the function $|G(t_N)|$ have to grow slower than the function $C_0 d_N^\gamma$. These conditions provide the mean square convergence of $G(t_N)$ to $G(t)$ (see Koshkin (1999)).

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

$$G(t_N) - G(t) = \nabla G(\vartheta_N)(t_N - t), \quad \vartheta_N \in (t_N, t,) \quad (22)$$

where ∇ is gradient w.r.t. t. From here according to Koshkin (1999)

$$|\mathsf{E}(G(t_N) - G(t))^2 - \mathsf{E}(\nabla G(\vartheta_N)(t_N - t))^2| = O(d_N^{-3/2}), \quad (23)$$

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, 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.

As the proposition (22) is valid only for bounded functions we apply here some procedure of regularization, called a piecewise smooth approximation (see Koshkin (1999)). For the first time, a stable piecewise smooth approximation of plug-in estimators and their mean square errors (*MSEs*) have been investigated in Penskaya (1990), stable estimators of the ratios (convergence in law, convergence of *MSE*), in Koshkin (1993), deviation moments of arbitrary orders of piecewise smooth approximations of plug-in estimators, in Koshkin (1999).

Using this procedure, we obtain the following stable approximation of G :

$$\Phi(G(t), \delta_N) = \widetilde{\Phi}(t, \delta_N) = \frac{G(t)}{1 + \delta_N |G(t)|^4}, \quad (24)$$

where $\delta_N > 0$ is a regularization parameter. As it is proved in Koshkin (1999), $\widetilde{\Phi}(t_N, \delta_N)$ satisfies both the above mentioned conditions and therefore is dominated by the power function of N. Moreover, $\widetilde{\Phi}(t_N, \delta_N)$ converges to $G(t)$ in the mean square sense, i.e., as $\mathsf{E} \parallel t_N - t \parallel \to 0$ and $\delta_N \to 0$, then

$$\lim_{N\to\infty} \mathsf{E}(\widetilde{\Phi}(t_N, \delta_N) - G(t))^2 = 0. \quad (25)$$

In special case the procedure of piecewise smooth approximation coincides with the Tychonoff regularization method (see Tychonoff and Arsenin (1979); Tychonoff and Ufimtsev (1988)). Indeed, formally, the estimator of piecewise smooth approximation can be obtained by the Tychonoff regularization method by minimizing the smoothing functional

$$Q = [\widetilde{\Phi}(t_N) - G(t_N)]^2 + \alpha\widetilde{\Phi}^2(t_N),$$

and it equals

$$\widetilde{\Phi}(t_N, \alpha) = \mathrm{argmin}_{\widetilde{\Phi}} Q = \frac{G(t_N)}{1 + \alpha}, \quad (26)$$

with the bias, variance and *MSE* of $\widetilde{\Phi}(t_n, \alpha)$ being given by the expressions

$$b(\widetilde{\Phi}(t_N)) = \mathsf{E}\widetilde{\Phi}(t_N) - G(t) = \frac{b(G(t_N)) - \alpha G(t)}{1 + \alpha},$$
$$\mathsf{D}\widetilde{\Phi}(t_N) = \frac{\mathsf{D}G(t_N)}{(1 + \alpha)^2},$$
$$u^2(\widetilde{\Phi}(t_N)) = \frac{u^2(G(t_N)) - 2\alpha b(G(t_N))G(t) + \alpha^2 G^2(t)}{(1 + \alpha)^2}. \quad (27)$$

Comparing (24) and (26) yields

$$\alpha = \delta_N |G(t)|^\tau. \quad (28)$$

The statistic $\psi_N(x_{k-\tau}^{k+\tau})$ in (21) is unstable when denominator is close to zero. So, we use the stable estimate

$$\widetilde{\psi}_N(x_{k-\tau}^{k+\tau}) = \frac{\psi_N(x_{k-\tau}^{k+\tau})}{1+\delta_N|\psi_N(x_{k-\tau}^{k+\tau})|^4}, \tag{29}$$

where the regularization parameter δ_N has to be found. This parameter can be found by optimizing the criterion of mean integrated square error (*MISE*) for estimating function $\psi(x_{k-\tau}^{k+\tau})$ with the weight function $\omega(\cdot)$, i.e.

$$MISE(\delta_N) = \int u^2\left(\widetilde{\psi}_N(x_{k-\tau}^{k+\tau})\right)\omega(x_{k-\tau}^{k+\tau})dx_{k-\tau}^{k+\tau},$$
$$u^2\left(\widetilde{\psi}_N(x_{k-\tau}^{k+\tau})\right) \triangleq \mathsf{E}\left(\widetilde{\psi}_N(x_{k-\tau}^{k+\tau}) - \psi(x_{k-\tau}^{k+\tau})\right)^2. \tag{30}$$

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

Calculation of the expectation of ratio (21) is laborious. According to (25), for the mean square convergence of regularized estimate $\widetilde{\psi}_N(x_{k-\tau}^{k+\tau})$ to $\psi(x_{k-\tau}^{k+\tau})$ it is necessary that $\delta_n \to 0$ as $n \to \infty$. Therefore, under the assumption of small δ_N we expand the ratio (29) w.r.t. parameter δ_N and approximately obtain

$$\widetilde{\psi}_N(x_{k-\tau}^{k+\tau}) \approx \psi_N(x_{k-\tau}^{k+\tau}) - \delta_N\psi_N^5(x_{k-\tau}^{k+\tau}). \tag{31}$$

Substituting (31) into *MISE* (30) and making use of Theorem 2 from Koshkin (1999), we receive

$$\begin{aligned}
&\int u^2\left(\widetilde{\psi}_N(x_{k-\tau}^{k+\tau})\right) f(x_{k-\tau}^{k+\tau})dx_{k-\tau}^{k+\tau} \\
&\approx \int H_1^2 u^2\left(\frac{\partial}{\partial x_k} f_N(x_{k-\tau}^{k+\tau})\right) f(x_{k-\tau}^{k+\tau})dx_{k-\tau}^{k+\tau} \\
&+2\int H_1H_2\mathrm{cov}\left(\frac{\partial}{\partial x_k} f_N(x_{k-\tau}^{k+\tau}), f(x_{k-\tau}^{k+\tau})\right) f(x_{k-\tau}^{k+\tau})dx_{k-\tau}^{k+\tau} \\
&+\int H_2^2 u^2\left(f(x_{k-\tau}^{k+\tau})\right) f(x_{k-\tau}^{k+\tau})dx_{k-\tau}^{k+\tau} \\
&= \int H_1^2u^2(f_N^{'}) + 2\int H_1H_2\mathrm{cov}(\cdot) + \int H_2^2u^2\left(f_N\right),
\end{aligned} \tag{32}$$

where $H_1 = \dfrac{1-5\delta\psi^4}{f}$, $H_2 = \dfrac{-\psi+5\delta\psi^5}{f}$, $f_N^{'} = \dfrac{\partial}{\partial x_k}f_N$, $\mathrm{cov}(\cdot) = \mathrm{cov}(f_N^{'}, f_N)$. Now, minimizing (28) w.r.t. δ_N, we find

$$\delta_{opt} = \frac{\int u^2(f_N^{'}) - 2\int\psi\mathrm{cov}(\cdot) + \int\psi^2u^2(f_N)}{5\int\psi^4u^2(f_N^{'}) - 10\int\psi^5\mathrm{cov}(\cdot) + 5\int\psi^6u^2(f_N)}. \tag{33}$$

The integrals in the numerator and denominator of δ_{opt} depend on unknown densities, therefore, they have to be estimated. The main parts of expansions of $u^2(\cdot)$ and $\mathrm{cov}(\cdot)$ under $n \to \infty$ equal to

$$\begin{aligned}
u^2(f_N^{'}) &\approx \frac{f}{Nh_N^3}\int(K^{(1)}(u))^2du \\
&+\frac{h_N^4}{4}(f^{(3)})^2\left(\int u^2K(u)du\right)^2, \\
\mathrm{cov}(f_N^{'}, f_N) &\approx \frac{f}{Nh_N^2}\int K^{(1)}(u)K(u)du \\
&+\frac{h_N^4}{4}f^{(3)}f^{(2)}\left(\int u^2K(u)du\right)^2, \\
u^2(f_N) &\approx \frac{f}{Nh_N}\int K^2(u))du \\
&+\frac{h_N^4}{4}(f^{(2)})^2\left(\int u^2K(u)du\right)^2,
\end{aligned}$$

where $f^{(i)} = \dfrac{\partial^i}{\partial x_k^i}f$, $i = 1, 2, 3$. Substituting these formulae into (33), we find δ_{opt}, in which it is necessary to estimate the following integrals:

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

It can be done by the smoothed cross-validation method, developed for such functionals in Dobrovidov and Rud'ko (2010).

7. COMPARISON OF THE ESTIMATORS

In the simulation, we base ourselves on the sample risk

$$\widehat{R} = \frac{1}{M}\sum_{j=1}^{M}\left(\frac{1}{N}\sum_{i=1}^{N}\left(S_i(j) - \widehat{S}_i(j)\right)^2\right)^{1/2}, \tag{34}$$

where M is the number of repeated experiments.

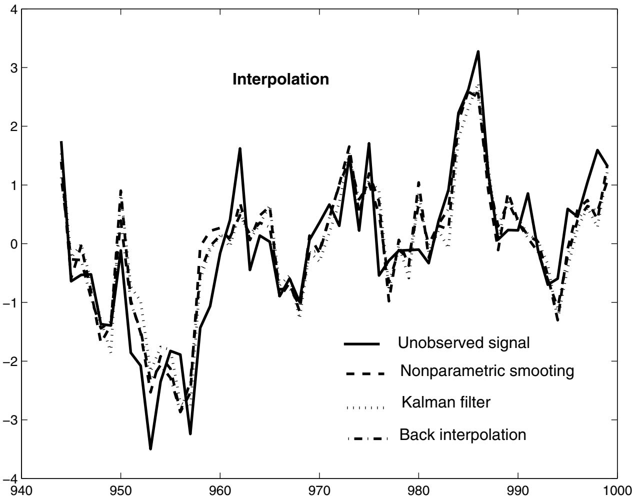

Fig. 1. Comparison of the non-parametric smoothing estimator with the optimal estimators

We consider three estimators:
— Kalman estimator $\widehat{S}_k$ (16) with risk R_K,
— optimal backward interpolation $\tilde{\tilde{S}}_k$ (18) with the risk R_{OI},
— non-parametric interpolation (19) with the risk R_{NI}.
Because $R_{OI} \leqslant R_K$ and $R_{OI} \leqslant R_{NI}$, for convenience of comparison, let us introduce the relative errors in percentage:

$$\varepsilon_K = \frac{R_K - R_{OI}}{R_{OI}} \times 100, \qquad \varepsilon_{NI} = \frac{R_{NI} - R_{OI}}{R_{OI}} \times 100.$$

The relative error shows how much an estimator is better or worse than another one. The simulation results are presented in Fig.1 by $n = 1000, \sigma^2 = 2, a = 0.7, b = 1, A = B = 1, \tau = 1$. The relative errors ε_K and ε_{NI} are given in Table 1. We see that non-parametric estimators can superior the optimal Kalman filtering estimators by the performance, but it is always inferior w.r.t. the optimal backward interpolation.

Table 1.

Relative excess of the empirical risk over the optimal smoothing risk

M	Optimal	Kalman ε_K	Non-par ε_{NI}
50	0%	7.48%	4.42%

8. CONCLUSION

The paper presents two interpolating algorithms for estimation of an unobservable signal disturbed by noise on the fixed time interval. In the case of a linear observation model, the simulation experiments illustrate the performance of the proposed non-parametric interpolation estimator in comparison with the optimal Kalman filtering estimator and with the optimal backward interpolation. It is shown that the quality of the regularized nonparametric algorithm naturally worse than backward interpolation with complete statistical information, but better than the Kalman filter. Improved quality is achieved by additional information to the right of the estimation point. In what follows it is supposed to build a similar algorithm for the dynamic model of observation (for example, autoregressive) taking into account both interpolation equations (2) and (3).

ACKNOWLEDGEMENTS

Authors thank the anonymous referees for their constructive suggestions, which have improved the earlier version of this work.

One of the authors Gennady M. Koshkin is thankful to Tomsk State University Competitiveness Improvement Program for providing excellent facilities for research.

REFERENCES

Briers, M., Douset, A., and Maskell, S. (2003). Smoothing algorithms for state-space models. *IEEE Amer. Statist. Assoc.*, 85, 66–72.

Chentsov, N.N. (1972). *Statistical Decision Rules and Optimal Inference*. Nauka, Moscow (in Russian).

Dobrovidov, A.V. (2008). Non-parametric signal interpolation. *Austrian Journal of Statistics*, 37(1), 21-29.

Dobrovidov, A.V., Koshkin, G.M., and Vasiliev, V.A. (2012). *Non-parametric State Space Models*. Kendrick Press, Inc., Heber, UT 84032, USA.

Dobrovidov, A.V., and Rud'ko, I.M. (2010). Bandwidth Selection in Nonparametric Estimators of Density Derivative by Smoothed Cross-Validation Method. *Automat. and Remote Control*, 71(2), 209-224.

Khazen, E.M. (1978). Reconstruction of the component of a multivariate Markov process under the observations of other its components. *Probl. Control Inform. Theory*, 7 (4), 263–275.

Koshkin, G.M. (1993) Stable estimation of ratios of random functions from experimental data. *Russian Physics Journal*, 36 (10), 1008–1015.

Koshkin, G.M. (1999) Deviation moments of the substitution estimate and its piecewise smooth approximations. *Siberian Math. Journal*, 40 (3), 515–527.

Liptser, R.Sh., and Shiryaev, A.N. (1977). *Statistics of Random Processes. 1: General Theory (1977). 2: Applications (1978).* Springer-Verlag, New York, 2nd edition.

Penskaya, M.Ya. (1990) On stable estimation of the function of parameter. *Statistical Methods of Estimation and Testing Hypotheses*, 44–55. Perm Univ. Press, Perm (in Russian).

Stratonovich, R.L. (1966). *Conditional Markovian Processes and their Application to the Optimal Control Theory.* Moscow Univ. Press, Moscow (in Russian).

Tychonoff, A.N., Arsenin, V.Ya. (1979). *Methods for Solving Incorrect Problems.* Nauka, Moscow (in Russian).

Tychonoff, A.N., Ufimtsev, M.V. (1988). *Statistical Processing the Results of an Experiment.* Moscow Univ. Press, Moscow (in Russian).