Non-linear Inverse Scattering via Sparsity Regularized Contrast Source Inversion

Martina T. Bevacqua, *Student Member, IEEE,* Lorenzo Crocco, *Senior Member, IEEE,*
Loreto Di Donato, *Member, IEEE* and Tommaso Isernia, *Senior Member, IEEE*

***Abstract*—Two Compressive Sensing inspired approaches for the solution of non-linear inverse scattering problems are introduced and discussed. Differently from the sparsity promoting approaches proposed in most of the papers published in the literature, the two methods here tackle the problem in its full non-linearity, by adopting a contrast source inversion scheme. In the first approach, the ℓ_1-norm of the unknown is added as a weighted penalty term to the usual contrast source cost functional, whereas the second approach enforces sparsity by constraining the solution to a convex set defined by the ℓ_1-norm of the unknown. Results with simulated data are given to assess the capabilities of the proposed approaches. Cases considering both a reduced number of data (with respect to Nyquist sampling) and/or an overcomplete dictionary are also successfully dealt with.**

***Index Terms*— Compressive Sensing, Compressed Measurements, Contrast-Source Inversion Method, Microwave Imaging, Non linear Inverse Problem, Stationary Wavelet Transform.**

I. Introduction

Compressive Sensing (CS) [1] is based on the concept of "sparsity" or "compressibility", i.e., the possibility to exactly (or anyway accurately) represent a function through a limited number of nonzero coefficients of some convenient basis. In a recovery problem, this may enable remarkable advantages with respect to techniques not exploiting this property, such as reducing the required measurements or achieving some kind of "super-resolution".

These perspectives are of outmost interest in inverse scattering and microwave imaging [2-8]. As a matter of fact, in many applications, such as subsurface sensing [3,4] and biomedical imaging [5], the unknown function can be considered as sparse (or compressible) in some suitable basis. On the other hand, the CS theory is well formulated and assessed only when the model relating the data to the unknowns is linear, whereas inverse scattering problems are generally nonlinear. Hence, this fundamental limitation has to be addressed to gain a full advantage of CS, without limiting its applicability to those cases that can be handled through linear scattering approximations. These include the Born [4-6] and Rytov [7] approximations for weak scatterers, as well as the approximation based on the more recent *virtual experiments* framework [3,8].

A possible way to extend the applicability of CS is to exploit distorted wave methods or distorted-iterated methods, in which the solution is iteratively achieved through a succession of linear inversion steps [9-11]. Because of the linearity of the models considered at each step, these latter methods naturally lend themselves to be paired with CS. However, as well known, the success of this scheme depends on the initial guess, which has to trigger a convergent sequence of linearizations. This is less critical for the method briefly (synthetically) suggested in [11], wherein one conversely needs a sufficient number of data to design the *virtual experiments*. Finally, distorted wave methods require to solve several forward problems at each step of the procedure, which has an obvious impact on the computational burden.

As an interesting alternative to the above possibilities, in this paper we propose and discuss the use of CS in conjunction with the contrast-source inversion (CSI) method [12,13]. CSI is one of the most popular and effective inversion schemes, which does not require to solve any forward problem during the iterative scheme, thus resulting in a computationally efficient procedure. Moreover, by paying the price of considering an enlarged number of unknowns (which seems to be an effective strategy also for the solution of other kinds of inverse problems, see [14]), it allows to deal with a mathematical problem involving just linear and quadratic equations.

In particular, we introduce and discuss two different approaches. In the first one, we add a penalty term (enforcing sparsity) to the standard cost functional defining the solution of the CSI. In the second approach, sparsity is enforced by introducing a ℓ_1-norm constraint, i.e., by limiting the search space of the unknown. As shown in the paper, the two methods provide indeed the expected advantages in terms of super-resolution capabilities and/or data reduction (with respect to the Nyquist criterion [15]). Moreover, the two

This is the postprint version of the following article: M. T. Bevacqua, L. Crocco, L. D. Donato and T. Isernia, "Non-Linear Inverse Scattering via Sparsity Regularized Contrast Source Inversion," in IEEE Transactions on Computational Imaging, vol. 3, no. 2, pp. 296-304, June 2017, doi: 10.1109/TCI.2017.2675708. Article has been published in final form at: https://ieeexplore.ieee.org/document/7866862.

strategies are flexible with respect to the use of different representation bases, ranging from orthonormal bases to truly redundant dictionaries [16], the latter being now widespread in signal and image processing [16,17,18].

The remainder of the paper is organized as follows. In Section II, we cast the inverse scattering problem and recall the CSI method. In Section III, we introduce the CS theory and the two proposed strategies. Finally, in Section IV, we assess their capabilities through numerical examples. Conclusions follow.

Throughout the paper we consider the canonical 2D TM scalar problem and we assume and drop the time harmonic factor $exp\{j\omega t\}$. Some of the ideas herein developed and thoroughly analyzed here had been first introduced, in a very synthetic form, in two conference communications [19,20].

II. Inverse Scattering Problem and The Adopted Contrast Source Inversion Method

Let us consider one or more unknown non-magnetic scatterers with support Σ and permittivity ε_x and conductivity σ_x, enclosed in the investigated domain Ω and embedded in a background medium of known electromagnetic features ε_b and σ_b. The scatterer is probed by means of a set of incident fields transmitted by an array of elementary sources located on a closed curve Γ, positioned in the far-field of Ω. The resulting scattered fields are measured by receiver antennas located on Γ. The scattering problem for the generic position r_t of the transmitting antennas can be described by the following equations:

$$E_s(r_m, r_t) = \mathcal{A}_e[W(r, r_t)] \tag{1}$$

$$W(r, r_t) = \chi(r)E_i(r, r_t) + \chi(r)\mathcal{A}_i[W(r, r_t)] \tag{2}$$

where $r\,\epsilon\,\Omega$, r_m denotes the receiver position on Γ, E_i, E_s and W are the incident field, scattered field and contrast source induced in the scatterer, respectively, and $\mathcal{A}_\mathrm{e}$ and $\mathcal{A}_\mathrm{i}$ are a short notation for the integral radiation operators.

The inverse scattering problem aims at retrieving the unknown contrast $\chi\ \forall\, r\,\epsilon\,\Omega$ from the scattered fields E_s measured on Γ for a set of (known) incident fields E_i. Several efforts are carried out in the literature to develop effective solution methods [21-25], due to the non-linearity and ill-posedness of the problem [26]. In particular, only a limited number of independent experiments (and measurements) is available and hence only a limited number of independent parameters can be recovered from scattered field data. Accordingly, the positions of N_T transmitting probes and those of N_R receiving ones are chosen in such a way to collect the available information in a non redundant.

The CSI method [12,13,23,24] tackles the problem in its full non-linearity by contemporarily looking for both the contrast χ and the contrast source W, considered as an *auxiliary* unknown. In particular, the problem's solution is iteratively built by minimizing a cost functional, which takes into account the data-to-unknown relationship and the physical model [12,13,23,24], i.e.:

$$\Phi(\chi, W) = \Phi_\Omega(\chi, W) + \Phi_\Gamma(W)$$

with:

$$\Phi_\Omega(\chi, W) = \sum_{t=1}^{N_T} \kappa^t{}_\Omega \| W - \chi E_i - \chi\mathcal{A}_i[W] \|_\Omega^2$$
$$\Phi_\Gamma(W) = \sum_{t=1}^{N_T} \kappa^t{}_\Gamma \| E_s - \mathcal{A}_e[W] \|_\Gamma^2 \tag{3}$$

where $\|\cdot\|$ is the ℓ_2 norm and $\kappa^t{}_\Omega$ and $\kappa^t{}_\Gamma$ are normalization coefficients, which read:

$$\kappa^t{}_\Omega = \|E_i\|_\Omega^{-2} \qquad \text{and} \qquad \kappa^t{}_\Gamma = \|E_s\|_\Gamma^{-2}. \tag{4}$$

Note the choice of the normalization term for the "state" equation (2) is not the usual one. As a matter of fact, this term is often set equal to the norm of the product of the contrast times the incident field [12,13]. However, such a choice entails the drawback of changing the metric at each step of the minimization procedure. Moreover, as explained below, the choice done here allows for a simpler updating scheme.

As it can be easily checked, functional (3) is a non quadratic functional of the unknowns. However, differently from the Newton-Kantorovich approach [25], the CSI formulation allows to deal with a cost functional, which is a low order polynomial in terms of the unknown parameters (see [27] for more details). Obviously, a price is paid in the fact that, differently from [25], the cost functional depends on both the contrast function and the contrast sources.

To perform such an optimization task, global optimization schemes are not viable. As a matter of fact, the problem usually depends on thousands of variables, so that, due the so called 'no free lunch theorem' [28], convergence to the ground truth of global optimization procedures is not granted in any reasonable time. Hence, local (gradient-based) minimization schemes are usually adopted. However, these latter only converge to the closest local minimum of the cost functional, unless proper regularization tools are introduced.

As proposed in [23,24], the minimization of (3) is pursued by means of a "quasi-Newton" iterative scheme:

$$\chi_{(k+1)} = \chi_k + \mu_k \mathrm{H}_k \nabla\chi_k$$
$$W_{(k+1)} = W_k + \mu_k \mathrm{H}_k \nabla \mathrm{W}_k \tag{5}$$

where k and $k+1$ denote the k^{th} and the $(k+1)^{th}$ iteration, respectively, $\nabla\chi$ and ∇W are the gradients of the functional Φ with respect to χ and W, respectively, μ_k is a scalar factor that has to be evaluated at each iteration in order to guarantee the maximum decrease of the functional along the gradient direction. Finally, H is defined accordingly to the Polak–Ribiere scheme [23,24].

Note we perform a joint updating of the two unknowns W and χ, which is a fundamental difference with respect to [12,13], where they are alternatively updated. Accordingly, the line minimization step μ_k is performed as:

$$\mu_k = \operatorname{argmin} \Phi(\chi_k + \mu_k \Delta\chi_k, W_k + \mu_k \Delta W_k) \tag{6}$$

where $\Delta\chi_k$ and ΔW_k represent the descent research directions considered at the k^{th} iteration. As (3) is a fourth degree polynomial with respect to the unknowns (which is a direct consequence of how we have defined $\kappa^t{}_\Omega$), solution of (6) implies the solution of a third degree algebraic equation, which is available in a closed form [29]. As a consequence, the line minimization step can be performed in a very efficient and precise manner. Moreover, the fact one can compute in a closed form all the different minima of the functional along the optimization direction could open the way to a kind of 'tunneling' of the optimization procedure, i.e., a path from an attraction basin of a local minima of (3) to a different and (hopefully) more convenient attraction basin. Such a possibility is precluded when alternately looking for χ and W.

III. New Regularized CSI Strategies Based On Compressive Sensing Theory

The CS theory provides useful tools for solving, in an accurate fashion, linear problems of the kind $y = Ax$ when the number M of equations is less than the number of unknowns N, but the unknown signal is known to be S-sparse[1]. In fact, as long as the unknown function can be expressed in a suitable basis Ψ ($x = \Psi s$), where the unknown sequence s can be assumed to be sparse, and suitable conditions hold true on M, N and S, at least three different strategies can be devised. First, the problem can be reduced to the well known basis pursuit denoising or LASSO problem [30], that is:

$$min\{\|s\|_{\ell_1}\} \quad s.t. \quad \|A\,\Psi s - y\|_{\ell_2} \le \sigma \tag{7}$$

where $\|\cdot\|_{\ell_p}$ denotes the ℓ_p-norm and σ is a positive parameter that depends on the required accuracy, the amount of noise on the data and the model error.

In a dual fashion, one can solve the following ℓ_1-norm constrained optimization problem [31]:

$$min\{\|A\,\Psi s - y\|_{\ell_2}\} \quad s.t. \quad \|s\|_{\ell_1} \le \delta \tag{8}$$

where δ is positive real number, which is related to the number and the amplitude of the contributing (i.e. nonzero) elements of s.

Last, but not least, one can consider the ℓ_1-norm penalized least squares problem:

$$\min\{\|A\Psi s - y\|_{\ell_2} + k^2\|s\|_{\ell_1}\} \tag{9}$$

where k is a positive parameter controlling the relative weight of the ℓ_1 regularization term.

In both (8) and (9), the aim is to match the data of the problem while conditioning the ℓ_1 norm of the unknown, (in such a way to enforce sparsity). This is realized in formulation (8) by constraining the unknown function inside a limited space defined by the ℓ_1-norm of the unknown, while a penalty term is used instead in formulation (9).

As generalization of LASSO to non linear problems would imply consideration of non convex constraints, in the following we limit ourselves to consider and discuss the extension of the approaches (8) and (9) to the CSI method.

A. ℓ_1-norm penalized CSI

By paralleling the approach (9), the solution of the inverse scattering problem can be obtained by adding a penalty term to the functional (3). Such a penalty term accounts for the sparsity of the contrast and the overall functional reads:

$$\min_{s,W} \Phi(s, W) + \Phi^P(s) = \min_{s,W} \; \Phi(s, W) + k^2\|s\|_{\ell_1} \tag{10}$$

where k is a non-negative weighting parameter to be determined. In (10), the inclusion of the penalty term Φ^P modifies the shape of the cost functional, and hence its attraction basins. In particular, depending on k, a number of local minima of (3) (corresponding to 'non sparse' profiles) will disappear when considering the functional (10).

The optimization of (10) within a conjugate gradient scheme requires the computation of the gradients of the different addenda. The expression of gradients for the first two terms is the same as in [23,24], whereas, as detailed in the Appendix, the gradient of the third term reads:

$$\nabla\Phi^P{}_s = k^2 \frac{s}{|s|} \tag{11}$$

Unfortunately, such an expression implies that it is not anymore possible to reduce (6) to the solution of an algebraic equation. As a consequence, the step length must be evaluated numerically, thus affecting the computational burden.

B. ℓ_1-norm constrained CSI

Inspired by (8), rather than adding a penalty term to the CSI functional, a ℓ_1-norm constraint limiting the space of search of the unknown can be included into the CSI scheme. More in detail, the inverse problem is solved by looking for the global minimum of the functional (3), among all the unknowns whose ℓ_1-norm (of the coefficients s of the representation basis Ψ) is lower than the pre-set parameter δ, i.e.,:

$$\min_{s,W} \Phi(s, W) \quad subject\ to \;\; \|\boldsymbol{s}\|_{\ell_1} \le \delta \;. \tag{12}$$

The inclusion of the ℓ_1-norm constraint allows to restrict the search space, so that the optimization procedure more easily falls into the right attraction basin, thus avoiding a number of false solutions. In fact, all local minima of (3) not belonging to the space defined by the ℓ_1-norm constraint are not anymore admissible. Hence, the local minimum achieved at the end of the procedure will more easily correspond to the actually global optimum. On the other hand, as the attraction regions of the local minima outside the constraining set may

[1] S is the number of nonzero coefficients of a proper basis.

intersect the set defined by (12), the optimization procedure may be trapped on the boundary of such a set. Hence, some additional action is eventually required in order to escape from these points (see below).

With respect to the considered CSI scheme, the sparsity constraint implies (at each iteration) a modification of the line minimization step along the search direction. In particular, the minimization step length μ_k has to be chosen such to minimize (6), while also guaranteeing that the solution at the next iteration belongs to a feasible set, i.e., $\|s_k + \mu_k \Delta s_k\|_{\ell_1} \leq \delta$. Saying it in a different fashion, μ_k has to be lower than the distance β between the current solution s_k and the boundary of the feasible solution set along the considered direction.

At each step of the iterative procedure, the distance β along the descent direction $\hat{\iota}_g$ can be evaluated as:

$$\beta: \quad \left\|s_k + \beta\hat{\iota}_g\right\|_{\ell_1} = \delta \tag{13}$$

so that at each iteration the solution of problem (6) can be compared with the distance β. Then, if the minimizer of (6) μ_k is lower than β, the solution is updated according to (5) and it is still inside the feasible set. Otherwise, μ_k is set equal to β. This means that the solution belongs to the boundary of the set. However, the overall problem herein at hand is non-linear, so that the solution does not necessary belong to the boundary of the set. As a consequence, whenever the procedure is trapped into a boundary point corresponding to an unsatisfactory fitting of data, some action is required in order to escape from such point.

By preserving some characteristics of the current trial solution, a possible strategy amounts to project the achieved current solution s_k inside the feasible set defined by the ℓ_1-norm constraint. This projection is performed in the following numerical analysis by multiplying s_k times a binary mask Π:

$$s_k = s_k \Pi \tag{14}$$

which is null in correspondence of the coefficients whose magnitude is lower than the 30% of magnitude of the maximum one.

As a final note it is worth remarking that both approaches can be particularized to the 'most convenient' representation basis. For instance, in many cases of actual interest the targets are piecewise constant so that the unknown contrast profiles are sparse in terms of the step functions. In these cases, the penalty term and the constraints can be rewritten in terms of $\|\mathcal{D}\chi\|_{\ell_1}$, where $\mathcal{D}$ is the discretized version of partial derivative, i.e., the vector containing the finite differences of the unknown function χ. Moreover, the analytical expression of the gradient of the penalty term $\Phi^P{}_s$ is $-\mathcal{D}\left(k\,\frac{\mathcal{D}\chi}{|\mathcal{D}\chi|}\right)$.

IV. Numerical Results

To assess and compare the performance of the CS based CSI approaches described in the previous Section in controlled conditions, we have carried out a numerical analysis using an example dealing with simulated data. In particular, we have considered the so called Austria profile, as this has been previously used as a testbed for CSI or CSI-based schemes [32-34].

The Austria target consists of a ring and two smaller cylinders, with permittivity and conductivity equal to 2 and 8 mS/m. The target is hosted in free space and the real and imaginary parts of the contrast function are shown in figs. 1(a)-(b), respectively. A multiview-multistatic illumination setup has been assumed, with $N_T = N_R = 24$ filamentary currents acting as primary sources and receivers. At the considered frequency ($400MHz$), this number of antennas is adopted to sample all the available information in a non-redundant fashion [15]. The receivers and transmitters are located on a circumference of radius R = 3 m surrounding the region of interest Ω, which is a square of side $L = 2$, discretized into $N_c = 64 \times 64$ cells. The scattered field data have been obtained by means of a full-wave forward solver based on CG-FFT procedure and corrupted with a random Gaussian noise with SNR= 20 dB. In all cases, the initial guess is given by the back-propagation solution.

To evaluate the accuracy of the retrieved contrast $\tilde{\chi}$, we consider the normalized mean square reconstruction error:

$$err = \frac{\|\chi - \tilde{\chi}\|_{\ell_2}}{\|\chi\|_{\ell_2}} \tag{15}$$

where χ is the actual profile.

Concerning the choice of the (positive) parameter k in penalized CSI, during the numerical analysis we have set it equal to $\frac{1}{N_c}$. In so doing, the penalty term is normalized with respect to the number cells.

As far as the choice of δ parameter in constrained CSI is concerned, this latter influences the dimension of the search space. In fact, if δ is 'too large', the constraint would not influence the minimization procedure (any solution would belong to the constraining set), whereas a 'too small' value may entail that even the actual solution could not belong to the set. As the 'right value' of δ is obviously unknown, we have adopted in the following an "*a posteriori*" criterion. More in detail, after evaluating the solution for different values of δ, we have selected the one corresponding to the lowest value of the CSI functional (3). Note this procedure can be eventually performed by using parallel computing, thus limiting the computational overhead.

In a first set of examples, we have exploited the fact that the profile at hand can be considered to be sparse in the space of step functions. Accordingly, the penalty term and the constraint in (10) and (12) are expressed by the discretized versions of partial derivatives with respect to the vertical and horizontal directions. In Fig. 1(c)-(d), we show the real and imaginary parts of the reconstructed contrast obtained with penalized CSI. In Figs. 1(g)-(h), we report the real and imaginary profiles retrieved by the constrained CSI. As it can be seen, the reconstructions are satisfactory and quite similar in all cases, as confirmed by the reconstruction errors which are equal to 15% and 13%, respectively (see Table 1).

To further evaluate the robustness of the methods and show the possible advantages offered by sparsity regularizations, we have reduced the number of antennas down to $N_T = N_R = 16$. This corresponds to a total number of 136 independent data. As it can be seen from figs. 1(e)-(f) and (i)-(j), the obtained results are still completely satisfactory.

To prove the flexibility of both approaches, we have also tested the use of a different representation for the unknown contrast function. In particular, the Stationary Wavelet Haar (SWH) basis of level 1 has been considered instead of the step functions. The SWH is a wavelet transform achieving translation invariance, a property that is missing in the discrete wavelet transform (DWT) [16]. Translation-invariance is achieved by removing the downsamplers and upsamplers in the DWT and by upsampling the filter coefficients by a factor of 2j at the (j−1) level of the algorithm. For this reason, the output of each level of the SWH contains the same number of samples as the input. In comparison with the previous case of step functions, this means that the unknown is now expressed by means of a redundant dictionary in which there are more columns than rows [16]. Saying it in other words, such a representation is not univocal, as the considered functions are not orthonormal (and not even orthogonal). The use of these overcomplete dictionaries is now widespread in signal processing and data analysis, as there are numerous practical examples in which a signal is not sparse in an orthonormal basis or incoherent dictionary, but it is instead sparse in terms of a truly redundant dictionary [16].

The results reached by both approaches are shown in figure 2. Note that in this case the parameter k has been multiplied by the number of columns of the dictionary in order to take into account its redundancy. As it can be seen, enforcing sparsity by means of the two techniques allows to achieve reliable results with a negligible reconstruction errors (see Table 2), also in case of a significant reduction of the number of measurements. By further reducing the number of antennas, the accuracy of the reconstructions gradually deteriorates. In fact, for instance, by considering the performance of the constrained CSI approach with $N_T = N_R = 14$ and $N_T = N_R = 12$, the obtained mean square errors are 25% and 37%, respectively. On the other hand, by considering the penalized approach with $N_T = N_R = 14$, the reconstruction error is 23%.

It is worth to remark that, as expected, when using the CSI scheme without any additional regularization, the algorithm is trapped into a meaningless solution. Conversely, the proposed sparsity enhanced CSI schemes allow to achieve results comparable with those already obtained in the literature for the same target [32,34], but using a reduced number of data, which is in line with the theoretical predictions of compressive sensing theory.

The above results are representative of a large number of examples that assess the capability of the two approaches to deal with targets having both complex shapes and large values of contrast. Of course, cases exist in which the use of ℓ_1 regularizations is not sufficient to overcome the occurrence of local minima. In these cases, additional regularization schemes should be considered to restore the correct behavior of the inversion scheme.

V. Conclusions

Compressive Sensing theory can be applied whenever the problem to be faced is linear and when the unknown is sparse. In inverse scattering problems, CS could offer advantages in terms of improvements in resolution and reduction of the number of antennas (and hence of the complexity of the measurements apparatus). Unfortunately, the inverse scattering problem is nonlinear, so that a straightforward application of CS tools is possible only in those special conditions in which the problem can be linearized. To overcome this remarkable limitation, in this paper we have explored the application of CS to non-linear inverse scattering, when no approximation of the scattering phenomena is introduced and no a priori information on the admissible values of contrast profile is enforced. The two proposed approaches are inspired to CS theory and can promote sparsity in different ways. They have been developed in conjunction with the well-known CSI method and they are flexible with respect to the adopted linear representation basis, ranging from the step functions to stationary wavelets (and many others), as proved in the numerical analysis section.

The numerical analysis suggests that enforcing the sparsity condition on the unknowns, both by means of a constrained and penalized scheme, can provide useful advantages in imaging and quantitative diagnostics. In particular, the results have shown that one can accurately image the target not only in case of noisy data, but also when a reduced numbers of probing antennas (with respect to the Nyquist criterion) is adopted, which is an extremely useful capability in the applications.

	$N_T = N_R$		*parameters*
	16	24	
Penalized CSI	19%	15%	$5 \cdot 10^{-8}$
Constrained CSI	30%	13%	520

Table 1. Mean square reconstruction errors in case of step functions.

	$N_T = N_R$		*parameters*
	16	24	
Penalized CSI	20%	14%	$1 \cdot 10^{-6}$
Constrained CSI	11%	9%	2700

Table 2. Mean square reconstruction errors in case of SWH.

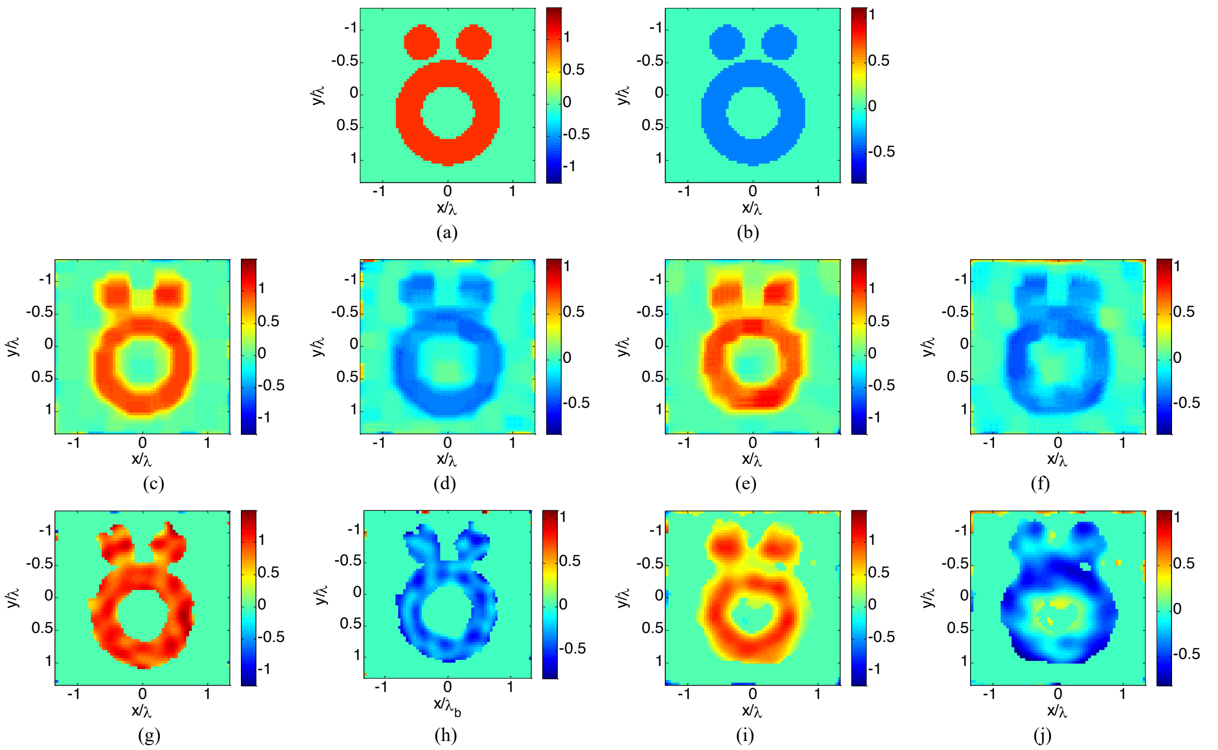

Figure 1. Numerical assessment of the proposed methods against the lossy Austria profile. (a) Real part and (b) imaginary part of the reference contrast. (c)-(d) Real and imaginary part of the retrieved profile with penalized CSI for k = $5{\cdot}10^{-8}$ (err=15%). (e)-(f) the same as (c)-(d) for a reduced amount of data: $N_T = N_R = 16$ (err=19%). (g)-(h) Real and imaginary part of the contrast retrieved profile by constrained CSI for δ = 520 (err=13%). (i)-(j) the same as (g)-(h) for a reduced amount of data: $N_T = N_R = 16$ (err=30%).

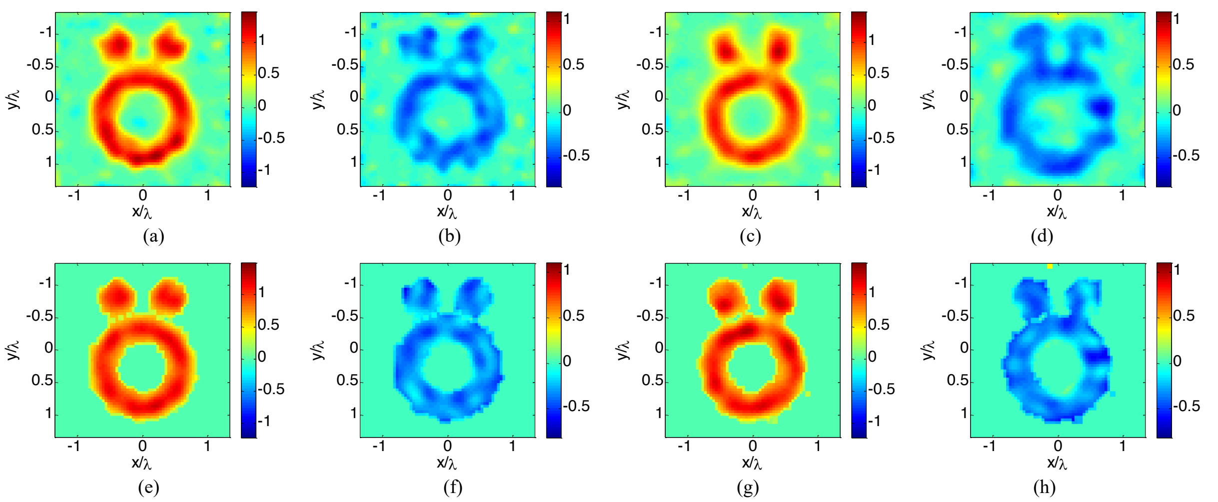

Figure 2. Numerical assessment of the proposed methods against the lossy Austria profile, using stationary wavelets trasform. (a)-(b). Real and imaginary part of the profile retrieved with penalized CSI for k =1 10^{-6} (err=14%). (c)-(d) the same as (a)-(b) for a reduced amount of data: $N_T = N_R = 16$ (err=20%). (e)-(f) Real and imaginary part of the profile retrieved with constrained CSI for δ = 2700 (err=9%). (g)-(h) the same as (e)-(f) for a reduced amount of data: $N_T = N_R = 16$ (err=11%).

APPENDIX

The expression of the gradient for the penalty function Φ^P defined in (10), within a conjugate gradient scheme, is analytically derived by following the approach in [23,24]. The gradient of Φ^P with respect to s is computed by using the following definition:

$$\Delta\Phi^P{}_s = \langle \nabla\Phi^P{}_s, \Delta s\rangle \tag{A.1}$$

where the variation Δs gives rise to $\Delta\Phi^P{}_s$, the variation of Φ^P.

Then, by considering the expression of $\Phi^P(s)$ in (10), it follows that:

$$\Delta\Phi^P{}_s = \frac{k^2}{2}\langle s^{\frac{1}{2}}, \Delta s\, s^{-\frac{1}{2}}\rangle + c.c. = \frac{k^2}{2}\langle s^{\frac{1}{2}}\left(s^{-\frac{1}{2}}\right)^*, \Delta s\,\rangle + c.c. \tag{A.2}$$

where $\langle\cdot,\cdot\rangle$ denotes the scalar product, while $c.c.$ stands for the complex conjugate of the first addendum. By comparing eq. (A.1) and (A.2) the analytically expression of the gradient is so identified as:

$$\nabla\Phi^P{}_s = k^2\ \frac{s}{|s|} \tag{A.3}$$

REFERENCES

[1] Donoho D., “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] H. F. Alqadah, M. Ferrara, H. Fan and J. T. Parker, "Single Frequency Inverse Obstacle Scattering: A Sparsity Constrained Linear Sampling Method Approach," in *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2062-2074, April 2012.

[3] Bevacqua M., Crocco L., Di Donato L., Isernia T., Palmeri R., "Exploiting Field Conditioning and Sparsity for Microwave Imaging of Non-weak Buried Targets", *Radio Science,* 2016.

[4] Ambrosanio M. and Pascazio V., “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.

[5] Bevacqua M., Scapaticci R., "A Compressive Sensing Approach for 3D Breast Cancer Microwave Imaging with Magnetic Nanoparticles as Contrast Agent," *IEEE Trans. on Medical Imaging*, vol.35, no.2, pp.665-673, Feb. 2016.

[6] Poli L., Oliveri G., Massa A., “Microwave Imaging Within the First-Order Born Approximation by Means of the Contrast-Field Bayesian Compressive Sensing,” *Antennas and Propagation, IEEE Transactions on*, 60(6):2865-2879, 2012.

[7] Oliveri G., Poli L., Rocca P., and Massa A., “Bayesian compressive optical imaging within the Rytov approximation”,*Opt. Lett.*, vol. 37, no. 10, pp. 1760–1762, 2012.

[8] Bevacqua M.T., Crocco, L., Di Donato, L., Isernia, T., "Microwave Imaging of Nonweak Targets via Compressive Sensing and Virtual Experiments," in *Antennas and Wireless Propagation Letters, IEEE*, vol.14, no., pp.1035-1038, 2015.

[9] Desmal, A., Bagci, H., "Shrinkage-Thresholding Enhanced Born Iterative Method for Solving 2D Inverse Electromagnetic Scattering Problem," *Antennas and Propagation, IEEE Transactions on*, 62(7): pp. 3878-3884, 2014.

[10] Azghani M., Kosmas P., Marvasti F., “Microwave Medical Imaging Based on Sparsity and an Iterative Method With Adaptive Thresholding,” *IEEE Trans. Medical Imaging*, vol.34, no.2, pp. 357-365, 2015.

[11] M. Bevacqua, R. Palmeri, L. Di Donato, L. Crocco, T. Isernia, "Microwave Imaging via Iterated Virtual Experiments", in *Antennas and Propagation (EuCAP), 2016 10th European Conference on*.

[12] van den Berg P. M. and Kleinman R. E., **“**A contrast source inversion method,**”** *Inv. Prob.*, vol. 13, pp. 1607**–**1620, 1997.

[13] A. Abubaker and P. M. Van Den Berg, "Total variation as a multiplicative constraint for solving inverse problems," in *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1384-1392, Sep 2001.

[14] Candes, E. J., T. Strohmer, and Vl. Voroninski, “Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming”, *Communications on Pure and Applied Mathematics* 66.8 1241-1274, 2013.

[15] Bucci O. M. and Isernia T., “Electromagnetic inverse scattering: Retrievable nformation and measurement strategies”, *Radio Science*, vol. 32, pp. 2123–2138, 1997.

[16] E. J Candes., Y. C Eldar, D. Needell, and P. Randall, “Compressed sensing with coherent and redundant dictionaries”, *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59-73, 2011.

[17] J. Yang, J. Wright, T. S. Huang and Y. Ma, "Image Super-Resolution Via Sparse Representation," in *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861-2873, Nov. 2010.

[18] M. Elad and M. Aharon, "Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries," in *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736-3745, Dec. 2006.

[19] Bevacqua, M., Isernia, T., Crocco, L., Di Donato, L., "A $(CS)^2$ approach to inverse scattering," in *Antenna Measurements & Applications (CAMA), 2014 IEEE Conference on*, vol., no., pp.1-3, 16-19 Nov.2014.

[20] Bevacqua, M.; Crocco, L.; Di Donato, L.; Isernia, T, "Exploiting compressive sensing for non-linear inverse scattering," in *Radio Science Conference (URSI AT-RASC), 2015 1st URSI Atlantic*, vol., no., pp.1-1, 16-24 May 2015.

[21] V. Pascazio and G. Ferraiuolo, "Statistical regularization in linearized microwave imaging through MRF-based MAP estimation: hyperparameter estimation and image computation," in *IEEE Transactions on Image Processing*, vol. 12, no. 5, pp. 572-582, May 2003.

[22] M. S. Tong, K. Yang, W. T. Sheng and Z. Y. Zhu, "Efficient Reconstruction of Dielectric Objects Based on Integral Equation Approach With Gauss–Newton Minimization," in *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4930-4937, Dec. 2013.

[23] D’Urso M., Isernia T., and Morabito A., “On the solution of 2-d inverse scattering problems via source-type integral equations,” *IEEE Trans. Geosci. Remote Sens.*, 48(3):1186–1198, 2010.

[24] Isernia, T., Pascazio, V., Pierri, R., "A nonlinear estimation method in tomographic imaging," in *Geoscience and Remote Sensing, IEEE Transactions on* , vol. 35, no. 4, pp. 910-923, 1997.

[25] N. Joachimowicz, J. J. Mallorqui, J. C. Bolomey, and A. Broquetas, “Convergence and stability assessment of Newton Kantorovich reconstruction algorithms for microwave tomography”, *IEEE Trans. Med. Imag.*, vol. 17, no 4, pp. 562-569, 1998.

[26] Bertero M. and Boccacci P., “Introduction to Inverse Problems in Imaging”, Institute of Physics, Bristol, UK, 1998.

[27] O. M. Bucci, N. Cardace, L. Crocco, T. Isernia, “Degree of nonlinearity and a new solution procedure in scalar two-dimensional inverse scattering problems”, *J. Opt. Soc. Am. A*, vol. 18, no. 8, pp. 1832-1843, 2001.

[28] Wolpert, David H., and William G. Macready. "No free lunch theorems for optimization." *Evolutionary Computation, IEEE Transactions on,* vol. 1, no. 1, pp. 67-82, 1997.

[29] G. Cardano, “Artis magnae sive de regulis algebraicis, liber unus,”1545 (Translated in “Ars Magna or the Rules of Algebra” New York, NY, USA: Dover, 1993).

[30] Chen S., Donoho D., M. Saunders D., “Atomic decomposition by basis pursuit,” *SIAM J. Sci. Comput*., vol. 20, no. 1, pp. 33–61, 1999.

[31] Tibshirani R., “Regression shrinkage and selection via the lasso,” *J. Roy. Stat. Soc. Ser*., vol 58, no.1, pp. 267-288, 1996.

[32] Chen X., "Subspace-Based Optimization Method for Solving Inverse-Scattering Problems," in *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 1, pp. 42-49, 2010.

[33] Xu K., Zhong Y., Song R., Chen X., Ran L., "Multiplicative-Regularized FFT Twofold Subspace-Based Optimization Method for Inverse Scattering Problems," in *Geoscience and Remote Sensing, IEEE Transactions on*, vol.53, no.2, pp.841-850, 2015.

[34] P. M. van den Berg, A. Abubakar, and J. T. Fokkema, Multiplicative regularization for contrast profile inversion, *Radio Science*, vol. 38, no. 2, 2003.