New Robust Sparse Convolutional Coding Inversion Algorithm for Ground Penetrating Radar Images

Matthieu Gallet, Ammar Mian, Guillaume Ginolhac, Esa Ollila, Nickolas Stelzenmuller

New Robust Sparse Convolutional Coding Inversion Algorithm for Ground Penetrating Radar Images

Matthieu Gallet, Ammar Mian, Guillaume Ginolhac, *Senior Member, IEEE*, Esa Ollila, *Senior Member, IEEE*, Nickolas Stelzenmuller

***Abstract*—In this paper, we propose two algorithms to enhance the interpretability of the hyperbola in B-scans obtained with a Ground Penetrating Radar (GPR). These hyperbolas are the responses of buried objects or cavities. To correctly detect and classify them, a denoising is typically necessary for GPR images as the signal-to-noise ratio is low, and the various interfaces naturally present in the earth have a strong response. Both algorithms are based on a sparse convolutional coding model plus a low rank component. It is solved through an Alternating Direction Method of Multipliers (ADMM) framework. In order to take into account the presence of outliers and the artifacts caused by the acquisition, the second algorithm is based on the Huber norm instead of the classic L_2-norm. These algorithms are tested on a real dataset labeled by geophysicists. The results show the denoising efficiency of this approach, and in particular the robustness of the second algorithm.**

***Index Terms*—Ground Penetrating Radar, Sparse Inversion, Convolutive Model, Robust methods**

I. Introduction

Ground Penetrating Radar (GPR) consists in transmitting an electromagnetic wave in the ground [1]–[3]. This wave can be reflected by buried objects or cavities to be then received by the radar. The linear displacement of the radar antenna during a measurement allows the system to create a 2D image which is called a radargram or B-scan. Since the radar is moving and several acquisitions are made, the response of an object has a form of hyperbola. Material interfaces, such as air-earth, rock-soil, among other naturally occurring layers in the earth also reflect the radar signal. These refelections are visibile in the B-scan, often appearing as quasi-horizontal lines. An example of B-scan with one main object, the response of the ground and the response of a layer is shown in Figure 1.

A typical objective in the analysis of a GPR image is to classify the underground objects. The main information about the object is contained in the shape of the hyperbola. For this task, several classification methods [4]–[11] have been proposed that are based on supervised approaches like deep neural networks. These methods require a relatively large amount of labeled data to achieve an acceptable classification accuracy. However, the accurate labeling of B-scans is often difficult, even for trained users of GPR systems, due to the low Signal-to-Noise Ratio (SNR) of the objects of interest. Moreover, the responses of the various interfaces naturally present complicate the interpretation of the shape of the hyperbola. The improvement of the quality of these images would allow a more accurate interpretation and classification of objects found in GPR images.

One strategy to improve image quality involves increasing signal diversity in order to better distinguish the contribution of the various sources. For example, a stepped frequency GPR is used in [12] which improves the SNR and increases the effective depth of the measurement. Higher SNR has also been reported from measurements using novel GPR systems such as those using polarization diversity (like in Polarimetric SAR) or a Multiple Input Multiple Ouput radar [13]. The later option greatly increases the resolution of the B-scans. Unfortunately these novel systems are often impractical given cost, weight, and size constraints. For example, drone-based measurements are increasingly common, and impose strict size and weight constraints on GPR systems. To improve the quality of the images, it is then necessary to resort to signal or image processing techniques. For example, Kalman filters are often used signal processing technique for detection/localization of buried landmines as in [14]. An improvement of this method using a particle filter is proposed in [15]. A landmine detection method based on hidden Markov model has been proposed in [16]. A main problem in B-scan images is the responses of the ground and the layers which are removed by a correlation-based approach in [17] or a PCA-based approach [18]. Statistical approaches have been also proposed in the literature. A seminal work [19] developed an energy detector. A more elaborated detector based on a GLRT is given in [20]. This method takes into account the correlation of the noise but it works column by column and assumes that the noise is Gaussian. A detector working on the full image and considering non-Gaussian noise has been developed in [21]. This method improves the object detection and allows to remove clutter. These methods have been shown to be effective in improving the SNR and detection performance, but they are often targeted to a specific type of object, such as landmines or pipes and do not allow a better classification since the information regarding the properties of the hyperbolas is lost.

In order to develop approaches that are applicable to a wide range of objects, we investigate inversion algorithms. In the GPR domain, most inversion methods are based on a precise physical model by computing for example the Green's function. These approaches consist then in applying a full waveform inversion algorithm which is based on a simple L_2 norm and a Thikonov regularization [22]–[24] or an interpolation [25]. Approaches based on compressive sensing [26] have also been proposed. However, all of these models require a complex computation of the direct model

Matthieu Gallet, Ammar Mian and Guillaume Ginolhac are with LISTIC (EA3703), University Savoie Mont-Blanc, FRANCE. Esa Ollila is with Aalto University, Finland. Nickolas Stelzenmuller is with Geolithe, FRANCE. This work has been done thanks to the facilities offered by the Univ. Savoie Mont Blanc - CNRS/IN2P3 MUST computing center.

and do not take into account layer suppression. For the latter, a spatial-averaging-based pre-processing step or PCA-based approaches as in [18] are often used before inversion. Methods using deep learning have also been developed in [27], [28] [29]–[31] but these types of approaches require a learning step which is non-robust to datasets of different modalities (frequencies of the radar, elevation, ...). In order to use the inversion approach but with a simple model for the dictionary, the sparse coding representation is a well-suited model to correct the shape of the hyperbola in B-scan images [32], [33]. Indeed, the dictionary could be easily constructed from a family of hyperbolas and even estimated from the data, as in [34]. But as for the previous algorithms, these methods do not take into account the responses of the ground and the layers. A seminal work combining the suppression of the clutter and the enhancement of the hyperbola was proposed in [35]. The method is based on the Robust Principal Component Analysis (R-PCA) initially proposed in [36] and extended to the convolutional model in [37]. The proposed algorithm is effective because it removes the contribution of the ground and the layers while the hyperbola is enhanced, allowing better interpretation of the B-scan image. But some drawbacks remain such as the construction of the dictionary which does not explicitly use the physics of the wave propagation. This can lead to physically unrealistic hyperbola which can reduce the performance of the final algorithm.

In this paper, we propose a realistic way to build our dictionary with limited size while keeping a good representation of all possible hyperbolas. One drawback of algorithm proposed in [35] is that it is not very robust to noise and other artifacts caused by the acquisition of the data (such as outliers, disalignment, ...). To take such effects into account, we propose a new inversion approach based on the Huber norm [38]. This new operator is well known to be more robust than the classical norm in optimization algorithms with regards to outliers and high noise levels. To the best of our knowledge, this is first time that this operator is used in a sparse convolutional coding estimation with the presence of a low rank contribution. In this paper, we develop an Alternating Direction Method of Multipliers (ADMM) framework to solve the optimization problem using L_2-norm and robust Huber norm. Both algorithms are presented and differences in their steps are underlined. In particular, we discuss the choice of the different hyper-parameters of the algorithms as well as how to improve its convergence.

Finally, both algorithms are tested using a dataset provided by the Geolithe company. This dataset consists of several B-scan images composed of different objects of interest. The main advantage of this dataset is that it is labeled by geo-physicists from the company, who provided the corresponding mask images locating and labeling the different hyperbolas. We use these data to demonstrate the effectiveness of the robust approach in removing noise and acquisition artifacts from the images.

We summarize the main contributions of the paper in the following: i) we propose a realistic way to build our dictionary with limited size while keeping a good representation of all possible hyperbolas. This dictionary is based on physical considerations and could be built for any dataset by using the RADAR parameters and a linescape on the permittivity and the radius of the object. ii) we develop a new robust inversion algorithm by replacing the L_2-norm of [35] by robust Huber norm. The ADMM framework is used to solve the optimization problem with new updates for the low-rank contribution and the coefficient maps. Note that this approach is non - supervised and do not need any training steps and labeled data. iii) The proposed algorithms shows a real improvement on a real dataset.

The article is structured as follows. Section II presents the data model and the procedure for building the dictionary dedicated to GPR images. In the Section III, the algorithm inspired by [35] is given. The main contribution of the paper is given in Section IV where the new optimization problem with Huber norm and the corresponding algorithm are described. Finally both algorithms are tested on a real dataset in the section V. Table I summarizes the notations used in throughout the paper.

TABLE I: Notations and definitions used in the paper

General:

- $\mathbb{R}$ is the set of real numbers and $\mathbb{C}$ is the set of complex numbers
- lowercase (resp. uppercase) bold letter denote vectors (resp. matrices)
- $(\bullet)^{\mathrm{T}}$ (resp. $(\bullet)^{\dagger}$) denotes the transpose (transpose conjugate) operator
- $|\bullet|$ is the absolute value (resp. modulus) for a real-valued (resp. complex-valued) scalar
- $\|\bullet\|_2$ is the L_2-norm, $\|\bullet\|_1$ is the L_1-norm, $\|\bullet\|_*$ is the nuclear norm and $\|\bullet\|_{\mathcal{H}}$ is the Huber norm defined at (15)
- $*$ (resp. $\circledast$) is the convolution operator for functions (resp. matrices)
- $\odot$ denotes the Hadamard product
- for a vector $\mathbf{x}$ of dimension d, $\mathrm{diag}(\mathbf{x})$ is an operator that returns a diagonal matrix of size $d \times d$.
- for a matrix $\mathbf{X}$, $\mathrm{vec}(\mathbf{X})$ yields its vectorization while $\mathrm{unvec}(\bullet)$ allows to get back the matrix
- $\mathcal{F}(\bullet)$ is the Fast Fourier Transform of a matrix. $\mathcal{F}^{-1}(\bullet)$ is the inverse transform
- given a multivariate function f with scalar output, $\nabla_{\mathbf{x}} f$ denotes the gradient of f with regards to the elements of vector $\mathbf{x}$
- $\delta(x)$ is the Dirac function equal to 1 for $x = 0$ and equal to 0 everywhere else
- $\mathbf{I}$ is the identity matrix and $\mathbf{1}_d$ is the vector of dimension d with all elements equal to 1.

Inversion model:

- N_t, N_x denote the number of samples in dimensions t and x. K is the number of atoms in the inversion model
- $\mathbf{Y}$, $\mathbf{N}$, $\mathbf{L}$, $\mathbf{H}_k$, $\mathbf{C}_k$, $\mathbf{S}_k$ and $\mathbf{M}$ are matrices of size $N_t \times N_x$
- the notation $\{\bullet_k\}$ is used to simplify $\{\bullet_k\}_{k=1,\dots,K}$

Physics:

- c is the speed of light, ϵ_r stands for relative permittivity, σ stands for conductivity and ω is the RADAR frequency.

Optimization and ADMM framework:

- λ is a sparsity favoring hyperparameter
- $\widetilde{\mathbf{L}}$ and $\widetilde{\mathbf{S}}_k$ are the dual variables of $\mathbf{L}$ and $\mathbf{S}_k$
- $\mathcal{L}$ is the augmented Lagrangian
- $\rho_{\mathbf{L}}$ and $\rho_{\mathbf{S}}$ are the augmented Lagrangian penalty parameters regarding $\mathbf{L}$ and $\mathbf{S}$
- $\bullet^{(i)}$ is the value of a parameter at iteration i of the global ADMM algorithm.
- for matrices $\mathbf{Y}$, $\mathbf{H}_k$, $\mathbf{C}_k$, $\mathbf{S}_k$, $\widetilde{\mathbf{S}}$ and $\widetilde{\mathbf{L}}$, we define $\mathbf{y}$, $\mathbf{h}_k$, $\mathbf{c}_k$, $\mathbf{s}_k$, $\widetilde{\mathbf{s}}_k$ and $\widetilde{\mathbf{l}}$ by applying the operation $\mathrm{vec}(\mathcal{F}(\bullet))$
- $\mathcal{S}_{\gamma}(\bullet)$ is the proximal soft thresholding operator (28)
- $\mathcal{T}_{\gamma}(\bullet)$ is the proximal singular value thresholding operator (34)
- $\mathcal{PH}_{\gamma,\delta}(\bullet)$ is the Huber proximal operator (25)

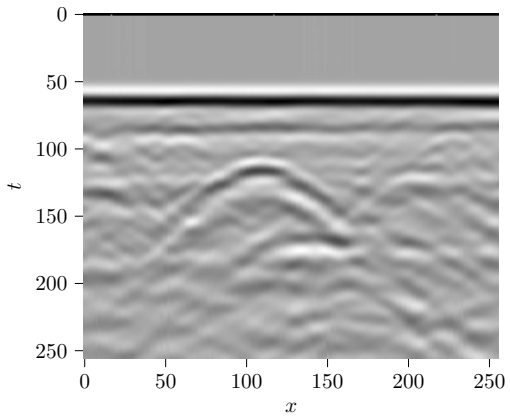


Fig. 1: Example of a B-scan $y(x,t)$ with t the delay of the wave and x the displacement of the radar.

II. Data Model of GPR

In this section we introduce a modeling of the measured GPR data by splitting the backscattering of buried objects from clutter and noise components. From this model, we obtain a formulation involving a sparse convolutional coding of probable targets backscattering. Finally, a physical modeling is proposed to construct a dictionary of possible target responses that will be subsequently used in section III.

A. Data model

A B-Scan radargram can be decomposed as follows:

$$y(x,t) = s(x,t) + cl(x,t) + n(x,t) \tag{1}$$

where:

- $n(x,t)$ is a noise component relative to the acquisition
- $s(x,t)$ is the backscattering of the different targets. Given the acquisition paradigm of GPR, a single target signal is hyperbola-shaped. In the presence of multiple targets, we have a sum of different hyperbolas [35]:

$$s(x,t) = \sum_{k=1}^{K} c_k(x,t) * h_k(x,t) \tag{2}$$

  where $h_k(x,t)$ is the image of a centered hyperbola and $c_k(x,t)$ the corresponding coefficient map.
- $cl(x,t)$ is a clutter component representing all the elements which cannot be represented by eq.(2) such as the response of the different layers in the ground.

Since the signal $y(x,t)$ is sampled in both dimensions, we obtain a matrix which can be written in the form:

$$\mathbf{Y} = \sum_{k=1}^{K} \mathbf{C}_k \circledast \mathbf{H}_k + \mathbf{L} + \mathbf{N}, \tag{3}$$

where the dimensions of the matrices are equal to those of the B-scan image, i.e. $N_x \times N_t$ and $\circledast$ is the matrix convolution product. The matrix $\mathbf{L}$ contains the responses of the clutter such as horizontal layers or acquisition artifacts. This matrix is assumed to be low-rank since those clutter signals are similar between themselves. The matrix $\mathbf{N}$ is mainly composed of the contributions of the different noises. Each $\mathbf{H}_k$ represents the hyperbola signal of each individual target in the scene and $\mathbf{C}_k$ is its corresponding mapping. In this scenario, $\mathbf{C}_k$ is expected to be very sparse, i.e., it has a few values where each target is located.

When the targets are not precisely known, this model allows to consider a large dictionary of possible target signatures $\{\mathbf{H}_k\}_{k=1,\dots,K}$ for which even fewer associated mappings $\{\mathbf{C}_k\}_{k=1,\dots,K}$ are non-zeros. We propose hereafter to construct such a dictionary of target signatures based upon a physical modeling of the interaction between a buried object and the emitted electromagnetic wave.

B. Construction of the dictionary $\{\mathbf{H}_k\}_{k=1,\dots,K}$

In [35], the dictionary is not completely related to the physical behavior of the buried objects, which leads to a dictionary with outlier elements. Figure 2 displays two elements of the dictionary developed in [35]. We notice that the element shown in the left image is a hyperbola with a physically-plausible shape, but the element on the right panel has an unrealistic shape that will never be present in real GPR images. To overcome this issue, we propose a new way to build our dictionary based on the physics of the GPR signal.

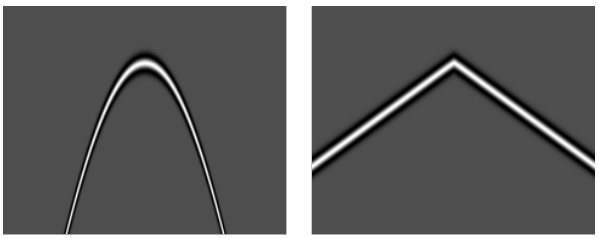


Fig. 2: Two elements of the dictionary developed in [35]: the left one has a correct shape whereas the right one is an outlier.

First, we recall the expression of the Ricker wavelet which is the transmitted signal by the GPR [39], [40]:

$$r(t|\omega_{max}) = (1 - \frac{\omega_{max}^2 t^2}{2}) \exp(-\frac{\omega_{max}^2 t^2}{4}), \tag{4}$$

where ω_{max} is the maximal pulsation (in $rad \cdot s^{-1}$) of the RADAR. By using the Ricker formulation (4), we are now able to construct the atom of our dictionary by using the following expression:

$$h(x,t) = A(x,t) \cdot (r(t) * \delta_{dirac}(t - g(x))), \tag{5}$$

where $A(x,t)$ is an attenuation value which ensures the finite dimension of the hyperbola and $g(x)$ is the hyperbola defined by:

$$g(x|x_0,t_0,a,p) = a\sqrt{p^2 + (x-x_0)^2} + t_0 - pa \tag{6}$$

where (x_0,t_0) is the top of the hyperbola. It is useful to connect (x,t) to the true dimensions of a GPR image and we propose the following interpretation:

- the number of pixels along the x axis and the distance covered by the radar in meters,
- the number of pixels along the t axis and the propagation time (one way) of the electromagnetic wave.

Thus the dimensions x and t become physical (respectively meter and second). We denote the sampling intervals in both dimensions (Δ_x, Δ_t). The choice of x_0 and t_0 can be made to maximize the size of the hyperbola in the matrix, and we choose $x_0 = \max(x)/2,\ t_0 = \max(t)/4$. A precise position does not matter since the convolution product allows us to correctly place the hyperbola in the final image. An example of Ricker wavelet and the corresponding attenuation matrix $A(x,t)$ are shown in Fig. 3.

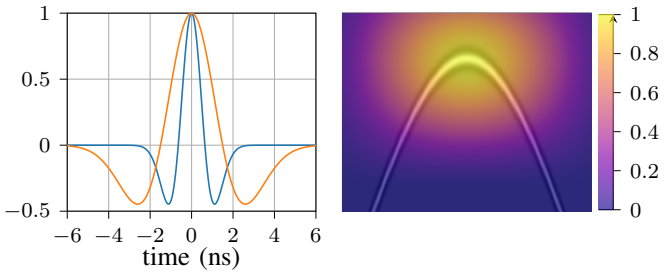


Fig. 3: Main elements of the dictionary construction: Ricker wavelet (left) $r(t|f_{max})$ with $f_{max} = 350$MHz (blue) and $f_{max} = 150$MHz (orange) and attenuation matrix right $A(x,t)$ (right).

To obtain all the atoms of the dictionary, we have to compute several hyperbola $g(x)$ of (6) by choosing different values of p and a. These parameters are linked to the physical properties of the materials crossed by the transmitted signal of the GPR and could be written as follows:

$$\begin{array}{rcl} p & = & t_0 \frac{v}{2} + R \\ a & = & \frac{2}{v} \end{array}, \tag{7}$$

where:

- R is the radius of the target illuminated by the RADAR,
- v is the wave speed and is expressed w.r.t to physical parameters of the traversed material:

$$v = \frac{c}{\left(\frac{\mu_r \epsilon_r}{2}\left[\sqrt{1+\left(\frac{c_r}{\omega}\right)^2}+1\right]\right)^{1/2}}, \tag{8}$$

with c being the speed light, ω the RADAR pulsation, μ_r the relative permeability of the material, ϵ_r the relative permittivity of the material and c_r the conductivity of the material.

Since the materials are mostly non-magnetic and non-conductive, μ_r and c_r will be the same for all experiments ($\mu_r = 1$ and $c_r = 0$)[1] which leads to a simplification of (8):

$$v = c/\sqrt{\epsilon_r} \tag{9}$$

Therefore, we are going to choose different relative permittivity ϵ_r besides R to build our dictionary. We propose two choices to find an interval of relative permittivity: either to have several relative permittivity ϵ_r depending of the scene illuminated by the RADAR or to compute an effective permittivity ϵ that takes into account a possible target of relative permittivity ϵ_i embedded in a material of relative permittivity ϵ_m. This latter is obtained by the Maxwell-Garnett model:

$$\epsilon = \epsilon_m \frac{2\delta_i(\epsilon_i - \epsilon_m) + \epsilon_i + 2\epsilon_m}{2\epsilon_m + \epsilon_i - \delta_i(\epsilon_i - \epsilon_m)} \tag{10}$$

where δ_i is the fraction of the inclusion of the target in the material.

To conclude, we resume the way to easily build a dictionary. First, we simply take some parameters of the RADAR like f_{max}, Δ_x and Δ_t. Finally, by varying ϵ (or simply ϵ_r), and R, we obtain a collection of hyperbolas giving us the dictionary $\{\mathbf{H}_k\}_{1,K}$. The interval of ϵ and R can be large but it is important to avoid having a dictionary that is too large, especially for computational complexity and memory issues. An example of four atoms are shown in Fig. (4). We notice that the convolution of $g(x)$ by the Ricker wavelet function $r(t)$ allows to obtain the triple polarity (-/+/-) which is a main characteristic of GPR images as shown in Fig. 1.

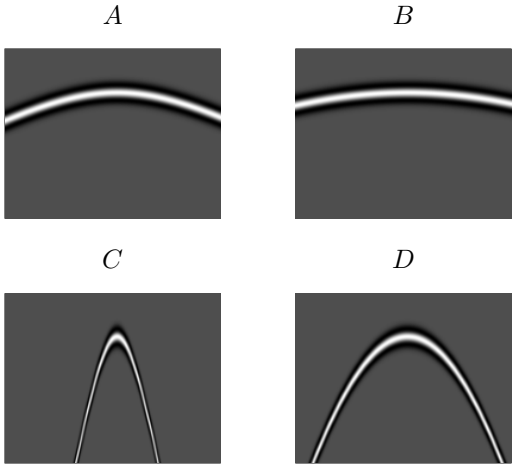


Fig. 4: Main elements of the physical dictionary proposed in the paper ($\omega_{max} = 2\pi f_{max} = 2\pi \cdot 350$MHz $\sim$ 2.2e9 rad.s^{-1}, $\Delta_x = 0.0101$ m and $\Delta_t = 0.105{\cdot}10^{-9}$s). A: $\epsilon_r = 1$, $R = 0.2$, B: $\epsilon_r = 1$, $R = 2$, C: $\epsilon_r = 100$, $R = 0.2$ and D: $\epsilon_r = 100$, $R = 2$

III. Classical Convolutive Inversion

In Section II we proposed a model of the GPR signal involving an useful signal, clutter elements and noise. The useful signal is described using a sparse coding of possible target signatures in dictionary related to physical parameters of the scene. We consider the problem of estimating the mapping $\{\mathbf{C}_k\}_{k=1,\ldots,K}$ and clutter matrix $\mathbf{L}$ allowing us to study the composition of a GPR image in terms of buried targets and layer composition.

[1]Because of the small conductivity to the target materials such as wood, silica and plastic

A. Posing the inversion problem

In order to invert the GPR model, the following optimization problem function is proposed in [35]:

$$(\{\mathbf{C}_k\},\mathbf{L}) = \underset{\{\mathbf{C}_k\},\mathbf{L}}{\operatorname{argmin}} \underbrace{\|\mathbf{Y} - \sum_{k=1}^{K} \mathbf{C}_k \circledast \mathbf{H}_k - \mathbf{L}\|_2^2}_{\text{data fidelity}} + \underbrace{\lambda \sum_{k=1}^{K} \|\mathbf{C}_k\|_1}_{\text{sparsity promotion}} + \underbrace{\kappa \|\mathbf{L}\|_*}_{\text{rank minimization}}, \tag{11}$$

where λ and κ are penalty parameters. This cost function is designed to take into account three effects:

- data fidelity term measures the discrepancy of the model to the measured image signal,
- a sparsity promoting term by using a L_1-norm on each coefficient map $\mathbf{C}_k$,
- a rank minimization term by using a nuclear-norm on the clutter matrix. While not strictly minimizing the rank, it is known to promote low-rank solutions while retaining convexity [41].

B. Solving using an ADMM formulation

In order to solve the sparse deconvolution problem (11) with additional variables, several approaches can be leveraged as discussed in [42]. The ADMM framework in particular is reportedly the most efficient approach compared to others such as FISTA [43, Sec. IV.B]. The idea behind the ADMM is to split the optimization problem into sub-problems involving only one unknown at a time [44], making it easier to handle problems having as many unknowns as in GPR inversion problem.

In order to apply this framework, we introduce auxiliary variables $\{\mathbf{S}_k\}_{k=1,\dots,K}$[2]. This permits the same formulation in the augmented Lagrangian of (13). In this case we discard κ since the interplay between λ and $\rho_{\mathbf{S}}$ parameters can be used to favor rank-minimization:

$$\begin{cases} \text{minimize} & \|\mathbf{L}\|_* + \lambda \sum_{k=1}^{K} \|\mathbf{S}_k\|_1 \\ s.t & \mathbf{Y} = \sum_{k=1}^{K} \mathbf{C}_k \circledast \mathbf{H}_k + \mathbf{L} \\ & \forall k,\, \mathbf{S}_k = \mathbf{C}_k \end{cases} \tag{12}$$

The augmented Lagrangian of this problem is as follows:

$$\begin{aligned} \mathcal{L}\left(\{\mathbf{C}_k\},\{\mathbf{S}_k\},\mathbf{L},\{\widetilde{\mathbf{S}}_k\},\widetilde{\mathbf{L}}\right) &= \frac{\rho_{\mathbf{S}}}{2}\|\sum_{k=1}^{K} \mathbf{C}_k - \mathbf{S}_k + \widetilde{\mathbf{S}}_k\|_2^2 + \\ \|\mathbf{L}\|_* + \frac{\rho_{\mathbf{L}}}{2}\|\mathbf{Y} - \sum_{k=1}^{K} \mathbf{C}_k \circledast \mathbf{H}_k - \mathbf{L} + \widetilde{\mathbf{L}}\|_2^2 + \lambda \sum_{k=1}^{K} \|\mathbf{S}_k\|_1 \\ - \frac{\rho_{\mathbf{S}}}{2}\|\sum_{k=1}^{K} \widetilde{\mathbf{S}}_k\|_2^2 - \frac{\rho_{\mathbf{L}}}{2}\|\widetilde{\mathbf{L}}\|_2^2, \end{aligned} \tag{13}$$

where $\{\widetilde{\mathbf{S}}_k\}_{k=1,\dots,K}$ and $\widetilde{\mathbf{L}}$ are the dual variables of $\{\mathbf{S}_k\}_{k=1,\dots,K}$ and $\mathbf{L}$, and, $\rho_{\mathbf{S}}$ and $\rho_{\mathbf{L}}$ are Lagrangian penalty parameters[3].

The ADMM framework solves separably each optimization problem for one variable (by assuming the remaining constants) and iterates until convergence. It is applied here by solving the augmented Lagrangian iteratively for variables $\{\mathbf{C}_k\}$, then $\{\mathbf{S}_k\}$ and $\mathbf{L}$ (by assuming the remaining ones constants). At each iteration, we also update the dual variables thanks to the residual of the equality constraints. The algorithm to solve (13) has been proposed in [35]. A summary is given in A with the notations used in this paper.

In next section we propose an improvement of this approach by replacing the L_2-norm by a robust norm to obtain better performance of denoising.

IV. Robust Convolutive Inversion

In previous section, we have solved the optimization problem (11), where the attach to the data is performed by exploiting a L_2-norm on the residual between the reconstructed image and the original one. This approach is efficient in the absence of outliers or when the original B-scan image is subjected to low-level of noise. But when some outliers, such as discontinuities in the radargram due to poor pre-processing, are present or when the measurement is very noisy this method performs poorly. Indeed, in those scenarios, the use of the L_2-norm will favor solutions that replicate those outliers and noise components. For such cases, it is possible to mitigate these effects by employing a robust cost function that finds a compromise between outlier rejection and closeness to the original problem. In this section, we propose to employ the well-known Huber norm [38] used in linear regression tasks. In particular, we consider the modified Huber function proposed in [45] to handle the case of complex-valued data in the Fourier domain when handling the sub-problem to find $\{\mathbf{C}_k\}$:

$$\mathcal{H}_\delta(x) = \begin{cases} |x|^2 & , \quad |x| \leq \delta \\ 2\delta|x| - \delta^2 & , \quad |x| > \delta \end{cases}, \tag{14}$$

where δ is a parameter allowing to perform a compromise between a cost close to the L_2-norm (high value of δ) and a cost reducing the effect of outliers (lower value of δ). In practice, it can be estimated using the empirical distribution of the image pixel values and selecting a quantile. Based on this function, we define the Huber norm of a vector $\mathbf{x} \in \mathbb{C}^p$:

$$\|\mathbf{x}\|_{\mathcal{H}}^2 = \sum_{i=1}^{p} \mathcal{H}_\delta(x_i). \tag{15}$$

The derivative function of (14) will also be needed and is equal to:

$$\Psi_\delta(x) = \begin{cases} x & , \quad |x| \leq \delta \\ \delta\,\mathrm{sign}(x) & , \quad |x| > \delta \end{cases}, \tag{16}$$

where $\mathrm{sign}(x)$ is equal to 0 when $x = 0$ and equal to $x/|x|$ otherwise.

[2]Note that the term attached to the data has been replaced by the additional constraint

[3]See [44, chap. 3] for details on the algorithm and definition of the terms used.

Using Huber's norm we have the following optimization problem:

$$\underset{\{\mathbf{C}_k\},\mathbf{L}}{\operatorname{argmin}} \|\mathbf{Y} - \sum_{k=1}^{K} \mathbf{C}_k \circledast \mathbf{H}_k - \mathbf{L}\|_{\mathcal{H}}^2 + \lambda \sum_{k=1}^{K} \|\mathbf{C}_k\|_1 + \kappa \|\mathbf{L}\|_*, \tag{17}$$

that we rewrite into the form given below in order to exploit the ADMM framework:

$$\begin{cases} \text{minimize} & \|\mathbf{Y} - \sum_{k=1}^{K} \mathbf{C}_k \circledast \mathbf{H}_k - \mathbf{L}\|_{\mathcal{H}}^2 \\ & + \kappa \|\mathbf{M}\|_* + \lambda \sum_k^K \|\mathbf{S}_k\|_1 \\ \text{subject to} & \mathbf{M} = \mathbf{L} \\ & \forall k,\, \mathbf{S}_k = \mathbf{C}_k \end{cases}, \tag{18}$$

where $\mathbf{M}$ is an auxiliary variable introduced to decouple the minimization of the rank of the clutter from the minimization of the Huber-norm. This implies that a new step for the variable $\mathbf{M}$ needs to be introduced to the ADMM iterations. The augmented Lagrangian of this new problem is as follows:

$$\begin{aligned} \mathcal{L}\left(\{\mathbf{C}_k\}, \{\mathbf{S}_k\}, \mathbf{L}, \mathbf{M}, \{\widetilde{\mathbf{S}}_k\}, \widetilde{\mathbf{L}}\right) &= \kappa \|\mathbf{M}\|_* + \lambda \sum_{k=1}^{K} \|\mathbf{S}_k\|_1 \\ &+ \|\mathbf{Y} - \sum_{k=1}^{K} \mathbf{C}_k \circledast \mathbf{H}_k - \mathbf{L}\|_{\mathcal{H}}^2 + \frac{\rho_{\mathbf{S}}}{2} \|\sum_{k=1}^{K} \mathbf{C}_k - \mathbf{S}_k + \widetilde{\mathbf{S}}_k\|_2^2 \\ &- \frac{\rho_{\mathbf{S}}}{2} \|\sum_{k=1}^{K} \widetilde{\mathbf{S}}_k\|_2^2 + \frac{\rho_{\mathbf{L}}}{2} \|\mathbf{M} - \mathbf{L} + \widetilde{\mathbf{L}}\|_2^2 - \frac{\rho_{\mathbf{L}}}{2} \|\widetilde{\mathbf{L}}\|_2^2. \end{aligned} \tag{19}$$

We describe below the derivation of the updates for each variable. Main contributions for this algorithm are in the updates of $\{\mathbf{C}_k\}$ and $\mathbf{L}$. A value of a variable at a given iteration i is denoted $\bullet^{(i)}$. A summary of the algorithm is presented in Algorithm 1.

Algorithm 1: Robust Inversion of GPR model with ADMM

Data: Image $\mathbf{Y} \in \mathbb{R}^{N_t, N_x}$, dictionary $\{\mathbf{H}_k \in \mathbb{R}^{N_t, N_x}\}_{k=1,\dots,K}$
Parameters: Sparsity parameter λ, Low-rank parameter κ, lagrangian penalties $\rho_{\mathbf{S}}$, $\rho_{\mathbf{L}}$, stopping criterion ξ
Result: Coefficients maps $\{\mathbf{C}_k \in \mathbb{R}^{N_t, N_x}\}_{k=1,\dots,K}$, clutter matrix $\mathbf{L} \in \mathbb{R}^{N_t, N_x}$
Initialize $\{\mathbf{C}_k^{(0)}\}_{k=1,\dots,K}$, $\{\mathbf{S}_k^{(0)}\}_{k=1,\dots,K}$, $\mathbf{L}^{(0)}$, $\mathbf{M}^{(0)}$, $\{\widetilde{\mathbf{S}}_k^{(0)}\}_{k=1,\dots,K}$ and $\widetilde{\mathbf{L}}^{(0)}$ with zeros
Precompute Fourier elements $\mathbf{y}$, $\{\mathbf{h}_k\}$ as in IV-A
$i \leftarrow 0$
while $\eta^{(i)} > \xi$ *as in* (30) **do**
  Compute $\mathbf{c}_k^{(i)}$, $\mathbf{s}_k^{(i)}$, $\widetilde{\mathbf{s}}_k^{(i)}$ and $\widetilde{\mathbf{l}}^{(i)}$ as in IV-A
  Update $\{\mathbf{C}_k^{(i+1)}\}$ with (22), $\{\mathbf{S}_k^{(i+1)}\}$ with (27), $\mathbf{M}^{(i+1)}$ with (29) and $\mathbf{L}^{(i+1)}$ with (26)
  $\forall k,\, \widetilde{\mathbf{S}}_k^{(i+1)} = \widetilde{\mathbf{S}}_k^{(i)} + \mathbf{S}_k^{(i+1)} - \mathbf{C}_k^{(i+1)}$
  $\widetilde{\mathbf{L}}^{(i+1)} = \widetilde{\mathbf{L}}^{(i)} + \sum_{k=1}^{K} \mathbf{C}_k^{(i+1)} \circledast \mathbf{H}_k + \mathbf{L}^{(i+1)} - \mathbf{Y}$
  $i \leftarrow i + 1$

A. Update $\{\mathbf{C}_k\}$

As this problem involves a heavy convolution, we can transpose this problem to the frequency domain with a Fast Fourier Transfrom (FFT) as in [43]. We vectorize the resulting matrices. For matrices $\mathbf{Y}$, $\mathbf{H}_k$, $\mathbf{C}_k$, $\mathbf{S}_k$ and $\widetilde{\mathbf{S}}$, of size $N_t \times N_x$, we define vectors $\mathbf{y}$, $\mathbf{h}_k$, $\mathbf{c}_k$, $\mathbf{s}_k$ and $\widetilde{\mathbf{s}}_k$ of size $N = N_t N_x$ by applying the operation $\operatorname{vec}(\mathcal{F}(\bullet))$. Note that these vectors are now complex-valued due to the FFT operation.

We also define the following variables $\mathbf{z}_k = \mathbf{s}_k + \widetilde{\mathbf{s}}_k$ and $\mathbf{x} = \mathbf{y} - \mathbf{l}$, and rewrite minimization of (19) w.r.t $\{\mathbf{C}_k\}$ into:

$$\begin{aligned} \{\mathbf{c}_k\}^{(i+1)} = \underset{\{\mathbf{c}_k\}}{\operatorname{argmin}} &\|\sum_{k=1}^{K} \mathbf{h}_k \odot \mathbf{c}_k - \mathbf{x}\|_{\mathcal{H}}^2 \\ &+ \frac{\rho_{\mathbf{S}}}{2} \|\sum_{k=1}^{K} \mathbf{c}_k - \mathbf{z}_k\|_2^2, \end{aligned}$$

where $\odot$ is the Hadamard product and $\| \bullet \|_2$ is the L_2-norm on complex vectors. Replacing the Hadamard product with a matrix product by concatenating the K different vectors to form the new KN vectors $\mathbf{c} = [\mathbf{c}_1^T, \dots, \mathbf{c}_K^T]^{\mathrm{T}}$ and $\mathbf{z} = [\mathbf{z}_1^T, \dots, \mathbf{z}_K^T]^{\mathrm{T}}$ and defining a $N \times KN$ matrix

$$\mathbf{H} = [\operatorname{diag}(\mathbf{h}_1), \dots, \operatorname{diag}(\mathbf{h}_K)],$$

so that the optimization problem becomes:

$$\mathbf{c}^{(i+1)} = \underset{\mathbf{c}}{\operatorname{argmin}} \|\mathbf{H}\mathbf{c} - \mathbf{x}\|_{\mathcal{H}}^2 + \frac{\rho_{\mathbf{S}}}{2} \|\mathbf{c} - \mathbf{z}\|_2^2, \tag{20}$$

We have to compute the gradient of this equation w.r.t. $\mathbf{c}$. For the second term, we have the same result as in the classical inversion (see [35] or A), but for the first term, the derivative is given by $\mathbf{H}^{\dagger} \Psi_{\delta}(\mathbf{H}\mathbf{c} - \mathbf{x})$ where the function Ψ_{δ} is applied element-wise. We introduce the new vector $\mathbf{r}_{\delta} = [\Psi_{\delta}(\mathbf{H}\mathbf{c} - \mathbf{x})]$. By using the structure of matrix $\mathbf{H}$ and using linear algebra identities, the total gradient is given by:

$$\nabla_{\mathbf{c}} \mathcal{L} = \mathbf{h}^{\dagger} \odot (\mathbf{1}_K \otimes \mathbf{r}_{\delta}) + \rho_{\mathbf{S}}(\mathbf{c} - \mathbf{z}), \tag{21}$$

where $\mathbf{h} = [\mathbf{h}_1^T, \dots, \mathbf{h}_K^T]^{\mathrm{T}}$. This equation does not admit an analytical solution and we use a simple gradient descent approach method to estimate $\mathbf{c}^{(i+1)}$:

$$\begin{aligned} \mathbf{c}_0^{(i+1)} &= \mathbf{c}^{(i)}, \\ \mathbf{c}_{j+1}^{(i+1)} &= \mathbf{c}_j^{(i+1)} - \frac{\zeta_{\mathbf{c}}}{j} \nabla_{\mathbf{c}_j^{(i+1)}} \mathcal{L}, \end{aligned} \tag{22}$$

where $\zeta_{\mathbf{c}}$ is the initial step-size of the gradient. While a stopping criterion can be used, experiments show that a satisfactory solution can be obtained with a few iterations of the gradient descent.

Finally, $\{\mathbf{C}_k^{(i+1)}\}$ is obtained by splitting $\mathbf{c}$ into the set of vectors $\{\mathbf{c}_k^{(i+1)}\}$ and by applying the operation $\mathcal{F}^{-1}(\operatorname{unvec}(\bullet))$.

B. Update $\mathbf{L}$

In this case, the optimization problem becomes:

$$\mathbf{L}^{(i+1)} = \underset{\mathbf{L}}{\operatorname{argmin}} \|\mathbf{Y} - \sum_{k=1}^{K} \mathbf{C}_k \circledast \mathbf{H}_k - \mathbf{L}\|_{\mathcal{H}}^2 + \frac{\rho_{\mathbf{L}}}{2} \|\mathbf{M} - \mathbf{L} + \widetilde{\mathbf{L}}\|_2^2 \tag{23}$$

This problem is simplified by using the new variables $\mathbf{A} = \sum_{k=1}^{K} \mathbf{C}_k \circledast \mathbf{H}_k - \mathbf{Y}$ and $\mathbf{V} = \mathbf{M} + \widetilde{\mathbf{L}}$:

$$\mathbf{L}^{(i+1)} = \underset{\mathbf{L}}{\operatorname{argmin}} \|\mathbf{A} + \mathbf{L}\|_{\mathcal{H}}^2 + \frac{\rho_{\mathbf{L}}}{2}\|\mathbf{V} - \mathbf{L}\|_2^2. \tag{24}$$

For this optimization problem we can use the Huber proximal [46] defined by:

$$\mathcal{PH}_{\gamma,\delta}(x) = \begin{cases} \frac{x}{\gamma+1} & \text{if } |x| < \delta(\gamma+1) \\ x - \delta\gamma\,\text{sign}(x) & \text{otherwise} \end{cases}. \tag{25}$$

Then using proximal identities [47] we have the update:

$$\mathbf{L}^{(i+1)} = -\mathbf{A} + \mathcal{PH}_{\rho_{\mathbf{L}}^{-1},\delta}(\mathbf{V} + \mathbf{A}), \tag{26}$$

where the proximal is applied element-wise.

C. Updates $\{\mathbf{S}_k\}$ and $\mathbf{M}$

In this case, we need to solve:

$$\{\mathbf{S}_k\}^{(i+1)} = \underset{\{\mathbf{S}_k\}}{\operatorname{argmin}}\, \lambda \sum_{k=1}^{K} \|\mathbf{S}_k\|_1 + \frac{\rho_{\mathbf{S}}}{2}\|\sum_{k=1}^{K} \mathbf{S}_k - \mathbf{C}_k + \widetilde{\mathbf{S}}_k\|_2^2.$$

This problem is well known and the solution is given by [44]:

$$\mathbf{S}_k^{(i+1)} = \mathcal{S}_{\lambda/\rho_{\mathbf{S}}}(\mathbf{C}_k - \widetilde{\mathbf{S}}_k) \tag{27}$$

where $\mathcal{S}_{\lambda/\rho_{\mathbf{S}}}$ is the soft thresholding operator defined for a real-valued vector $\mathbf{u}$ and scalar value γ by:

$$\mathcal{S}_{\gamma}(\mathbf{u}) = \text{sign}(\mathbf{u}) \odot \max(0, |\mathbf{u}| - \gamma), \tag{28}$$

where the sign and max operations are applied element-wise.

For $\mathbf{M}$, the derivation is also very similar and leads to:

$$\mathbf{M}^{(i+1)} = \mathcal{T}_{\kappa/\rho_{\mathbf{L}}}(\mathbf{L} - \widetilde{\mathbf{L}}). \tag{29}$$

D. Algorithm tuning

Convergence and stopping criterion: The optimization problem is convex over all the variables and constraints and the convergence is ensured by the ADMM framework. Nevertheless, this convergence can be slow and can be improved by employing an over-relaxation technique with a relaxation term [44], [48] $\alpha > 1$ in the update $\mathbf{C}_k$.

To stop the algorithm, we use a distance between two iterations of the ADMM over their reconstructed image thanks to $\mathbf{L}$ and $\{\mathbf{C}_k\}_{1,K}$:

$$\eta^{(i+1)} = \|\mathbf{L}^{(i+1)} - \mathbf{L}^{(i)} + \sum_{k=1}^{K} \mathbf{C}_k^{(i+1)} \circledast \mathbf{H}_k^{(i+1)} - \mathbf{C}_k^{(i)} \circledast \mathbf{H}_k^{(i)}\|_2. \tag{30}$$

Choice of ρ_S, ρ_L, κ, $\zeta_{\mathbf{c}}$ and λ: The values of $\rho_{\mathbf{S}}$ and $\rho_{\mathbf{L}}$ are obtained by design of experiments techniques (see section V). These parameters can remain fixed throughout the iterations of the algorithm. It also possible to update them at each iteration by using the primal and dual residuals to improve convergence rate as discussed in [37], [44]. The low rank parameter κ is fixed to 1 because the associated Lagrangian $\rho_{\mathbf{L}}$ allows to control the low rank character. The gradient step $\zeta_{\mathbf{c}}$ is found by a cross validation approach. Actually, all these parameters are mainly chosen to ensure the convergence of the ADMM and to do not need to be next changed even if the dataset is different.

The more important parameter is obviously the sparsity parameter λ as it controls the sparsity of the solution. It can be chosen as a compromise between high-response hyperbolas with associated low-noise and closeness to the original image. In practice, a cross-validation approach on a well-known scene can be used to determine it, or it can be adjusted with regards to geology expertise. In section V, we propose a cross-validation approach but we also show that the performance of the proposed algorithm remains quasi constant over an interval of λ.

Solution without $\mathbf{L}$: If the layers responses are removed by a pre-processing then one may omit the low-rank matrix $\mathbf{L}$ from the model. In this case, we can discard the corresponding update and the dual variable $\widetilde{\mathbf{L}}$ from the equations. Moreover in the update of $\mathbf{C}_k$ in Section IV-A, we only have $\mathbf{x} = \mathbf{y}$.

Computational complexity of the algorithm: The updates of the variables $\{\mathbf{S}_k\}$, $\{\tilde{\mathbf{S}}_k\}$ and $\tilde{\mathbf{L}}$ are relatively light since they are mainly based on thresholding methods and simple matrix operations (Hadamard product, addition,...). The update of $\{\mathbf{C}_k\}$ is based on the computation of the gradient (21) which is based on a Hadamard product which is also light. This step is a little bit more complex since it needs a certain number of iterations. But this number can be chosen small (3 or 5 are enough to ensure the convergence of the ADMM) which limits the complexity cost of this step. The update of $\mathbf{L}$ in (26) is also light since the proximal is just a thresholding applied element-wise. Finally, the main complexity cost comes from the update of $\mathbf{M}$ in (29) which needs a SVD of a matrix of the size of the image. To have a better idea, we give an overview of the computation time. For the computation time, we have for 100 iterations with 3 iterations of gradient descent and a image of $(N_x, N_y) = (233, 128)$: $28.1s \pm 1.39s$. Compared to $24.8s \pm 1.35s$ of the classical inversion algorithm of [35], the increase in computation time is reasonable. This study has been done with a laptop equipped with an 8-core i5-1135G7 @ 2.40GHz processor and 16Gb of RAM.

Another issue to investigate is the RAM memory needed for the updates. For this topic, the main issue comes from the gradient computation in (21) since it involves the matrix product with the matrix $\mathbf{H}$ which is the resultant of the diagonalization of the atoms of the dictionary (i.e. for k atoms of $N \times N$ pixels this gives us a matrix of size $(k \times N^2, N^2)$). For example, if $k = 40$ and $N = 128$ in float32 it takes almost 37.3Go for a (655360×16384) matrix. However as only the values of the diagonals of the sub-blocks of the matrix are non null, the matrix calculation $\mathbf{H}^{\dagger}\mathbf{H}$ becomes simply $\{\mathbf{h}_k\}_{k=1,\dots,K} \odot \{\mathbf{h}_k\}_{k=1,\dots,K}$ and gives us a matrix of size $(k \times N^2)$. In a similar way, with $k = 40$ and $N = 128$ in float32 it takes only 2.36Mo.

V. Numerical experiments

In this section we first describe experiments done on B-scans acquired in the field to showcase the benefits of the proposed convolutive inversion, as well as the use of the Huber norm. We also analyze the effect of hyperparameters tuning on the inversion performance.

Parameter	IRADAR__00H	IRADAR_032	IRADAR_005	IRADAR_832
Antenna frequency	350 MHz	200MHz	200MHz	350MHz
Elevation	0cm	50cm	0cm	150cm
Objects	Wooden shelter, Concrete, Guns, Metal fut, Void	Wooden shelter, Concrete, Guns, Metal Fut, Void	Wooden shelter, Concrete, Guns	Wooden shelter, Metallic mine, Plastic mine

TABLE II: Details of the images used in the study

In the following, the classical algorithm developed in [35] and presented in the section III will be referred to as L_2-GPR. Main details of the updates and the algorithm are given in appendix A. The proposed algorithm described in the section IV with Huber norm will be referred to as Hub-GPR. As a comparison between different methods, we compute Receiver Operator Curve (ROC) plots. In addition to the algorithms proposed in this paper, we have tested a simple SVD (to remove the spatially coherent noise) and a simple L_2 recovery (without the matrix L) combined with the SVD. Both methods will be referred by SVD-GPR and L_2-SVD-GPR.

The code is available online: https://github.com/Matthieu-Gallet/GPR-robust-inversion. The implementation of the algorithms was done under Python 3.6. We made use of the packages pyDOE[4] for the design of experiments in Section V-B and GPRPy[5] for preprocessing the GPR images.

A. Dataset description and Dictionary parameters

The full dataset provided by Geolithe is composed of 1000 radargrams of $(N_x, N_y) = (4000, 800)$ pixels. These radargrams were acquired from a prepared site, in which various objects were buried at known locations and depths. The objects vary in size (10 cm to 2 m) and material (wood, plastic, metal). The test site includes two 46m x 7m trenches, one composed of sand and the other of a mixture of sand, gravel and earth, including stones up to 10cm in diameter. GPR acquisitions were performed with two GSSI hyperstacking antennas. In addition to the GPR acquisitions performed by sliding the antennas along the ground(a typical utilization, with the highest SNR), acquisitions were also performed with the antennas held a fixed distances offset from the surface, up to 150 cm, in order to simulate a GPR acquisition by drone. The radargrams were annotated by geophysicists at Geolithe by tracing the signature of the object in the radargram. This pixel-wise annotation produces a labeled mask, in which all image pixels all labeled by class (e. g. object A, object B, no object, etc.) An example of a radargram with an annotation mask is shown in Fig. 6a. Although the extent of an object signature in the radargram is somewhat subjective (e. g. the bottom boundaries of the hyperboles signatures typically fade into the background noise of the radargram), the known position, type, and depth of the object allow us to use these annotations as a reliable ground truth. Many other characteristics are given for each image as illustrated in Table II. We build our dictionary using the following parameters provided by the GSII RADAR:

- $\omega_{max} = 2\pi f_{max}$ where f_{max} is either 200 or 300 MHz
- $\Delta_x = 0.0101$ m and $\Delta_t = 0.105{\cdot}10^{-9}$s

[4]https://pythonhosted.org/pyDOE/
[5]https://github.com/NSGeophysics/GPRPy

We finally built a representative dictionary of $K = 30$ atoms by using for values of $R \in \{0.01, 0.1, 1\}$ and $\epsilon \in \{5, 6.46, 8.34, 10.77, 13.91, 17.97, 23.21, 29.97, 38.71, 50\}$.

B. Study of hyperparameters $\rho_{\mathbf{S}}$ and $\rho_{\mathbf{L}}$

In order to check the effects of hyperparameters, we perform an experimental cross-validation study. We first used the inversion of the L_2-GPR algorithm (the behavior is the same for Hub-GPR algorithm). In particular, we want to check the efficiency of the convolutional approach and the fit provided by our dictionary and verify if the choice of $\rho_{\mathbf{S}}$ and $\rho_{\mathbf{L}}$ is critical. We selected a representative radargram for this study. The original radargram is shown in Fig. 6a, and the mask image showing the true hyperbola is given in Fig. 6b. The acquisition of this radargram was made with an elevation of 75cm. We propose to study the robustness of our approach w.r.t. these hyperparameters $(\rho_{\mathbf{S}}, \rho_{\mathbf{L}})$. To do this, we use the Latin hypercubes approach [49], which allows us to carry out a design of experiment by judiciously sampling the parameter space. We transformed the uniform distribution of the Latin hypercubes samples to a log-uniform distribution in order to cover the large scale of both $\rho_{\mathbf{S}}, \rho_{\mathbf{L}}$ (between 1 to 10^5). This approach allows minimizing the number of simulations by reducing the interactions between the parameters. The resulting image given in Fig. 5 allows us to conclude that to avoid a large error, it is sufficient to choose a value of $(\rho_{\mathbf{S}}, \rho_{\mathbf{L}})$ smaller than 10^3. To conclude, a narrow choice for $(\rho_{\mathbf{S}}, \rho_{\mathbf{L}})$ is not needed to reach correct performance.

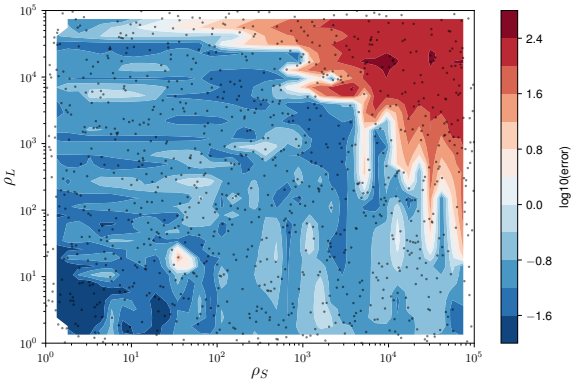
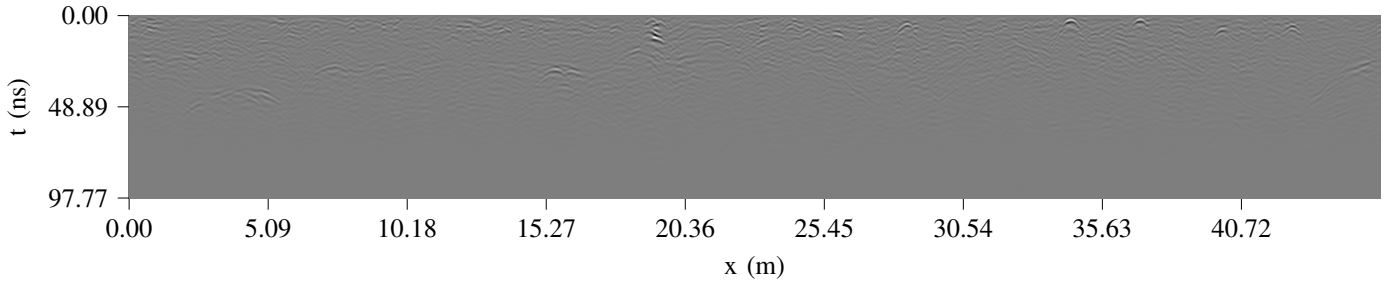


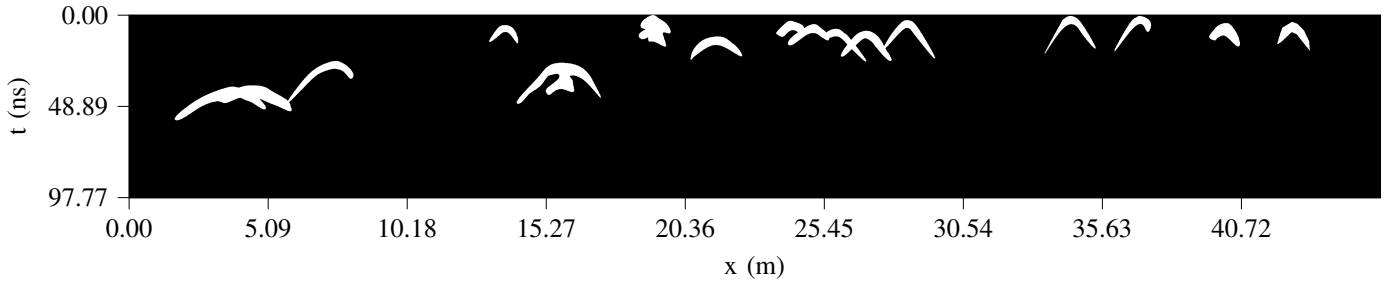
Fig. 5: MSE w.r.t. the hyperparameters $\rho_{\mathbf{S}}$ and $\rho_{\mathbf{L}}$. The dots display the sampling points used.

C. Performance of Hub-GPR algorithm

In this section, we investigate the potential benefits of using a Huber norm instead of the L_2-norm in the optimization



(a) Original radargram



(b) masks of expertised hyperbola by geologists

Fig. 6: Original radargram IRADAR_00H (6a) with the hyperbolas obtained by true objects labeled by Geolithe (6b). The red boxes show the locations of the labeled hyperbolas.

problem. In order to have a better conclusion, we also tests different methods which are summarized in the following:

- Original image,
- SVD-GPR algorithm in which the first rank contribution is removed,
- L_2-SVD-GPR algorithm where the sparse recovery is applied on the image preprocessed by the SVD (as before the first rank mode is removed),
- L_2-GPR algorithm of section III,
- Hub-GPR algorithm of section IV.

Figures 7a, 7b, 7c and 7d respectively show the original radargram along with the results of the SVD-GPR, L_2-SVD-GPR, L_2-GPR and Hub-GPR algorithms. The parameter values are $\rho_{\mathbf{S}} = 500$ and $\rho_{\mathbf{L}} = 250$. The threshold for the Huber norm δ in (14) is set to 750. We choose the number of iterations of the gradient descent in (22) equal to 5, and with a step size $\zeta_{\mathbf{c}}$ equal to $5 \cdot 10^{-4}$. The number of iterations of the ADMM is 100. The sparsity parameter λ could vary between 0.2 and 0.8. The masks in red allow to clearly identify the objects of interest. Compared to the original radargram in 6a, the result of all algorithms allows better detection of hyperbolas. The noise is also strongly reduced. This reduction is more visible with the Hub-GPR algorithm. Moreover, it seems that several hyperbolas are better detected in the image obtained by the latter.

For a quantitative study, we calculate ROC plots (true positive rate versus false positive rate parameterized by the detection threshold)[6] on the outputs of the considered methods. The energy is next computed on each pixel for all resulting images before performing the detection step. For fairness, we have tested several values for λ for the three last methods and chosen the best.

The results for ROC plots are shown in Fig. 8. In all radargrams, we verify that the Hub-GPR algorithm gives the best result. The method seems to work well whatever the elevation (and thus the SNR) and the frequency of the RADAR. We also notice that the Hub-GPR algorithm improves the detection whatever the material of the object to be detected (wood, metal, plastic, ...).

D. Robustness to the sparsity parameter

In this section, we propose to study the robustness of the Hub-GPR algorithm to the choice of the sparsity parameter λ. For each image IRADAR, we compute the ROC plots for $\lambda = 0.2, 0.4, 0.6, 0.8$. The results are shown in Fig. 9. We first notice that the results keep close for all values of λ. Actually, the best value is around 0.4 and the performance decrease for 0.8.

E. Robustness to the dictionary

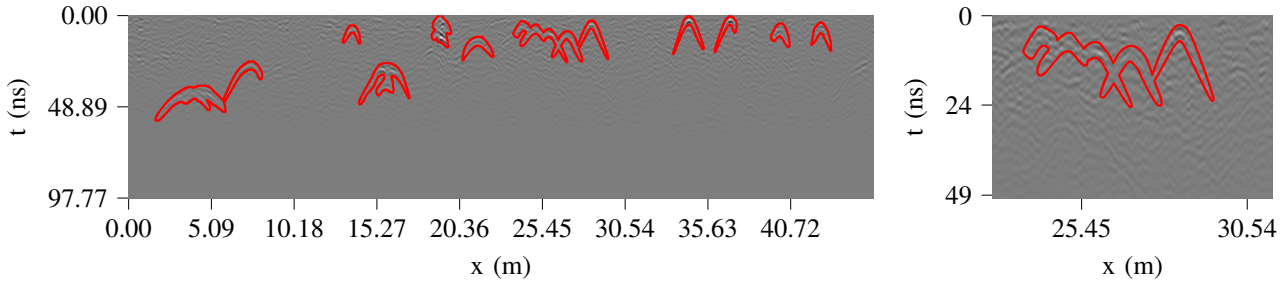
To do that, we made a simulation with different sizes of dictionaries:

- small: 18 atoms
- medium: 36 atoms
- big: 65 atoms

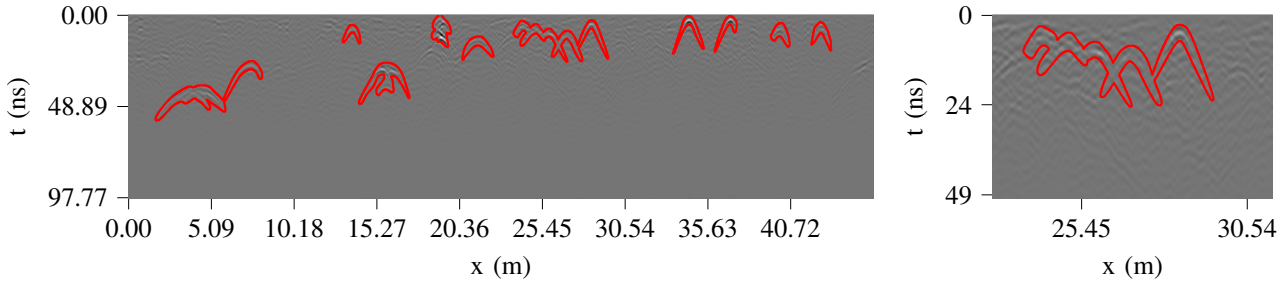
We experiment on a crop[7] of the same image of the subsection V.C. The cropped image corresponds to the left part of the image split in two.

[6]ROC are computed by using the scikit-learn method roc_auc_score with the masks and the image obtained by the studied algorithm.

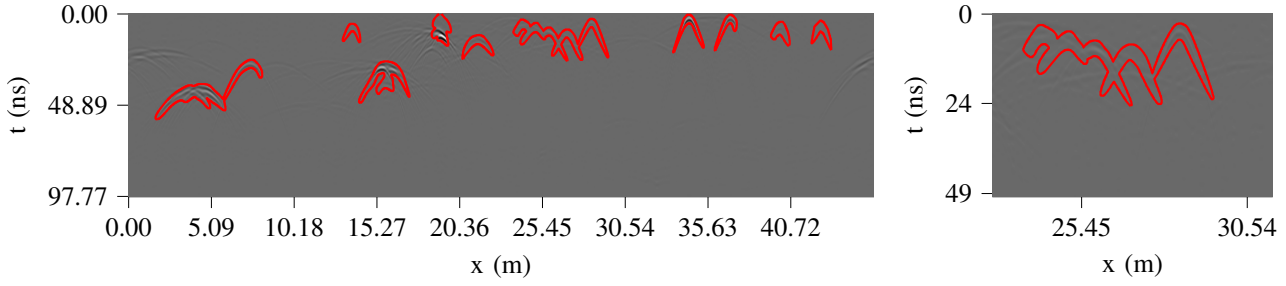
[7]To save time and test many sparsity options.



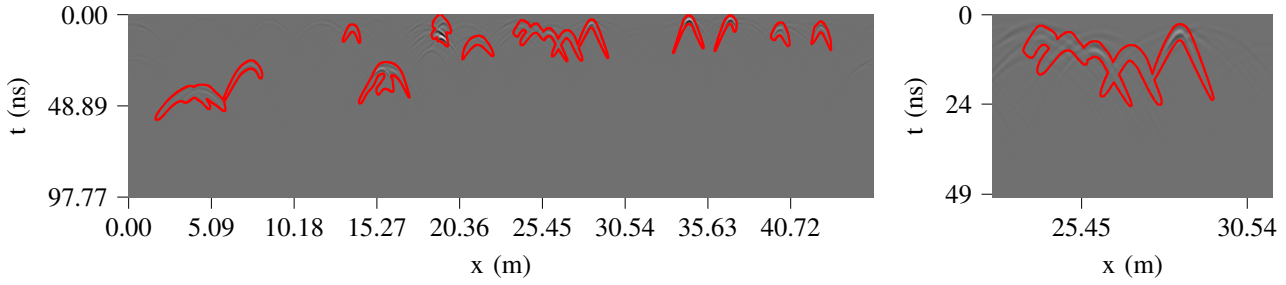
(a) Denoising with SVD-GPR



(b) Denoising with L_2-SVD-GPR



(c) Denoising with L_2-GPR



(d) Denoising with Hub-GPR

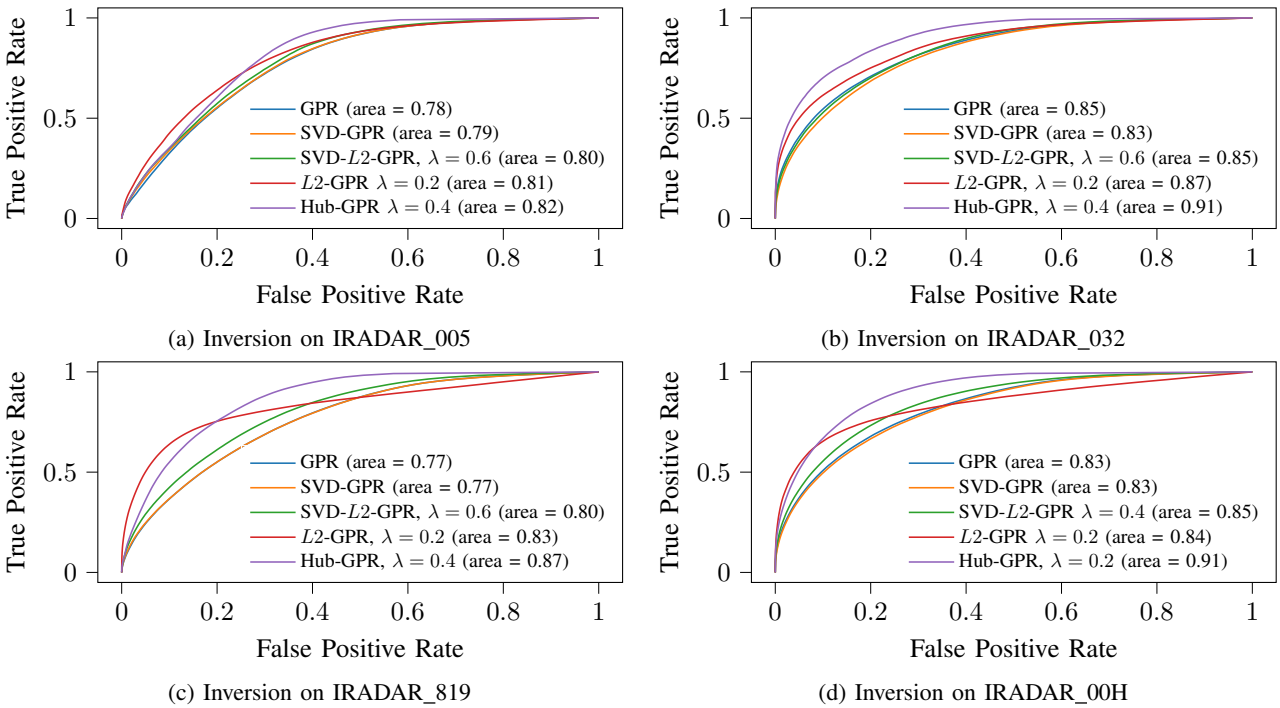
Fig. 7: Results of the inversion algorithms for the IRADAR_00H radargram. $\rho_{\mathbf{S}} = 500$ and $\rho_{\mathbf{L}} = 250$. The threshold for the Huber norm δ is 750. The number of iterations of the gradient descent is 5, and with a step size $\zeta_{\mathbf{c}}$ equal to $5 \cdot 10^{-4}$. The number of iterations of the ADMM is 500.

The results shown in Fig. 10 showcase the effectiveness of the approach even with lower or bigger (when we are not sure about the objects present in the scene) number of atoms.

F. Robustness to noise of Hub-GPR algorithm

In this section, we measure the robustness of both approaches to noise. We will consider the original radargram acquired with GSSI software. The original as well as the labels mask, are shown in Fig. 11. For the rest of the experiments, we set $\rho_{\mathbf{S}}$ to 500 and $\rho_{\mathbf{L}}$ to 250. The threshold for Huber norm δ in (14) is fixed to 750. We choose the number of iterations of the gradient descent in (22) equal to 5, and step size $\zeta_{\mathbf{c}}$ equal to $5 \cdot 10^{-4}$. The total number of iterations of the algorithm is set to 20. This saves computation time, and experiments show that the solution is not so different with this reduced number of iterations. The sparsity parameter is equal to 0.4 as it is a value that gives good performance for the L_2-GPR and Hub-GPR algorithms.

We consider two types of noise: one additive and one multiplicative. For both noises, we take a Gaussian noise N_g



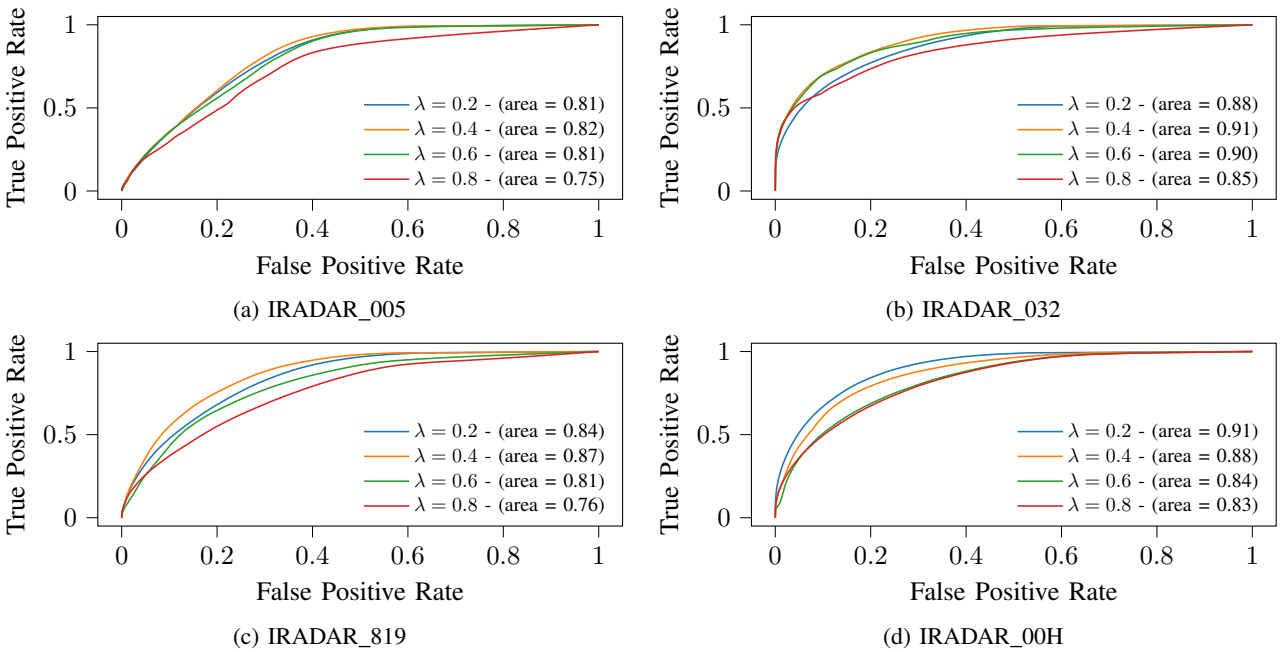
(a) Inversion on IRADAR_005

(b) Inversion on IRADAR_032

(c) Inversion on IRADAR_819

(d) Inversion on IRADAR_00H

Fig. 8: ROC plots for SVD-GPR, L_2-SVD-GPR, L_2-GPR and Hub-GPR algorithms on IRADAR_005, IRADAR_032, IRADAR_819 and IRADAR_00H radargrams. For each method, λ is chosen to give the best area. $\rho_{\mathbf{S}} = 500$ and $\rho_{\mathbf{L}} = 250$. The threshold for the Huber norm δ is 750. The number of iterations of the gradient descent is 5, and with a step size $\zeta_{\mathbf{c}}$ equal to $5 \cdot 10^{-4}$. The number of iterations of the ADMM is 500.



(a) IRADAR_005

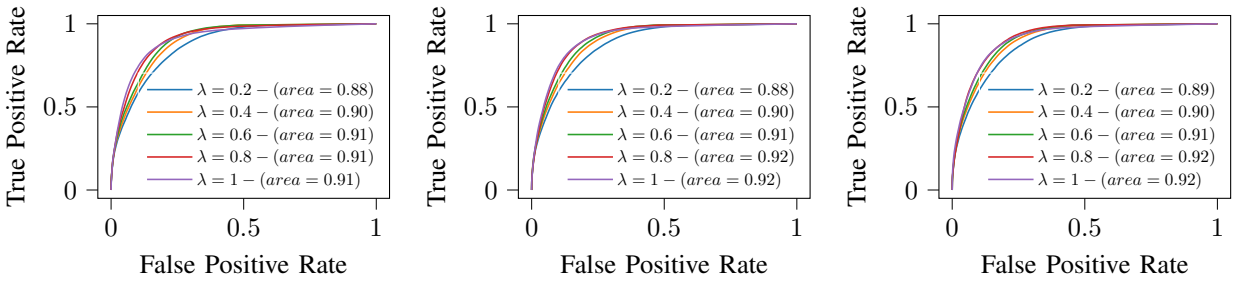
(b) IRADAR_032

(c) IRADAR_819

(d) IRADAR_00H

Fig. 9: ROC plots for Hub-GPR algorithm on IRADAR_005, IRADAR_032, IRADAR_819 and IRADAR_00H radargrams. For each radargram, several λ are tested. $\rho_{\mathbf{S}} = 500$ and $\rho_{\mathbf{L}} = 250$. The threshold for the Huber norm δ is 750. The number of iterations of the gradient descent is 5, and with a step size $\zeta_{\mathbf{c}}$ equal to $5 \cdot 10^{-4}$. The number of iterations of the ADMM is 500.

of zero mean and variance between 0.01 and 10. For additive noise the noisy image is given by: $I_{noisy} = I + N_g$, and for multiplicative noise: $I_{noisy} = I + I * N_g$. The quantitative study is done on both the original un-processed and the pre-processed images. We compute several metrics: the classical MSE (Mean Square Error), the PSNR (Peak Signal to Noise



(a) Large dictionary (65 atoms) (b) Medium dictionary (36 atoms) (c) Small dictionary (18 atoms)

Fig. 10: Results of Inversion with Huber norm on a crop of image 00H

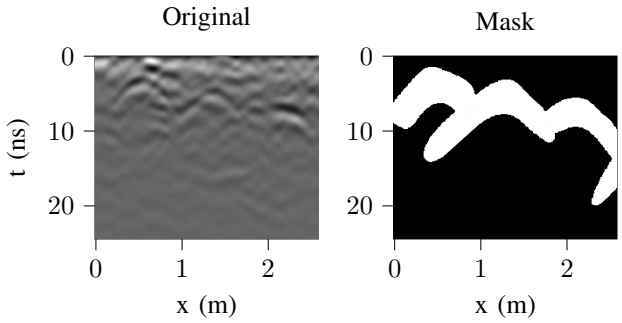


Fig. 11: Original radargram (left) and image mask (right) localizing the labeled hyperbolas.

Ratio) and the SSIM (Structural Similarity Index) [50]. These metrics with respect to the variance of the noise are given in Figs 12 and 13. For all cases, the robust method gives better performance for all metrics, and shows its resilience to noise whether it is additive or multiplicative.

VI. Conclusion

We developed an inversion algorithm adapted to GPR images by using a convolutional model with a dictionary that is directly linked with to the physics of the GPR. We then generalized the approach by replacing the $L2$-norm in the problem of optimization by a Huber norm in order to cope with heavy-tailed noise, outliers and/or artifacts (misalignment, ...) often encountered in GPR images. Results of the new robust GPR inversion method showed the advantage of this new algorithm. Furthermore, our studies showed that the method is not overly sensitive to the choice of hyperparameters. All of the experiments were conducted on a real dataset provided by Geolithe, and labeled by geophysicists, which showcased the real world applicability.

Future works will focus on the construction of a method that can avoid gradient descent step and thus save computation times. Majoration-Minization framework [51] will be investigated to solve this issue. Moreover, this convolutional approach will be extended to dictionary-learning approaches [42] in order to be able to learn the dictionary directly from the data.

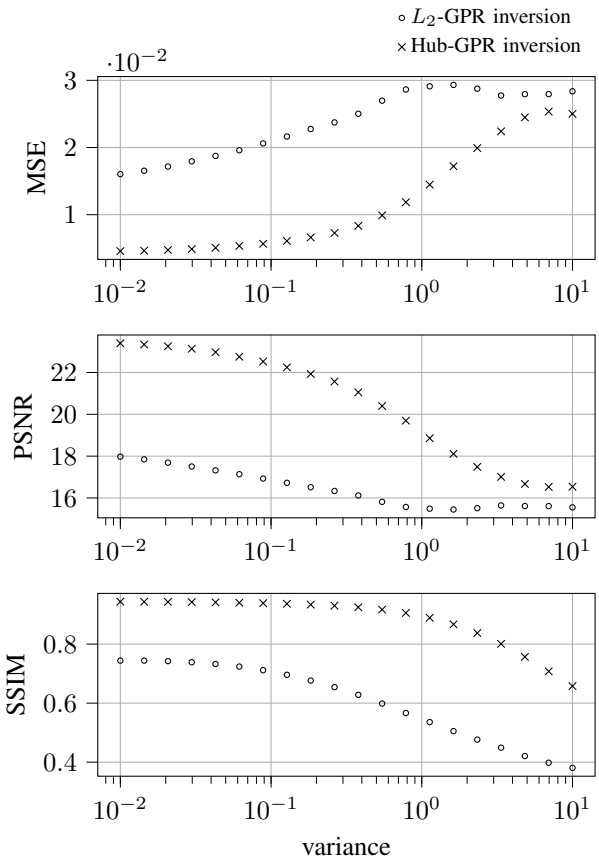


Fig. 12: MSE (top), PSNR (middle) and SSIM (bottom) w.r.t. noise variance with the original image and for the multiplicative noise. $\lambda = 0.4$ and the number of iterations is 20.

Appendix

The ADMM for the classical inversion problem (13) is summarized in Algorithm 2. We will next quickly describe the method to obtain the updates of $\{\mathbf{C}_k\}_{k=1,\dots,K}$ and $\mathbf{L}$.

A. Update of $\{\mathbf{C}_k\}_{k=1,\dots,K}$

By using the notations in the frequency domain and the same transformations as IV-A, we obtain from (13) the new

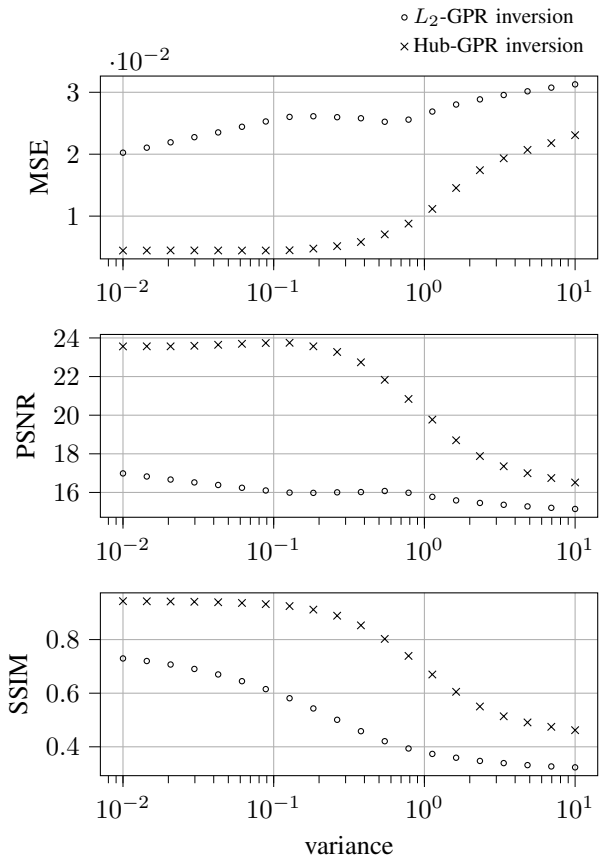


Fig. 13: MSE (top), PSNR (middle) and SSIM (bottom) w.r.t. noise variance with the original image and for the additive noise. $\lambda = 0.4$ and the number of iterations is 20.

optimization problem:

$$\mathbf{c}^{(i+1)} = \underset{\mathbf{c}}{\operatorname{argmin}} \frac{\rho_{\mathbf{L}}}{2}\|\mathbf{H}\mathbf{c} - \mathbf{x}\|_2^2 + \frac{\rho_{\mathbf{S}}}{2}\|\mathbf{c} - \mathbf{z}\|_2^2, \tag{31}$$

where $\mathbf{x} = \mathbf{y} - \mathbf{l} + \widetilde{\mathbf{l}}$ and $\mathbf{H}$, $\mathbf{z}$, $\mathbf{c}$ are as defined in Section IV-A. Optimization problem (31) has been shown in [43] to have a closed-form solution:

$$\mathbf{c}^{(i+1)} = (\rho_{\mathbf{L}}\mathbf{H}^{\dagger}\mathbf{H} + \rho_{\mathbf{S}}\mathbf{I})^{-1}(\rho_{\mathbf{L}}\mathbf{H}^{\dagger}\mathbf{y} + \rho_{\mathbf{S}}\mathbf{z}), \tag{32}$$

where $(\bullet)^{\dagger}$ is the conjugate transpose operator. For numerical stability, we can use the Sherman-Morrison formula instead of the inversion of $(\rho_{\mathbf{L}}\mathbf{H}^{\dagger}\mathbf{H} + \rho_{\mathbf{S}}\mathbf{I})$. And finally, $\{\mathbf{C}_k^{(i+1)}\}$ is obtained by splitting $\mathbf{c}$ into the set of vectors $\{\mathbf{c}_k^{(i+1)}\}$ and by applying the operation $\mathcal{F}^{-1}(\mathrm{unvec}(\bullet))$.

As for the robust inversion, we obtain the coefficients values by reconstructing each $\mathbf{c}_k$ and performing the operation $\mathcal{F}^{-1}(\mathrm{unvec}(\bullet))$ on each coefficient map.

B. Update of $\mathbf{L}$

We need to solve the following optimization problem:

$$\mathbf{L}^{(i+1)} = \underset{\mathbf{L}}{\operatorname{argmin}} \|\mathbf{L}\|_* + \frac{\rho_{\mathbf{L}}}{2}\|\sum_{k=1}^{K}\mathbf{C}_k \circledast \mathbf{H}_k - \mathbf{Y} + \mathbf{L} + \widetilde{\mathbf{L}}\|_2^2,$$

Algorithm 2: Classical inversion of GPR model with ADMM

Data: Image $\mathbf{Y} \in \mathbb{R}^{N_t,N_x}$, dictionary $\{\mathbf{H}_k \in \mathbb{R}^{N_t,N_x}\}_{k=1,\dots,K}$
Parameters: Sparsity λ, lagrangian penalties $\rho_{\mathbf{S}}$, $\rho_{\mathbf{L}}$, relaxation α, stopping criterion ξ
Result: Coefficients maps $\{\mathbf{C}_k \in \mathbb{R}^{N_t,N_x}\}_{k=1,\dots,K}$, clutter matrix $\mathbf{L} \in \mathbb{R}^{N_t,N_x}$
Initialize $\{\mathbf{C}_k^{(0)}\}_{k=1,\dots,K}$, $\{\mathbf{S}_k^{(0)}\}_{k=1,\dots,K}$, $\mathbf{L}^{(0)}$, $\{\widetilde{\mathbf{S}}_k^{(0)}\}_{k=1,\dots,K}$ and $\widetilde{\mathbf{L}}^{(0)}$ with zeros
Precompute Fourier elements $\mathbf{y}$, $\{\mathbf{h}_k\}$ as in IV-A
$i \leftarrow 0$
while $\eta^{(i)} > \xi$ *as in* (30) **do**
  Compute $\mathbf{c}_k^{(i)}$, $\mathbf{s}_k^{(i)}$, $\widetilde{\mathbf{s}}_k^{(i)}$ and $\widetilde{\mathbf{l}}^{(i)}$ as in IV-A
  Update $\{\mathbf{C}_k^{(i+1)}\}$ with (32), $\{\mathbf{S}_k^{(i+1)}\}$ with (27) and $\mathbf{L}^{(i+1)}$ with (33)
  $\forall k$, $\widetilde{\mathbf{S}}_k^{(i+1)} = \widetilde{\mathbf{S}}_k^{(i)} + \mathbf{S}_k^{(i+1)} - \mathbf{C}_k^{(i+1)}$
  $\widetilde{\mathbf{L}}^{(i+1)} = \widetilde{\mathbf{L}}^{(i)} + \sum_{k=1}^{K}\mathbf{C}_k^{(i+1)} \circledast \mathbf{H}_k + \mathbf{L}^{(i+1)} - \mathbf{Y}$
  $i \leftarrow i + 1$

the solution of which is given by [36]:

$$\mathbf{L}^{(i+1)} = \mathcal{T}_{\rho_{\mathbf{L}}^{-1}}(\mathbf{Y} - \widetilde{\mathbf{L}} - \sum_{k=1}^{K}\mathbf{C}_k \circledast \mathbf{H}_k), \tag{33}$$

where $\mathcal{T}_{\rho_{\mathbf{L}}^{-1}}$ is the proximal operator of thresholding the singular values [52]. For a matrix $\mathbf{M}$ with Singular Value Decomposition (SVD), $\mathbf{M} \overset{\text{SVD}}{=} \mathbf{U}\,\mathrm{diag}(\mathbf{d})\mathbf{V}^{\mathrm{T}}$, we have:

$$\mathcal{T}_{\gamma}(\mathbf{M}) = \mathbf{U}\,\mathrm{diag}(\mathcal{S}_{\gamma}(\mathbf{d}))\mathbf{V}^{\mathrm{T}}. \tag{34}$$

REFERENCES

[1] David J Daniels, *Ground Penetrating Radar*, IEE, 2004.

[2] Harry M. Jol, Ed., *Ground Penetrating Radar: Theory and Applications*, Elsevier, 2009.

[3] Andrea Benedetto, Fabio Tosti, Luca Bianchini Ciampoli, and Fabrizio D'Amico, "An overview of ground-penetrating radar signal processing techniques for road inspections," *Signal Processing*, vol. 132, pp. 201–209, 2017.

[4] Lance E. Besaw and Philip J. Stimac, "Deep convolutional neural networks for classifying GPR B-scans," in *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XX*. May 2015, vol. 9454, pp. 385–394, SPIE.

[5] Xisto L. Travassos, Sérgio L. Avila, and Nathan Ida, "Artificial Neural Networks and Machine Learning techniques applied to Ground Penetrating Radar: A review," *Applied Computing and Informatics*, vol. 17, no. 2, pp. 296–308, Jan. 2020, Publisher: Emerald Publishing Limited.

[6] Haoqiu Zhou, Xuan Feng, Yan Zhang, Enhedelihai Nilot, Minghe Zhang, Zejun Dong, and Jiahui Qi, "Combination of Support Vector Machine and H-Alpha Decomposition for Subsurface Target Classification of GPR," in *2018 17th International Conference on Ground Penetrating Radar (GPR)*, June 2018, pp. 1–4, ISSN: 2474-3844.

[7] W. Shao, A. Bouzerdoum, S. L. Phung, L. Su, B. Indraratna, and C. Rujikiatkamjorn, "Automatic classification of GPR signals," in *Proceedings of the XIII Internarional Conference on Ground Penetrating Radar*, June 2010, pp. 1–6.

[8] Mostafa Elsaadouny, Jan Barowski, and Ilona Rolfes, "ConvNet Transfer Learning for GPR Images Classification," in *2020 German Microwave Conference (GeMiC)*, Mar. 2020, pp. 21–24, ISSN: 2167-8022.

[9] Maha Almaimani, Dalei Wu, Yu Liang, Li Yang, Dryver Huston, and Tian Xia, *Classifying GPR Images Using Convolutional Neural Networks*, Jan. 2018.

[10] Zhi-Kang Ni, Jun Pan, Cheng Shi, Shengbo Ye, Di Zhao, and Guangyou Fang, "DL-Based Clutter Removal in Migrated GPR Data for Detection of Buried Target," vol. 19, pp. 1–5, 2022, Conference Name: IEEE Geoscience and Remote Sensing Letters.

[11] R. Yelf and W. Al-Nuaimy, "Classification system for GPR parameters," in *Proceedings of the Tenth International Conference on Grounds Penetrating Radar, 2004. GPR 2004.*, June 2004, vol. 1, pp. 407–410.

[12] A.C. Gurbuz, J.H. McClellan, and W.R. Scott, "A compressive sensing data acquisition and imaging method for stepped frequency GPRs," *Signal Processing, IEEE Transactions on*, vol. 57, no. 7, pp. 2640–2650, July 2009.

[13] Zhaofa Zeng, Jing Li, Ling Huang, Xuan Feng, and Fengshan Liu, "Improving target detection accuracy based on multipolarization MIMO GPR," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 53, no. 1, pp. 15–24, Jan 2015.

[14] A.M. Zoubir, I.J. Chant, C.L. Brown, B. Barkat, and C. Abeynayake, "Signal processing techniques for landmine detection using impulse ground penetrating radar," *Sensors Journal, IEEE*, vol. 2, no. 1, pp. 41–51, Feb 2002.

[15] W. Ng, T. Chan, H.C. So, and K.C. Ho, "Particle filtering based approach for landmine detection using Ground Penetrating Radar," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 11, pp. 3739–3755, Nov 2008.

[16] A. Manandhar, P.A. Torrione, L.M. Collins, and K.D. Morton, "Multiple-instance hidden markov model for gpr-based landmine detection," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 53, no. 4, pp. 1737–1745, April 2015.

[17] V. Kovalenko, A.G. Yarovoy, and L.P. Ligthart, "A novel clutter suppression algorithm for landmine detection with GPR," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 11, pp. 3740–3751, Oct 2007.

[18] Jing Li, Cai Liu, Zhaofa Zeng, and Lingna Chen, "Gpr signal denoising and target extraction with the ceemd method," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 8, pp. 1615–1619, 2015.

[19] H. Brunzell, "Detection of shallowly buried objects using impulse radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 2, pp. 875–886, 1999.

[20] K.C. Ho and P.D. Gader, "A linear prediction land mine detection algorithm for hand held ground penetrating radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 6, pp. 1374–1384, 2002.

[21] Q. Hoarau, G. Ginolhac, A. M. Atto, and J. M. Nicolas, "Robust adaptive detection of buried pipes using GPR," *Signal Processing*, vol. 132, pp. 293–305, Mar. 2017.

[22] P. Meincke, "Linear gpr inversion for lossy soil and a planar air-soil interface," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 12, pp. 2713–2721, 2001.

[23] Deshan Feng, Cen Cao, and Xun Wang, "Multiscale full-waveform dual-parameter inversion based on total variation regularization to on-ground gpr data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9450–9465, 2019.

[24] Xun Wang and Deshan Feng, "Multiparameter full-waveform inversion of 3-d on-ground gpr with a modified total variation regularization scheme," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 3, pp. 466–470, 2021.

[25] Jianping Wang, Pascal Aubry, and Alexander Yarovoy, "Efficient implementation of gpr data inversion in case of spatially varying antenna polarizations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 2387–2396, 2018.

[26] Michele Ambrosanio and Vito Pascazio, "A compressive-sensing-based approach for the detection and characterization of buried objects," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 7, pp. 3386–3395, 2015.

[27] Xingkun He, Can Wang, Rongyao Zheng, Zhibin Sun, and Xiwen Li, "Gpr image denoising with nsst-unet and an improved bm3d," *Digital Signal Processing*, vol. 123, pp. 103402, 2022.

[28] Qiqi Dai, Yee Hui Lee, Hai-Han Sun, Genevieve Ow, Mohamed Lokman Mohd Yusof, and Abdulkadir C. Yucel, "Dmrf-unet: A two-stage deep learning scheme for gpr data inversion under heterogeneous soil conditions," *IEEE Transactions on Antennas and Propagation*, pp. 1–1, 2022.

[29] Paolo Bestagini, Federico Lombardi, Maurizio Lualdi, Francesco Picetti, and Stefano Tubaro, "Landmine detection using autoencoders on multipolarization gpr volumetric data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 182–195, 2021.

[30] Christian Maas and Jörg Schmalzl, "Using pattern recognition to automatically localize reflection hyperbolas in data from ground penetrating radar," *Computers and Geosciences*, vol. 58, pp. 116–125, 2013.

[31] Bin Liu, Yuxiao Ren, Hanchi Liu, Hui Xu, Zhengfang Wang, Anthony G. Cohn, and Peng Jiang, "Gprinvnet: Deep learning-based ground-penetrating radar data inversion for tunnel linings," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 10, pp. 8305–8325, 2021.

[32] Wenbin Shao, Abdesselam Bouzerdoum, and Son Lam Phung, "Sparse representation of gpr traces with application to signal classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 7, pp. 3922–3930, 2013.

[33] Fok Hing Chi Tivive, Abdesselam Bouzerdoum, and Canicious Abeynayake, "Gpr signal classification with low-rank and convolutional sparse coding representation," in *2017 IEEE Radar Conference (RadarConf)*, 2017, pp. 1352–1356.

[34] Fabio Giovanneschi, Kumar Vijay Mishra, Maria Antonia Gonzalez-Huici, Yonina C. Eldar, and Joachim H. G. Ender, "Dictionary learning for adaptive gpr landmine classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 10036–10055, 2019.

[35] Guillaume Terrasse, Jean-Marie Nicolas, Emmanuel Trouvé, and Émeline Drouet, "Sparse decomposition of the GPR useful signal from hyperbola dictionary," in *2016 24th European Signal Processing Conference (EUSIPCO)*, 2016, pp. 2400–2404.

[36] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, jun 2011.

[37] Brendt Wohlberg, "Convolutional sparse representations with gradient penalties," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 6528–6532.

[38] P.J. Huber and E.M. Ronchetti, *Robust Statistics*, Wiley, 2009.

[39] Aditya Ramadhan, Erfansyah Ali, and A.A. Pramudita, "The effect ricker wavelet of duty cycle adjustment on gpr detection result," *JMECS (Journal of Measurements, Electronics, Communications, and Systems)*, vol. 8, no. 1, pp. 17–22, 2021.

[40] Yanghua Wang, "Frequencies of the Ricker wavelet," *Geophysics*, vol. 80, no. 2, pp. A31–A37, 02 2015.

[41] Benjamin Recht, Weiyu Xu, and Babak Hassibi, "Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization," in *2008 47th IEEE Conference on Decision and Control*, 2008, pp. 3065–3070.

[42] Cristina Garcia-Cardona and Brendt Wohlberg, "Convolutional dictionary learning: A comparative review and new algorithms," *IEEE Transactions on Computational Imaging*, vol. 4, no. 3, pp. 366–381, 2018.

[43] Brendt Wohlberg, "Efficient convolutional sparse coding," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 7173–7177.

[44] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[45] Esa Ollila, "Multichannel sparse recovery of complex-valued signals using huber's criterion," in *2015 3rd International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, 2015, pp. 26–30.

[46] P. L. Combettes and J.-C. Pesquet, "Proximal thresholding algorithm for minimization over orthonormal bases," *SIAM Journal on Optimization*, vol. 18, no. 4, pp. 1351–1376, 2007.

[47] Neal Parikh and Stephen Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, jan 2014.

[48] J. Eckstein, "J. parallel alternating direction multiplier decomposition of convex programs," *J Optim Theory Appl*, vol. 80, pp. 39–62, 1994.

[49] Jessica Franco, *Planification d'expériences numériques en phase exploratoire pour la simulation des phénomènes complexes*, Theses, Ecole Nationale Supérieure des Mines de Saint-Etienne, Sept. 2008.

[50] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[51] Ying Sun, Prabhu Babu, and Daniel P Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3576–3590, 2016.

[52] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.