Nonlinear Test Statistic to Improve Signal Detection in Non-Gaussian Noise

François Chapeau-Blondeau, *Member, IEEE*

***Abstract*—We compare two simple test statistics that a detector can compute from multiple noisy data in a binary decision problem based on a maximum *a posteriori* probability (MAP) criterion. One of these statistics is the standard sample mean of the data (linear detector), which allows one to minimize the probability of detection error when the noise is Gaussian. The other statistic is even simpler and consists of a sample mean of a two-state quantized version of the data (nonlinear detector). Although simpler to compute, we show that this nonlinear detector can achieve smaller probability of error compared to the linear detector. This especially occurs for non-Gaussian noises with heavy tails or a leptokurtic character.**

***Index Terms*—Detection, non-Gaussian noise, nonlinear statistic, threshold nonlinearity.**

I. Introduction

LINEAR procedures are useful in signal processing because they usually allow a thorough theoretical treatment and can often be proved optimal when dealing with Gaussian noise. Yet linear procedures also come with inherent limitations. Nonlinear procedures are potentially richer but are also generally more difficult to theoretically tackle. Still, there are classes of nonlinear systems that may prove simple enough for a theoretical analysis (and for practical implementation), while offering improvement over linear processes in specific situations. This may be the case with threshold nonlinearities. Here, we study a specific problem of signal detection, and we compare the performance of a standard linear detector with that of a simple two-state threshold nonlinearity. This nonlinear detector is simple to implement and represents each data point by a parsimonious single bit. These properties are especially useful for applications in a number of existing and future multisensor networks or distributed intelligent systems, so as to maximize the speed and efficacy of processing with limited resources for data handling, storage, communication, and energy supply [1]. While simpler to compute, we show that with Gaussian noise, the nonlinear detector comes close to the performance of the optimal linear detector. With non-Gaussian noise, the nonlinear detector can improve the performance over a standard linear detector, and this especially occurs for generalized Gaussian noises with heavy tails or a leptokurtic character.

Manuscript received February 24, 2000. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. H. Messer-Yaron.

The author is with the Laboratoire d'Ingénierie des Systèmes Automatisés (LISA), Université d'Angers, Angers, France.

Publisher Item Identifier S 1070-9908(00)05874-0.

II. Linear Detection

A random signal s is assumed to take the constant value s_0 (hypothesis H_0) or $s_1 > s_0$ (hypothesis H_1), respectively, with probabilities P_0 and $P_1 = 1 - P_0$. The signal s is additively corrupted by a stationary white noise η of mean zero, variance σ_η^2, cumulative distribution function $F_\eta(u)$, and probability density function $f_\eta(u) = dF_\eta(u)/du$ (not necessarily Gaussian). The noisy signal $s + \eta = x$ is sampled so as to provide N observations x_k for $k = 1$ to N. The data samples x_k are therefore independent random variables with either mean s_0 or s_1 and variance σ_η^2. We want to use the data $\mathbf{x} = (x_1, \cdots x_N)$ to decide whether $s = s_0$ or s_1.

The maximum *a posteriori* probability (MAP) criterion uses the likelihood ratio [2]

$$\lambda = \frac{\Pr\{s = s_1|\mathbf{x}\}}{\Pr\{s = s_0|\mathbf{x}\}} = \frac{p_x(\mathbf{x}|s_1)P_1}{p_x(\mathbf{x}|s_0)P_0} \tag{1}$$

with the conditional density $p_x(\mathbf{x}|s_1) = \Pi_{k=1}^N f_\eta(x_k - s_1)$, and a similar expression for $p_x(\mathbf{x}|s_0)$. Equivalently, one can use the loglikelihood ratio

$$\ln(\lambda) = \ln(P_1/P_0) + \sum_{k=1}^N \ln[f_\eta(x_k - s_1)] - \sum_{k=1}^N \ln[f_\eta(x_k - s_0)]. \tag{2}$$

Whenever $\ln(\lambda) > 0$, the decision is that $s = s_1$ (decision D_1). Otherwise it is $s = s_0$ (decision D_0). The performance of the detection can be assessed by the overall probability of error

$$P_{\mathrm{er}} = \Pr\{\mathrm{D}_1|\mathrm{H}_0\}P_0 + \Pr\{\mathrm{D}_0|\mathrm{H}_1\}P_1. \tag{3}$$

The MAP criterion leads to a very simple test statistic (the quantity the detector has to compute from the data to base its decision) when the noise η is Gaussian [2]. In this case, the MAP criterion leads to decide $s = s_1$ when

$$\overline{x} > \frac{s_0 + s_1}{2} + \frac{\sigma_\eta^2/N}{s_1 - s_0} \ln(P_0/P_1) = x_T \tag{4}$$

and to decide $s = s_0$ when $\overline{x} < x_T$, with the statistic given by the sample mean of the data

$$\overline{x} = \frac{1}{N}\sum_{k=1}^N x_k, \tag{5}$$

and this procedure leads to a minimum of $P_{\rm er}$ with value

$$P_{\rm er} = \frac{1}{2}\left[1 + P_1 \ \mathrm{erf}\ \left(\sqrt{N}\frac{x_T - s_1}{\sqrt{2}\sigma_\eta}\right) - P_0 \ \mathrm{erf}\ \left(\sqrt{N}\frac{x_T - s_0}{\sqrt{2}\sigma_\eta}\right)\right]. \tag{6}$$

When the noise η is non-Gaussian, the MAP procedure in general does not lead to a simple test statistic, as simple as the sample mean $\overline{x}$ of (5) or simpler, in order to make an optimal use of the complete data set $\mathbf{x} = (x_1, \cdots x_N)$ to base the detection. Yet, it is often desirable, especially for fast real-time processing, to maintain a simple test statistic like $\overline{x}$ on which to base the detection. One is thus faced with the question: "Given the sample mean $\overline{x}$ of (5) to represent the data set, what is the best use of $\overline{x}$ on which to base the detection?" This question receives a simple answer in the case of a large data set (i.e., when N is large), because in this case, the Gaussian condition is recovered. Irrespective of the distribution of the noise η corrupting the data x_k, provided η can be assigned a finite variance σ_η^2, the sample mean $\overline{x}$, thanks to the central limit theorem, gets normally distributed with variance σ_η^2/N and either mean s_0 or s_1.

Thus, with an arbitrarily distributed noise η and N large, the optimal use of $\overline{x}$ for the detection is again to apply the decision scheme of (4), and this procedure reaches the probability of error (6), which is the minimal probability of error that can be expected when basing the detection on the statistic $\overline{x}$.

III. Nonlinear Detection

We will now introduce another test statistic, even simpler to compute than $\overline{x}$. We consider that the noisy signal $s + \eta = x$ is not observed directly but through a two-state nonlinearity

$$y = \mathrm{sign}(s + \eta - \theta) \tag{7}$$

with threshold $\theta = (s_0 + s_1)/2$. It is then the binary output $y = \pm 1$ that is sampled to yield the binary data set $(y_1, \cdots y_N)$. We next compute the sample mean of the binarized data

$$\overline{y} = \frac{1}{N}\sum_{k=1}^{N} y_k \tag{8}$$

and ask what is the best use of $\overline{y}$ that can be implemented to detect whether $s = s_0$ or s_1, and what is the performance of this detection?

The MAP detection based on the statistic $\overline{y}$ relies on the likelihood ratio

$$\lambda_y = \frac{\Pr\{s = s_1|\overline{y}\}}{\Pr\{s = s_0|\overline{y}\}} = \frac{p_y(\overline{y}|s_1)P_1}{p_y(\overline{y}|s_0)P_0} \tag{9}$$

with the conditional densities $p_y(\overline{y}|s_0)$ and $p_y(\overline{y}|s_1)$, which we shall now address.

For a fixed s and the noise η distributed according to F_η, the random signal y, defined by (7), has mean

$$\mathrm{E}(y) = 1 - 2F_\eta(\theta - s) \tag{10}$$

and variance

$$\mathrm{var}(y) = 4F_\eta(\theta - s)[1 - F_\eta(\theta - s)]. \tag{11}$$

<u>If</u> $\sigma_0 > \sigma_1$
  Let $\sigma_2 = \sqrt{\sigma_0^2 - \sigma_1^2}$
  <u>If</u> $(y_1 - y_0)^2 + 2\sigma_2^2 \ln[(\sigma_0 P_1)/(\sigma_1 P_0)] > 0$
    Let $y_2 = \sqrt{(y_1 - y_0)^2 + 2\sigma_2^2 \ln[(\sigma_0 P_1)/(\sigma_1 P_0)]}$
    Let $y_{T1} = (\sigma_0^2 y_1 - \sigma_1^2 y_0 - \sigma_0\sigma_1 y_2)/\sigma_2^2$
    Let $y_{T2} = (\sigma_0^2 y_1 - \sigma_1^2 y_0 + \sigma_0\sigma_1 y_2)/\sigma_2^2$
    <u>If</u> $y_{T1} < \overline{y} < y_{T2}$ **decide** $s = s_1$ <u>Else</u> **decide** $s = s_0$ <u>End If</u>
    $P_{\rm er} = \frac{1}{2}\left[\mathrm{erf}\left(\frac{y_{T2} - y_0}{\sqrt{2}\,\sigma_0}\right) - \mathrm{erf}\left(\frac{y_{T1} - y_0}{\sqrt{2}\,\sigma_0}\right)\right] P_0 + \left[1 + \frac{1}{2}\mathrm{erf}\left(\frac{y_{T1} - y_1}{\sqrt{2}\,\sigma_1}\right) - \frac{1}{2}\mathrm{erf}\left(\frac{y_{T2} - y_1}{\sqrt{2}\,\sigma_1}\right)\right] P_1$
  <u>Else</u>
    decide $s = s_0$
    $P_{\rm er} = P_1$
  <u>End If</u>
<u>Else If</u> $\sigma_0 < \sigma_1$
  Let $\sigma_2 = \sqrt{\sigma_1^2 - \sigma_0^2}$
  <u>If</u> $(y_1 - y_0)^2 + 2\sigma_2^2 \ln[(\sigma_1 P_0)/(\sigma_0 P_1)] > 0$
    Let $y_2 = \sqrt{(y_1 - y_0)^2 + 2\sigma_2^2 \ln[(\sigma_1 P_0)/(\sigma_0 P_1)]}$
    Let $y_{T1} = (\sigma_1^2 y_0 - \sigma_0^2 y_1 - \sigma_0\sigma_1 y_2)/\sigma_2^2$
    Let $y_{T2} = (\sigma_1^2 y_0 - \sigma_0^2 y_1 + \sigma_0\sigma_1 y_2)/\sigma_2^2$
    <u>If</u> $y_{T1} < \overline{y} < y_{T2}$ **decide** $s = s_0$ <u>Else</u> **decide** $s = s_1$ <u>End If</u>
    $P_{\rm er} = \left[1 + \frac{1}{2}\mathrm{erf}\left(\frac{y_{T1} - y_0}{\sqrt{2}\,\sigma_0}\right) - \frac{1}{2}\mathrm{erf}\left(\frac{y_{T2} - y_0}{\sqrt{2}\,\sigma_0}\right)\right] P_0 + \frac{1}{2}\left[\mathrm{erf}\left(\frac{y_{T2} - y_1}{\sqrt{2}\,\sigma_1}\right) - \mathrm{erf}\left(\frac{y_{T1} - y_1}{\sqrt{2}\,\sigma_1}\right)\right] P_1$
  <u>Else</u>
    decide $s = s_1$
    $P_{\rm er} = P_0$
  <u>End If</u>
<u>Else If</u> $\sigma_0 = \sigma_1$
  Let $y_T = \frac{y_0 + y_1}{2} + \frac{\sigma_0^2}{y_1 - y_0}\ln(P_0/P_1)$
  <u>If</u> $\overline{y} > y_T$ **decide** $s = s_1$ <u>Else</u> **decide** $s = s_0$ <u>End If</u>
  $P_{\rm er} = \frac{1}{2}\left[1 + P_1\,\mathrm{erf}\left(\frac{y_T - y_1}{\sqrt{2}\,\sigma_0}\right) - P_0\,\mathrm{erf}\left(\frac{y_T - y_0}{\sqrt{2}\,\sigma_0}\right)\right]$
<u>End If</u>

Fig. 1. Detection algorithm based on the nonlinear test statistic $\overline{y}$.

We now consider the case of N large. Then, thanks to the central limit theorem, the statistic $\overline{y}$ gets normally distributed with mean $y_0 = 1 - 2F_\eta(\theta - s_0)$ and variance $\sigma_0^2 = 4F_\eta(\theta - s_0)[1 - F_\eta(\theta - s_0)]/N$ when $s = s_0$ and mean $y_1 = 1 - 2F_\eta(\theta - s_1)$ and variance $\sigma_1^2 = 4F_\eta(\theta - s_1)[1 - F_\eta(\theta - s_1)]/N$ when $s = s_1$. The conditional densities in (9) are then given by

$$p_y(\overline{y}|s_0) = \frac{1}{\sigma_0\sqrt{2\pi}}\exp\left[-\frac{(\overline{y} - y_0)^2}{2\sigma_0^2}\right] \tag{12}$$

and

$$p_y(\overline{y}|s_1) = \frac{1}{\sigma_1\sqrt{2\pi}}\exp\left[-\frac{(\overline{y} - y_1)^2}{2\sigma_1^2}\right]. \tag{13}$$

The loglikelihood ratio derived from (9) follows as

$$\ln\,(\lambda_y) = \ln\,\left(\frac{\sigma_0 P_1}{\sigma_1 P_0}\right) + \frac{(\overline{y} - y_0)^2}{2\sigma_0^2} - \frac{(\overline{y} - y_1)^2}{2\sigma_1^2}. \tag{14}$$

The MAP test is then to decide $s = s_1$ when $\ln\,(\lambda_y) > 0$ and decide $s = s_0$ otherwise. This leads to conditions on the test statistic $\overline{y}$ that are expressed in the algorithm of Fig. 1.

We have performed comparisons of the test statistics $\overline{y}$ and $\overline{x}$ based on the probability of detection error $P_{\rm er}$ for different

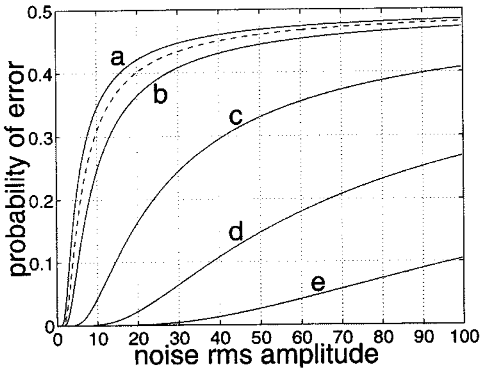

Fig. 2. Probability of detection error $P_{\rm er}$ as a function of the root mean-squared (RMS) amplitude σ_η of the zero mean white noise η. The solid lines are $P_{\rm er}$ from the algorithm of Fig. 1 with the nonlinear statistic $\overline{y}$ and for various probability densities of η according to (15) with (a) $\alpha = 2$, (b) $\alpha = 1$, (c) $\alpha = 1/2$, (d) $\alpha = 1/3$, and (e) $\alpha = 1/4$. The dashed line is $P_{\rm er}$ from (6), with the detection scheme of (4), which is the optimal detection obtainable with the linear statistic $\overline{x}$, and which achieves $P_{\rm er}$ of (6) irrespective of the probability density of η. The other parameters are $s_0 = 0$, $s_1 = 1$, $P_0 = 0.5$, and $N = 100$.

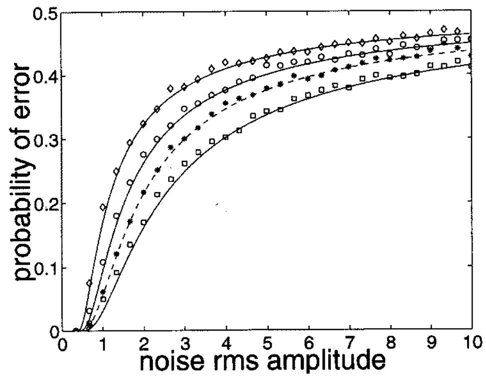

Fig. 3. Probability of detection error $P_{\rm er}$ as a function of the RMS amplitude σ_η of the zero mean white noise η. The smooth lines are the theoretical $P_{\rm er}$ from Fig. 1 or (6). The discrete data points are the corresponding numerical estimations of $P_{\rm er}$ over 10^4 trials. The solid lines are $P_{\rm er}$ from the algorithm in Fig. 1 with the nonlinear statistic $\overline{y}$ and for various probability densities of η according to (15) with: ($\circ$), $\alpha = 2$ (η Gaussian), ($\square$) $\alpha = 1$ (η Laplacian), ($\diamond$) $\alpha = \infty$ (η uniform). The dashed line and (*) is $P_{\rm er}$ from (6) with the detection scheme of (4). The other parameters are $s_0 = 0$, $s_1 = 1$, $P_0 = 0.5$, and $N = 10$.

distributions of the noise η. Typical results are shown in Fig. 2. When the noise η is Gaussian, the detection scheme of (4) based on the linear statistic $\overline{x}$ represents the best detection (lowest $P_{\rm er}$) based on the complete data set $\mathbf{x}$. Therefore, in this case, the nonlinear statistic $\overline{y}$ cannot yield a lower $P_{\rm er}$, but its performance still comes close to that of the best detector, as visible in Fig. 2, while requiring only a more parsimonious single bit per data point x_k. When η is non-Gaussian, the detection of (4) represents the best detection (lowest $P_{\rm er}$) when basing the detection on $\overline{x}$ only. In this case, the nonlinear statistic $\overline{y}$, although simpler to compute, can yield lower $P_{\rm er}$, as shown in Fig. 2. Our analysis establishes that this superiority of $\overline{y}$ over $\overline{x}$ especially occurs for non-Gaussian noises η having heavy tails or a leptokurtic character. This is true when η belongs to the family of generalized Gaussian densities [3] $f_\eta(u) = f_{\rm red}(u/\sigma_\eta)/\sigma_\eta$, where the standardized density

$$f_{\rm red}(u) = A\exp(-|bu|^\alpha) \tag{15}$$

with $A = (\alpha/2)[\Gamma(3/\alpha)]^{1/2}/[\Gamma(1/\alpha)]^{3/2}$ and $b = [\Gamma(3/\alpha)/\Gamma(1/\alpha)]^{1/2}$ is parameterized by the positive exponent α. For $\alpha = 2$, one recovers the Gaussian density. For $0 < \alpha < 2$, one obtains leptokurtic densities with tails thicker than the Gaussian, but yet having all their moments finite, especially the variance required in the detection. In the range $0 < \alpha < 2$, the nonlinear detection with $\overline{y}$ systematically outperforms the linear detection with $\overline{x}$, as illustrated by Fig. 2, which shows conditions where $\overline{y}$ can sometimes achieve a probability of error $P_{\rm er}$ ten times smaller than that of $\overline{x}$.

Fig. 3 offers a validation of the nonlinear detection of Fig. 1 through the numerical evaluation of $P_{\rm er}$ as the relative frequency of error made over a large number of detection trials. The results of Fig. 3 are obtained on data sets of size $N = 10$. They show that the scheme of Fig. 1 and the associated expressions for $P_{\rm er}$, although valid in principle in the large N limit, also constitute very good approximation for small N, and therefore, the superiority observed for $\overline{y}$ over $\overline{x}$ is robustly preserved for small data sets.

IV. Conclusion

A MAP detetector based on the nonlinear statistic $\overline{y}$ has been studied. The scheme is intrinsically appealing, as it uses a parsimonious single bit for each data point. Compared to the standard linear statistic $\overline{x}$, its performance measured by the probability of error $P_{\rm er}$ comes close for Gaussian noise and is better for non-Gaussian leptokurtic noises. The linear statistic $\overline{x}$ operates on a continuous (analog) representation of the data or with practical hardware on a 16-, 12-, or 8-bit representation, to be contrasted with the one bit per data point of the nonlinear statistic $\overline{y}$. Thus, if some notion of hardware requirement is included in the evaluation of the performance, the interest of the nonlinear detector becomes even more manifest. The superiority of $\overline{y}$ over $\overline{x}$ would also be reflected with alternative detection strategies based on a Neyman-Pearson criterion or the minimization of a cost function. Other forms of detection problems, (for instance, involving multiple hypotheses or other types of signals) may also be candidates for receiving improvement from nonlinear schemes like the one considered here [4].

References

[1] H. C. Papadopoulos, G. W. Wornell, and A. V. Oppenheim, "Low-complexity digital encoding strategies for wireless sensors networks," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 1998, pp. 3273–3276.

[2] R. N. McDonough and A. D. Whalen, *Detection of Signals in Noise*. New York: Academic, 1995.

[3] J. R. Hernández, M. Amado, and F. Pérez-González, "DCT-Domain watermarking techniques for still images: Detector performance analysis and a new structure," *IEEE Trans. Image Processing*, vol. 9, pp. 55–68, Jan. 2000.

[4] F. Chapeau-Blondeau, "Stochastic resonance and optimal detection of pulse trains by threshold devices," *Digital Signal Process.*, vol. 9, pp. 162–177, 1999.