ESTIMATING THE STANDARD DEVIATION OF SOME ADDITIVE WHITE GAUSSIAN NOISE ON THE BASIS OF NON SIGNAL-FREE OBSERVATIONS

Dominique Pastor

GET - ENST Bretagne, CNRS TAMCIC (UMR 2872),
Technopôle Brest-Iroise, CS 83818, 29285 Brest Cedex - FRANCE
email: dominique.pastor@enst-bretagne.fr

ABSTRACT

Consider n-dimensional observations where random signals are present or absent in independent and additive white Gaussian noise (AWGN) with standard deviation σ_0. On the basis of recent results in statistical decision theory, this paper presents a new algorithm for estimating σ_0 when the signals are less present than absent and have unknown probability distributions. The bias, the consistency and the minimum attainable mean square estimation error of the estimator we propose are still unknown. However, experimental results are very promising. When the Minimum-Probability-of-Error decision scheme for the non-coherent detection of modulated sinusoidal carriers in independent AWGN is tuned with the estimate instead of the true value σ_0, the Binary Error Rate obtained tends rapidly to the optimal error probability after a few hundred observations.

1. INTRODUCTION

In many signal processing applications, observations are modelled by n-dimensional random vectors that result from the random presence of signals in independent and additive white Gaussian noise (AWGN). When the noise standard deviation is unknown, it may be necessary to perform an estimate of it in order to adjust some further processing. For instance, Constant False Alarm Rate (CFAR) systems, standardly used in radar processing, perform an estimate of the noise standard deviation in order to detect radar targets with a false alarm rate close to some pre-specified value.

To perform an estimate of the noise standard deviation, one way is to rely on the physics of the problem as in radar processing. In contrast to this approach, we propose an estimator based on statistics only.

This paper is organized as follows. Section 2 presents a theoretical result on the basis of preliminary pieces of terminology and notations. This theoretical result is a simplified version of a more general limit theorem established in [5]. This simplified version is sufficient to derive an estimator of the noise standard deviation in section 2.1. This estimator applies to sets of non signal-free observations where signals have unknown probability distributions and unknown probabilities of presence (or priors) less than or equal to one half. The bias, the consistency and the mean square error of this estimator are still unknown. Nevertheless, experimental results presented in section 3 are very promising. In fact, when the estimate performed by the algorithm we propose is used instead of the true value of the noise standard deviation for the non-coherent detection of modulated sinusoidal carriers, the Binary Error Rate (BER) rapidly tends to the optimal error probability after a few hundred observations. Concluding remarks and perspectives are given in section 4.

2. A THEORETICAL RESULT FOR ESTIMATING THE NOISE STANDARD DEVIATION

In what follows, only one probability space $(\Omega, \mathcal{M}, P)$ is considered and every random vector or variable is assumed to be defined for every $\omega \in \Omega$ by setting this random vector or variable to 0 on any negligible subset where it could be undefined. As usual, if a property P holds true almost surely, we write P (a-s).

Let S stand for the set of all the sequences of n-dimensional real random vectors. Given a positive real value σ_0, an element $X = (X_k)_{k\in\mathbf{N}}$ of S will be called an *n-dimensional white Gaussian noise* (WGN) with standard deviation σ_0 if the random vectors X_k, $k = 1, 2, \ldots$, are mutually independent and identically Gaussian distributed with null mean vector and covariance matrix $\sigma_0^2 \mathbf{I}_n$.

We define the *minimum amplitude* of an element $\Lambda = (\Lambda_k)_{k\in\mathbf{N}}$ of S as the supremum $\mathsf{a}(\Lambda)$ of the set of those $\alpha \in [0, \infty]$ such that, for every natural number k, $\|\Lambda_k\|$ is larger than or equal to α (a-s):

$$\mathsf{a}(\Lambda) = \sup \left\{ \alpha \in [0, \infty] : \forall k \in \mathbf{N}, \|\Lambda_k\| \geq \alpha \text{ (a-s)} \right\}. \quad (1)$$

If f is some map of S into $\mathbf{R}$, we say that the limit of f is $\ell \in \mathbf{R}$ when $\mathsf{a}(\Lambda)$ tends to ∞ and write that $\lim_{\mathsf{a}(\Lambda)\to\infty} f(\Lambda) = \ell$ if, for any positive real value η, there exists some $\alpha_0 \in (0, \infty)$ such that, for every $\alpha \geq \alpha_0$ and every $\Lambda \in$ S such that $\mathsf{a}(\Lambda) \geq \alpha$, we have that $|f(\Lambda) - \ell| \leq \eta$.

Let $L^2(\Omega, \mathbf{R}^n)$ stand for the set of those n-dimensional real random vectors Y such that $E[\|Y\|^2] < \infty$. We will hereafter deal with the set $\ell^\infty(\mathbf{N}, L^2(\Omega, \mathbf{R}^n)) \subset \mathrm{S}$ of those elements $\Lambda = (\Lambda_k)_{k\in\mathbf{N}}$ of S such that $\Lambda_k \in L^2(\Omega, \mathbf{R}^n)$ for every $k \in \mathbf{N}$ and $\sup_{k\in\mathbf{N}} E[\|\Lambda_k\|^2]$ is finite.

In what follows, ${}_0F_1$ is the generalized hypergeometric function ([3, p. 275]); given $\rho \in [0, \infty)$, $\xi(\rho)$ is the unique positive solution for x in the equation ${}_0F_1(n/2; \rho^2 x^2/4) = e^{\rho^2/2}$; Θ stands for the map defined for every $x \in [0, \infty)$ by $\Theta(x) = \int_0^x t^n e^{-t^2/2} dt / \int_0^x t^{n-1} e^{-t^2/2} dt$. The following result is a particular case of a more general limit theorem established in [5]. Given any random vector Y and any real number τ, $\mathcal{I}(\|Y\| \leq \tau)$ stands for the indicator function of the event $\{\|Y\| \leq \tau\}$.

Proposition 2.1 *Let $U = (U_k)_{k\in\mathbf{N}}$ be some element of* S *such that, for every $k \in \mathbf{N}$, $U_k = \varepsilon_k \Lambda_k + X_k$ where $\Lambda = (\Lambda_k)_{k\in\mathbf{N}} \in \ell^\infty(\mathbf{N}, L^2(\Omega, \mathbf{R}^n))$, $X = (X_k)_{k\in\mathbf{N}}$ is some n-dimensional WGN with standard deviation σ_0 and $\varepsilon = (\varepsilon_k)_{k\in\mathbf{N}}$ is a sequence of random variables valued in $\{0, 1\}$ respectively.*

Assume that

(A1) *for every $k \in \mathbf{N}$, Λ_k, X_k and ε_k are mutually independent;*

(A2) *the random vectors U_k, $k \in \mathbf{N}$, are mutually independent;*

(A3) *the random variables ε_k, $k \in \mathbf{N}$, are mutually independent;*

(A4) *the priors $P(\{\varepsilon_k = 1\})$, $k \in \mathbf{N}$, are less than or equal to one half.*

Given some natural number m and any pair (σ, T) of positive real numbers, define the random variable $\Delta_m(\sigma, T)$ by

$$\Delta_m(\sigma, T) = \left| \frac{\sum_{k=1}^{m} \|U_k\| \mathcal{I}(\|U_k\| \leq \sigma T)}{\sum_{k=1}^{m} \mathcal{I}(\|U_k\| \leq \sigma T)} - \sigma \Theta(T) \right|. \quad (2)$$

Then, σ_0 is the unique positive real number σ such that, for every $\beta_0 \in (0, 1]$,

$$\lim_{\mathsf{a}(\Lambda)\to\infty} \left\| \overline{\lim_{m}} \Delta_m(\sigma, \beta \xi(\mathsf{a}(\Lambda)/\sigma)) \right\|_\infty = 0 \quad (3)$$

uniformly in $\beta \in [\beta_0, 1]$.

In this statement, U models a sequence of observations where, for every given $k \in \mathbf{N}$, Λ_k stands for some possible random signal and ε_k models the possible occurrence of Λ_k in the background of AWGN modelled by X. The assumption that $\Lambda \in \ell^\infty(\mathbf{N}, L^2(\Omega, \mathbf{R}^n))$ corresponds to the practical case of interest where the energies of the signals are finite and bounded.

2.1. The algorithm

Even though proposition 2.1 and the theoretical results of [5] concern signals whose minimum amplitude is large enough, one of the conclusions of [6] is that this constraint can certainly be relaxed in practical applications. Hence, in the present paper, we propose to perform an estimate of the noise standard deviation by taking $\mathsf{a}(\Lambda) = 0$ since the null value is a trivial bound for the norm of any signal. By so proceeding, we significantly extend the algorithm proposed in [6] since we discard any hypothesis regarding the probability distributions and the norms of the signals and simply assume that these signals are less present than absent.

The algorithm is then the following one. We keep the notations of the foregoing section. Suppose that we have m observations $U_1, \ldots, U_m$. Let $L \in \mathbf{N}$ and set $\beta_\ell = \ell/L$ for every $\ell \in \{1, \ldots, L\}$. Proposition 2.1 suggests estimating σ_0 by a possibly local minimum of

$$\sup_{\ell\in\{1,\ldots,L\}} \Delta_m(\sigma, \beta_\ell \xi(\mathsf{a}(\Lambda)/\sigma)).$$

Since we take $\mathsf{a}(\Lambda) = 0$ and know that $\xi(0) = \sqrt{n}$ ([4]), we start by computing a minimum $\hat{\sigma}_0$ of

$$\sup_{\ell\in\{1,\ldots,L\}} \left\{ \left| \frac{\sum_{k=1}^{m} \|U_k\| \mathcal{I}(\|U_k\| \leq \beta_\ell \sigma \sqrt{n})}{\sum_{k=1}^{m} \mathcal{I}(\|U_k\| \leq \beta_\ell \sigma \sqrt{n})} - \sigma \Theta(\beta_\ell \sqrt{n}) \right| \right\}. \quad (4)$$

Any minimization routine for scalar bounded non-linear functions is suitable. For instance, the experimental results presented in the next section were obtained with the `MATLAB` routine `fminbnd.m` based on parabolic interpolation ([8]). The search interval $[\sigma_{\min}, \sigma_{\max}]$ is computed as follows. Sort the observations $U_1, \ldots, U_m$, $k = 1, \ldots, m$, by increasing norm. Let $U_{[k]}$, $k = 1, \ldots, m$, be the resulting sequence. The right endpoint of the search interval is then $\sigma_{\max} = \|U_{[m]}\|/\sqrt{n}$. As far as the left endpoint is concerned, choose a real number Q close to 1 but less than or equal to $1 - \frac{m}{4(m/2-1)^2}$. A typical choice is $Q = 0.95$, provided that $m \geq 24$. Set $h = 1/\sqrt{4m(1-Q)}$ and $k_{\min} = m/2 - hm$. The left endpoint is then $\sigma_{\min} = \|U_{[k_{\min}]}\|/\sqrt{n}$. Because of the limited length of this paper, the reader is asked to refer to [5] and [6] for justifications as to this construction of the search interval.

It would be natural to stop at this stage. In fact, we go a step further and our final estimate of the noise standard deviation is

$$\tilde{\sigma}_0 = \sqrt{\frac{\sum_{k=1}^{m} \|U_k\|^2 \mathcal{I}(\|U_k\| \leq \hat{\sigma}_0 \sqrt{n})}{\sum_{k=1}^{m} \mathcal{I}(\|U_k\| \leq \hat{\sigma}_0 \sqrt{n})}}. \quad (5)$$

The rationale is the following. Since ξ is a non-decreasing map and $\xi(0) = \sqrt{n}$, it follows from [4] that $\sigma_0\sqrt{n}$ is the smallest threshold that should be used to detect, with a probability of error less than or equal to $1/2$, any signal less present than absent. It is to be expected that most of the observations with norms less than or equal to this threshold derive from noise alone. On the other hand, if an observation is equal to some signal plus noise and has a norm less than or equal to this threshold, the norm of the signal can reasonably be expected to be very small and, thus, to hardly influence the estimate $\tilde{\sigma}_0$.

3. EXPERIMENTAL RESULTS

The bias, the consistency and the minimum attainable mean square estimation error of the estimator proposed above are still unknown. Therefore, in this section we will restrict ourselves to some experimental results.

We keep the notations used so far. Given any non-negative real number A, let $\tilde{\mathcal{T}}$ be the test $\mathcal{I}(\|\cdot\| \geq \tilde{\sigma}_0\xi(A/\tilde{\sigma}_0))$. Given $k \in \mathbf{N}$, the decision of this test is that ε_k is 0 if $\|U_k\| \leq \tilde{\sigma}_0\xi(A/\tilde{\sigma}_0)$ and that ε_k is 1, otherwise. If $\tilde{\sigma}_0$ is a reasonably good estimate of the noise standard deviation, the performance of test $\tilde{\mathcal{T}}$ can be expected to be close to that of the thresholding test with threshold height $\sigma_0\xi(A/\sigma_0)$; given any real number x, by thresholding test with threshold height $x \in \mathbf{R}$, we mean the test $\mathcal{T}_x$ defined for every $u \in \mathbf{R}^n$ by $\mathcal{T}_x(u) = 1$ if $\|u\| \geq x$ and $\mathcal{T}_x(u) = 0$ if $\|u\| < x$. To detect the presence of any signal with norm larger than or equal to A and prior less than or equal to one half, it follows from [4, Theorem VII.1] that the probability of error $\mathcal{P}_e\{\mathcal{T}_{\sigma_0\xi(A/\sigma_0)}\}$ of $\mathcal{T}_{\sigma_0\xi(A/\sigma_0)}$ satisfies the following inequalities

$$\mathcal{P}_e\{\mathcal{L}\} \leq \mathcal{P}_e\{\mathcal{T}_{\sigma_0\xi(A/\sigma_0)}\} \leq V(A/\sigma_0). \tag{6}$$

In (6), $\mathcal{P}_e\{\mathcal{L}\}$ stands for the probability of error of the Minimum-Probability-of-Error (MPE) decision scheme $\mathcal{L}$, that is the likelihood ratio test with the smallest possible probability of error among all possible hypothesis binary tests; the map V is defined for every $x \in [0, \infty)$ and the reader can refer to [4] for the general expression of this map. The inequalities above then become equalities in the least favourable situation where the signal is uniformly distributed on the sphere AS^{n-1} centred at the origin with radius A and has prior equal to one half ([4, Theorem VII.1]). Therefore, if every Λ_k, $k \in \mathbf{N}$, is uniformly distributed on AS^{n-1} and has prior less than or equal to $1/2$, the probability of error $\mathcal{P}_e\{\tilde{\mathcal{T}}\}$ of $\tilde{\mathcal{T}}$ should not significantly exceed $V(A/\sigma_0)$. If every Λ_k has prior equal to one half, $\mathcal{P}_e\{\tilde{\mathcal{T}}\}$ should even be close to $V(A/\sigma_0)$. We do not know the theoretical value of $\mathcal{P}_e\{\tilde{\mathcal{T}}\}$. Hence, we approximate it by the Binary Error Rate (BER) obtained by a Monte-Carlo simulation and compare this BER to $V(A/\sigma_0)$. This Monte-Carlo simulation is carried out in the case of two-dimensional real random observations ($n = 2$); we choose some $p \in (0, 1/2]$ and every Λ_k is uniformly distributed on AS^1 and has a probability of presence equal to p. We then have $V(x) = \frac{1}{2}e^{-\frac{x^2}{2}}\int_0^x e^{-\frac{t^2}{2}} t I_0(xt)dt + \frac{1}{2}e^{-\frac{t^2}{2}}$ where I_0 is the zeroth-order modified Bessel function of the first kind ([4]). The two components of every Λ_k can be regarded as the in-phase and quadrature components of a sinusoidal carrier. In other words, we consider the "non-coherent detection of modulated sinusoidal carriers", a problem particularly relevant for telecommunication and radar processing ([7, p. 65]).

The BER of $\tilde{\mathcal{T}}$ is computed as follows. Independent trials of m observations each are carried out until two conditions are fulfilled. First, at least M trials must be performed. Inasmuch as the decision about the presence or the absence of signals is made on the observations used for estimating σ_0, the accuracy of the estimate affects m decisions at one go. This effect is then reduced by fixing a minimum number of trials. Second, the total number N_e of errors made by test $\tilde{\mathcal{T}}$ for detecting the presence or the absence of signals must be above or equal to some specified number N. If j is the first trial number larger than or equal to M for which the total number of errors N_e becomes larger than or equal to N, the BER of test $\tilde{\mathcal{T}}$ is then defined as the ratio $N_e/(j \times m)$.

The simulation is achieved with $\sigma_0 = 1$. The pre-specified number of errors is $N = 400$ and the minimum number of trials is $M = 100$. We choose $L = m$ and $Q = 0.95$ on the basis of preliminary trials. The comparison between the BER of $\tilde{\mathcal{T}}$ and $V(A/\sigma_0)$ is achieved for $A = \{0.5, 1, 1.5, \ldots, 5\}$ and $p \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. The results are those of figures 1, 2 and 3 for different values of m. Test $\tilde{\mathcal{T}}$ yields performance close to that of $\mathcal{T}_{\sigma_0\xi(A/\sigma_0)}$. Figures 4 and 5 present the average value and the standard deviation of $\tilde{\sigma}_0$ calculated on the basis of 100 trials for $m = 100$ observations.

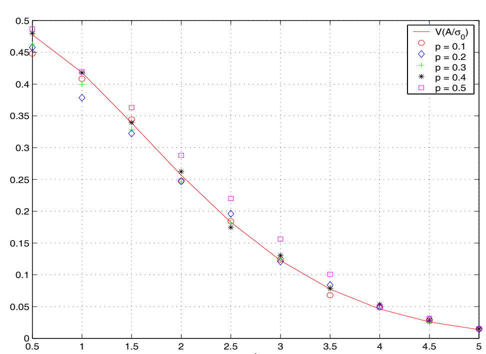

Fig. 1. $V(A/\sigma_0)$ vs BERs of $\tilde{\mathcal{T}}$ for $m = 50$

4. PERSPECTIVES AND EXTENSIONS

We have presented an algorithm for estimating the standard deviation of some background of AWGN when observations derive from signals less present than absent in this background. According to experimental results, this algorithm is very promising. Further theoretical developments should address the

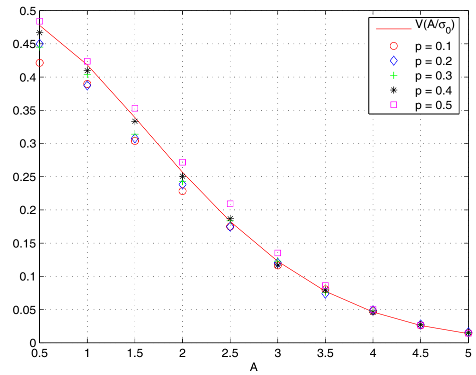

Fig. 2. $V(A/\sigma_0)$ vs BERs of $\tilde{T}$ for $m = 100$

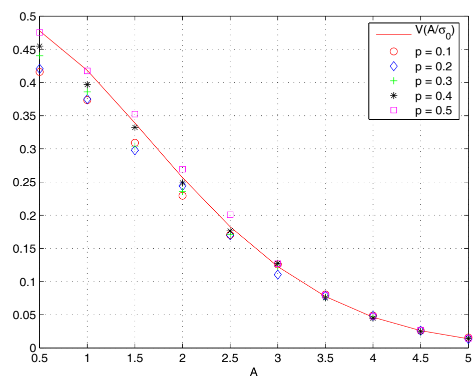

Fig. 3. $V(A/\sigma_0)$ vs BERs of $\tilde{T}$ for $m = 200$

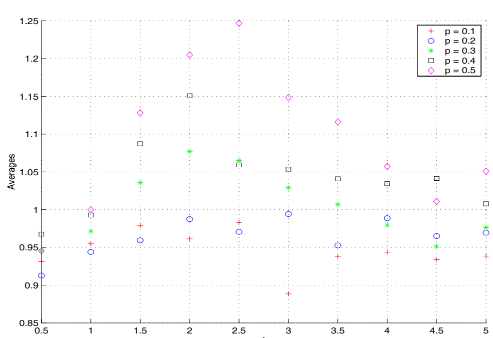

Fig. 4. Average value of $\tilde{\sigma}_0$

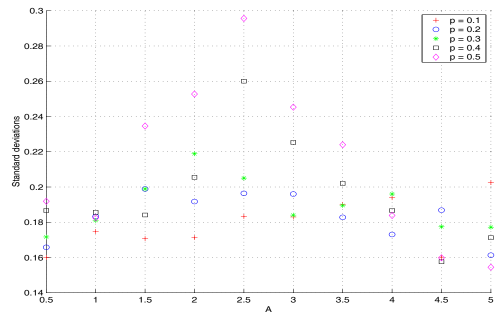

Fig. 5. Standard deviation of $\tilde{\sigma}_0$

bias, the consistency and the mean square estimation error of this algorithm. Possible links between the approach proposed in this paper and that introduced in [2] is also an issue to address.

From a more practical point of view, the estimator we propose can contribute to the detection of radar and sonar targets in complement to standard approaches. It can also be expected to avoid the use of a Voice Activity Detector for tuning standard Wiener filtering or spectral substraction for denoising speech signals corrupted by AWGN.

5. REFERENCES

[1] M. Abramowitz, I. Stegun, *Handbook of Mathematical Functions*, Ninth printing. Dover Publications Inc., New York, 1972.

[2] D. Donoho, I. Johnstone, "Ideal Spatial Adaptation by Wavelet Shrinkage", *Biometrika*, 81, pp. 425-455, 1994.

[3] N. N. Lebedev, *Special Functions and their Applications*, Prentice-Hall, Englewood Cliffs, 1965.

[4] D. Pastor, R. Gay, A. Groenenboom, "A Sharp Upper-Bound for the Probability of Error of the Likelihood Ratio Test for Detecting Signals in White Gaussian Noise", *IEEE Transactions on Information Theory*, 48, pp. 228-238, 2002.

[5] D. Pastor, "On the detection of signals with unknown distributions and priors in white Gaussian noise", *Collection des Rapports de Recherche de l'ENST Bretagne*, RR-2006001-SC, 2006.

[6] D. Pastor, "Un théorème limite et un test pour la détection non paramétrique de signaux dans un bruit blanc gaussien de variance inconnue", *20ème colloque GRETSI sur le traitement du signal et des images, GRETSI'05*, Louvain-La-Neuve, 2005.

[7] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd Edition, Springer-Verlag, New York, 1994.

[8] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical recipes in C, The Art of Scientific Computing*, second edition, University Press, Cambridge, 1992.