MAXIMUM A POSTERIORI LIKELIHOOD DETECTION USING LEMPEL-ZIV ALGORITHM

T. T. Kadota

AT&T Bell Laboratories
Murray Hill, NJ 07974

RESUME

On etudie un schéma semi-déterministe pour la détection d'un signal déterministe en présence de bruit blanc normal et d'une perturbation transitoire. Le schéma estime la perturbation par la méthode de la vraisemblance maximum a posteriori et ensuite forme le rapport de vraisemblance apartir de cette estimation. On procède en deux étapes: d'abord la présence ou l'absence de la perturbation est décidée, et, en cas de presence, la perturbation est estimée. On donne deux examples: dans le premier la perturbation a une forme connue et une époque d'apparition aléatoire; dans le second c'est un bruit gaussien avec une amplitude partiellement Rayleigh. L'algorithme de Lempel-Ziv est proposé comme procédure simplifiée pour décider la présence de la perturbation. Cette alternative a des avantages quand la perturbation est rare et le filtre adapté est approprié la plupart du temps.

SUMMARY

A semi-deterministic scheme is presented for detecting a deterministic signal in white Gaussian noise and a random transient disturbance. This scheme estimates the disturbance via the maximum a posteriori likelihood method and then forms the likelihood ratio using this estimate. The estimation proceeds in two stages: first, presence or absence of the disturbance is decided, and, if present, then the disturbance itself is estimated. Two examples are given to illustrate the scheme: in the first, the disturbance has a known shape with a random appearance time; in the second, it is a Gaussian noise with a partially Rayleigh distributed amplitude. The Lempel-Ziv algorithm is proposed as a simpler alternative to the computationally burdensome procedure for deciding the presence of the disturbance. Such an alternative is advantageous when the disturbance is rare and, hence, the matched filter is appropriate most of the time.

I. INTRODUCTION

Consider a problem of detecting a deterministic signal in white Gaussian noise and a transient disturbance which has a known waveform but unpredictable and infrequent occurrence. Stated more precisely,

$$dx(t) = \begin{cases} s(t)dt + f(t-u)dt + dw(t)\,, & (H_1) \quad (1a) \\ & 0 < t \le T\,, \\ f(t-u)dt + dw(t)\,, & (H_0) \quad (1b) \end{cases}$$

$dx(\cdot)$ is the observable data,

$s(\cdot)$ is the deterministic signal,

$f(\cdot)$ is a known waveform with $f(t) = 0\,, \quad t \le 0 \quad \text{or} \quad t > T\,,$

u is a random appearance time with the probability of appearance ε and, once it appears, it is uniformly distributed on $(0, T)$, namely,

$$P(u \ge T) = 1 - \varepsilon\,, \quad P(t < u \le t + dt) = \varepsilon\, dt/T\,, \qquad (2)$$

$dw(\cdot)$ is white Gaussian noise with power spectral density σ^2 , i.e. ,

$$E \mid dw(t) \mid^2 = \sigma^2 dt .$$

The likelihood ratio for this detection problem is easily derived as follows:

$$\frac{dP_1}{dP_0} = \frac{dP_1/dP_w}{dP_0/dP_w} ,$$

$$\frac{dP_1}{dP_w}(x) = E_u \exp\left\{\frac{1}{\sigma^2}\left[(s+f_u,x) - \frac{1}{2}\|s+f_u\|^2\right]\right\} , \quad (3)$$

$$\frac{dP_0}{dP_w} = \frac{dP_1}{dP_w}\Bigg|_{s=0} ,$$

thus

$$\log \frac{dP_1}{dP_0}(x) = \frac{1}{\sigma^2}\left[(s,x) - \frac{1}{2}\|s\|^2\right]$$

$$+ \log \frac{1-\varepsilon+\frac{\varepsilon}{T}\int_0^T \exp\left\{\frac{1}{\sigma^2}\left[(f_u,x-s) - \frac{1}{2}\|f_u\|^2\right]\right\}du}{1-\varepsilon+\frac{\varepsilon}{T}\int_0^T \exp\left\{\frac{1}{\sigma^2}\left[(f_u,x) - \frac{1}{2}\|f_u\|^2\right]\right\}du} \quad (4)$$

where P_1 and P_0 are probability measures induced by the right-hand sides of (1a) and (1b) respectively, and P_w is the Wiener measure corresponding to white Gaussian noise. E_u denotes the expectation with respect to u and

$$x = \{dx(t) ,\ \ 0 < t \le T\} ,\quad s = \{s(t) ,\ \ 0 \le t \le T\} ,$$

$$f_u = \{f(t-u) ,\ \ 0 \le t \le T\} ,\quad (s,x) = \int_0^T s(t)dx(t) .$$

The first term of (4) is the familiar log likelihood ratio for detecting a deterministic signal in white Gaussian noise which yields the matched filter as the optimum processor. The second term is due to the possible occurrence of the transient disturbance. According to the Neyman-Pearson theorem, use of this log likelihood ratio gives the maximum detection probability for a given false-alarm probability.

It is noteworthy that addition of such a simple disturbance, even with rare occurrence ($\varepsilon << 1$), should introduce a considerable complication in processing the data. One might wonder if he should choose the simple matched filter which is the appropriate processor most of the time. This brings up the question of how seriously one should take the completely probabilistic optimality criteria such as the Neyman-Pearson.

In the absence of the disturbance, i.e., when white Gaussian noise is the only noise in the data, the matched filter is traditionally operated at high output S/N (signal-to-noise ratio), which is achieved either by a high input S/N or a long observation time (if the signal has a long duration). Thus, the output is sufficiently greater when the signal is present than when it is absent so that the decision on the signal presence is nearly deterministic. In other words, although the matched filter is derived as the optimum processor according to the probabilistic criterion, its performance is deterministically satisfactory.

When the transient disturbance is introduced, the situation changes greatly. Because of its rare occurrence, the data on hand most likely contain no disturbance and, hence, use of (4) rather than the simple matched filter results in inferior detection performance. In addition, the data-processing complexity due to the correction term in (4) makes this detection statistic unattractive, specially because the correction term is unnecessary most of the time. On the other hand, when the disturbance does appear and, especially if it resembles the signal and is large in comparison, the correction term may not be adequate to avoid the false-alarm. Thus, use of the completely probabilistic criteria such as the Neyman-Pearson in this case leads to unsatisfactory results.

In this paper we propose a mixed strategy where the disturbance is treated as a semi-deterministic object while white Gaussian noise is as a random object. That is, the disturbance is estimated using the maximum a posteriori likelihood method and then the classical likelihood ratio is formed using this estimate. The estimation proceeds in two stages: first to determine if the disturbance is absent, i.e., $u \ge T$, and, if not, then to determine where it is in $[0, T)$. This estimation - detection procedure is illustrated by the next two examples.

II. EXAMPLE A

In the classical formulation where the data x is a finite sequence, the likelihood function is the joint probability density of x viewed as a function of the parameter u. When x becomes an infinite sequence or a function, the natural generalization is the Radon-Nikodym derivative with respect to some reference measure which does not contain the parameter. Recall that the probability density is the Radon-Nikodym derivative with respect to Lebesque measure. The natural choice for our reference measure is the Wiener measure P_w. Then, the likelihood functions under two hypotheses H_1 and H_0 of (1a) and (1b) are

$$\frac{dP_1 \mid u}{dP_w} = \psi(x,s)\exp\left\{\frac{1}{\sigma^2}\left[(f_u,x-s) - \frac{1}{2}\|f_u\|^2\right]\right\} ,\ \text{under } H_1 \quad (7a)$$

$$\frac{dP_0 \mid u}{dP_w} = \exp\left\{\frac{1}{\sigma^2}\left[(f_u,x) - \frac{1}{2}\|f_u\|^2\right]\right\} \ \text{under } H_0 \quad (7b)$$

where

$$\psi(x,x) = \exp\left\{\frac{1}{\sigma^2}\left[(s,x) - \frac{1}{2}\|s\|^2\right]\right\}$$

and $P_1 \mid_u$ and $P_0 \mid_u$ denote the probability measures induced by (1a) and (1b) respectively when u is regarded as a fixed constant. Upon combination with the "a priori" probability of u given by (2), the maximum a posterior likelihood estimation of u proceeds in two stages. First we compare the following four likelihoods:

$$(1-\varepsilon)\frac{dP_1 \mid_{u\ge T}}{dP_w} = (1-\varepsilon)\psi(x,s) ,\quad (1-\varepsilon)\frac{dP_0 \mid_{u\ge T}}{dP_w} = 1-\varepsilon , \quad (8)$$

$$\varepsilon\frac{dP_1 \mid_{u<T}}{dP_w} = \frac{\varepsilon}{T}\psi(x,s)\int_0^T \exp\left\{\frac{1}{\sigma^2}\left[(f_u,x-s) - \frac{1}{2}\|f_u\|^2\right]\right\}dt , \quad (9a)$$

$$\varepsilon \cdot \frac{dP_0 \mid_{u<T}}{dP_w} = \varepsilon\frac{1}{T}\int_0^T \exp\left\{\frac{1}{\sigma^2}\left[(f_u,x) - \frac{1}{2}\|f_u\|^2\right]\right\}dt . \quad (9b)$$

If either one in (8) is the largest, the maximum a posteriori likelihood estimate of u, denoted by $\hat{u}(x-s)$ (or $\hat{u}(x)$), is greater than or equal to T and hence $f_{\hat{u}} = 0$. Thus, the detection statistic is given by

$$\sigma^2 \log \frac{\dfrac{dP_1 \mid u}{dP_w}\Big|_{u\ge T}}{\dfrac{dP_0 \mid u}{dP_w}\Big|_{u\ge T}} = (s,x) - \frac{1}{2}\|s\|^2 . \quad (10)$$

That is, we decide that the disturbance is absent during the observation period and, therefore, we use the matched-filter output as the detection statistic.

On the other hand, if (9a) (or (9b)) is the largest, $\hat{u}(x-s)$ (or $\hat{u}(x)$) is taken as the value of u that maximizes (7a) (or (7b)), and the detection statistic is given by

$$\sigma^2 \log \frac{\dfrac{dP_1 \mid_u}{dP_w} \Big|_{u=\hat{u}}}{\dfrac{dP_0 \mid_u}{dP_w} \Big|_{u=\hat{u}}} = (s, x - f_{\hat{u}}) - \frac{1}{2} \| s \|^2 \tag{11}$$

where $\hat{u} = \hat{u}(x-s)$ if (9a) is the largest (or $\hat{u} = \hat{u}(x)$ if (9b) is the largest). That is, we decide that the disturbance is present and estimate the appearance time, which is the only unknown parameter, and subtract the estimated disturbance from the data before matched-filtering.

Observe that the disturbance has been treated as a semi-deterministic object while the white Gaussian noise is treated as a genuinely probabilistic object. As such object, the disturbance (actually the appearance time) is estimated and the probability distribution associated with it is used in this (maximum a posteriori likelihood) estimation as the a priori probability. Once the disturbance is estimated, it is treated as a known object. That is, it is simply removed from the data and the detection problem is reduced to the classical one. We note that whether it is the decision on the presence-or-absence of the disturbance or the estimation of the appearance time after it is decided to be present, there are two likelihood functions to be considered corresponding to two hypotheses. This is characteristic of the estimation in the detection context.

III. EXAMPLE B

Instead of using the appearance time, we may use the amplitude of the transient disturbance to characterize its occurrence as follows:

$$dx(t) = \begin{cases} s(t)dt + rz(t)dt + dw(t)\,, & (H_1) \quad (12a) \\ & 0 < t \le T\,, \\ rz(t)dt + dw(t)\,, & (H_0) \quad (12b) \end{cases}$$

where $\{z(t),\ 0 \le t \le T\}$ is a Gaussian process with mean 0 and covariance function R, and

$$P(r=0) = 1 - \varepsilon\,, \quad P(0 < r < \xi) = \varepsilon \int_0^{\xi} \frac{\eta}{\alpha^2} \exp\left[-\frac{\eta^2}{2\alpha^2}\right] d\eta \tag{13}$$

and r is independent of w and z. The disturbance in this example is a Gaussian noise and its rare occurrence is described by the high probability $1-\varepsilon$ that the amplitude is 0 and, once it occurs, its effect is characterized by the probability distribution of the amplitude, which is taken to be Rayleigh. This model is of special interest as $rz(t)dt + dw(t)$ for each t has a probability density which is Gaussian near the origin but exponentially decaying far away from the origin. Such a probability density was shown to be the least favorable density in the "contaminated Gaussian" class and the limiter-correlator was shown to be the "locally optimum" detector (detection statistic) [1].

The two likelihood functions, corresponding to (7a) and (7b), in this example are given respectively by

$$\frac{dP_1 \mid_{z,r}}{dP_w} = \psi(x,s) \exp\left\{\frac{1}{\sigma^2}\left[r\,(z, x-s) - \frac{1}{2} r^2 \| z \|^2\right]\right\} \tag{14a}$$

$$\frac{dP_0 \mid_{z,r}}{dP_w} = \exp\left\{\frac{1}{\sigma^2}\left[r(z,x) - \frac{1}{2} r^2 \| z \|^2\right]\right\} \tag{14b}$$

and the four likelihoods to be compared are

$$(1-\varepsilon)\,\psi\,(x,s)\,, \quad 1-\varepsilon\,, \quad E_{z,r>0} \frac{dP_1 \mid z,r}{dP_W}\,, \quad E_{z,r>0} \frac{dP_0 \mid_{z,r}}{dP_w} \tag{15}$$

where $E_{z,r>0}$ denotes the expectation with respect to the process z and the random variable r under the constraint $r > 0$. Evaluation of the expectation with respect to z is carried out via the Karhunen-Loève expansion of z:

$$z(t) = \sum_j \zeta_j \phi_j(t)\,, \quad 0 \le t \le T\,, \tag{16}$$

where

$$\int_0^T R\,(t,t')\,\phi_j(t')dt' = \sigma^2 \lambda_j \phi_j(t)\,, \quad 0 \le t \le T\,, \tag{17}$$

and ζ_j, $j = 1, 2, \ldots$ are independent Gaussian variables with mean 0 and variance $\sigma^2 \lambda_j$. Thus, for example,

$$E_{z,r>0} \exp\left\{\frac{1}{\sigma^2}\left[r(z,x) - \frac{1}{2} r^2 \| z \|^2\right]\right\}$$

$$= E_{r>0} \prod_j (1+r^2\lambda_j)^{-\frac{1}{2}} \exp\left[\frac{1}{2\sigma^2} \frac{r^2\lambda_j}{1+r^2\lambda_j} (\phi_j, x)^2\right]$$

$$\left[\frac{2\pi\sigma^2\lambda_j}{1+r^2\lambda_j}\right]^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1+r^2\lambda_j}{2\sigma^2\lambda_j}\left[\zeta - \frac{r\lambda_j}{1+r^2\lambda_j}(\phi_j, x)\right]^2\right\} d\zeta \tag{18}$$

$$= \varepsilon \int_0^{\infty} \left[\prod_j (1+q\lambda_j)\right]^{-\frac{1}{2}} \exp\left[\frac{1}{2\sigma^2} \sum_j \frac{q\lambda_j}{1+q\lambda_j} (\phi_j, x)^2\right] p\,(q)dq \tag{19}$$

where $p\,(q)$ is the probability density of $q = r^2$ which is exponentially distributed with parameter $2\alpha^2$.

If either $(1-\varepsilon)\,\psi(x,s)$ or $1-\varepsilon$ is the largest among the four members of (15), we decide the disturbance to be absent and the detection statistic is the right-hand side of (10), namely, the output of the matched filter minus the bias $\| s \|^2/2$. Otherwise, the likelihood function (14a) or (14b) must be maximized with respect to the probability measures induced by z and r to obtain the maximum a posteriori likelihood estimates of z and r. The maximization regarding z is carried out again via the Karhunen-Loève expansion, namely, by maximizing with respect to the probability density of ζ_j for each $j = 1, 2, \ldots$.

Suppose the last member of (15) is the largest. By examining the integrand of (18) and through (16), it is easily seen that the maximum a posteriori likelihood estimate of z given r is

$$\hat{z}(x \mid r) = \sum_j \frac{r\lambda_j}{1+r^2\lambda_j} (\phi_j, x)\,\phi_j\,. \tag{20}$$

The maximum a posteriori likelihood estimate of r, denoted by $\hat{r}$, is obtained as $\hat{r} = (\hat{q})^{½}$ where $\hat{q}$ is the solution of the equation

$$\sum_j \frac{\lambda_j}{(1+\lambda_j q)^2} (\phi_j, x)^2 = \left[\frac{\sigma}{\alpha}\right]^2. \tag{21}$$

Then the maximum a posteriori likelihood estimate of $rz(\cdot)$ is given by

$$\hat{r}\,\hat{z}(x \mid r=\hat{r}) = \sum_j \frac{\hat{r}^2\lambda_j}{1+\hat{r}^2\lambda_j} (\phi_j, x)\phi_j\,. \tag{22}$$

In the case where the third member of (15) is the largest, the maximum a posteriori likelihood estimate of $rz(\cdot)$ is obtained by simply replacing x in (22) with $x-s$ where $\hat{r}$ is the same functional but of $x-s$ in this case.

By substituting each of (22) and its "$x-s$" version into (14a) and (14b) and by taking the logarithm of the ratio of the two, we obtain the detection statistic

$$\sigma^2 \log \frac{\left.\dfrac{dP_1 \mid_{z,r}}{dP_w}\right|_{\substack{z=\hat{z}\\ r=\hat{r}}}}{\left.\dfrac{dP_0 \mid_{z,r}}{dP_w}\right|_{\substack{z=\hat{z}\\ r=\hat{r}}}}$$

$$= \begin{cases} (s,\, x - \hat{r}\hat{z}\,(x-s \mid r=\hat{r})) - \frac{1}{2}\, \| s \|^2 , & (23a) \\ \text{where } \hat{r} = \hat{r}\,(x-s) & \\ \qquad \text{if } E_{z,\, r>0} \dfrac{dP_1 \mid_{z,r}}{dP_w} \text{ is the largest,} & \\ (s,\, x - \hat{r}\hat{z}\,(x \mid r=\hat{r})) - \frac{1}{2}\, \| s \|^2 & (23b) \\ \text{where } \hat{r} = \hat{r}(x) & \\ \qquad \text{if } E_{z,\, r>0} \dfrac{dP_0 \mid_{z,r}}{dP_w} \text{ is the largest,} & \\ (s,\, x) - \frac{1}{2}\, \| s \|^2 \text{ otherwise.} & \end{cases}$$

By substituting (20) into (23b), we find the input to the matched filter to be

$$x - \hat{r}\,\hat{z}\,(x \mid r=\hat{r})\Big|_{\hat{r}=\hat{r}(x)} = \sum_j \frac{1}{1+\hat{r}^2(x)\lambda_j}\, (\phi_j, x)\phi_j$$

which is in the form of a generalized process. Then the left-hand side of (21) is seen as the norm-square of $x - \hat{r}\,\hat{z}(x \mid r=\hat{r})$ relative to the kernel R. Thus, when it is decided that the disturbance is present, the nonlinear device sets the norm (relative to R) of the input data to a constant, namely, σ/α. In other words, instead of the input itself being clamped (or limited) at every instant of time, as in [1], the norm of the input is clamped and this norm-clamped output of the nonlinear device is matched-filtered.

IV. DETERMINATION OF DISTURBANCE PRESENCE BY LEMPEL-ZIV ALGORITHM

We observe in (8) - (9b) and (15) - (19) that determining the presence of the transient disturbance by comparing the four likelihoods is computationally too burdensome, especially in the case of (19). Hence we propose as a simpler, though ad hoc, alternative the Lempel-Ziv parsing algorithm [2]. In order to apply the algorithm, the instantaneous output $y\,(t)$ of the matched filter must be discretized and quantized first. Define two sequences $(y_1, \ldots, y_n)$ and $(v_1, \ldots, v_n)$ by

$$y_i = s\,(\frac{i-1}{n}\,T) \int_{\frac{i-1}{n}T}^{\frac{i}{n}T} dx(t)\,, \qquad v_i = \begin{cases} 0 & \text{if } y_i < 0\,, \\ 1 & \text{otherwise}\,, \end{cases}$$

The parsing algorithm can be best described by an example. Suppose the given data result in the following binary sequence (with length n = 15):

0 1 0 0 0 1 0 0 0 0 0 1 1 1 0

Then we place a comma whenever a new subsequence (phrase) is encountered. Thus, we have

0,1,0 0,0 1,0 0 0,0 0 1,1 1,0 $\qquad c\,(n) = 7$

and the number of commas, denoted by $c\,(n)$, is 7. It is not too difficult to imagine that $c(n)$ would be smaller for a "regular" sequence than for a "random" sequence. For example, the sequence of all 0's has $c\,(n)$ increasing as $\sqrt{n}$ for a large n and, hence, $c\,(n) \log c\,(n)$ grows as $\sqrt{n}\,\log n$, while $c\,(n) \log c(n)$ for a sequence taken from independent and identical distributions is known to grow as n [2]. Therefore, a binary sequence resulting from the "signal-plus-noise" data would yield a smaller $c\,(n)$ than the one resulting from the "noise-alone" data. Thus, if the disturbance resembles the signal and, hence, is much more regular than the noise, presence of the disturbance, with or without the signal, should be exhibited by the decline in $c\,(n)$ from its value for the noise-alone data. Note the disturbances which differ significantly from the signal (or even orthogonal) are filtered out by the matched filter and, hence, there is no need for concern.

Numerical illustration using Example A is shown in Table I where the simulation using data-sequence of length 1000 is repeated 1000 times. The signal is taken to be a sinusoid with the period 100 and S/N = 40 at the matched-filter output. The disturbance appears at the mid-point of the observation interval, i.e., i = 501, but has a much larger amplitude, 5 times the signal, and slightly higher frequencies, 1.05 times the signal (1.05f) in the first case and 1.1 times (1.1f) in the second. In the first case, the matched-filter output is significantly increased from the "noise-alone" level whether the data contain the signal or the disturbance. Hence the threshold decision device would register "signal-present". However, the $c(n)$ value for the disturbance is much lower than the one for the signal. Hence, the disturbance can be detected by examining the $c\,(n)$. Note that it cannot detect the presence of the signal. In the second case, the matched-filter output for the disturbance is negligible compared to the one for the signal because the disturbance is sufficiently different from the signal and, hence, is filtered out. Nevertheless, the $c\,(n)$ values exhibit definite difference comparable to the previous case. Although these numerical results are only preliminary, we can conclude the following: the Lempel-Ziv algorithm used in conjunction with the matched filter (i) is effective when the signal is weak but has a long duration (i.e., a long observation time) and the disturbance resembles the signal and is larger in comparison; (ii) is not effective in detecting the signal whether the disturbance is present or not. Its main advantage is its simplicity and universality, not requiring the detailed knowledge of the disturbance.

		Min	Mean	Max	S.D.
No Signal &	(s, x)	−.0170	.0001	.0224	.0064
No Dist.	$c(n)$	168	171.9	175	1.0
Signal &	(s, x)	.0230	.0401	.0624	.0064
No. Dist.	$c\,(n)$	166	170.3	173	1.2
No Signal &	(s, x)	.0480	.0651	.0874	.0064
Dist. (1.05f)	$c(n)$	158	164.3	169	1.9
Signal &	(s, x)	.0880	.1054	.1274	.0064
Dist. (1.05f)	$c\,(n)$	151	159.2	165	2.1
No Signal &	(s, x)	−.0174	−.0003	.0220	.0064
Dist. (1.1f)	$c\,(n)$	159	166.1	171	1.7
Signal &	$c\,(n)$	.0226	.0397	.0620	.0064
Dist. (1.1f)	$c\,(n)$	158	164.7	169	1.8

Table I. Numerical Simulation of Example A,
(s, x) = Matched-Filter Output,
S.D. = Standard Deviation,
f = Signal Frequency.

REFERENCES

[1] R. D. Martin, S. C. Schwartz, "Robust Detection of a Known Signal in Nearly Gaussian Noise", IEEE Trans., IT-17, Jan. 1971, pp. 50-56.

[2] A. Lempel, J. Ziv, "On the Complexity of Finite Sequences", IEEE Trans., IT-22, Jan. 1976, pp. 75-81.