# Application of Probability Filter to Signal Processing Problems

**D. P. Vetrov and D. A. Kropotov**

*Dorodnicyn Computing Center, Russian Academy of Sciences, ul. Vavilova 40, GSP-1, Moscow, 119991 Russia*
*e-mail: vetrovd@yandex.ru; dkropotov@yandex.ru*

**Abstract**—A method of signal filtering based on the maximum likelihood principle is presented. This method provides elimination of random distortions caused by noises, conserving local specific features of the signal. The concept of probability filtering ensures processing of both discrete and piecewise-continuous signals.

**DOI:** 10.1134/S1054661806030151

## 1. INTRODUCTION

The problem of digital signal processing occurs in a wide range of applied problems of data analysis, prediction, and decision making. In this paper a particular problem of elimination of interference caused by partial loss of signal is considered. The existing methods for signal filtering allow, in principle, the elimination of interference; however, their action is based on some signal smoothing [2]. As a result, local specific features of the useful signal which are often of basic interest for subsequent study are lost. There is the need to develop and implement filtering methods that eliminate significant noise, on the one hand, and have minimal impact on the shape of the noiseless signal, on the other. A possible tool for such signal processing is probability theory, in particular, the maximum likelihood principle. Treating each point of the signal as a continuous random value with a parametrically specified density allows evaluation of the degree of likelihood that this point belongs to the useful signal. In this case, the point with maximum likelihood is conserved in the signal, and the preceding points are discarded as noise. Note that similar ideas are used in structural signal analysis, in particular, in constructing hidden Markov models [1]. In this paper, the problems of applying these methods to signal filtering are discussed.

The development of a probability filter is considered in detail in Section 2. In Section 3, certain properties of the proposed filtering algorithm as applied to continuous signals are considered, and in Section 4 the problem of applying the filter to discrete signals is studied. Then, a special method for determining the first point of the useful signal and a symmetric filtering procedure based on the method of dynamic programming are proposed. The results of practical experiments and comparison with other filtering methods are given in Section 7.

Received August 18, 2005

## 2. DEVELOPMENT OF CONTINUOUS PROBABILITY FILTER

Consider the following problem. A certain input noisy signal $x(i)$ is given on a time interval $[0, T]$. It is required to construct a filter forming the output signal $y(i)$ in such a way that the filtered signal does not contain substantial inhomogeneities related to partial loss of the original signal during measurement. If there are no such inhomogeneities, the signal should not be changed, i.e., $y(i) = x(i)$, $i \in [0, T]$.

We consider the input signal $x(i)$ as an implementation of a random process with independent increments $\{\zeta(t), t \in [0, T]\}$ defined on the probability space $\{\Omega, \mathbb{A}, \mathbb{P}\}$. Assume that the random process $\zeta(t)$ can be represented in the form

$$\begin{aligned}\zeta(t_1) - \zeta(t_0) &= \mathrm{A}(t_0, t_1)(\xi(t_1) - \xi(t_0)) \\ &\quad + \mathrm{B}(t_0, t_1) + \eta(t_1), \\ &\quad \forall 0 \le t_0 < t_1 \le T.\end{aligned} \tag{1}$$

Here, $\xi(t)$ and $\eta(t)$ are random processes acting on the same probability space and time interval as $\zeta(t)$, and the functions $\mathrm{A}(t_0, t_1)$ and $\mathrm{B}(t_0, t_1)$ are certain mappings of the Cartesian product $[0, T] \times [0, T]$ on the real line. The process $\xi(t)$ represents the process of Brownian motion, and $\eta(t)$ is a process with unknown properties reflecting the presence of noise in the signal. The only constraint on the process $\eta(t)$ is that it acts only at some points of the space $[0, T]$, i.e.,

$$\mathbb{P}\{\eta(t) = 0\} = p > 0. \tag{2}$$

Representation of the random process in the form (1) is related to the following. As a rule, the process of measurement of the input signal is influenced by a large number of random factors, each of them introducing a slight correction. Such random factors occur as a consequence of imperfection of the measuring device, inhomogeneity of the medium of measurement, and other circumstances. In this regard, formalization of such noise by a random process of Brownian motion seems natural, while significant inhomogeneities which are objects of filtering represent an independent random process with unknown properties. Requirement (2) to this process limits the influence of the significant noise and provides a signal distortion only on some subsets of the time interval.

Under the assumption that the random process $\eta(t)$ does not act on the interval $[t_0, t_1]$ and taking into account that $\xi(t)$ is the Brownian motion, we obtain from formula (1)

$$\xi(t_1) - \xi(t_0) = \frac{\zeta(t_1) - \zeta(t_0) - \mathrm{B}(t_0, t_1)}{\mathrm{A}(t_0, t_1)} \sim \mathbb{N}(0, t_1 - t_0). \tag{3}$$

Consider the mean value and dispersion of the deviation of the original random process. On the one hand, taking into account the properties of the Brownian motion and expression (1), we conclude that

$$\mathbb{E}(\zeta(t_1) - \zeta(t_0)) = \mathrm{B}(t_0, t_1),$$

$$\mathbb{D}(\zeta(t_1) - \zeta(t_0)) = \mathrm{A}^2(t_0, t_1)(t_1 - t_0).$$

On the other hand, taking into account that increments of the random process $\zeta(t)$ are independent, the following chain of equalities holds:

$$\mathbb{E}(\zeta(t_1) - \zeta(t_0)) = \mathbb{E}\int_{t_0}^{t_1} d\zeta = \int_{t_0}^{t_1} \mathbb{E}d\zeta = \int_{t_0}^{t_1} d\mathbb{E}\zeta(\tau)$$

$$= \{\varepsilon(t) \triangleq \mathbb{E}\zeta(t)\} = \int_{t_0}^{t_1} d\varepsilon(\tau) = \int_{t_0}^{t_1} \dot{\varepsilon}(\tau)d\tau$$

$$= \{\mu(\tau) \triangleq \dot{\varepsilon}(\tau)\} = \int_{t_0}^{t_1} \mu(\tau)d\tau,$$

$$\mathbb{D}(\zeta(t_1) - \zeta(t_0)) = \mathbb{D}\int_{t_0}^{t_1} d\zeta = \int_{t_0}^{t_1} \mathbb{D}d\zeta = \int_{t_0}^{t_1} d\mathbb{D}\zeta(\tau)$$

$$= \int_{t_0}^{t_1} d\mathrm{K}(\tau)\tau = \int_{t_0}^{t_1} \sigma^2(\tau)d\tau.$$

The original random process $\zeta(t)$ represents the process of Brownian motion $\xi(t)$ with the dispersion $\tau$ normalized by a time-dependent value. Therefore, the dispersion of the process $\zeta(t)$ equals $\mathrm{K}(\tau)\tau$. Thus,

$$\mathrm{B}(t_0, t_1) = \int_{t_0}^{t_1} \mu(\tau)d\tau \triangleq \mathrm{M}_{t_1, t_0}, \tag{4}$$

$$\mathrm{A}^2(t_0, t_1)(t_1 - t_0) = \int_{t_0}^{t_1} \sigma^2(\tau)d\tau \triangleq \Sigma^2_{t_1, t_0}. \tag{5}$$

The values $\mu(t)$ and $\sigma(t)$ are referred to below as centroidal and variational fields, respectively.

Substituting expressions (4) and (5) into formula (2), we finally obtain

$$\frac{\zeta(t_1) - \zeta(t_0) - \mathrm{M}_{t_1, t_0}}{\Sigma_{t_1, t_0}} \sim \mathbb{N}(0, 1). \tag{6}$$

Therefore, if there are no substantial inhomogeneities, the deviation of the random process $\zeta(t)$, centered and normalized in a proper way, satisfies the standard normal distribution. This naturally results in the filtering method based on application of the maximum likelihood principle. Assume that it is known that, at a point $t_0$, the signal is not distorted, i.e., $\eta(t_0) = 0$. Then, determination of the next point of the undistorted signal is reduced to maximization of the probability density of the normal distribution, i.e.,

$$t_1 = \underset{t > t_0}{\operatorname{argmax}} \mathrm{p}\{\zeta(t) = x(t), \zeta(t_0) = x(t_0) | \eta(t) = 0, t \in [t_0, t]\}. \tag{7}$$

Thus, we come to a sequential procedure for signal filtering. At the first stage, the first undistorted point of the input signal is determined. A possible procedure for solving this problem is described in Section 5. Then, other undistorted points of the signal are determined sequentially. Thus obtained points are called the *admissible points*. At the final stage, the filtered signal $y(t)$ is interpolated in the intervals between the admissible points. An alternative version of this filtering procedure based on the dynamic programming method is described in Section 6.

## 3. PROPERTIES OF CONTINUOUS PROBABILITY FILTER

Consider a signal $x(t)$ on the interval $[0, T]$ with not more than a denumerable number of points of discontinuity of the first kind or removable points of discontinuity. Assume that the point $x(t_0)$ is the first admissible point. Then, for a point $t_1$, the measure of likelihood that it belongs to the useful signal equals

$$l\{y(t_1) = x(t_1) | y(t_0) = x(t_0)\} = -\frac{(x(t) - x(t_0) - \mathrm{M}_{t, t_0})^2}{2\Sigma^2_{t, t_0}} - \ln(\sqrt{2\pi}\Sigma_{t, t_0}), \tag{8}$$

where $\mathrm{M}_{t, t_0}$ and $\Sigma_{t, t_0}$ are calculated, respectively, by the formulas

$$\mathrm{M}_{t, t_0} = \int_{t_0}^{t_1} \mu(\tau)d\tau,$$

$$\Sigma_{t, t_0} = \sqrt{\int_{t_0}^{t_1} \sigma^2(\tau) d\tau}.$$

The following theorem illustrates the local properties of the probability filter in the continuous case.

**Theorem 1.** (i) Assume that $\tau$ is a removable point of discontinuity of the signal $x(t)$ and the signal satisfies the Lipschitz condition with exponent $L$ everywhere except for this point. Then, the filtered signal $y(t)$ is continuous at the point $\tau$ and $y(\tau) = \lim_{t \to \tau} x(t)$.

(ii) Assume that the signal $x(t)$ is continuous at the point $\tau$ and satisfies in a neighborhood the Lipschitz condition with exponent $L$. Then, there is a $\delta$-neighborhood of the point $\tau$ in which the filtered signal coincides with the original signal, $y(t) = x(t)$.

**Proof.** First, let us prove the second proposition. Consider the function of positive argument $l(t) = -\frac{(x(\tau+t) - x(\tau) - M_{\tau, \tau+t})^2}{2\Sigma^2_{\tau, \tau+t}} - \ln(\sqrt{2\pi}\Sigma_{\tau, \tau+t})$. It will be shown that it attains its maximum at the zero point. Indeed, $\forall t > 0$ $l(t) < \infty$ since $\Sigma_{\tau, \tau+t}$ is positive. In this case, $\lim_{t \to +0} l(t) = \infty$, since, for all sufficiently small $t$, using the formula of the mean value and the Lipschitz condition, we obtain

$$-\frac{(x(\tau+t) - x(\tau) - M_{\tau, \tau+t})^2}{2\Sigma^2_{\tau, \tau+t}} \geq -\left(\frac{C_1 t}{C_2 t}\right)^2 \geq C > -\infty.$$

In this case, the second term becomes infinite, since $\lim_{t \to +0} \Sigma_{\tau, \tau+t} = 0$. Therefore, the likelihood function attains its maximum at the point $t = +0$, and, therefore, $y(\tau + 0) = x(\tau + 0) = x(\tau)$. The condition of continuity of the function at a point implies its continuity in a neighborhood of this point. Proposition 2 is proved.

Consider the signal $z(t)$,

$$z(t) = \begin{cases} x(\tau), & \tau - \delta < t < \tau + \delta \\ x(t), & 0 \leq t \leq \tau - \delta, \tau + \delta \leq t \leq T. \end{cases} \tag{9}$$

Examine the behavior of the function $l\{y(\tau + k\delta) = x(\tau + k\delta) | y(\tau - \delta) = x(\tau - \delta)\}$, where $-1 < k < 1$. Obviously, for small $\delta$, there exists a constant $C$ such that $|z(\tau + k\delta) - z(\tau - \delta)| > C$. Taking into account that $\lim_{x \to +0} x^{-1} \exp(-c/x) = \lim_{y \to +\infty} y \exp(-cy) = 0$, we obtain

$$\lim_{\delta \to +0} \frac{\exp\left(-\frac{(z(\tau+k\delta) - z(\tau-\delta) - M_{\tau-\delta, \tau+k\delta})^2}{2\Sigma^2_{\tau-\delta, \tau+k\delta}}\right)}{\sqrt{2\pi}\Sigma_{\tau-\delta, \tau+k\delta}}$$

$$= \lim_{\delta \to +0} \frac{\exp\left(-\frac{(z(\tau+k\delta) - z(\tau-\delta) - \mu(\xi_2)(k+1)\delta)^2}{2\sigma(\xi_1)^2(k+1)^2\delta^2}\right)}{\sqrt{2\pi}\sigma(\xi_1)(k+1)\delta}$$

$$= \lim_{\delta \to +0} \frac{\exp\frac{-C^2}{2\sigma(\xi_1)^2(k+1)^2\delta^2}}{\sqrt{2\pi}\sigma(\xi_1)(k+1)\delta} = 0.$$

Therefore,

$$\lim_{\delta \to +0} l\{y(\tau + k\delta) = x(\tau + k\delta) | y(\tau - \delta) = x(\tau - \delta)\} = -\infty.$$

On the other hand, for sufficiently small $\delta$, taking into account the Lipschitz condition, we have $|z(\tau + \delta) - z(\tau - \delta)| = 2L\theta\delta$, where $\theta \in [0, 1]$. From this we obtain

$$\lim_{\delta \to +0} l\{y(\tau + \delta) = x(\tau + \delta) | y(\tau - \delta) = x(\tau - \delta)\}$$

$$= \lim_{\delta \to +0} \ln \frac{\exp\left(-\frac{(z(\tau+\delta) - z(\tau-\delta) - M_{\tau-\delta, \tau+\delta})^2}{2\Sigma^2_{\tau-\delta, \tau+\delta}}\right)}{\sqrt{2\pi}\Sigma_{\tau-\delta, \tau+\delta}}$$

$$= \lim_{\delta \to +0} \ln \frac{\exp\left(-\frac{(2L\theta\delta - \mu(\xi_2)2\delta)^2}{2\sigma(\xi_1)^2 4\delta^2}\right)}{\sqrt{2\pi}\sigma(\xi_1)2\delta}$$

$$= \lim_{\delta \to +0} \ln \frac{\exp\left(-\frac{\delta^2}{2\sigma(\xi_1)^2 4\delta^2}(2L\theta - 2\mu(\xi_2))^2\right)}{\sqrt{2\pi}\sigma(\xi_1)2\delta}$$

$$= \lim_{\delta \to +0} \ln \frac{C}{\delta} = +\infty.$$

Since $\lim_{\delta \to +0} z(t) = x(t)$, the following is proved: $l\{y(\tau + 0) = x(\tau + 0) | y(\tau - 0) = x(\tau - 0)\} > l\{y(\tau) = x(\tau) | y(\tau - 0) = x(\tau - 0)\}$. Similarly, it can be shown that $l\{y(t) = x(t) | y(\tau - 0) = x(\tau - 0)\} < l\{y(\tau + 0) = x(\tau + 0) | y(\tau - 0) = x(\tau - 0)\}$ $\forall t > \tau + \delta$. Without loss of generality, we may assume that the linear interpolation is made between admissible points. Then, taking into account that $\lim_{t \to +0} x(t) = \lim_{t \to -0} x(t)$, we easily obtain that $y(\tau) = \lim_{t \to \tau} x(t)$. Therefore, the filtered signal $y(t)$ is continuous. The theorem is completely proved.

The results of this theorem open the way to practical application of the probability filter in processing signals with continuous time. The problem of determination of the next admissible point arises in the case of a discontinuous signal. Therefore, any signal with a finite number of points of discontinuity of the first kind (or removable points of discontinuity) on any finite time interval can be processed by the filter. Note that the classical signal filtering theory involves the integration of the original signal and the functions characterizing the filter for each time instant, which is, generally speaking, a more complicated problem.

According to formula (7), finding the next admissible point, we must examine the signal up to its end, which can result in high computational complexity and does not allow processing of infinite or real-time signals. Denote $A_{t_0}(\tau) = \sup\limits_{t_0 < t \le t_0+\tau} l\{y(t) = x(t)|y(t_0) = x(t_0)\}$. Then, the following theorem is valid.

**Theorem 2.** Assume that the signal $x(t)$ has the discontinuity of the first kind at the point $t_0$. To determine the next admissible point, it is sufficient to examine the signal no further than the interval of length $\tau$, where $\tau$ is the minimum number for which the following inequality is satisfied:

$$\int_{t_0}^{t_0+\tau} \sigma^2(s)ds \ge \frac{\exp(-2A_{t_0}(\tau))}{2\pi}. \tag{10}$$

In particular, if, after the point $t_0$, the signal is continuous, then $\tau$ is the root of the equation

$$\int_{t_0}^{t_0+\tau} \sigma^2(s)ds = \frac{\exp(-2A_{t_0}(\tau))}{2\pi}. \tag{11}$$

**Proof.** Show that, if formula (10) is valid for a certain $t$, then it is valid for all $\tau > t$. Indeed, the left-hand side of the formula obviously does not decrease with increasing $\tau$, since $\sigma^2(s)$ is nonnegative. The function $A_{t_0}(\tau)$, as a function of the maximum, is also nondecreasing. Therefore, the right-hand side of formula (10) is a nonincreasing function and, therefore, the equality is satisfied for all $\tau$ starting from some number $t$.

Now, we prove the assertion of the theorem. By virtue of formula (8), $l\{y(t_0 + t) = x(t_0 + t)|y(t_0) = x(t_0)\}$ does not exceed $-\ln(\sqrt{2\pi}\Sigma_{t_0, t_0+t})$. Therefore, if there exists at least one point $\tau < t$ such that $l\{y(t_0 + \tau) = x(t_0 + \tau)|y(t_0) = x(t_0)\} \ge -\ln(\sqrt{2\pi}\Sigma_{t_0, t_0+t})$, then, certainly, the point $x(t_0 + t)$ is not the next admissible point. In other words, the sufficient condition of termination is the inequality $A_{t_0}(t) \ge -\ln(\sqrt{2\pi}\Sigma_{t_0, t_0+t})$. After simple transformations, we obtain expression (10).

In order to prove the final part of the theorem, it is sufficient to show that the function $A(t)$ is continuous. Then, taking into account that the other functions involved in the inequality are continuous and monotonic, we obtain from the theorem of passing the zero point by a continuous function that Eq. (11) has a unique solution.

Consider the likelihood function $l\{y(t) = x(t)|y(t_0) = x(t_0)\}$. It is easily seen that, since the centroidal and variational fields are continuous, it is sufficient to examine the continuity of the numerator $(x(t) - x(t_0) - M_{t_0, t})^2$. By the condition of the theorem, $x(t)$ is continuous for $t > t_0$. This implies that the numerator of the likelihood function is continuous. Therefore, the function itself is continuous, and the operation of finding the supreme value leaves the function in the class of continuous functions.

The theorem is completely proved.

This theorem means that the probability filtering is applicable to processing continuous real-time signals.

## 4. DESIGN OF A DISCRETE PROBABILITY FILTER

In this section, we assume that the signal is given at discrete time instants. Assume that there is a noisy signal known at instants $t_1, \ldots, t_T$. Consider the case of the homogeneous grid, i.e., $t_i = t_1 + h(i - 1)$, and denote the original signal by $x[i]$, where $i = 1, 2, \ldots, T$.

Similar to Section 2, the task is to construct a filter forming the output signal $y[i]$ with the minimum possible susceptibility to random distortions. If there are no such distortions, the signal should not be changed, i.e., $y[i] = x[i]$. Assume that there is a point $x[i_0]$ which is known to be the first point of the useful signal, i.e., $\forall l \le i < i_0$, $y[i] \ne x[i]$, $y[i_0] = x[i_0]$. Following the results obtained in Section 2, determine the measure of likelihood that this point belongs to the useful signal under the condition that the point $x[i_0]$ belongs to the useful signal in the following way:

$$l\{y[i] = x[i]|y[i_0] = x[i_0]\} = -\frac{(x[i] - x[i_0] - M_{i, i_0})^2}{2\Sigma_{i, i_0}^2} - \ln(\sqrt{2\pi}\Sigma_{i, i_0}). \tag{12}$$

Here, as previously, M and $\Sigma$ are expressed in terms of the centroidal and variational fields, respectively,

$$M_{i, i_0} = \sum_{k=i_0}^{i} \mu[k], \tag{13}$$

$$\Sigma_{i, i_0} = \sqrt{\sum_{k=i_0}^{i} \sigma^2[k]}. \tag{14}$$

If the useful point of the signal is fixed, the next point is found by maximization of likelihood function (12) with respect to the time instants. Thus obtained points are called *admissible*. The other points of the signal $x[i]$ are treated as noise and ignored. Thus, the useful signal $y[i]$ is represented by a sequence of admissible points and interpolated signal between them.

For realization of the above filtering algorithm, it is necessary to evaluate the centroidal and variational fields. The centroidal field $\mu[i]$ which is the derivative of the expectation of the random process under study can be interpreted as our notion of the tendency of variation of the useful signal value on some segment, and the variational field characterizes our a priori assumptions on the local measure of variation of the useful signal at the point. To determine the values of the functions $\mu[i]$ and $\sigma[i]$, the idea concerning the approximate shape of the signal or preliminary processing of the signal under study by low-frequency filters, for example, by a moving average, can be used. Assume that $z[i]$ is the signal obtained by smoothing the original signal. Then, the centroidal and variational fields can be determined by the formulas

$$\mu[i] = z[i+1] - z[i], \quad i = 1, \dots, T-1,$$

$$\mu[T] = 0,$$

$$\sigma[i] = \lambda + \delta|\mu[i]|, \quad i = 1, \dots, T,$$

where the real parameters $\lambda$ and $\delta$ are chosen with due regard for the particular shape of the signal and the noise level.

Let us prove the analogue of Theorem 2 for the case of discrete signals. Denote $A_{i_k}[t] = \max\limits_{i_k < i \le i_k + t} l\{y[i] = x[i] | y[i_0] = x[i_0]\}$. Then, the following theorem is valid.

**Theorem 3.** To determine the next admissible point of the signal, it is sufficient to check $t$ points, where $t$ is the minimum number for which the following inequality is satisfied:

$$\sum_{i = i_k}^{i_k + t} \sigma[i]^2 \ge \frac{\exp(-2A_{i_k}[t])}{2\pi}. \tag{15}$$

**Proof.** We show that, if formula (15) is satisfied for some $t$, then it is also satisfied for all $\tau > t$. Indeed, the left-hand side of the formula obviously is nondecreasing with increasing $t$. The function of maximum $A_{i_k}[t]$ is also nondecreasing. Hence, the right-hand side of formula (15) is a nonincreasing function, and, therefore, the inequality is satisfied for all $\tau$ starting from a number $t$.

Now, let us prove the assertion of the theorem. By virtue of formula (12), the value $l\{y[i_k + t] = x[i_k + t] | y[i_k] = x[i_k]\}$ does not exceed $-\ln(\sqrt{2\pi}\Sigma_{i_k, i_k + t})$. Therefore, if there is at least one point $\tau < t$ such that $l\{y[i_k + \tau] = x[i_k + \tau] | y[i_k] = x[i_k]\} \ge -\ln(\sqrt{2\pi}\Sigma_{i_k, i_k + t})$, then the point $x[i_k + t]$ is certainly not the next admissible point. In other words, the sufficient condition of termination is the inequality $A_{i_k}[t] \ge -\ln(\sqrt{2\pi}\Sigma_{i_k, t})$. After simple transformations, we obtain expression (15). *The theorem is proved.*

This theorem provides filtering of infinite and real-time signals. In the last case, formula (15) allows determination of the filtering delay.

## 5. DETERMINATION OF INITIAL POINT OF THE SIGNAL

Consider now in more detail the problem of finding the first admissible point. The first point of the original signal often serves as the first admissible point. However, situations are possible in which a group of initial points of the original signal turns out to be noise and the problem of finding the first admissible point becomes nontrivial. Here, a procedure based on augmenting the original signal with one point and further selection among different initial points is assumed.

More formally, the procedure of finding the initial point is as follows.

A. Choose the interval of search for the initial point $[0, T_1]$.

B. Take the first point of the original signal $x[1]$ as the first admissible point $x[i_0]$.

C. Using the centroidal field, augment the original signal with the point $x[0] = x[i_0] - \sum\limits_{i=1}^{i_0} \mu[i]$ and filter the signal on the interval $[0, T_1]$ from the initial point $x[i_0]$.

D. Calculate the quality criterion of the obtained filtering $F(i_0)$.

E. $i_0 = i_0 + 1$. If $i_0 > T_1$, go to step C.

As the quality criterion of the filtering from different initial points, we can take the number of points declared to be noise. The smaller number of points filtered for a fixed initial point, the more probable it is that this point is an admissible point of the signal.

It follows from Theorem 3 that, to search for an admissible point, it is sufficient to take $T_1$ as the minimum $t$ satisfying the condition

$$\sum_{i=1}^{t} \sigma[i]^2 \ge \frac{\exp(-2A_1[t])}{2\pi}.$$

## 6. USING THE DYNAMIC PROGRAMMING METHOD FOR OPTIMAL PROBABILITY SIGNAL FILTERING

The approach to signal filtering described above has an essential disadvantage. The sequence of admissible points is constructed "from left to right" and, at each step, the set of admissible points is augmented by one point with the best local characteristics. Thus, the proposed method is the realization of the so called "greedy" algorithm, and it possesses all the disadvantages entailed. In particular, such filtering is, generally speaking, an asymmetric procedure. If the signal is examined from end to beginning, the filtered signal can differ from the signal obtained by filtering from beginning to end. Moreover, locally choosing the most likely signals at each step, the whole signal is not necessarily the most likely one. To solve the problem of optimal (in the sense of maximum likelihood) signal filtering, the well-known Bellman principle [3] can be used. Although the strict proof of validity of the principle of local optimal character of the filtered signal is a separate nontrivial problem, intuitively, its applicability seems quite obvious. Otherwise, the optimal sequence on this segment would be formed by other points and its general likelihood would be higher. Pass to a formal description of the method. We assume that the points $x[1]$ and $x[T]$ are points of the useful signal. The likelihood of the most probable trajectory from the point 1 to the point $i$ is taken as the Bellman function

$$V(i) = P_{i_0 i_1} + P_{i_1 i_2} + \dots + P_{i_{k-1} i_k},$$

where

$$P_{ij} = \frac{L_{ij} + L_{ji}}{2},$$

$$L_{ij} = \frac{\sum_{s=1}^{j-i} 1\{(y[i+s] = \alpha x[j]) + (1-\alpha)x[i] \mid y[i] = x[i]\}}{j-i}.$$

Here, $\alpha = s/(j-i)$. This formula allows commensurable values for different values of the difference $j - i$. Denote by $R(i)$ the number of the last admissible points preceding $x[i]$ (under the condition that $x[i]$ is an admissible point). Then, the filtering algorithm based on the principle of dynamic programming can be represented as follows. During the first run, the Bellman function $V(i)$ is calculated for each point of the signal. Assuming that $V(1)$ is zero, for each next point, we obtain the formula

$$V(i) = \max_{1 \le j < i} (V(j) + P_{ji}).$$

For each $i$, we find $R(i) = \arg\max_{1 \le j < i} (V(j) + P_{ji})$. Passing all points of the signal, we obtain a sequence of admissible points in the inverse order, namely, $\{T, R(T), R(R(T)), \dots, 1\}$.

This algorithm finds the most likely (in the sense of the above measure of likelihood) trajectory of the useful signal on the whole time interval, rather than locally for each particular instant. Obviously, this procedure is symmetric and the obtained signal is optimal. The disadvantage of the dynamic programming method as applied to the probability signal filtering is a rather low processing rate and impossibility of real-time filtering (i.e. the method is applicable to finite signals which are fully known at the beginning of filtering).

## 7. RESULTS

The above technique has been used for analysis of discrete signals in solving the problem of determining narcotic intoxication by the reaction of the pupil of an eye to a light burst. The necessary data were provided by Iritech Inc. Figure 1 shows the shape of a typical pupilogram (a sequence of pupil sizes for 2.5 seconds after the exciting pulse). The ratio of the diameter of the pupil and the diameter of the iris is shown along the ordinate axis. The qualitative form of the pupilogram is well known to specialists [4]. Unfortunately, during taking of data there often arises noise due to camera defocusing, movement of the head, blinking of the patient, and so on. This results in multiple distortions to be processed. Figures 2 and 3 show the results of application of the probability and ordinary low-frequency filters to a typical noisy signal. Some parts of the pupilograms (circumscribed by ellipses in the figures) are of special importance for further analysis, and their shape should be minimally distorted during filtering. It is well seen in the figures that the probability filtering provides a more precise determination of the corresponding parts of the pupilograms.

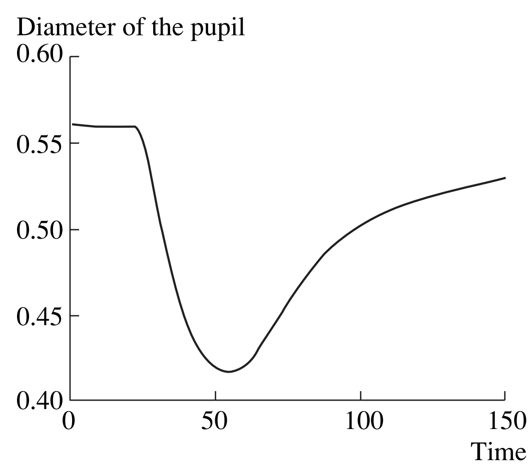


**Fig. 1.** General form of a pupilogram.

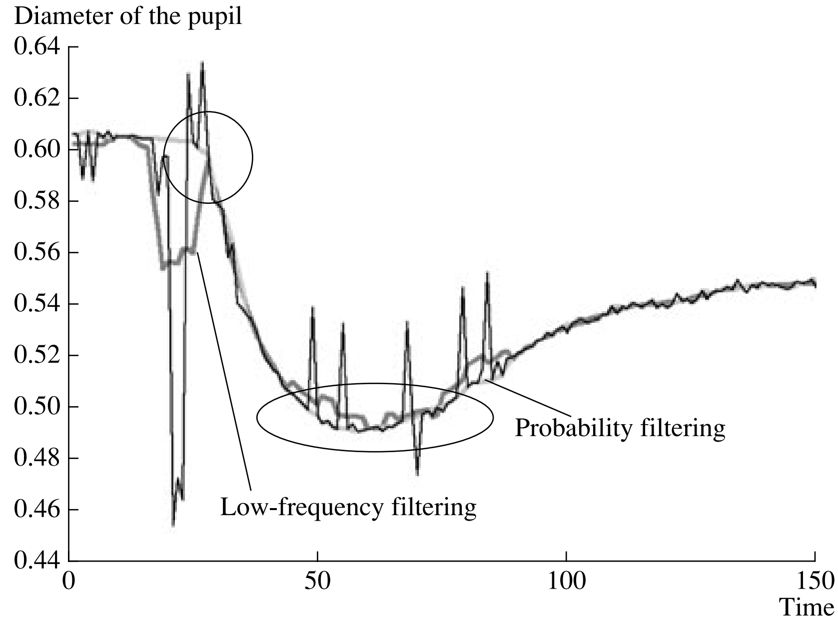


**Fig. 2.** Application of the probability and linear low-frequency filters to a noisy signal.

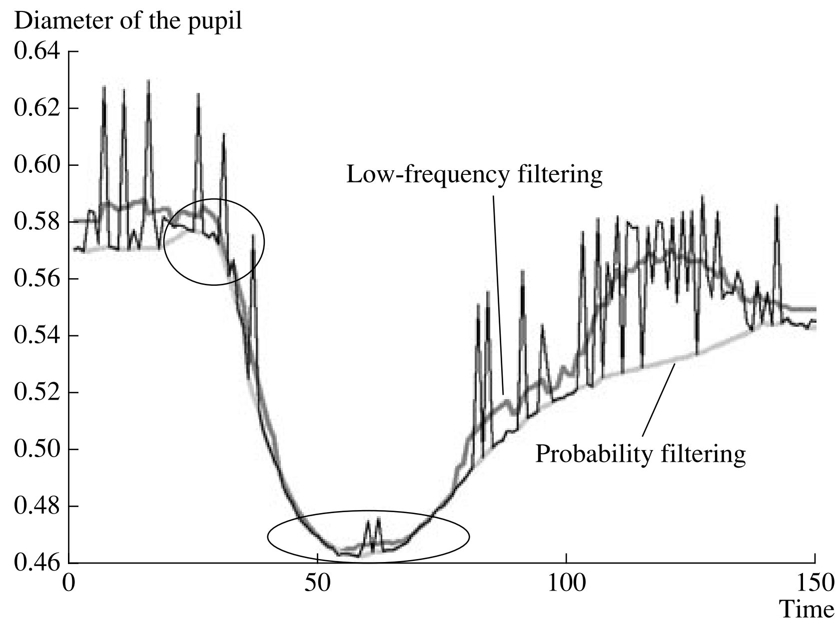


**Fig. 3.** Application of the probability and linear low-frequency filters to a noisy signal.

A good quality of filtering for problems similar to the one described is reached by rejecting traditional characteristics of the filters—pulse and threshold responses. Depending on the values of the pulse and threshold, the response can vary. This provides elimination of random noise, conserving the local specific features of the signal.

## ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research, project nos. 03-01-00580, 04-01-08045, and 04-01-00161; by INTAS, grant YS 04-83-2942; by the Goal-Oriented Program DMS RAS no. 2; and by a grant from the President of the Russian Federation, NSH-1721.2003.1.

REFERENCES

1. V. V. Mottle and I. B. Muchnik, *Hidden Markov Models in Signal Structural Analysis* (Fizmatlit, Moscow, 1999).
2. S. W.Smith, *The Scientist and Engineer's Guide to Digital Signal Processing* (Techn. Publs., California, 1997).
3. R. Bellman, *Dynamic Programming* (Princeton University Press, Princeton, NJ, 1957).
4. T. I. Dolmatova, N. D. Graevskaya, D. Kim, et al., "Screening Method of Binocular Pupilometry in Monitoring of Functional State of Athletes," Teoriya i Praktika Fizicheskoi Kul'tury, No. 6 (2001).

**Dmitrii P. Vetrov.** Born 1981. Post-graduate student at Moscow State University, to graduate in 2006. Works at the Dorodnicyn Computing Center, Russian Academy of Sciences. Scientific interests: pattern recognition, machine learning, mathematical statistics, expert systems. Nine publications.

**Dmitrii A. Kropotov.** Born 1981. Post-graduate student at Moscow State University, to graduate in 2006. Works at the Dorodnicyn Computing Center, Russian Academy of Sciences. Scientific interests: pattern recognition, machine learning, mathematical statistics, expert systems. Eight publications.