Robust Adaptive Detection of Buried Pipes using GPR

Q. Hoarau*, G. Ginolhac*, A. M. Atto*, J.M. Nicolas† and J.P. Ovarlez‡

*LISTIC, Université de Savoie Mont-Blanc, FRANCE
Email:quentin.hoarau@univ-smb.fr
†LTIC, Telecom ParisTech, FRANCE
‡ONERA & SONDRA, CentraleSupelec, FRANCE

Abstract—The Ground Penetrating Radar (GPR) consists in an electromagnetic signal which is transmitted at different positions through the ground in order to obtain an image of the subsoil. In particular, the GPR is used to detect buried objects like pipes. Their detection and localisation are intricate for three main reasons. First, the noise is important in the resulting image due to the presence of several rocks and/or layers. Second, the wave speed and the response of the pipe depend on the characteristics of the different layers. Finally, the signal attenuation could be important because of the depth of pipes. In this paper, we propose to derive an adaptive detector where the steering vector is parametrised by the wave speed in the ground and the noise follows a Spherically Invariant Random Vector (SIRV) distribution in order to obtain a robust detector. To estimate the covariance matrix, we propose to use robust maximum likelihood-type estimators called $M$-estimators. To handle the large size of data, we consider regularised versions of such $M$-estimators. Simulations will allow to estimate the relation Probability of False Alarm (PFA)-Threshold. Application on real datasets will show the relevancy of the proposed analysis for detecting buried objects like pipes.

I. INTRODUCTION

Ground Penetrating Radar (GPR) imaging involves transmitting an electromagnetic wave at several spatial positions and receiving subsoil retro-diffusion waves to form images [1], [2]. In particular, GPR can be used to detect buried objects like landmine, pipes, etc. Most of GPR acquisition devices transmit a simple Ricker wavelet of small duration leading to a large bandwidth. In most applications, the GPR is composed of a monostatic antenna with a given sub-aperture, though more complex radar systems can be used. In this paper, we will consider a monostatic Radar system which transmits a Ricker wavelet. Several papers are devoted to landmine detection, another type of buried objects close to the ground level, leading to a poor SNR scenario, but with high object response. For this, a Kalman filter based technique has been used in [3], later improved using a particle filter in [4]. For deeper objects (like pipes), new strategies need to be found. In this paper, we propose a new approach built from the Synthetic Aperture Radar (SAR) [5] configuration where the raw data is also composed of several hyperbola which are the responses of all scatterers located in the ground, coherently combined to form an image of the ground. But in SAR the speed of the electromagnetic wave is known whereas in GPR applications, the wave speed is unknown as multiple ground layers could be present and be located between the radar and the buried objects. The problem of subsurface layers is also encountered in the domain of Through-the-Wall SAR [6] where the clutter can be removed for example by a subspace projection [7]. This last method could be inefficient in our case as the subspace operation removes part of the hyperbola and thus affects the response of the buried pipe.

For the derivation of the method, we use the radar detection/estimation framework initially given in [8]. The proposed detector depends on the response of the buried pipe, denoted by analogy with the SAR the steering vector [9]. This steering vector is built from the theoretical hyperbolas and the known transmitted signal. It is parametrised by the position in the subsoil and the dielectric constant which are unknown. We consider that a noise corrupts the received signal, which is a combination of electronic noise, clutter, and also small buried objects like rocks. If we consider that the noise is modeled by a multivariate Gaussian vector of zero mean and of an unknown covariance matrix, we can derive the corresponding "2-Step" Generalised Likelihood Ratio Test (GLRT) [10] called Adaptive Normalised Matched Filter (ANMF). Unfortunately, the Gaussian distribution seems to be not adequate in this configuration: the clutter or the presence of small objects, denoted in the following outliers, tends to create signals with strong heterogeneity. In this case, the Gaussian distribution is not adapted and family of statistical distributions is better suited for this kind of modeling, the Real Elliptic Symmetric (RES) [11] or a sub-family of RES, the Spherical Invariant Random Vector (SIRV) [12]. When the statistic of the SIRV texture is not known but considered as unknown deterministic parameter, the MLE of the covariance matrix is found to be the Tyler’s estimator [13]. This estimator is well-known to be robust to strong heterogeneity and outliers but a recent paper [14] shows that it could be suffering from bad estimation for some specific outliers. The authors recommend to use another estimate of the covariance matrix in the family of the M-estimators [15], [11]: the Huber’s estimator [16], [17], [18] which is the combination of the SCM and the Tyler’s estimator. In this paper, we will compare the detectors built from SCM and Huber’s estimators.

A major problem in adaptive detection comes from the estimation of the covariance matrix, which needs a set of
consists usually in choosing secondary data, and for correct estimation a good choice consists usually in choosing \( K = 2N \) where \( N \) is the tested data size [19]. However, in our configuration, the data size is really high and it is impossible to have a sufficient set of secondary data. Strategies exist to reduce the required amount of secondary data. In most cases, the drawback is the choice of the regularisation parameter. For the SCM, a strategy is proposed in [22], and adapted to array processing in [23]. For Huber's estimator, the existence has been proven in [24], but no criteria has been defined to estimate the optimal value. We propose here to estimate it by choosing the value ensuring the most stable false alarm rate versus data heterogeneity. This analysis will be performed by means of simulations.

To summarise, we propose in this paper the derivation of a robust adaptive detector in order to detect buried pipes in presence of subsurface layers and outliers, while using a small amount of secondary data. In most cases, the covariance matrix is known to have a particular structure which allows to reduce the number of estimated parameters [20], [21]. But the most common approach is to regularise the estimation: this is really high and it is impossible to have a sufficient set of secondary data. However, in our configuration, the data size [19].}

II. SIGNAL MODELING

A. GPR Signal modeling

In a general configuration, we consider a GPR moving along a \( v \)-axis, parallel to the ground (\( y \)-axis) and at a height \( h \) from it. At every position \( u_m, m \in [1, M] \), the radar emits a signal \( e(t) \) in the ground. This signal is reflected by ground scatterers characterized by their reflection coefficients \( a_p \in [-1, 1] \) where \( p \in [1, P] \). The echo of the signal received by the radar at position \( u_m \) is:

\[
 r_m(t) = \sum_{p=1}^{P} a_p e(t - \tau_m(y_p, z_p, \epsilon)),
\]

where \( \tau_m \) is the time taken by the signal to travel from the radar to the scatterer and back, and depends on the value of the permittivity \( \epsilon \). In the following, we consider a monolayer ground model for the computation of \( \tau_m \), though more complex models can be used if needed.

GPR systems return digital data, we thus consider that the imaged scene is divided into \( P = N_h \times N_z \) pixels each considered as a potential scatterer, and observe the received signals over discrete time samples \( t_i, i \in [1; NT] \). We create a vector \( a = [a_1, \ldots, a_p, \ldots, a_P]^T \) containing all the coefficients arranged in lexicographic order, and build a set of \( M \) matrices.

![Fig. 1: Sampling process for the vector \( x \): red lines represent the samples along the different columns of the image, which are then assembled in the vector \( x \).](image)

\[
 H_m \in \mathbb{R}^{N_T \times P}, \quad \text{with } [H_m]_{ip} = e(t_i - \tau_m(y_p, z_p, \epsilon)).
\]

We express the sampled version of the received signal \( r_m \) by the radar at position \( u_m \) as \( r_m = H_m a \). Then, by concatenating all the \( M \) matrices \( H_m \) into \( H = [H_1^T \ldots H_M^T]^T \), and the \( M \) vectors \( r_m \) into \( r = [r_1^T \ldots r_M^T]^T \), we get the full GPR B-scan image modeling:

\[
 r = H a.
\]

B. Detection problem

Because the GPR antenna emits in a wide beam pattern, buried objects appear as hyperbolic shapes on the scan image. Thus, trying to detect a buried pipe equivalent to checking the presence of a reflection hyperbola around the sampled position. To do so, we sample the pixels around the position \((y, z)\) while following the shape of the hyperbola as summarised in fig. 1.

This sampling can be formalised in matrix form. From the signal \( r \), the signal \( x_{\epsilon', y, z} \in \mathbb{R}^N \) to be tested is extracted by means of a selection matrix unique to each position \((y, z)\)

\[
 T_{\epsilon', y, z} \in \{0; 1\}^{N \times MN_T}, \quad \text{where } N \text{ is the number of samples in the vector } x_{\epsilon', y, z}.
\]

This matrix is built as follows: if the \( i^{th} \) component of the tested vector \( x \) is set to be the \( j^{th} \) component of \( r \) then \( [T_{\epsilon', y, z}]_{ij} = 1 \), and if not \( [T_{\epsilon', y, z}]_{ij} = 0 \).

\[
 x_{\epsilon', y, z} = T_{\epsilon', y, z} r.
\]

This signal \( x_{\epsilon', y, z} \) is then tested under two hypotheses: either it contains only noise \( n \), or it contains the sum of a set pattern \( p \), characterized by a steering vector with amplitude \( a_{\epsilon', y, z} \) and additive noise \( n \). The steering vector \( p \) represents the ideal theoretical reflection hyperbola. The detection problem is expressed as:

\[
 \begin{align*}
 H_0 : & \quad x_{\epsilon', y, z} = n \\
 H_1 : & \quad x_{\epsilon', y, z} = a_{\epsilon', y, z} p + n.
\end{align*}
\]

In the following, in order to simplify notations, we will denote \( x = x_{\epsilon', y, z} \) and \( a = a_{\epsilon', y, z} \).

Along with the vector \( x \) we sample a set of secondary data \( \{x_k\}_{k=1}^{K} \) following the same fashion, for positions located at the same depth \( z \) as \( x \), while respecting a safe zone along the \( y \)-axis to prevent any signal being inadvertently sampled in this secondary set. It will be used in later sections to estimate the noise \( n \) characteristics.
C. Noise modeling

To derive the detector corresponding to the solution of problem (4), we have to propose a model for the noise \( n, n_k \in \mathbb{R}^N \). Here, we will consider two cases: partially homogeneous Gaussian and SIRV noise to take into account the heterogeneity of the data.

Partially homogeneous Gaussian modeling: In this case, the noise \( n \) follows a zero-mean Gaussian process \( \mathcal{N}(0, \mathbf{R}) \) where \( \mathbf{R} \) is the covariance matrix. We assume that this covariance matrix is unknown. Moreover we assume that the secondary data \( \{ x_k \}_{1,K} \) follow a zero-mean Gaussian process \( \mathcal{N}(0, \sigma^2 \mathbf{R}) \) where \( \mathbf{R} \) is the same covariance matrix and where \( \sigma \) is an unknown deterministic parameter. This last parameter allows to take into account different powers between the data under test and the secondary data. Nevertheless, this assumes that the power for all secondary data is the same which is not realistic in many applications. Therefore, we propose in the following paragraph another kind of modeling.

SIRV modeling: We consider that the powers associated to each \( x_k \) and the data under test are different. In such a situation, it is common to model this power variation of clutter by a SIRV [25], a non-homogeneous Gaussian random vector with random power. The SIRVs [12] \( n, n_k \) are respectively the product of positive scalar random variables \( \kappa, \kappa_k \), called the texture, and \( N \)-dimensional vectors \( g, g_k \) which follow independent Gaussian process \( \mathcal{N}(0, \mathbf{R}) \): \( n = \sqrt{\kappa} g \) and \( n_k = \sqrt{\kappa_k} g_k \). In the following, for identifiability issue in the SIRV case, we consider \( \text{Tr}(\mathbf{R}) = N \).

III. DETECTION

To derive the detector of the detection problem (4), we can resort to the GLRT framework. When the covariance matrix is assumed to be known, the derivation of the LRT leads to the same detector for both Gaussian and SIRV noise models [26][27][28], here parametrised by the dielectric constant \( \epsilon' \), which is then compared to a threshold to decide between hypotheses \( H_1 \) or \( H_0 \):

\[
\Lambda = \max_{\epsilon' \in \mathbb{R}^+} \frac{|p^T \mathbf{R}^{-1} x|^2}{(p^T \mathbf{R}^{-1} p)(x^T \mathbf{R}^{-1} x)} \leq \frac{\eta}{\eta}.
\]

In this section, we present different estimators of the covariance matrix. The problem is that the number of secondary data available \( K \) will be inferior to the size data \( N \) causing the inversion in (5) to be inconsistent. Therefore, it is necessary to develop a strategy to propose an efficient estimator using a small support of secondary data. Since, we do not assume any structure (low-rank, Toeplitz) for the covariance matrix, we propose here to use shrinkage estimators of the covariance matrix which have known a great interest these last years in the signal processing community. First, we consider the Gaussian case, then the SIRV case.

a) Gaussian case: In this case, the most effective method is to replace the true covariance matrix by its Maximum Likelihood Estimators (MLE), the Sample Covariance Matrix (SCM). In the under-sampled case, the seminal paper [22] proposes the following estimator:

\[
\hat{\mathbf{R}} = \beta \frac{1}{K} \sum_{k=1}^{K} x_k x_k^T + \alpha \mathbf{I}_N,
\]

where \( \beta \) and \( \alpha \) have to be estimated. Optimal values are given in [23].

b) SIRV case: In the over sampled case \( (K > N) \), a well-known and efficient estimate of the covariance matrix (MLE for deterministic texture) is the Tyler’s estimator. It is known to be robust to the statistic of the noise and the presence of outliers in the secondary data. Nevertheless, in a recent paper [14], the authors prove that the Tyler’s estimator could be suffering of data contamination for some specific configurations. In that case, one can resort to M-estimators [15][16][11] and use, for example, the Huber’s estimator. As for the Gaussian case, we can build an under-sampled version of this estimator [24]:

\[
\hat{\mathbf{R}} = \frac{1}{K} \sum_{k=1}^{K} u \left( x_k^T \hat{\mathbf{R}}^{-1} x_k \right) x_k x_k^T + \alpha \mathbf{I}_N,
\]

where \( u(t) = \begin{cases} 1, & \text{for } t \leq c^2 \\ c^2/t, & \text{for } t > c^2 \end{cases} \) and where \( c > 0 \) can be set from the percentage of data considered as Gaussian distributed. Iterations associated to this fixed point equation allow, for any \( \alpha \) and for any initial value \( \hat{\mathbf{R}}_0 \) [24] to reach the unique solution. For the best of our knowledge, we do not know any paper which proposes an optimal value for \( \alpha \).

IV. SIMULATIONS

It is useful to know the distribution of the detector with a particular covariance matrix estimator in order to set the threshold \( \eta \) for a given PFA level. Unfortunately, this distribution is difficult to obtain when regularised estimators of the covariance matrix are used in the ANMF. Therefore, we propose to estimate the relation PFA-threshold by means of Monte Carlo trials for the ANMF built from the SCM and Huber’s estimator. Moreover, we will estimate the optimal regularisation parameter value for Huber’s estimator by finding the value that keeps a constant false alarm behaviour as a function of data heterogeneity.

A. Statistical simulations

We consider the SCM and Huber’s adaptive detector: \( \hat{\Lambda}_{SCM} \) and \( \hat{\Lambda}_{Hub} \), built by replacing the true covariance matrix in the NMF of (5) by the regularised version of the estimators. The simulation consists in Monte Carlo trials, to estimate the PFA as a function of noise heterogeneity and threshold value. It was set up with a SIRV noise, the Gaussian vectors \( g, g_k \sim \mathcal{N}(0, \mathbf{R}) \) where \( [\mathbf{R}]_{ij} = \rho^{|i-j|} \) is a Toeplitz covariance matrix of parameter \( \rho = 0.9 \) and where the texture is Gamma-distributed of shape parameter \( \nu \) and scale parameter \( 1/\nu \), with \( \nu = 2 \) is leading to a near-Gaussian noise and \( \nu = 0.1 \) is leading to a more impulsive noise.

Fig. 2a shows the PFA versus \( \nu \) curves for Huber’s estimator with varying \( \alpha \) values. Low \( \alpha \) values appear to give the most
Fig. 2: (a) Huber’s estimator PFA versus $\nu$ for varying alpha values given for threshold value $\eta = -15\text{dB}$, (b) PFA versus threshold for all estimators for $\nu = 2$ and $\nu = 0.1$. Parameters are $N = 153$ (corresponding 51 sampled columns with 3 samples per column), $K = 40$, and results are averaged over 100000 Monte-Carlo trials.

![Huber - PFA vs $\nu$ vs $\alpha$](image1)

![PFA vs Threshold - $\nu = 2, 0.1$](image2)

Fig. 3: Adaptive detector applied to real data using the ANMF. Approximate pipes locations are indicated by red squares. (a) Original image (Pipes indicated by red squares) (b) White Noise hypothesis (c) SCM (d) Huber ($\alpha = 0.05$). $N = 153$, $K = 40$.

![Original image](image3)

![Detection image (White Noise)](image4)

![Detection image (SCM)](image5)

![Detection image (Huber)](image6)

stable PFA values across the texture parameter range. Fig. 2b shows the PFA versus threshold graphs for Huber’s estimator with $\alpha = 0.05$ and the SCM with optimal $\alpha$ and $\beta$ values from [23], along with both limit case Oracle ($\hat{R} = R$) and White noise ($\hat{R} = I_N$), for $\nu = 0.1$ and $\nu = 2$. In the nearly-Gaussian case, the estimators achieve similar results, and the Huber’s detector gives better PFA levels in the more impulsive case. Also we can observe that Huber’s regularised detector has a more stable constant false alarm rate behaviour than the SCM.

One must note that having the lowest PFA values does not guarantee the best detection result. This should be observed...
with the proper set of simulations, such as observing the Probability of Detection as a function of the SNR, which have not been conducted here due to lack of space.

B. Application on real data

In this section we apply the two regularised detectors, SCM and Huber’s, as well as the white noise limit case ($R = I_x$) to real data. The image (fig. 3a) has a value $M$ was set with four possible values: 3, 5, 7, and 15. The reflected signal is a Ricker wavelet of peak frequency 500 MHz. To speed up the computation, the permittivity value is set with four possible values: 3, 5, 10, and 15. Other parameters are kept identical to the previous statistical simulations.

The tests on fig. 3 reveal good overall results from all detectors. On the image, one secondary pipe is enhanced by the detectors, and the other is still weak and close to noise level. Result image similarities indicate that the noise model is closer to Gaussian than impulsive, but also that there is no performance loss in considering a more complex noise model. In comparison to the white noise hypothesis, all detectors greatly succeed in removing the layer interface response in the top part of the image, proving the relevancy of the adaptive approach, and confirming that a pre-processing of the images is not necessary for the adaptive detection approach. This is a good thing, as such techniques could be partially destructive to the signals of interest by removing the top part of the reflection hyperbolas. Huber’s detector has a loss of performance at the bottom of the image, which is due to the image being padded with zeros. This issue may be solved using another regularisation model. Finally, the mean estimated permittivity value for the three detectors in this image is $\epsilon' = 5$.

V. CONCLUSION

In this paper, we developed a robust adaptive detector in order to detect and localise buried pipes, based on the regularised version of ANMF. In the derivation, the noise has been modeled by a SIRV process in order to take into account of the heterogeneity of the data. Simulations allowed to observe the impact of the regularisation on the PFA-threshold relation. Application on real data showed the interest of the approach, especially in detecting pipes having a weak response level.

REFERENCES