

# Parametrically Optimal, Robust and Tree-Search Detection of Sparse Signals

A. T. Burrell<sup>1</sup>, P. Papantoni-Kazakos<sup>2</sup>

<sup>1</sup>Oklahoma State University, Computer Science Department, Stillwater, USA; <sup>2</sup>University of Colorado Denver, Electrical Engineering Department, Denver, USA.

Email: [tburrell@okstate.edu](mailto:tburrell@okstate.edu), [titsa.papantoni@ucdenver.edu](mailto:titsa.papantoni@ucdenver.edu)

Received June 30<sup>th</sup>, 2013; revised July 30<sup>th</sup>, 2013; accepted August 13<sup>th</sup>, 2013

Copyright © 2013 A. T. Burrell, P. Papantoni-Kazakos. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## ABSTRACT

We consider sparse signals embedded in additive white noise. We study parametrically optimal as well as tree-search sub-optimal signal detection policies. As a special case, we consider a constant signal and Gaussian noise, with and without data outliers present. In the presence of outliers, we study outlier resistant robust detection techniques. We compare the studied policies in terms of error performance, complexity and resistance to outliers.

**Keywords:** Sparse Signals; Detection; Robustness; Outlier Resistance; Tree Search

## 1. Introduction

In recent years, some refreshed interest has been given to *sparse signals*, by the signal processing community [1,2], while the effective probing/transmission of such signals; previously denoted *bursty*, has been addressed by both tree-search [3,4], and random access algorithms [5]. The revisited investigation of sparse signals has focused on linear transformations [1,2], while the term *robustness* has been used loosely in [1].

In this paper, we focus on the detection of sparse signals embedded in white Gaussian noise with the possible occasional occurrence of data outliers. We study both optimal and sub-optimal detection techniques, when data outliers are considered both absent and present. In the latter case, we consider *robust* detection techniques, where robustness is here precisely defined as referring to outlier resistant operations [6,7]. We compare our techniques in terms of error performance, complexity and resistance to outliers.

The organization of the paper is as follows: In Section 2, we state the fundamental general problem, present assumptions and notation, and determine, as well as partially analyze, the general optimal detector. In Section 3, we present and analyze the optimal detector for the case of white Gaussian noise and constant signal, when no outliers are considered in the design. In Section 4, we present and analyze the robust (outlier resistant) detector.

In Section 5, we present and evaluate tree-search sub-optimal detectors. In Section 6, we include discussion and conclusions.

## 2. Problem Statement and General Solution

We consider a sequence of observations generated by mutually independent random variables, a small percentage of which represent signal embedded in noise, while the remaining percentage represent just noise. Let it be known that the percentage of observations representing signal presence is bounded from above by a given value  $\alpha$ . We assume that the random variables representing the signal are identically distributed, and that so are those representing the noise. We denote by  $x_1, \dots, x_n$ , a sequence of  $n$  such observations, while we denote by  $X_1, \dots, X_n$ , the sequence of mutually independent random variables whose realization is the sequence  $x_1, \dots, x_n$ . We also denote by  $f_1(\cdot)$  either the probability distribution (for discrete variables) or the probability density (for absolutely continuous variables) function (pdf) of the variables which represent signal presence, while we denote by  $f_0(\cdot)$  the pdf of the variables which represent just noise.

Given the observation sequence  $x_1, \dots, x_n$  and assuming  $f_1(\cdot)$ ,  $f_0(\cdot)$ ,  $\alpha$  known, the objective is to identify the locations of the signal presence; that is, which ones of the  $x_1, \dots, x_n$  observations originated from the  $f_1(\cdot)$  pdf.

Our approach to the problem solution will be Maximum Likelihood (ML); which is equivalent to that of the Bayesian minimization of error probability approach, when all signal locations and their number are equally probable [7]. That is, given the sequence  $x_1, \dots, x_n$ , the optimal detector decides in favor of the  $i_1, \dots, i_m$ ;  $0 \leq m \leq \alpha n$  signal locations if:

$$\sum_{1 \leq j \leq m} \log f_1(x_k) + \sum_{k \neq \{i_j\}} \log f_0(x_k) = \max \left( \sum_{(j)} \log f_1(x_i) + \sum_{k \neq \{j\}} \log f_0(x_k) \right) \quad (1)$$

Let us define:

$$g(x) \equiv \log \left( \frac{f_1(x)}{f_0(x)} \right) \quad (2)$$

Via some straight forward modifications of the expression in (1), we may then express the optimal ML detector as follows:

*Optimal ML Detector*

1) Given the sequence  $x_1, \dots, x_n$ , compute all  $g(x_j)$ ;  $1 \leq j \leq n$

2) If  $g(x_k) \leq 0$ ; for all  $k$ ,  $1 \leq k \leq n$ , then, decide that no observation contains signal.

3) If  $\exists$  a set of integers  $\{i_1, \dots, i_m\}$ ;  $1 \leq m \leq \alpha n$ :  $g(x_k) > 0$ ; for all  $k \in \{i_1, \dots, i_m\}$  and  $g(x_k) \leq 0$ ; for all  $k \notin \{i_1, \dots, i_m\}$ , then decide that the observations containing the signal are all those with indices in the set  $\{i_1, \dots, i_m\}$ .

4) If  $\exists$  a set of integers  $\{i_1, \dots, i_m\}$ ;  $m > \alpha n$ :  $g(x_k) > 0$ ; for all  $k \in \{i_1, \dots, i_m\}$  and  $g(x_k) \leq 0$ ; for all  $k \notin \{i_1, \dots, i_m\}$ , then decide that the observations containing the signal are those whose indices  $k$  are contained in the set  $\{i_1, \dots, i_m\}$  and whose  $g(x_k)$  values are the  $\alpha n$  highest in the set.

Considering the log likelihood ratio in (2), let  $h_i(w)$  and  $H_i(w)$ ;  $i = 0, 1$ , denote, respectively, the pdf and the cumulative distribution function of the random variable  $g(X)$  at the value point  $w$ , given that the pdf of  $X$  is  $f_i$ ,  $i = 0, 1$ . Let  $P_d(0)$  denote the probability of correct detection induced by the optimal ML detector, given that no observation contains signal. Let, instead,  $P_d(\{i_1, \dots, i_m\})$  denote the probability of correct detection, given that the indices of the observations containing signal are given by the set  $\{i_1, \dots, i_m\}$ . Then, assuming absolutely continuous  $\{X_i\}$  random variables; without lack in generality, we obtain the following expressions; without much difficulty, where  $\alpha n$  is assumed an integer; for simplicity in notation:

$$P_d(0) = [H_0(0)]^n$$

$$P_d(\{i_1, \dots, i_m\}) = [1 - H_1(0)]^m [H_0(0)]^{m-n}; 1 \leq m \leq \alpha n$$

$$P_d(\{i_1, \dots, i_m\}) = \alpha n \int_0^\infty dw h_1(w) [1 - H_1(w)]^{\alpha n - 1} [H_0(w)]^{(1-\alpha)n}; m = \alpha n \quad (3)$$

**Remarks**

It is important to note that the optimal detector presented above assumes no knowledge as to any structure of the sparse signal and requires  $n$ -size memory, as well as ordering of the positive  $g(x_k)$  values, inducing complexity of order  $n \log n$ . If, on the other hand, a structure of the signal is known a priori and is such that it appears as a bursty batch, then, the sequential algorithm in [7-9] that monitors changes in distribution should be deployed instead; it requires no memory and its complexity is of order  $n$ .

**3. Constant Signal and White Gaussian Additive Noise**

In this section, we consider the special case where the signal is a known constant  $\theta > 0$ , and the noise is zero mean white Gaussian with standard deviation  $\sigma$ . After some simple straight forward normalizations, the optimal ML detector of Section II takes here the following form:

*Optimal ML Detector*

1) Given the sequence  $x_1, \dots, x_n$ , compute all  $(x_i - \theta/2)$ ;  $1 \leq j \leq n$

2) If  $(x_k - \theta/2) \leq 0$ ; for all  $k$ ,  $1 \leq k \leq n$ , then, decide that no observation contains signal.

3) If  $\exists$  a set of integers  $\{i_1, \dots, i_m\}$ ;  $1 \leq m \leq \alpha n$ :  $(x_k - \theta/2) > 0$ ; for all  $k \in \{i_1, \dots, i_m\}$  and  $(x_k - \theta/2) \leq 0$ ; for all  $k \notin \{i_1, \dots, i_m\}$ , then decide that the observations containing the signal are all those with indices in the set  $\{i_1, \dots, i_m\}$ .

4) If  $\exists$  a set of integers  $\{i_1, \dots, i_m\}$ ;  $m > \alpha n$ :  $(x_k - \theta/2) > 0$ ; for all  $k \in \{i_1, \dots, i_m\}$  and  $(x_k - \theta/2) \leq 0$ ; for all  $k \notin \{i_1, \dots, i_m\}$ , then decide that the observations containing the signal are those whose indices  $k$  are contained in the set  $\{i_1, \dots, i_m\}$  and whose  $(x_k - \theta/2)$  values are the  $\alpha n$  highest in the set.

As to the probabilities of correct detection, defined in Section 2, they take the following form here, where  $\Phi(x)$  and  $\Phi(x)$  denote, respectively, the pdf and the cumulative distribution function of the zero mean and unit variance Gaussian random variable and where  $\alpha n$  is assumed again to be an integer :

$$P_d(\{i_1, \dots, i_m\}) = \left[ \Phi \left( \frac{\theta}{2\sigma} \right) \right]^n; 0 \leq m \leq \alpha n$$

$$\begin{aligned}
 &P_d(\{i_1, \dots, i_m\}) \\
 &= \alpha n \int_{\frac{-\theta}{2\sigma}}^{\infty} dw \varphi(w) [\Phi(-w)]^{\alpha n - 1} \left[ \Phi\left(w + \frac{\theta}{\sigma}\right) \right]^{(1-\alpha)n} \\
 &= \alpha n \int_{-\infty}^{\frac{\theta}{2\sigma}} dw \varphi(w) [\Phi(w)]^{\alpha n - 1} \left[ \Phi\left(-w + \frac{\theta}{\sigma}\right) \right]^{(1-\alpha)n}; \quad m = \alpha n
 \end{aligned} \tag{4}$$

Remarks

It is interesting to note here that if it is known that the signal may appear as a set bursty batches of unknown sizes, then the re-initialization sequential algorithm in [8] will sequentially detect the beginning and the ending of each batch with minimal complexity, no memory requirements and with accuracy increasing with the signal-to-noise ratio and the size of each batch. Let then  $T_n$  denote the value of the algorithm which detects the beginning of such a batch, upon the processing of the  $n^{\text{th}}$  datum  $x_n$  from its beginning. Let  $W_n$  denote the value of the algorithm which detects the ending of the batch, upon processing of the  $n^{\text{th}}$  datum  $y_n$  from the beginning of its initialization. The whole algorithmic system operates then as follows, where  $\delta_0$  and  $\delta_1$  are two positive thresholds pre selected to satisfy power and false alarm trade offs:

1) Process the observed sequence  $x_1, \dots, x_n$  sequentially starting with the algorithm  $\{T_n\}$  whose values are updated as follows:

$$\begin{aligned}
 T_0 &\equiv 0 \\
 T_{n+1} &= \max\left(0, T_n + x_{n+1} - \frac{\theta}{2}\right)
 \end{aligned}$$

Stop the first time  $n$ , such that  $T_n \geq \delta_0$  and declare  $n$  as the time when the signal batch begins.

Then, switch immediately to the algorithm  $\{W_n\}$  whose values are updated as follows, where time zero denotes the time when the algorithm begins and where  $y_n$  denotes the  $n^{\text{th}}$  observed datum after the latter beginning:

$$\begin{aligned}
 W_0 &\equiv 0 \\
 W_{n+1} &= \max\left(0, W_n - y_{n+1} + \frac{\theta}{2}\right)
 \end{aligned}$$

Stop the first time  $n$ , such that  $W_n \geq \delta_1$  and declare that the signal batch has ended.

We now express a Corollary which will be useful in the computation of bounds for the probability of correct detection in (4). The expressions in the Corollary are derived from recursive relationships produced via integration by parts and can be proven easily by induction.

Corollary

The following equations hold:

$$\int_{-\infty}^{\lambda} dw \varphi(w) \Phi^k(w) = (k+1)^{-1} \Phi^{k+1}(\lambda) \tag{5}$$

$$\int_0^{\lambda} dw \varphi(w) \Phi^k(w) = (k+1)^{-1} [\Phi^{k+1}(\lambda) - 2^{-(k+1)}] \tag{6}$$

$$\begin{aligned}
 &\int_{-\infty}^{\lambda} dw \varphi(w) \Phi^k(w) \Phi^{n-k}(-w) \\
 &= \frac{(n-k)! k!}{(n+1)!} \sum_{m=k+1}^{n+1} \binom{n+1}{m} \Phi^{n+1-m}(-\lambda) \Phi^m(\lambda)
 \end{aligned} \tag{7}$$

$$\begin{aligned}
 &\int_0^{\lambda} dw \varphi(w) \Phi^k(w) \Phi^m(-w) \\
 &= \sum_{i=0}^m \frac{m! k!}{(k+1+i)!(m-i)!} * (\Phi^{k+1+i}(\lambda) \Phi^{m-i}(-\lambda) - 2^{-(k+m+1)})
 \end{aligned} \tag{8}$$

Lemma 1 below utilizes the results in the Corollary, to express two lower bounds for the probability of correct detection in (4). The bound in (9) is relatively tight for low signal-to-noise ratio values  $\theta/\sigma$ . The bound in (10) is relatively tight for high signal-to-noise ratio  $\theta/\sigma$ , instead.

Lemma 1

The probability of correct detection in (4) increases monotonically with increasing value of the signal-to-noise ratio  $\theta/\sigma$ , converging to the value 1 as  $\theta/\sigma$  reaches asymptotically large values. This probability is bounded from below as follows, assuming that  $\alpha n$  is an integer; for simplicity in notation:

$$\begin{aligned}
 &P_d(\{i_1, \dots, i_m\}) > \frac{[(1-\alpha)n]! [\alpha n - 1]!}{n!} \\
 &\times \sum_{m=\alpha n}^{n+1} \binom{n+1}{m} \Phi^m\left(\frac{\theta}{2\sigma}\right) \Phi^{n+1-m}\left(-\frac{\theta}{2\sigma}\right); \quad m = \alpha n
 \end{aligned} \tag{9}$$

$$\begin{aligned}
 &P_d(\{i_1, \dots, i_m\}) > \sum_{i=0}^{(1-\alpha)n} \frac{[(1-\alpha)n]! [\alpha n - 1]!}{[(1-\alpha)n - i]! [\alpha n + i]!} \\
 &\times \left( \Phi^{\alpha n + i}\left(\frac{\theta}{2\sigma}\right) \Phi^{(1-\alpha)n - i}\left(-\frac{\theta}{2\sigma}\right) - 2^{-(n+1)} \right); \quad m = \alpha n
 \end{aligned} \tag{10}$$

Proof

Considering  $\theta/\sigma$  asymptotically large, we first approximate  $\Phi(-w + \theta/\rho)$  by 1, in Expression (4). We then, use expression (5) and consider again asymptotically large values of  $\theta/\sigma$ . The result proves that the probability in (4) converges to 1, as  $\theta/\sigma$  approaches infinity.

To derive the bound in (9), we first bound  $\Phi(-w + \theta/\rho)$  from below by  $\Phi(-w)$  in the integrand of expression (4). Then, we use Equation (7) on the resulting integral expression.

To derive the bound in (10), we bound  $\Phi(-w + \theta/\rho)$  from below by: 1)  $\Phi(\theta/\sigma)$ ; for negative  $w$  values and 2) by  $\Phi(-w)$ ; for positive  $w$  values. We then use Expressions (5) and (8) from the Corollary.

Lemma 2 below states a probability of correct detec-

tion result for the case where the percentage of signal-including observations and the signal-to-noise ratio are both very small.

Lemma 2

Let  $\alpha \rightarrow 0$  and  $\theta/\sigma \rightarrow 0$ . Then, the probability of correct detection in (4) is of order  $\alpha n \Phi^n(\theta/\sigma)$ .

Proof

We break the integral in (4) into two parts: the part from  $-\infty$  to 0 and the part from 0 to  $\theta/2\sigma$ . For the first part, we expand  $[\Phi(-\omega + \theta/\rho)]^{(1-\alpha)n}$  via Taylor series expansion to first order  $\theta/\sigma$  approximation. For the second part, we approximate the integral by  $\theta/2\sigma$  times the value of the integrand at zero. Then, we approximate  $1 - \alpha \approx 1$ . As a result, we then obtain:

$$P_d(\{i_1, \dots, i_m\}) \approx \alpha n \Phi^n\left(\frac{\theta}{\sigma}\right) \left\{ \frac{\theta}{2\sigma} \varphi(0) 2^{\alpha n} + \int_{-\infty}^0 dw \varphi(w) \Phi^{\alpha n}(w) \right. \\ \left. + n \left(\frac{\theta}{\sigma}\right) \int_{-\infty}^0 dw \varphi^2(w) \Phi^{\alpha n}(w) \right\}; \quad m = \alpha n \quad (11)$$

We note that, in general, the probability of correct detection induced by the ML optimal detector is of the order  $[\Phi(\theta/2\sigma)]^n$ , increasing exponentially to 1; with increasing signal-to-noise ratio  $\theta/\sigma$ , while it is also decreasing exponentially to 0; with increasing sample size.

#### 4. The Outlier Resistant Detector

In this section, we consider the case where extreme occasional outliers may be contaminating the Gaussian environment of Section 3. Then, instead of white and Gaussian, the noise environment is modeled as white with pdf belonging to a class  $F$  of density functions, defined as follows, for some given value  $\varepsilon$  in  $(0, 0.5)$ , where  $\varepsilon$  represents the outlier contamination level:

$F = \{f; f = (1-\varepsilon)f_0 + \varepsilon h, f_0$  is the Gaussian zero mean and standard deviation  $\sigma$  pdf,  $h$  is any pdf}.

The outlier resistant robust detector is then found based on the *least favorable* density  $f^*$  in class  $F$  above, where the Kullback-Leibler number between  $f^*$  and its shifted by location parameter  $\theta$  version attains the infimum among the Kullback-Leibler numbers realized by all pdfs in  $F$  [6,7]. As found in [7], the log likelihood ratio in (2) is a truncated version of that used in Section 3. As a result, for  $\theta > 0$ , the ML robust detector is operating as follows:

*Robust ML Detector*

1) Given the sequence  $x_1, \dots, x_n$ , compute all  $[z(x_i) - \theta/2]; 1 \leq j \leq n$ , where,

$$z(x) = \begin{cases} d; & x \geq d \\ x; & -d + \theta < x < d \\ -d + \theta; & x \leq -d + \theta \end{cases} \quad (12)$$

$$d : (1-\varepsilon) \left\{ \Phi\left(\frac{-d+\theta}{\sigma}\right) + \exp\left\{\frac{\theta d}{\sigma^2} - \frac{\theta^2}{2\sigma^2}\right\} \Phi\left(\frac{d}{\sigma}\right) \right\} = 1 \quad (13)$$

2) If  $[z(x_k) - \theta/2] \leq 0$ ; for all  $k, 1 \leq k \leq n$ , then, decide that no observation contains signal.

3) If  $\exists$  a set of integers  $\{i_1, \dots, i_m\}; 1 \leq m \leq \alpha n$ :

$[z(x_k) - \theta/2] > 0$ ; for all  $k \in \{i_1, \dots, i_m\}$  and

$[z(x_k) - \theta/2] \leq 0$ ; for all, then decide that the observations containing the signal are all those with indices in the set  $\{i_1, \dots, i_m\}$ .

4) If  $\exists$  a set of integers  $\{i_1, \dots, i_m\}; m > \alpha n$ :

$[z(x_k) - \theta/2] > 0$ ; for all  $k \in \{i_1, \dots, i_m\}$  and

$[z(x_k) - \theta/2] \leq 0$ ; for all  $k \notin \{i_1, \dots, i_m\}$ , then decide that the observations containing the signal are those whose indices  $k$  are contained in the set  $\{i_1, \dots, i_m\}$  and whose  $[z(x_k) - \theta/2]$  values are the  $\alpha n$  highest in the set.

We will denote by  $P_{do}^r(\{i_1, \dots, i_m\})$  the probability of correct detection induced by the robust ML detector, given that the noise is Gaussian containing no outliers and given that the signal occurs at the observation indices  $\{i_1, \dots, i_m\}$ . Then, we can derive the expressions below, with some extra caution, assuming again that  $\alpha n$  is an integer:

$$P_{do}^r(\{i_1, \dots, i_m\}) = \Phi^n\left(\frac{\theta}{2\sigma}\right); \quad 0 \leq m < \alpha n \\ P_{do}^r(\{i_1, \dots, i_m\}) = \alpha n \int_{\frac{-\theta}{2\sigma}}^{\frac{d-\theta}{\sigma}} dw \varphi(w) [\Phi(-w)]^{\alpha n - 1} \\ \times \left[ \Phi\left(w + \frac{\theta}{\sigma}\right) \right]^{(1-\alpha)n} + \Phi^{\alpha n}\left(\frac{-d+\theta}{\sigma}\right) \Phi^{(1-\alpha)n}\left(\frac{d}{\sigma}\right); \quad m = \alpha n \quad (14)$$

Comparing Expressions (4) and (14), we notice that the robust detector induces lower probability of correct detection at the nominal Gaussian model; for the case of  $m = \alpha n$ , where the difference of the two probabilities decreases monotonically with decreasing contamination level  $\varepsilon$ . As we will see in the sequel, this loss of performance of the robust detector at the nominal Gaussian model is at the gain of resistance to outliers.

Let there exist a small positive value  $\zeta$ , such that the noise per observation is zero mean Gaussian; with probability  $1 - \zeta$ , and is an infinite positive value  $y$ ; with probability  $\zeta$ . We express below the probabilities

$P_{d\zeta}(\{i_1, \dots, i_m\})$  and  $P_{d\zeta}^r(\{i_1, \dots, i_m\})$  induced by this outlier model and the optimal ML detector in Section 3 versus the robust detector, respectively.

$$P_{d\zeta}(\{i_1, \dots, i_m\}) = P_{d\zeta}^r(\{i_1, \dots, i_m\}) = (1-\zeta)^{n-m} \times \left[ (1-\zeta) \Phi\left(\frac{\theta}{2\sigma}\right) + \zeta \right]^m \Phi^{n-m}\left(\frac{\theta}{2\sigma}\right); 0 \leq m < \alpha n \quad (15)$$

$$P_{d\zeta}(\{i_1, \dots, i_m\}) = (1-\zeta)^{1+(1-\alpha)n} \alpha n \times \int_{\frac{-\theta}{2\sigma}}^{\infty} dw \varphi(w) \left[ (1-\zeta) \Phi(-w) + \zeta \right]^{\alpha n - 1} \left[ \Phi\left(w + \frac{\theta}{\sigma}\right) \right]^{(1-\alpha)n} + \zeta^{\alpha n} (1-\zeta)^{(1-\alpha)n}; m = \alpha n \quad (16)$$

$$P_{d\zeta}^r(\{i_1, \dots, i_m\}) = (1-\zeta)^{1+(1-\alpha)n} \alpha n \int_{\frac{-\theta}{2\sigma}}^{\frac{d-\theta}{\sigma}} dw \varphi(w) \times \left[ (1-\zeta) \Phi(-w) + \zeta \right]^{\alpha n - 1} \left[ \Phi\left(w + \frac{\theta}{\sigma}\right) \right]^{(1-\alpha)n} + (1-\zeta)^{(1-\alpha)n} \times \left[ (1-\zeta) \Phi\left(\frac{-d+\theta}{\sigma}\right) + \zeta \right]^{\alpha n} \Phi^{(1-\alpha)n}\left(\frac{d}{\sigma}\right); m = \alpha n \quad (17)$$

Comparison between Expressions (16) and (17) reveals that the robust detector attains higher probability of correct detection than the detector in Section 3; in the presence of the extreme outliers, where the difference of this performance increases with increasing  $\zeta$  value.

**Remarks**

If it is known that the signal may appear as a set of bursty batches of unknown sizes and protection against data outliers is needed, then, the robust re-initialization sequential algorithm in [9] will sequentially detect the beginning and the ending of each batch with minimal complexity, no memory requirements and with accuracy increasing with the signal-to-noise ratio and the size of each batch. Let then  $T_n^r$  denote the value of the robust algorithm which detects the beginning of a signal batch, upon the processing of the  $n^{\text{th}}$  datum  $x_n$  from its beginning. Let  $W_n^r$  denote the value of the robust algorithm which detects the ending of the batch, upon processing the  $n^{\text{th}}$  datum  $y_n$  from the beginning of its initialization. The whole algorithmic system operates then as follows, where  $\delta_0^r$  and  $\delta_1^r$  are two positive thresholds pre selected to satisfy power and false alarm trade offs and where  $z(x)$  is as in (12):

1) Process the observed sequence  $x_1, \dots, x_n$  sequentially starting with the algorithm  $\{T_n^r\}$  whose values are updated as follows:

$$T_0^r \equiv 0$$

$$T_{n+1}^r \equiv \max\left(0, T_n^r + z(x_{n+1}) - \frac{\theta}{2}\right)$$

Stop the first time  $n$ , such that  $T_n^r \geq \delta_0^r$  and declare  $n$  as the time when the signal batch begins.

Then, switch immediately to the algorithm  $\{W_n^r\}$  whose values are updated as follows, where time zero denotes the time when the algorithm begins and where  $y_n$  denotes the  $n^{\text{th}}$  observed datum after the latter beginning:

$$2) W_0^r \equiv 0$$

$$W_{n+1}^r \equiv \max\left(0, W_n^r - z(y_{n+1}) + \frac{\theta}{2}\right)$$

Stop the first time  $n$ , such that  $W_n^r \geq \delta_1^r$  and declare that the signal batch has ended.

**5. Suboptimal Tree—Search Detectors**

In this section, we consider the special case where the  $\alpha n$  components of the sparse signal are spread relatively evenly across the  $n$  members of the observation set. Then, we wish to devise a detector whose objective is to identify the presence of isolated signal—including observations within clusters of signal-absent observations. In this case, we may draw from the information theoretic concepts of noiseless source coding to devise tree-search-type detectors for sparse signals, as was done for the transmission/probing of bursty signals [3,4]. In particular, referring to the notation and model in Section 2, where  $f_1(\cdot)$  and  $f_0(\cdot)$  are the pdfs of signal-including versus signal-absent observations, respectively and where

$g(x) \equiv \log(f_1(x)/f_0(x))$ , we define a tree-search detector as follows, considering for simplicity in notation that the size of the observation set is a power of 2:

*Tree—Search Detector*

1) Given the sequence  $x_1, \dots, x_N$ , compute all  $g(x_j)$ ;  $1 \leq j \leq N = 2^n$ .

2) Utilize a sequence  $\{\beta_k\}$  of given algorithmic constants as:

a) If  $\sum_{k=1}^N g(x_k) \leq \beta_n$ , then, decide that at most a single

signal component is contained in the sequence  $x_1, \dots, x_N$  and stop.

b) If  $\sum_{k=1}^N g(x_k) > \beta_n$ , then, create the two partial sums

$$\sum_{k=1}^{\frac{N}{2}} g(x_k) \quad \text{and} \quad \sum_{k=\frac{N}{2}+1}^N g(x_k).$$

c) Test each of the two sums in b) against the constant  $\beta_{n-1}$  and go back to steps a) and b).

3) In general, the observation set  $x_1, \dots, x_N$  is sequentially subdivided in powers of 2 number of portions, until the subdivision stops. If, during the algorithmic

process, the observations with indices  $\{i_1, \dots, i_m\}$ ;  $m = 2^l$  are tested, then,

a) If  $\sum_{k=1}^m g(x_{i_k}) \leq \beta_l$ , then, decide that at most a single signal component is contained in the sequence  $x_1, \dots, x_N$  and stop.

b) If  $\sum_{k=1}^m g(x_{i_k}) > \beta_l$ , then, create the two partial sums  $\sum_{k=1}^{\frac{m}{2}} g(x_{i_k})$  and  $\sum_{k=\frac{m}{2}+1}^m g(x_{i_k})$ .

c) Test each of the two sums in b) against the constant  $\beta_{l-1}$  and go back to steps a) and b).

When the signal—including observations are uniformly distributed across all observations, the operational complexity of the tree—search detector is of order  $[\log m]N$ ; where  $m$  is the number of signal-including components and  $N$  is the size of the observation set. The error performance of the tree—search detector may be studied for general pdfs, asymptotically: that is when the observation set is asymptotically large and each signal—including observation component is isolated in the middle of an asymptotically large population of signal—absent observations. Then, the central limit theorem provides probability of correct detection expressions which are functions of the Kullback—Leibler numbers between the pdfs  $f_1(\cdot)$  and  $f_0(\cdot)$ . However, in this section, we will focus on the constant signal and additive Gaussian noise model of Section 3, where the function  $g(x)$  in the description of the tree—search detector equals  $x - \theta/2$ . In the latter case, we express the induced probabilities of correct detection in Lemma 3, below.

Lemma 3

Let a constant signal  $\theta$  be additively embedded in white zero mean Gaussian noise with standard deviation  $\sigma$ . Let  $m = 2^l$  be the number of signal components, given that they are spread uniformly across a total of  $N = 2^n$  observations, where  $n > l$ . Let the constants  $\{\beta_k\}$  used by the tree—search detector be such that:  $\beta_k < 2\beta_{k-1}$ ; for all  $2 \leq k \leq n$ . Then, the probability,  $P_d(l, n)$ , of correct detection induced by the tree-search detector is given by the following expression, where this probability is conditioned on the above uniform signal spreading assumption (see Equation (18)).

where,

$$c_{in} \equiv \left[ \beta_{n-l} + (2^{n-l-1} - 1)\theta \right] \sigma^{-1} 2^{-\frac{n-l}{2}} \quad (19)$$

$$d_{in} \equiv \left[ \beta_{n-l+1} + (2^{n-l-1} - 1)2\theta \right] \sigma^{-1} 2^{-\frac{n-l}{2}} \quad (20)$$

Proof

For the proof, we consider that a single signal component per  $2^{n-l}$  observations is present. Then, we express  $P_d(l, n)$  as the probability that two independent sums of such  $2^{n-l}$  observations are each smaller than  $\beta_{n-l}$ , while their sum is larger than  $\beta_{n-l+1}$ .

It is relatively easy to conclude that the probability of correct decision in (18) is increasing with increasing signal-to-noise ratio  $\theta/\sigma$ , as well as with increasing difference  $n - l$ . It is also relatively simple to conclude that the  $\beta_{n-l}$  and  $\beta_{n-l+1}$  values should be of  $2^{-(n-l)}$  order; for asymptotically large values of the difference  $n - l$ . We may select the specific values of the constants  $\beta_{n-l}$  and  $\beta_{n-l+1}$  based on a maximization of correct detection criterion, as stated in Lemma 4 below.

Lemma 4

Let a constant signal  $\theta$  be additively embedded in white zero mean Gaussian noise with standard deviation  $\sigma$ . Let the signal  $\theta$  occur with probability  $q$  per observation, independently across all  $N = 2^n$  observations. Let  $P_d(n-l, q)$  denote then the probability of correctly deciding between at most one and at least two signal components in  $2^{n-l}$  observations, via the use of the tree-search constant  $\beta_{n-l}$ . The constant  $\beta_{n-l}$  may be selected as that which maximizes the probability  $P_d(n-l, q)$ , where the latter is given by the following expression.

$$P_d(n-l, q) = (1-q)^{2^{n-l}} \Phi(c_{n/l0}) + 2^{n-l} q (1-q)^{2^{n-l}-1} \Phi(c_{n/l1}) + \sum_{k=2}^{2^{n-l}} \binom{2^{n-l}}{k} q^k (1-q)^{2^{n-l}-k} \Phi(-c_{n/lk}) \quad (21)$$

where,

$$c_{clk} \triangleq \lambda_{n-l} + (2^{n-l-1} - k)\mu_{n-l} \quad (22)$$

$$\lambda_{n-l} \triangleq \frac{\beta_{n-l}}{\sigma 2^{-\frac{n-l}{2}}} \quad (23)$$

$$\mu_{n-l} \triangleq \frac{\theta}{\sigma 2^{-\frac{n-l}{2}}} \quad (24)$$

For signal-to-noise ratio  $\theta/\sigma$  asymptotically small, for  $n - l \geq 2$  and  $q < 2^{-\frac{1}{2^{n-l}-1}}$ , the constant  $\beta_{n-l}$  is given by Equation (25) below, where it can be then shown that  $\beta_{n-l+1} < 2\beta_{n-l}$ .

$$P_d(l, n) = \left[ \Phi(c_{in})(\Phi(c_{in}) - \Phi(d_{in} - c_{in})) - \int_{d_{in} - c_{in}}^{c_{in}} dx \varphi(x) \Phi(d_{in} - x) \right]^{2^{l-1}} \quad (18)$$

$$\frac{\beta_{n-l}}{\sigma} = \frac{2\sigma}{\theta} \frac{2(1-q)^{2^{n-l}-1} - 1}{4(2^{n-l}-1) + 2p - 1} \quad (25)$$

$$\cong_{n-l \gg 1} 2^{-(n-l)} \frac{\sigma}{\theta} \left[ 2(1-q)^{2^{n-l}} - 1 \right]$$

Proof

Expression (25) is derived in a straight forward fashion. When  $\theta/\sigma$  is asymptotically small, we use first order Taylor expansion approximations with respect to the  $\mu_{n-l}$  (defined in (24)) of the  $\Phi(\cdot)$  functions in (21). Subsequently, we find the maximizing  $\lambda_{n-l}$  (defined in (23)) value of the resulting expression, as given by (25).

We note that we may “robustify” the tree-search detector at the Gaussian nominal model, by using, instead,  $g(x) = z(x) - \theta/2$ ; for  $z(x)$  as in (12). The error performance of the robust tree-search detector may be then studied asymptotically. We will not include such asymptotic study in this paper.

## 6. Discussion and Conclusion

In this paper, we investigate the problem of detecting sparse signals that are embedded in noise. The effective solution of the problem requires clear modeling of any a priori knowledge available to the designer. We first present and analyze the problem solution when only the highest percentage of signal-including observations is a priori known. In the latter case, we consider the special case of a constant signal embedded in additive white Gaussian noise and presented both parametric and outlier resistant solutions. Secondly, we present an efficient solution to the problem when it is a priori known that the signal occurs in bursty batches. Finally, we present a tree-search solution approach for the case when the sparse signal is known to be uniformly distributed within the observations set. We analyze all our approaches in terms of probability of correct detection performance and complexity.

## REFERENCES

- [1] E. J. Candes, J. Romberg and T. Tao, “Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information,” *IEEE Transactions on Information Theory*, Vol. 52, No. 2, 2006, pp. 489-509. [doi:10.1109/TIT.2005.862083](https://doi.org/10.1109/TIT.2005.862083)
- [2] J. A. Tropp and A. C. Gilbert, “Signal Recovery from Random Measurements via Orthogonal Matching Pursuit,” *IEEE Transactions on Information Theory*, Vol. 53, No. 2, 2007, pp. 4655-4666. [doi:10.1109/TIT.2007.909108](https://doi.org/10.1109/TIT.2007.909108)
- [3] J. F. Hayes, “An Adaptive Technique for Local Distribution,” *IEEE Transactions on Communications*, Vol. 26, No. 8, 1978, pp. 1178-1186. [doi:10.1109/TCOM.1978.1094204](https://doi.org/10.1109/TCOM.1978.1094204)
- [4] J. I. Capetanakis, “Tree Algorithms for Packet Broadcast Channels,” *IEEE Transactions on Information Theory*, Vol. 25, No. 5, 1979, pp. 505-515. [doi:10.1109/TIT.1979.1056093](https://doi.org/10.1109/TIT.1979.1056093)
- [5] A. T. Burrell and P. Papantoni-Kazakos, “Random Access Algorithms in Packet Networks: A Review of Three Research Decades,” *International Journal of Communications Network and System Sciences (IJCNS)*, Vol. 5, No. 10, 2012, pp. 691-707. [doi:10.4236/ijcns.2012.510072](https://doi.org/10.4236/ijcns.2012.510072)
- [6] P. Huber, “A Robust Version of the Probability Ratio Test,” *The Annals of Mathematical Statistics*, Vol. 36, 1965, pp. 1753-1758. [doi:10.1214/aoms/1177699803](https://doi.org/10.1214/aoms/1177699803)
- [7] D. Kazakos and P. Papantoni-Kazakos, “Detection and Estimation,” Computer Science Press, 1990.
- [8] A. T. Burrell and P. Papantoni-Kazakos, “Extended Sequential Algorithms for Detecting Changes in Acting Stochastic Processes,” *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 28, No. 5, 1998, pp. 703-710. [doi:10.1109/3468.709621](https://doi.org/10.1109/3468.709621)
- [9] A. T. Burrell and P. Papantoni-Kazakos, “Robust Sequential Algorithms for the Detection of Changes in Data Generating Processes,” *Journal of Intelligent and Robotic Systems*, Vol. 60, No. 1, 2010, pp. 3-17.