Group sparsity penalized contrast source solution method for 2-D non-linear inverse scattering

Yarui Zhang, Marc Lambert, Aurélia Fraysse, Dominique Lesselier

Group Sparsity Penalized Contrast Source Solution Method for 2-D Non-Linear Inverse Scattering

YARUI ZHANG [1], MARC LAMBERT [1], AURÉLIA FRAYSSE[2], AND DOMINIQUE LESSELIER [2] (Senior Member, IEEE)

[1]Université Paris-Saclay, CentraleSupélec, CNRS, Laboratoire de Génie Electrique et Electronique de Paris, 91192 Gif-sur-Yvette, France

[2]Laboratoire des signaux et systèmes, Université Paris-Saclay, CNRS, CentraleSupélec, 91190 Gif-sur-Yvette, France

CORRESPONDING AUTHOR: Y. ZHANG (e-mail: yarui.zhang@geeps.centralesupelec.fr)

ABSTRACT A group sparsity penalized CSI in the wavelet domain is proposed to alleviate ill-posedness within the framework of a contrast-source inversion (CSI) method. It is then applied to the retrieval of a large inhomogeneous dielectric scatterer from time-harmonic single-frequency data. As dependency exists between wavelet coefficients at different scales, referred to as the parent-child relationship, it enables to yield the wavelet quad tree structure. Therefore, wavelet coefficients can be regarded not only as pixel-wise sparse, but also group-wise sparse. Focus is put on using the dual-tree complex wavelet transform ($\mathbb{C}$WT) to properly achieve the sought-after group-wise sparse representation of the spatial distribution of the contrast. It provides a $\ell_{2,1}$ norm which is added to the standard cost functional to enforce group sparsity onto the wavelet coefficients of the spatially-varying contrast. The replication strategy is combined with the proximal method in order to solve the overlapping group penalized problem. Simulations from synthetic data in different configurations with in particular different signal-to-noise ratios illustrate pros and cons of the proposed method. The approach is shown to overcome the standard CSI method in demanding situations. Comparisons with the discrete wavelet transform (DWT) as usually performed and the ℓ_1 norm confirm the advantage of the proposed methodology.

INDEX TERMS Contrast source inversion, dual-tree complex wavelet transform, group sparsity, wavelet tree, $\ell_{2,1}$ norm.

I. INTRODUCTION

INVERSE scattering problems (ISPs) [1] are about the retrieval of the characteristics of an unknown or partially unknown scatterer, such as its geometry or the distribution of its physical parameters, from the knowledge of the fields it scatters when probed by known sources (the calculation of such fields results from the solution of a direct scattering problem). Those are relevant to various applications [2], including non-destructive testing [3], medical imaging [4], and remote sensing [5].

A wide range of solution techniques has been applied to tackle the ISPs, which may be grouped into two categories, deterministic methods and stochastic methods. Some deterministic methods like Rytov [6] and Born [7] approximations employ first-order approximations to linearize the

problem. This type of method requires less computational time, but they are mostly used for weak-enough scatterers. To more rigorously deal with high contrast scatterers, a number of iterative methods like the contrast source inversion method (CSI) [8], the distorted Born iterative method (DBIM) [9], and the subspace-based optimization method (SOM) [10], and their many variants, have been developed. In the iterative process which those methods always entail, the optimal solution is sought step by step according to some search direction and the step length is determined at each iteration.

Compared with first-order-approximation-based methods, the latter suffer from higher computational burden yet enjoy a much wider range of applications. However, deterministic methods are essentially local optimization methods. They are

very efficient in finding local optimal solutions, but hardly have the ability to get global optimal solutions in multi-extremal problems unless provided with a reasonable initial guess, i.e., ill-posedness in Hadamard's sense remains an issue.

Stochastic approaches, e.g., colony optimization method [11] and genetic algorithms [12], to mention but a few among many, have a better capability to find global optimal solutions, yet they are generally less popular in the community due to their often large computational burden—one will not dwell further on those, and focus onto the non-linearized deterministic approaches, and ways one can overcome ill-posedness by benefiting from proper representations of the sought scatterer and from suitable priors.

To quickly summarize, the inverse problem is essentially an optimization problem in which one wishes to minimizes the discrepancy between obtained results and expected ones. However, ill-posed inverse problems usually call for additional penalty terms to regularize the optimization faced with.

To that intent, various regularization schemes can be applied [13], generally inserted as prior information or constraints, such as positivity [8], total variation (TV) [14], or sparsity [15]. In [16], the TV regularization is implemented within the CSI framework as a multiplicative regularization term that leads to a better reconstruction quality and improves robustness to noise. In the DBIM algorithm, Tikhonov regularization [17] is employed by default in order to circumvent the instability of the problem. Edge-preserving regularization has also been introduced into ISPs [18], [19] to improve the reconstruction quality with respect to noisy corrupted data while preserving the edges of the images. In [19], Huber regularization is incorporated with CSI method as a smoothness constraint.

Using sparsity as prior information is becoming increasingly popular due to its potential to effectively solve the inverse problem and its robustness to noise. A direct way to find the sparse solution in an optimization problem is to minimize the number of elements of the entries, which can be realized by adding a ℓ_0 penalty term into the cost function the minimum of which is sought. However, the optimization of the ℓ_0 norm [20] that counts the number of non-zeros elements of a vector is well known as a NP-hard problem and thus it is not computationally practical. Usually, such an approach can be replaced by indirect approaches including the use of ℓ_1 norm [21], the solution of which has been verified to be very close to the sparsest solution.

Sparsity has been widely investigated and incorporated within classical computational methods in order to solve ISPs [22]. In [23], the fast iterative shrinkage-thresholding algorithm (FISTA) [24] is combined with a modified gradient method (MGM) to enforce the sparsity in both non-sparse and sparse domain, where the discrete wavelet transform (DWT) [25] is employed to reach a sparse representation of the unknown. In [23], an iterative algorithm incorporating

a Tikhonov functional with a sparsity promoting ℓ_1 term is proposed, which allows a sharp reconstruction in the sparse domain. Iterative shrinkage thresholding algorithm (ISTA) [24] is implemented within the framework of a boundary integral method in [26] to solve the optimization problem with zeroth/first-norm penalty term. The proposed method is proved to be able to yield a sharper and more accurate reconstruction with benefiting from a faster convergence speed. In [27], a joint ℓ_1-ℓ_2 norm regularized BIM approach in the wavelet domain is proposed, where appropriate weights adaptively updated at each iteration are allocated to wavelet coefficients at different levels. This scheme enables the performance to be independent of the initial weights. In [28] and [29], DBIM is incorporated with ℓ_2-ℓ_1 and ℓ_1-ℓ_0 joint norm, respectively, solved by various optimization techniques [30], [31]. In [32], two sparsity-enhanced approaches within the CSI framework are proposed, the ℓ_1 norm being used as weighted penalty term and constraint. Therein, the stationary wavelet transform (SWT) [33] is adopted to project the unknown onto a redundant dictionary. Compressive sensing (CS) based reconstruction algorithms have also been exploited [34], [35]. In [36], the commonly used DWT offers a sparse representation of the profile to be reconstructed, then the linearized inverse scattering problem within the first-order Born approximation framework is solved with the Bayesian version CS-based procedure.

The aforementioned ℓ_0, ℓ_1 and ℓ_2 norms regularize each element individually regardless of its spatial position. When the elements are not only sparse but also have potential structure features, using these norms as regularization terms does not enable to extract the structure information. Mixed norms [37] such as $\ell_{2,1}$ or $\ell_{\infty,1}$ are usually employed to impose both sparsity and structural prior information in the optimization problem [38]. In [39], as the row vectors of the contrast source share a common non-zero support, a two-step sparse recovery process is investigated where the $\ell_{2,1}$ norm is to enforce joint sparsity on the contrast source.

In the present contribution, a group sparsity penalized approach within the framework of CSI is proposed. The $\ell_{2,1}$ norm is added to the cost functional of CSI to impose group sparsity on the contrast in the transformed domain. At each iteration, the contrast is decomposed by the dual-tree complex wavelet transform ($\mathbb{C}$WT) [40], which both makes it possible to represent the contrast in a sparse manner and yields the structure information of the contrast. Herein, the structure information provided by dual-tree $\mathbb{C}$WT lies in the wavelet tree structure [41], led by the parent-child relationship [42] between wavelet coefficients across different scales.

The contribution is organized as follows. The forward problem is described in Section II. The proposed methods are developed in Section III. Numerical illustrations are discussed in Section IV. Conclusion and perspectives are is in Section V. Preliminary bricks of the analysis are found in summaries of conferences [43] and [44], the first with

focus on the multi-scale wavelet-based method, the second with focus on how soft-thresholding can help to sparsify the reconstruction.

II. FORWARD PROBLEM

The scenario considered here is of a time-harmonic two-dimensional electromagnetic scattering problem in the transverse magnetic (TM) polarization with time convention $e^{-i\omega t}$, ω angular frequency. The inhomogeneous linear isotropic scatterer is embedded within a homogeneous linear isotropic medium $\mathcal{D}$ with permittivity of air ϵ_0 and permeability μ_0. $\epsilon_r(\mathbf{r})$ and $\sigma(\mathbf{r})$ denote the relative permittivity and conductivity of the scatterer with $\mathbf{r} \in \mathcal{D}$ as a given observation point. TM waves generated by N_s ideal line sources located outside the scatterer at positions $\mathbf{r}_s$ illuminate it. For each illumination, the scattered fields are collected by N_r ideal receivers located at positions $\mathbf{r}_r$ along a circle of observation $\mathcal{S}$, not intersecting the scatterer, and not necessarily jointive with the sources.

The scattered electric field $E^{\text{diff}}(\mathbf{r}_r, \mathbf{r}_s)$ collected by a receiver placed at $\mathbf{r}_r$ and associated with the source placed at $\mathbf{r}_s$ satisfies the integral equation

$$E^{\text{diff}}(\mathbf{r}_r, \mathbf{r}_s) = \int_D G(\mathbf{r}_r, \mathbf{r}')J(\mathbf{r}', \mathbf{r}_s)d\mathbf{r}', \ \mathbf{r} \in \mathcal{D} \quad (1)$$

with

$$J(\mathbf{r}, \mathbf{r}_s) = \chi(\mathbf{r})E^{\text{tot}}(\mathbf{r}, \mathbf{r}_s) \quad (2)$$

$G(\mathbf{r}, \mathbf{r}')$ is the Green's function which represents the electromagnetic response to a line source radiating in free space. In the case of two dimensions, it is given by $G(\mathbf{r}, \mathbf{r}') = \frac{-i\omega\mu_0}{4} H_0^{(1)}(k_B\|\mathbf{r} - \mathbf{r}'\|)$, where $H_0^{(1)}$ is the zero-order Hankel function of the first kind.

The contrast function is defined as

$$\chi(\mathbf{r}) = k^2(\mathbf{r}) - k_B^2 \quad (3)$$

where $k^2(\mathbf{r}) = \omega^2\epsilon_0\epsilon_r(\mathbf{r})\mu_0 + i\omega\mu_0\sigma(\mathbf{r})$, $k_B^2 = \omega^2\epsilon_0\mu_0$. $J(\mathbf{r}, \mathbf{r}_s)$ and $E(\mathbf{r}, \mathbf{r}_s)$ are the equivalent current and the total electric field induced within the object by the incident wave.

The total electric field can be obtained according to

$$E^{\text{tot}}(\mathbf{r}, \mathbf{r}_s) = E^{\text{inc}}(\mathbf{r}, \mathbf{r}_s) + \int_{\mathcal{D}} G(\mathbf{r}, \mathbf{r}')J(\mathbf{r}', \mathbf{r}_s)d\mathbf{r}', \ \mathbf{r} \in \mathcal{D} \quad (4)$$

where $E^{\text{inc}}(\mathbf{r}, \mathbf{r}_s)$ is the incident field.

Using a pulse-basis point matching method of moments (MoM), the domain $\mathcal{D}$ is discretized in N small square pixels so that the electric field and the contrast can be considered as constants within each pixel. The discretized version of the previous equations stands as

$$E_{i_r, i_s}^{\text{diff}} = \sum_{i=1}^{N} G_{S; i_r, i}J_{i, i_s} \quad (5)$$

$$E_{i, i_s}^{\text{tot}} = E_{i, i_s}^{\text{inc}} + G_{D; i, i}J_{i, i_s} \quad (6)$$

with $i_s = 1, \ldots, N_s$, $i_r = 1, \ldots, N_r$, $i = 1, \ldots, N$, and $J_{i, i_s} = \text{diag}(\chi_i)E_{i, i_s}^{\text{tot}}$ the contrast source. Subscripts D and S of the operators indicate the location of the point $\mathbf{r}$, and the operators are identical in all other aspects:

$$G_{D, S}J_{i, i_s} = k_B^2 \int_{\mathcal{D}} G(\mathbf{r}, \mathbf{r}')J(\mathbf{r}', \mathbf{r}_s)d\mathbf{r}' \quad \mathbf{r} \in \mathcal{D} \text{ or } \mathbf{r} \in \mathcal{S} \quad (7)$$

The forward problem is defined as the calculation of $E_{i_r, i_s}^{\text{diff}}$ from the knowledge of χ_i and the inverse scattering problem is to retrieve χ_i from $E_{i_r, i_s}^{\text{diff}}$, which is nonlinear and ill-posed.

III. INVERSION ALGORITHM

Due to the non-linearity and ill-posedness of ISPs, classical iterative methods generally yield local minima, as already underlined. Sparsity regularization is proposed here to incorporate a prior information on the unknown into the CSI method. In this section, the notion of group sparsity and how to achieve it are first introduced. The dual-tree $\mathbb{C}$WT is then explained, as is used as the sparsity-inducing transform as well as to provide the structural dependencies of the data. Finally, the implementation of the group sparsity regularizer into the CSI framework is considered.

A. SPARSITY INDUCING NORMS
1) SPARSITY THROUGH A SINGLE NORM

A vector x of dimension M can be considered as *sparse* if most of its elements are zero, or almost zero. For an optimization problem, sparsity can be directly achieved by adding a regularization term that penalizes the number of non-zero elements of the vector. Referring to

$$x = \arg\min_x \{f + \|x\|_q\} \quad (8)$$

$q = 0$ leads to the penalization on x using the ℓ_0 norm defined as $\|x\|_0 = \#\{i = 1, \ldots, M | x_i \neq 0\}$.

Solving optimization problems with the non-convex ℓ_0 norm is as difficult as trying out all possible non-zeros supports of coefficients. A tractable approach to solving the ℓ_0 minimization problem is to replace it with a reasonable and convex approximation. When $q = 1$, the second term in (8) is the ℓ_1 norm, which is a natural approximation of ℓ_0 norm, defined as $\|x\|_1 = \sum_{i=1}^{M} |x_i|$.

Due to the presence of the ℓ_1 norm, the whole cost function becomes non-differentiable so that the existence of the gradient is not always guaranteed. One possible manner to deal with this kind of optimization problem is to use the subgradient technique [45], which allows to approximate the gradient of the non-differentiable function at non-derivable points. What one adopts in the present work is another well-known approach called proximal algorithm [46]. For the optimization problem $\min_x \{f(x) + g(x)\}$, x can be updated at iteration k according to

$$x^k = \text{prox}_{tg(\cdot)}\left(x^{k-1} - t\nabla f\left(x^{k-1}\right)\right) \quad (9)$$

wherein $\text{prox}_{tg(\cdot)}(x)$ is the *proximal operator* of $g(\cdot)$ at point x with t the descent step size. When $g(x) = \|x\|_1$, the corresponding $\text{prox}_{tg(\cdot)}(x)$ is the soft-thresholding operator.

Sparsity Through the Mixed Norm: In the previous formulation with ℓ_1 norm, every element is regarded as independent from the others. In some situations, however, the elements of a vector share the same support, or dependency relationships exist between different elements, which allows them to be divided into subgroups. It is then easy to see that the elements within a given group can be selected or deleted simultaneously, benefiting from this prior structure information. This is the so-called sparse group Lasso—least absolute shrinkage and selection operator—problem, which can be exploited using the mixed sparsity inducing norms, e.g., the $\ell_{2,1}$ or $\ell_{\infty,1}$ norms.

In this work, the $\ell_{2,1}$ norm is used. Let x_{g_i} be the sub-vectors of x indexed by g_i, $i = 1, \ldots, s$ where s is the number of subgroups. The mixed $\ell_{2,1}$ norm is defined as

$$\|x_g\|_{2,1} = \sum_{i=1}^{s} \|x_{g_i}\|_2 \tag{10}$$

For a non-overlapping group Lasso problem, the proximal operator of $\ell_{2,1}$ has a closed-form solution [37]

$$x'_{g_i} = x_{g_i} \times \left(1 - \frac{\lambda}{\|x_{g_i}\|_2}\right)_+, \quad \forall i \tag{11}$$

As for the overlapping group Lasso problem where an element can appear in different sub-groups, the problem becomes more complicated since the standard group Lasso optimization strategies cannot be used directly.

Much attention has been paid to exploiting the overlapping situations, especially for those with a hierarchical tree structure. In [47], an approach to deal with tree structured group Lasso problems is offered, in which the Moreau-Yosida regularization is employed and its analytical solution is given. In [48], the dual approach is used, and the proximal operator of the norm associated with overlapping hierarchical groups proved to be exactly computable, which can be regarded as the composition of elementary proximal operators. In [49], [50], [51], replication strategies are proposed which consist in decoupling the overlapping groups from each other by replicating the elements appearing repeatedly. With the decoupled groups, this problem can be tackled by using the same proximal operator as in the non-overlapping case.

B. DUAL-TREE COMPLEX WAVELET TRANSFORM

Dual-tree $\mathbb{C}$WT is an enhancement of the discrete wavelet transform (DWT), with some important additional properties. The DWT of a signal x is the response obtained by passing it through a filter bank. There are two filters involved: a scaling filter, which is the low-pass filter, and a wavelet filter, which is the high-pass filter. The transformed coefficients of a 2-D DWT can be divided into approximation and detail coefficients. More precisely, three subbands of detail coefficients are offered, henceforth denoted as *HL*, *LH* and *HH* which define the horizontal features, vertical features and diagonal features of the original image, respectively.
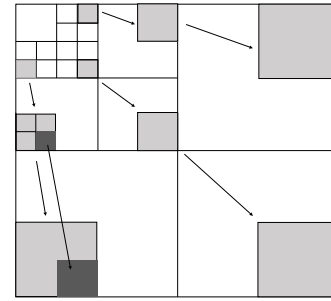

FIGURE 1. Wavelet quadtree structure of transformed 2-D image. For a detail coefficient located at (x, y) at the decomposition level *L*, its four *children* will be in the next decomposition level *L* − 1, located at $(2x - 1, 2y - 1)$, $(2x - 1, 2y)$, $(2x, 2y - 1)$, $(2x, 2y)$, resp.

The DWT coefficients have two main properties. First, detail transform coefficients are naturally sparse, which enables near-optimal image processing based on simple thresholding operation. Second, coefficients across different scales can be regarded as statistically dependent. If a wavelet coefficient is large/small, the coefficients at the same orientation in the next adjacent scale are likely to be large/small. This is the so-called parent-child relationship. Due to the above characteristic and its multi-scale structure, this dependency can be modeled as a set of trees, the roots of which are wavelet coefficients from the coarser scale of the decomposition. Fig. 1 illustrates the wavelet coefficients of a given image with the decomposition level $L = 3$. Denote three subbands of detail coefficients at level *L* as $\{HL, LH, HH\}_L$. Each coefficient in $\{HL, LH, HH\}_L$ possesses four *children* coefficients at the same orientation in $\{HL, LH, HH\}_{L-1}$. Hence, a quadtree structure can be established based on this relationship between wavelet coefficients across different scales.

However, DWT with real wavelets suffers from some fundamental shortcomings. First, the wavelet coefficients tend to oscillate positively and negatively around the singularities, and as a result, even a small shift will perturb the wavelet coefficient, which complicates the wavelet based processing. Moreover, any wavelet coefficient processing (like thresholding, filtering) will lead to artifacts in the reconstructed image. And classical 2-D DWT does not allow to efficiently represent non-horizontal or non-vertical singular features.

Inspired by the Fourier transform based on complex-valued oscillating sinusoids, one goes with the dual-tree $\mathbb{C}$WT [40] using two real DWTs, one which produces the real part of the transform, the other the imaginary part (Fig. 2). Additionally, the filter pairs of two DWTs have to be carefully designed so that the imaginary part of the complex wavelet is approximately the Hilbert transform of its real part.

Dual-tree $\mathbb{C}$WT has good shift invariance and directional selectivity. In the 2-D case, it gains additional properties. Differently from DWT, dual-tree $\mathbb{C}$WT has six directions that allow to capture more oriented singularities of an image without the checkerboard artifact that appears in the diagonal
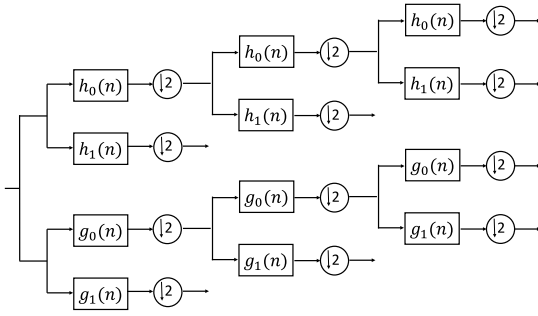
FIGURE 2. Analysis filter bank (FB) for 1-D dual-tree $\mathbb{C}$WT. ($h_0(n)$, $h_1(n)$) and ($g_0(n)$, $g_1(n)$) are low-pass/high-pass filter pairs for the upper and lower FB, followed by a down-sampling operator.

wavelet of DWT. Moreover, it has been proved in [52] that the near shift invariance of the dual-tree $\mathbb{C}$WT leads to better persistence of the magnitudes around the edgy regions, which ensures stronger dependence in inter-scale neighborhoods, in other words, which strengthens the relationship between *parent* and *children* wavelet coefficients.

Notice that the 2-D dual-tree $\mathbb{C}$WT is four times more expensive than DWT with real wavelets. The implementation is still very efficient since only additions and subtractions for respective subbands of all DWTs are required.

C. CONTRAST SOURCE INVERSION

The CSI method is one of the most used methods to tackle inverse scattering problems, as already noted. First, based on (5) and (6), the state and data equation stand as

$$J_{i,i_s} = \operatorname{diag}(\chi_i)\left[E_{i,i_s}^{\text{inc}} + G_{D,i,i}J_{i,i_s} \right] \tag{12}$$

$$E_{i,i_s}^{\text{diff}} = \sum_{i=1}^{N} G_{S;i_r,i}J_{i,i_s} \tag{13}$$

The cost functional is a linear combination of normalized mismatches in the data and object equations:

$$F\left(J_{i,i_s}, \chi_i\right) = \frac{\sum_{i_s}\sum_{i_r}\left\| \xi_{i_r,i_s} - \sum_i G_{S;i_r,i}J_{i,i_s} \right\|^2}{\sum_{i_s}\sum_{i_r}\left\| \xi_{i_r,i_s} \right\|^2}$$
$$+ \frac{\sum_{i_s}\sum_i\left\| \operatorname{diag}(\chi_i)E_{i,i_s}^{\text{inc}} - J_{i,i_s} + \operatorname{diag}(\chi_i)G_{D;i,i}J_{i,i_s} \right\|^2}{\sum_{i_s}\sum_i\left\| \operatorname{diag}(\chi_i)E_{i,i_s}^{\text{inc}} \right\|^2} \tag{14}$$

ξ_{i_r,i_s} being the measured scattered fields.

The contrast sources J_{i,i_s} and the contrast χ_i are alternatively reconstructed with a conjugate gradient iterative method. At each iteration, the contrast sources are first updated by minimizing the entire cost functional, then the contrast χ_i is determined by minimizing the error in the state equation. The minimization at each iteration is linear. A positivity constraint is imposed on the contrast via the projection method as described in [8]. The initial guess for the contrast source and for the contrast function is obtained by back-propagation.
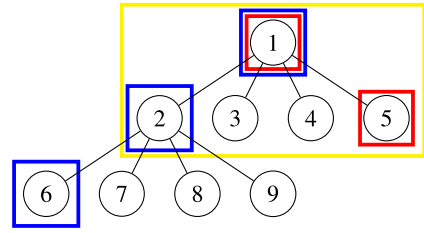

FIGURE 3. Grouping arrangement of wavelet coefficients.

Employing the dual-tree $\mathbb{C}$WT, the contrast χ_i is decomposed in the wavelet domain at each iteration. Sub-groups can be constituted based on the dependency between *parent* and *children* coefficients. Thus, the group sparsity regularization can be applied on the contrast in the wavelet-domain, which both promotes the sparsity on wavelet coefficients of χ_i, and the persistence of magnitudes across scales.

With this regularization, the cost functional becomes

$$F\left(J_{i,i_s}, \beta_i\right) = \frac{\sum_{i_s}\sum_{i_r}\left\| \xi_{i_r,i_s} - \sum_i G_{S;i_r,i}J_{i,i_s} \right\|^2}{\sum_{i_s}\sum_{i_r}\left\| \xi_{i_r,i_s} \right\|^2}$$
$$+ \frac{\sum_{i_s}\sum_i\left\| (\mathcal{W}^*\beta_i)\left(E_{i,i_s}^{\text{inc}} + G_{D;i,i}J_{i,i_s} \right) - J_{i,i_s} \right\|^2}{\sum_{i_s}\sum_i\left\| (\mathcal{W}^*\beta_i)E_{i,i_s}^{\text{inc}} \right\|^2}$$
$$+ \lambda\left\| \beta_g \right\|_{2,1} \tag{15}$$

Define $\mathcal{W}$ as the dual-tree $\mathbb{C}$WT operator and $\mathcal{W}^*$ its inverse. $\beta = \mathcal{W}\chi$ is the wavelet coefficients of χ, β_g is the set of sub-vectors of β associated with all sub-groups.

In Fig. 3, given that the elements surrounded by boxes of the same color belong to the same group, three groups appear: $\{1, 5\}$ (in red), $\{1, 2, 6\}$ (in blue) and $\{1, 2, 3, 4, 5\}$ (in yellow), representing three ways to group the wavelet coefficients.

For a wavelet coefficient A, the sub-groups can be constituted as follows:

- $\{A,\ parent\ of\ A\}$ (in red)
- $\{A,\ all\ ancestors\ of\ A\}$ (in blue)
- $\{A,\ all\ children\ of\ A\}$ (in yellow)

All three grouping choices form overlapping sub-groups. Thus, the regularized cost functional must be tackled with specific strategies. Herein, a replication approach proposed in [51] is adopted. Denote f_{d} as the normalized data error in (15) and f_{o} the normalized object error. Enforcing the replication strategy, the previous cost functional becomes

$$F(J, \beta) = f_{\text{d}}(J) + f_{\text{o}}(J, \beta) + \lambda_1\left\| z_g \right\|_{2,1} + \frac{\lambda_2}{2}\left\| z - G_{\beta z}\beta \right\|_2^2 \tag{16}$$

z is a "flattened" version of β, where all overlapping groups in β become non-overlapping by replicating the crossed elements. $G_{\beta z}$ is the mapping matrix that yields z from β.

In equation (16), two hyperparameters λ_1 and λ_2 have to be determined. Following [23], they are chosen adaptively according to the magnitudes of wavelet coefficients, and

Algorithm 1 CSI With Group Sparsity Regularization

1: Initialize J and β by back-propagation:

2: $J_{i_s}^{\mathrm{bp}} = \dfrac{\|G_S^* \xi_{i_s}\|^2}{\|G_S G_S^* \xi_{i_s}\|^2} G_S^* \xi_{i_s}$

3: $\beta^{\mathrm{bp}} = \mathcal{W}\left\{ \dfrac{\sum_{i_s=1}^{N_s} J_{i_s}^{\mathrm{bp}} E_{i_s}^{\mathrm{tot*}}}{\sum_{i_s=1}^{N_s} \|E_{i_s}^{\mathrm{tot}}\|^2} \right\}$

4: **for** $k = 1 : iter_{\max}$ **do**

5:     Calculate data error $\rho_{i_s,k} = \xi_{i_s} - G_S J_{i_s,k-1}$ and

6:     object error $r_{i_s,k} = (\mathcal{W}^* \beta_{k-1}) E_{i_s,k-1}^{\mathrm{tot}} - J_{i_s,k-1}$

7:     **Update the contrast sources**

8:     Calculate the gradient $g_{i_s,k}^J$

$$= -\frac{G_S^* \rho_{i_s,k}}{\sum_{i_s} \sum_{i_r} \|\xi_{i_r,i_s}\|^2} - \frac{r_{i_s,k} - G_D^* \beta_{k-1} r_{i_s,k}}{\sum_{i_s} \sum_i \|(\mathcal{W}^* \beta_i) E_{i,i_s}^{\mathrm{inc}}\|^2}$$

9:     Calculate the search direction $v_{i_s,k}$

$$= g_{i_s,k}^J + \frac{\Re\left\{ \sum_{i_s} \left\langle g_{i_s,k}^J, g_{i_s,k}^J - g_{i_s,k-1}^J \right\rangle \right\}}{\sum_{i_s} \left\langle g_{i_s,k-1}^J, g_{i_s,k-1}^J \right\rangle} v_{i_s,k-1}$$

10:     Determine the step size $\alpha_k^J =$

$$\frac{-\Re\left\{ \sum_{i_s} \left\langle g_{i_s,k}^J, v_{i_s,k} \right\rangle \right\}}{\dfrac{\sum_{i_s} \|G_S v_{i_s,k}\|^2}{\sum_{i_s} \sum_{i_r} \|\xi_{i_r,i_s}\|^2} + \dfrac{\sum_{i_s} \|v_{i_s,k} - (\mathcal{W}^* \beta_{k-1}) G_D v_{i_s,k}\|^2}{\sum_{i_s} \sum_i \|(\mathcal{W}^* \beta_i) E_{i,i_s}^{\mathrm{inc}}\|^2}}$$

11:     Update $J_{i_s,k} = J_{i_s,k-1} + \alpha_k^J v_{i_s,k}$

12:     **Update the contrast in the wavelet domain**

13:     Update the total field $E_{i_s,k}^{\mathrm{tot}} = E_{i_s}^{\mathrm{inc}} + G_D J_{i_s,k}$

14:     Update $\beta_{k-0.5} = \mathcal{W}\left\{ \dfrac{\sum_{i_s=1}^{N_s} J_{i_s,k} E_{i_s,k}^{\mathrm{tot*}}}{\sum_{i_s=1}^{N_s} \|E_{i_s,k}\|^2} \right\}$

15:     Update z_{g_k}

$$= (G_{\beta z} \beta_{k-0.5})_g \times \left(1 - \frac{\lambda_1 / \lambda_2}{\|(G_{\beta z} \beta_{k-0.5})_g\|_2} \right)_+$$

16:     Update $\beta_k = \beta_{k-0.5} - \nabla_\beta \left(\dfrac{\lambda_2}{2} \|z - G_{\beta z} \beta\|_2^2 \right)$

17:     Update hyperparameters

18:     Enforce positivity constraint on $\chi_k = \mathcal{W}^* \beta_k$

19: **end for**

updated at each iteration. Let us denote $P_{i\%}(\centerdot)$ the i-th percentile of an vector, λ_1 is set to $P_{q\%}(\beta)$ and λ_2 is associated with λ_1 where the ratio of λ_2 to λ_1 is set to $P_{q\%}(\|z_g\|_{2,1})$. q, which remains the only hyperparameter to be chosen, is fixed via a "trial-error" scheme according to the applied grouping strategy.

The optimization procedure is sketched in Algorithm 1.

IV. NUMERICAL ILLUSTRATIONS

In this section, numerical experiments on different synthetic models are considered. For all models, the region of interest $\mathcal{D}$ is of 2 m × 2 m, and is discretized into N square cells. Cell sizes for the direct and inverse problem are different in order to avoid "inverse crime". For the direct problem, $\mathcal{D}$ is discretized into 50×50 cells, and for the inverse one, $N = 64 \times 64$. N_r receivers and N_s transmitters are evenly distributed on a circle of radius $r = 3$ m. Save otherwise

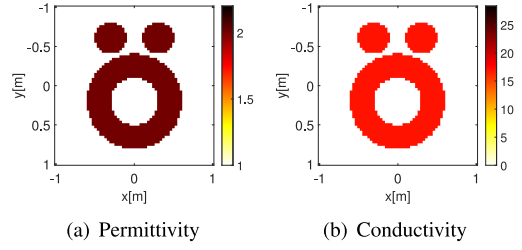(a) Permittivity   (b) Conductivity

FIGURE 4. "Austria" model—Relative permittivity $\epsilon(r)$ and conductivity $\sigma(r)$ (mS/m). The axes of the figures in the article are the same and will not be repeated due to space limitation.
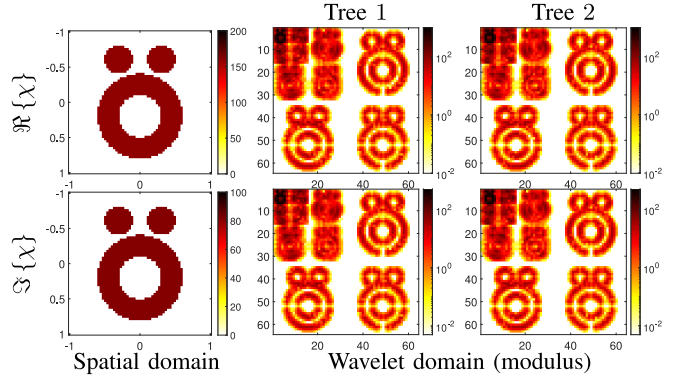

FIGURE 5. Contrast of "Austria" model in the spatial domain (column 1) and the wavelet domain (column 2, 3) using the dual-tree $\mathbb{C}$WT with a tree pair (Tree 1, Tree 2). Logarithmic color scale is used in column 2 and 3.

specified, the observed data are corrupted by a random white Gaussian noise with SNR = 20 dB. The embedding medium is of $\epsilon_b = 1$.

A. RECONSTRUCTION OF THE SYNTHETIC AUSTRIA PROFILE

The synthetic "Austria" model is made of two disks and one ring. The disks of radius 0.2m are centered at $(0.3, -0.6)$ m and $(-0.3, -0.6)$ m. The ring has an exterior radius of 0.6m and an inner radius of 0.3m, and is centered at $(0, 0.2)$ m. The scatterer has uniform permittivity and conductivity 2 and 17mS/m. Fig. 4 shows real and imaginary parts of the model. Fig. 5 depicts the contrast χ in the spatial domain and its wavelet coefficients obtained with the dual-tree $\mathbb{C}$WT. The frequency of operation is 600 MHz.

At each iteration, the contrast χ is decomposed by the dual-tree $\mathbb{C}$WT. The filter bank (FB) of the first stage of the dual-tree $\mathbb{C}$WT is required to be different from those of the succeeding stages. Let us denote $l_{\max}$ the maximum length of the decomposition filters and N_x and N_y the row and column dimensions of the image respectively. The level of decomposition L then should be a positive integer holding that:

$$L \leq \log_2\left(\frac{\min(N_x, N_y)}{l_{\max}} \right) + 1 \tag{17}$$

Hereafter, Farras filters [53] are used for the first stage and Kingsbury Q-shift 6-tap filters [54]. So here, $N_x = N_y = 64$,
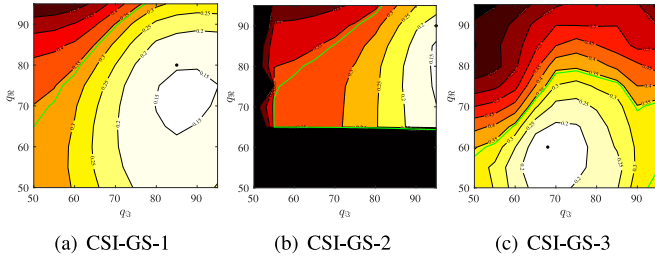
FIGURE 6. Parametric study for the choice of $q_{\Re}$ and $q_{\Im}$. The final error Err(χ) is plotted as a function of $q_{\Re}$ and $q_{\Im}$ and compared to the one of CSI (green line).

TABLE 1. Choice of q for different methods.

	CSI-GS-1	CSI-GS-2	CSI-GS-3
$q_{\Re}$	80	90	60
$q_{\Im}$	85	95	68

TABLE 2. Relative error with different $N_s \times N_r$ – SNR = 20 dB.

$N_s \times N_r$	Error	CSI	CSI-GS-1	CSI-GS-2	CSI-GS-3
32×32	Err(χ)	0.34	**0.15**	0.16	0.16
	Err($\Re\{\chi\}$)	0.27	**0.14**	**0.14**	0.15
	Err($\Im\{\chi\}$)	0.61	**0.21**	0.24	0.23
18×26	Err(χ)	0.42	**0.17**	0.18	0.18
	Err($\Re\{\chi\}$)	0.35	**0.15**	0.16	0.16
	Err($\Im\{\chi\}$)	0.70	**0.24**	0.26	0.25
16×24	Err(χ)	0.51	**0.18**	0.25	0.20
	Err($\Re\{\chi\}$)	0.47	**0.17**	0.24	0.19
	Err($\Im\{\chi\}$)	0.67	**0.21**	0.25	0.24

TABLE 3. Relative error with different noise level – $N_s \times N_r$ = 32 × 32.

Noise (in dB)	Error	CSI	CSI-GS-1	CSI-GS-2	CSI-GS-3
15 dB	Err(χ)	0.55	**0.16**	0.18	0.18
	Err($\Re\{\chi\}$)	0.47	**0.14**	0.16	0.16
	Err($\Im\{\chi\}$)	0.86	**0.23**	0.25	0.27
10 dB	Err(χ)	0.89	**0.21**	0.28	0.23
	Err($\Re\{\chi\}$)	0.76	**0.18**	0.25	0.20
	Err($\Im\{\chi\}$)	1.39	**0.31**	0.37	0.34

$l_{\max} = 10$ (due to the use of the Kingsbury Q-shift 6-tap filters) and according to (17) the largest possible decomposition level available is $L = 3$ which will be used in the following. The latter allows a finer description of the reconstructed obstacle thanks to the largest number of detail coefficients to be retrieved.

As seen in Fig. 3, three grouping strategies are adopted to introduce the structure information of the wavelet quad tree.

- CSI-GS-1 (red): the group size of {A, *parent of A*} is 1 or 2.
- CSI-GS-2 (blue): the group size varies according to the level of decomposition, ranging from 1 to L.
- CSI-GS-3 (yellow): the group is chosen as {A, all *children of A*}, the size of which being 1 or 5.

As explained at the end of Section III-C, the complex-valued hyperparameter $q = q_{\Re} + iq_{\Im}$ has to be *a priori* fixed. The choice is made by running the inversion for a set of ($q_{\Re}, q_{\Im}$) (varying from 50 to 95). In Fig. 6 the final error Err(χ) is plotted as a function of $q_{\Re}$ and $q_{\Im}$ and compared to the one of CSI (green line). It exhibits that CSI-GS methods outperformed the CSI one for a large range of $q_{\Re}$ and that their effectiveness does not depend too much on the chosen value. It can be noticed that the range for $q_{\Re}$ and $q_{\Im}$ is narrower for CSI-GS-2 (Fig. 6(b)) than for CSI-GS-1 (Fig. 6(a)) and CSI-GS-3 (Fig. 6(c)). The choices of $q_{\Re}$ and $q_{\Im}$ for different methods are in Table 1 and are drawn as point in each subfigure of Fig. 6.

Results are shown in Fig. 7. To analyze the advantages of the proposed methods when reduced information is provided, the algorithms are tested under three settings of sources and receivers. When there are 32 sources and 32 receivers, the real part of the model retrieved using CSI is roughly reconstructed. The general shape of the disks and rings can be recognized but with obvious discontinuities. As for the imaginary part, only the external boundary of the profile is preserved with also discontinuities whereas the internal pixels cannot be found.

With the group sparsity guided regularization, CSI-GS-1, CSI-GS-2 and CSI-GS-3 allow better reconstructions. The model is smoother and the edges are well preserved. The reconstructed permittivity and conductivity both have relatively smaller values. When the number of sensors is reduced to $N_s \times N_r = 18 \times 26$ and 16×24, more and more discontinuities appear in the results of CSI. Still, the proposed methods are capable of better reconstructing the model.

It should be noted that the results obtained by CSI-GS-3 are not as smooth as with the other two methods. Rectangle artifacts appear in both reconstructed permittivity and conductivity. This is due to the grouping method. The supplementary group sparsity regularization can be seen as an image restoration step. The pixels are supposed to be restored from their ancestors or children at adjacent scales. The first two grouping methods consist in connecting each pixel to its ancestor(s), which makes it possible to smooth pixels with the same ancestor(s). However, this implicit regularization disappears in the third grouping method as it places each pixel and its four children in the same group and adjacent pixels then do not share the same children.

To evaluate the quality of the reconstruction, the relative error of the reconstructed contrast is computed as

$$\text{Err}(x) = \frac{\|\tilde{x} - x\|_2^2}{\|x\|_2^2} \quad \{x = \chi, \Re(\chi), \Im(\chi)\} \quad (18)$$

The errors of the above reconstructed results are in Table 2.

The proposed methods provide a better quality of reconstruction. When the number of sensors is reduced, the relative error obtained by CSI clearly increases whereas the quality of reconstruction remains relatively stable with them.

To evaluate the robustness, noises of higher levels are considered with SNR = 15 dB and 10 dB. Simulations are conducted with CSI and the proposed methods with the same parameter settings. Relative errors are in Table 3.
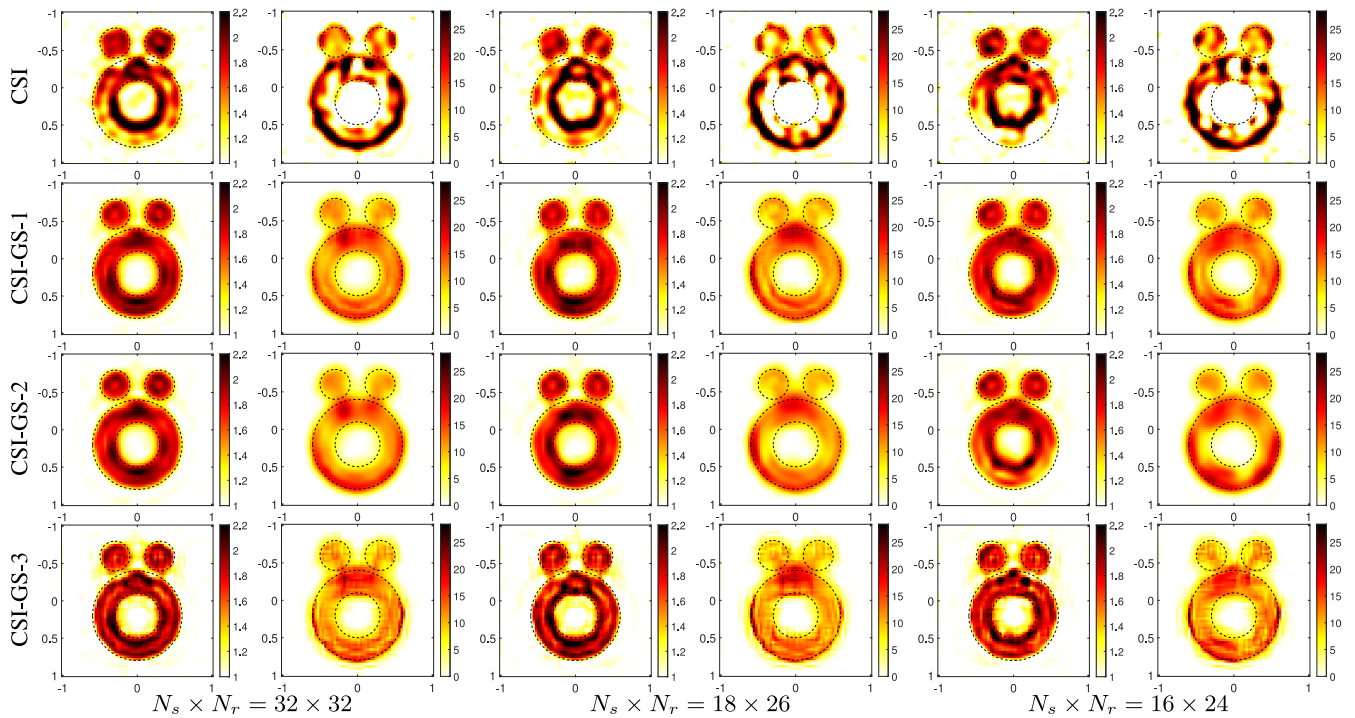
FIGURE 7. Reconstructed $\epsilon(r)$ (column 1, 3, 5) and $\sigma(r)$ (column 2, 4, 6) of "Austria profile" for various combinations of receiver and source numbers (N_r and N_S respectively).

TABLE 4. Relative error with other common wavelets – $N_S \times N_r = 32 \times 32$ with CSI-GS-1.

Error	Haar	D4	D6	D8	C6	C12
$Err(\chi)$	0.24	0.20	0.22	0.21	0.20	**0.19**
$Err(\Re\{\chi\})$	0.18	0.17	0.18	0.17	0.17	**0.16**
$Err(\Im\{\chi\})$	0.48	**0.33**	0.39	0.38	0.36	**0.33**

TABLE 5. Relative error with DWT – $N_S \times N_r = 32 \times 32$ with D4.

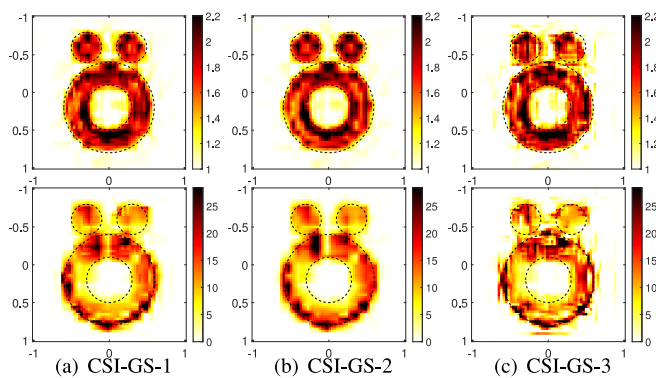| Error	CSI-GS-1	CSI-GS-2	CSI-GS-3
$Err(\chi)$	**0.20**	0.21	0.25
$Err(\Re\{\chi\})$	**0.17**	0.18	0.21
$Err(\Im\{\chi\})$	**0.33**	0.34	0.42

FIGURE 8. DWT-reconstructed $\epsilon(r)$ (top) and $\sigma(r)$ (bottom) with the decomposition basis D4. $N_S \times N_r = 32 \times 32$.
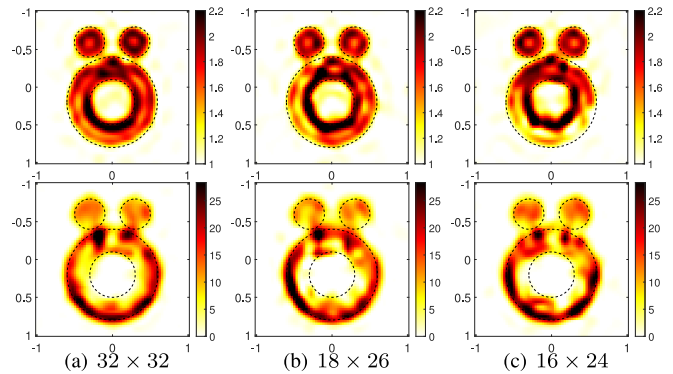

FIGURE 9. Reconstructed $\epsilon(r)$ (top) and $\sigma(r)$ (bottom) with ℓ_1-penalized CSI with different $N_S \times N_r$. The dual-tree $\mathbb{C}$WT is used.

When the noise is of 15 dB, the relative error of χ with CSI, CSI-GS-1, CSI-GS-2, and CSI-GS-3 increases by 22%, 1%, 2% and 2%, compared with a noise of 20 dB. For a 10 dB noise, the increase is of 55%, 6%, 12% and 7%. To conclude, the quality of reconstruction degrades if higher and higher noise level for all algorithms. Yet, the proposed methods are much more robust compared with CSI.

To further illustrate the advantages of the dual-tree $\mathbb{C}$WT, experiments are conducted with DWT using the proposed methods. Knowing that a large set of wavelets exists, some of the more common ones are used in combination with CSI-GS-1 since the latter generally gives better results. Table 4 shows the obtained relative errors using Haar, Daubechies 4 (D4), Daubechies 6 (D6), Daubechies 8 (D8), Coiflets 6 (C6) and Coiflets 12 (C12), respectively. It can be observed that the relative errors obtained using DWT with different wavelet bases do not vary much, so in the following only D4 will be used.
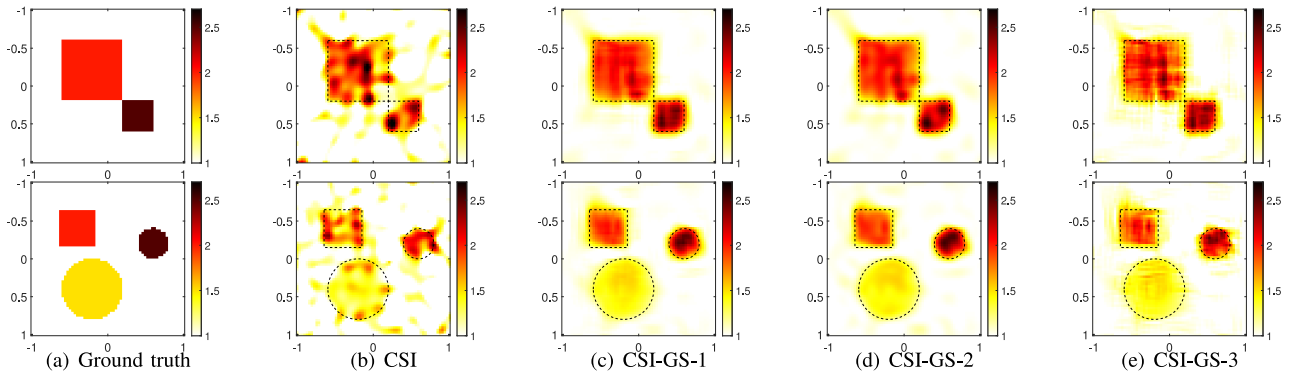
(a) Ground truth	(b) CSI	(c) CSI-GS-1	(d) CSI-GS-2	(e) CSI-GS-3

FIGURE 10. Ground truth and reconstructed $\epsilon(r)$ of Model 2 (top) and Model 3 (bottom). $N_S \times N_r$ is 9×12 for Model 2 and 6×18 for Model 3.

TABLE 6. Relative error with ℓ_1-penalized CSI.

	32×32	18×26	16×24
Err(χ)	0.22	0.27	0.35
Err($\Re\{\chi\}$)	0.18	0.24	0.34
Err($\Im\{\chi\}$)	0.37	0.43	0.38

TABLE 7. Relative error Err(χ) for models 2 and 3.

	CSI	CSI-GS-1	CSI-GS-2	CSI-GS-3
Model 2	0.34	**0.12**	0.13	0.14
Model 3	0.45	0.16	**0.14**	0.18

TABLE 8. Average CPU time (in s) for one iteration on an Intel Core i5-8365U CPU (1.60 GHz) with 16 GByte memory.

CSI	CSI+ℓ_1	CSI-GS-1		CSI-GS-2		CSI-GS-3	
		$\mathbb{C}$WT	DWT	$\mathbb{C}$WT	DWT	$\mathbb{C}$WT	DWT
0.24	0.30	0.28	0.28	0.28	0.31	0.30	0.30

The results for the three approaches (CSI-GS-1, CSI-GS-2 and CSI-GS-3) using D4 are then shown in Fig. 8. Suffering from lack of directionality, only details of three basic directions (horizontal, vertical and diagonal) are well retrieved. Also, the weaker persistence between *parent* and *children* coefficients reduces the quality of the retrieval. The relative error is given in Table 5. Compared with the methods with dual-tree $\mathbb{C}$WT, using DWT leads to growths of relative error of 5%, 5% and 9%.

To enforce sparsity, the most popular regularization term is the ℓ_1 norm. To further demonstrate the benefit of the structure information provided by the $\ell_{2,1}$ norm, simulations are led with ℓ_1-penalized CSI, in which the group sparsity regularization term $\lambda \|\beta_g\|_{2,1}$ in (15) is replaced by $\lambda' \|\beta\|_1$.

To tackle the new cost functional with ℓ_1 norm, lines 15-16 in the Algorithm dedicated to the $\ell_{2,1}$ norm are replaced by $\beta_k = \mathbb{S}_{\lambda'}\{\beta_{k-0.5}\}$ with $\mathbb{S}$ soft-thresholding operator. Hyperparameter λ' is determined like λ_2. $\lambda_1 = P_{q\%}(\beta)$, q is chosen in accord with Table 1.

The results of ℓ_1-penalized CSI are shown in Fig. 9. The permittivity shape can be roughly identified, though obvious discontinuities appear when the number of sensors is reduced. As for the conductivity, only edges are found. The relative error is in Table 6.

The proposed methods are tested on two more models:

- Model 2 consists of two squares centered at $(-0.2, -0.2)$ m and $(0.4, 0.4)$ m, of size 0.8m and 0.4m, with permittivity of 2 and 2.5. Frequency is 500 MHz.
- Model 3 contains one square centered at $(-0.4, -0.4)$ m and 0.5 m-sized with permittivity of 2, and two disks, one of radius 0.4 m centered at $(-0.2, 0.4)$ m with permittivity of 1.5, one of radius 0.2 m centered at $(0.6, -0.2)$ m with permittivity of 2.5. Frequency is 600 MHz.

Fig. 10 shows their ground truth and the results by CSI and the proposed methods. CSI even fails to get the main shape. The proposed methods appear to perform well, as manifested by relative errors in Table 7.

As for respective CPU times of the methods, they are illustrated on Model 1, with maximum iteration number of 500, on an Intel Core i5-8365U CPU (1.60 GHz) with 16 GByte memory, refer to Table 8.

V. CONCLUSION

In this contribution, a group sparsity penalized contrast source inversion (CSI) approach has been proposed. The $\ell_{2,1}$ norm of the wavelet coefficients of the contrast is added to the cost functional, which promotes, on the one hand, the pixel-wise sparsity, on the other hand, the group-wise sparsity. The latter one is based on the parent-child relationship between wavelet coefficients across different scales. The dual-tree $\mathbb{C}$WT is used to decompose the contrast instead of the classic DWT. Three grouping strategies have been investigated to build overlapping sub-groups. The replication approach and proximal method are combined with the CSI algorithm to solve the $\ell_{2,1}$-penalized cost functional.

Simulations on different synthetic models have been carried out to test the proposed methods. When the collected data are not sufficient for CSI to yield satisfactory results, they reconstruct the model well. Robustness to noise is also tested with different signal-to-noise ratios. In addition, the benefit of the dual-tree $\mathbb{C}$WT and $\ell_{2,1}$ norm has been further demonstrated through simulations with DWT and the ℓ_1 norm.

Incorporating the structure information of the unknown into classical optimization methods is of good potential. The parent-child dependency of wavelet coefficients can be modeled by Hidden Markov Tree (HMT) by constructing state transition matrices, which describe the probability that a wavelet coefficient is large or small when its parent is large or small [55]. Also, the hierarchical structure of wavelet coefficients can be combined within the Bayesian inference framework [56].

REFERENCES

[1] X. Chen, *Computational Methods for Electromagnetic Inverse Scattering*. Singapore: Wiley, 2018.

[2] M. Pastorino and A. Randazzo, *Microwave Imaging Methods and Applications*. London, U.K.: Artech House, 2018.

[3] R. Zoughi, *Microwave Non-Destructive Testing and Evaluation*. Dordrecht, The Netherlands: Kluwer Acad., 2000.

[4] N. K. Nikolova, "Microwave imaging for breast cancer," *IEEE Microw. Mag.*, vol. 12, no. 7, pp. 78–94, Dec. 2011.

[5] L. Tsang, J. A. Kong, and R. T. Shin, *Theory of Microwave Remote Sensing*. New York, NY, USA: Wiley, 1985.

[6] A. J. Devaney, "Inverse-scattering theory within the Rytov approximation," *Opt. Lett.*, vol. 6, no. 8, pp. 374–376, 1981. [Online]. Available: http://www.osapublishing.org/ol/abstract.cfm?URI=ol-6-8-374

[7] A. J. Devaney, "Inversion formula for inverse scattering within the Born approximation," *Opt. Lett.*, vol. 7, no. 3, pp. 111–112, 1982. [Online]. Available: http://www.osapublishing.org/ol/abstract.cfm?URI=ol-7-3-111

[8] P. M. Van den Berg and A. Abubakar, "Contrast source inversion method: State of art," *Progr. Electromagn. Res.*, vol. 34, no. 11, pp. 189–218, 2001.

[9] W. C. Chew and Y. M. Wang, "Reconstruction of two-dimensional permittivity distribution using the distorted Born iterative method," *IEEE Trans. Med. Imag.*, vol. 9, no. 2, pp. 218–225, Jun. 1990.

[10] X. Chen, "Subspace-based optimization method for solving inverse-scattering problems," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 42–49, Jan. 2010.

[11] M. Pastorino, "Stochastic optimization methods applied to microwave imaging: A review," *IEEE Trans. Antennas Propag.*, vol. 55, no. 3, pp. 538–548, Mar. 2007.

[12] A. Qing, C. K. Lee, and L. Jen, "Electromagnetic inverse scattering of two-dimensional perfectly conducting objects by real-coded genetic algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 665–676, Mar. 2001.

[13] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems* (Mathematics and its Applications). Dordrecht, The Netherlands: Kluwer Acad., 2000, vol. 375.

[14] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D, Nonlinear Phenom.*, vol. 60, nos. 1–4, pp. 259–268, 1992. [Online]. Available: https://www.sciencedirect.com/science/article/pii/016727899290242F

[15] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Found. Trends Mach. Learn.*, vol. 4, no. 1, pp. 1–106, 2012. [Online]. Available: https://doi.org/10.1561/2200000015

[16] P. M. van den Berg, A. L. van Broekhoven, and A. Abubakar, "Extended contrast source inversion," *Inverse Problems*, vol. 15, no. 5, pp. 1325–1344, 1999. [Online]. Available: https://doi.org/10.1088/0266-5611/15/5/315

[17] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Math. Doklady*, vol. 4, no. 3, pp. 1035–1038, 1963.

[18] P. Lobel, L. Blanc-Féraud, C. Pichot, and M. Barlaud, "A new regularization scheme for inverse scattering," *Inverse Problems*, vol. 13, no. 2, pp. 403–410, 1997. [Online]. Available: https://doi.org/10.1088/0266-5611/13/2/013

[19] Y. Qin, T. Rodet, M. Lambert, and D. Lesselier, "Joint inversion of electromagnetic and acoustic data with edge-preserving regularization for breast imaging," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 349–360, Mar. 2021.

[20] R. A. Horn and C. R. Johnson, *Norms for Vectors and Matrices*. Cambridge, U.K.: Cambridge Univ. Press, 1985, pp. 257–342.

[21] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution," *Commun. Pure Appl. Math.*, vol. 59, no. 6, pp. 797–829, 2006. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.20132

[22] H. Zaimaga and M. Lambert, "Sparsity-enforced microwave inverse scattering using soft shrinkage thresholding," in *Proc. 24th Eur. Signal Process. Conf. (EUSIPCO)*, Budapest, Hungary, Aug. 2016, pp. 350–354.

[23] U. Taşkin and Ö. Özdemir, "Sparsity regularized nonlinear inversion for microwave imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2220–2224, Dec. 2017.

[24] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009. [Online]. Available: https://doi.org/10.1137/080716542

[25] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, 3rd ed. Orlando, FL, USA: Academic, 2008.

[26] A. Desmal and H. Bağcı, "Shrinkage-thresholding enhanced Born iterative method for solving 2D inverse electromagnetic scattering problem," *IEEE Trans. Antennas Propag.*, vol. 62, no. 7, pp. 3878–3884, Jul. 2014.

[27] Y. Sanghvi, H. Bisht, U. K. Khankhoje, V. M. Gadre, and S. V. Kulkarni, "Iteratively reweighted $l_1 − l_2$ norm minimization using wavelets in inverse scattering," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 37, no. 4, pp. 680–687, 2020. [Online]. Available: http://www.osapublishing.org/josaa/abstract.cfm?URI=josaa-37-4-680

[28] D. W. Winters, B. D. Van Veen, and S. C. Hagness, "A sparsity regularization approach to the electromagnetic inverse scattering problem," *IEEE Trans. Antennas Propag.*, vol. 58, no. 1, pp. 145–154, Jan. 2010.

[29] M. Azghani, P. Kosmas, and F. Marvasti, "Microwave medical imaging based on sparsity and an iterative method with adaptive thresholding," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 357–365, Feb. 2015.

[30] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc. Ser. B, Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005. [Online]. Available: http://www.jstor.org/stable/3647580

[31] M. Azghani and F. Marvasti, "Iterative algorithms for random sampling and compressed sensing recovery," in *Proc. 10th Int. Conf. Sampling Theory Appl.*, 2013, pp. 182–185.

[32] M. T. Bevacqua, L. Crocco, L. D. Donato, and T. Isernia, "Non-linear inverse scattering via sparsity regularized contrast source inversion," *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 296–304, Jun. 2017.

[33] G. P. Nason and B. W. Silverman, *The Stationary Wavelet Transform and some Statistical Applications*. New York, NY, USA: Springer-Verlag, 1995, pp. 281–299. [Online]. Available: https://doi.org/10.1007/978-1-4612-2544-7_17

[34] G. Oliveri, M. Salucci, N. Anselmi, and A. Massa, "Compressive sensing as applied to inverse problems for imaging: Theory, applications, current trends, and open challenges," *IEEE Antennas Propag. Mag.*, vol. 59, no. 5, pp. 34–46, Oct. 2017.

[35] A. Massa, P. Rocca, and G. Oliveri, "Compressive sensing in electromagnetics—A review," *IEEE Antennas Propag. Mag.*, vol. 57, no. 1, pp. 224–238, Feb. 2015.

[36] N. Anselmi, M. Salucci, G. Oliveri, and A. Massa, "Wavelet-based compressive imaging of sparse targets," *IEEE Trans. Antennas Propag.*, vol. 63, no. 11, pp. 4889–4900, Nov. 2015.

[37] M. Kowalski, "Sparse regression using mixed norms," *Appl. Comput. Harmonic Anal.*, vol. 27, no. 3, pp. 303–324, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1063520309000608

[38] D. Tomassi, D. Milone, and J. D. B. Nelson, "Wavelet shrinkage using adaptive structured sparsity constraints," *Signal Process.*, vol. 106, pp. 73–87, Jan. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168414002953

[39] H. Zaimaga, A. Fraysse, and M. Lambert, "Sparse reconstruction algorithms for nonlinear microwave imaging," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Kos, Greece, Aug. 2017, pp. 713–717.

[40] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 123–151, Nov. 2005.

[41] R. G. Baraniuk, "Optimal tree approximation with wavelets," in *Proc. Wavelet Appl. Signal Image Process. VII*, vol. 3813, 1999, pp. 196–207. [Online]. Available: https://doi.org/10.1117/12.366780

[42] J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Trans. Image Process.*, vol. 10, pp. 1647–1658, 2001.

[43] Y. Zhang, M. Lambert, A. Fraysse, and D. Lesselier, "A wavelet-based contrast source inversion method," in *Proc. IEEE 19th Int. Symp. Antenna Technol. Appl. Electromagn. (ANTEM)*, Winnipeg, MB, Canada, Aug. 2021, pp. 1–2.

[44] Y. Zhang, M. Lambert, A. Fraysse, and D. Lesselier, "Use of sparsity in nonlinear electromagnetic imaging: wavelet-based contrast source method," in *Proc. 34th Gener. Assembly Sci. Symp. Int. Union Radio Sci. (URSI GASS)*, Rome, Italy, 2021, pp. 1–4.

[45] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions* (Springer Series in Computational Mathematics). Berlin, Germany: Springer-Verlag, 1985. [Online]. Available: https://cds.cern.ch/record/104965

[46] R. T. Rockafellar, *Convex Analysis* (Princeton Mathematical Series). Princeton, NJ, USA: Princeton Univ. Press, 1970.

[47] J. Liu and J. Ye, "Moreau-Yosida regularization for grouped tree structure learning," in *Advances in Neural Information Processing Systems*, vol. 23, J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, Eds. Red Hook, NY, USA: Curran Assoc., Inc., 2010. [Online]. Available: https://proceedings.neurips.cc/paper/2010/file/d490d7b4576290fa60eb31b5fc917ad1-Paper.pdf

[48] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, "Proximal methods for hierarchical sparse coding," *J. Mach. Learn. Res.*, vol. 12, pp. 2297–2334, Jul. 2011.

[49] L. Jacob, G. Obozinski, and J.-P. Vert, "Group Lasso with overlap and graph Lasso," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 433–440. [Online]. Available: https://doi.org/10.1145/1553374.1553431

[50] N. S. Rao, R. D. Nowak, S. J. Wright, and N. G. Kingsbury, "Convex approaches to model wavelet sparsity patterns," in *Proc. 18th IEEE Int. Conf. Image Process.*, Brussels, Belgium, Sep. 2011, pp. 1917–1920.

[51] C. Chen and J. Huang, "The benefit of tree sparsity in accelerated MRI," *Med. Image Anal.*, vol. 18, no. 6, pp. 834–842, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841513001758

[52] H. Choi, J. Romberg, R. Baraniuk, and N. Kingsbury, "Hidden Markov tree modeling of complex wavelet transforms," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 1. Istanbul, Turkey, 2000, pp. 133–136.

[53] A. Abdelnour and I. Selesnick, "Symmetric nearly orthogonal and orthogonal nearly symmetric wavelets," *Arabian J. Sci. Eng.*, vol. 29, no. 2C, pp. 3–16, 2004.

[54] N. Kingsbury, "A dual-tree complex wavelet transform with improved orthogonality and symmetry properties," in *Proc. Int. Conf. Image Process.*, vol. 2. Vancouver, BC, Canada, 2000, pp. 375–378.

[55] B. Deka, S. Datta, and S. Handique, "Wavelet tree support detection for compressed sensing MRI reconstruction," *IEEE Signal Process. Lett.*, vol. 25, no. 5, pp. 730–734, May 2018.

[56] L. Wang, A. Mohammad-Djafari, N. Gac, and M. Dumitru, "Bayesian 3D X-ray computed tomography with a hierarchical prior model for sparsity in Haar transform domain," *Entropy*, vol. 20, no. 12, p. 977, 2018. [Online]. Available: https://www.mdpi.com/1099-4300/20/12/977

YARUI ZHANG received the B.E. degree from the School of Electronic Engineering, Xidian University, Xi'an, China, in 2018, and the M.S. degree in control, signal and image processing from Université Paris–Saclay, Gif-sur-Yvette, France, in 2019. She is currently pursuing the Ph.D. degree in signal and image processing with the Group of Electrical Engineering, Université Paris–Saclay, Paris. Her research interests include inversion and imaging, signal and image processing, and inverse scattering problems.

MARC LAMBERT received the Doctorat en Sciences and Habilitation à Diriger des Recherches degrees from the Université Paris Sud, Orsay, France, in 1994 and 2001, respectively. Since 1995, he has been with Chargé de Recherche, Centre National de la Recherche Scientifique (CNRS), Paris, France, and has carried out his research work with the Laboratoire des Signaux et Systèmes, Joint Laboratory of Supélec, CNRS and Université Paris Sud, Gif-sur-Yvette, France, until 2014. In 2015, he joined the Group of Electrical Engineering Paris. His research focuses on solutions of direct and inverse scattering problems in both electromagnetics and acoustics, and their applications to the characterization of complex objects buried in complex environments from limited datasets, most of his attention being on the numerous theoretical and computational issues which this characterization entails.

AURÉLIA FRAYSSE was born in Créteil, France, in 1978. She received the graduation degree from University Paris 12 in 2001, and the Ph.D. degree in mathematics from University Paris 12 in 2005.

In 2006, she was a Research Assistant with Ecole Nationale Supérieure des Télécommunications (Telecom Paris Tech). She is currently an Associate Professor with IUT Cachan and a Researcher with the Laboratoire des Signaux et Systèmes, Université Paris–Saclay, CNRS, CentraleSupélec.

DOMINIQUE LESSELIER (Senior Member, IEEE) was born in Lons-le-Saunier, France, in August 1953. He received the Engineering degree from Ecole Supérieure d'Electricité, Paris, France, in 1975, and the Doctorat d'Etat et Sciences Physiques degree from the Université Pierre et Marie Curie, Paris, France, in 1982.

Since October 1981, he has been with the Centre National de la Recherche Scientifique (CNRS). He is currently Director of Research CNRS Emeritus. From 1982 to 1983, he was a Visiting Scholar with the Department of Electrical Engineering, University of California at Los Angeles, Los Angeles, CA, USA. He is a member of the Laboratoire des Signaux et Systèmes, Université Paris–Saclay, CNRS, CentraleSupélec, Gif-sur-Yvette, France. From 2006 to 2009, he was the Director of the Groupement de Recherche CNRS GDR Ondes, where he managed a network of scientists involved in the science of waves. His research focuses on the development of solution imaging methods of inverse problems under many guises, from mathematics to simulations to applications and vice versa. He was the recipient of the R. W. P. King Award in 1982 from the IEEE Antennas and Propagation Society. From 2005 to 2016, he was on the International Advisory Panel of Inverse Problems, after serving on its Editorial Board from 1997 to 2004. From 2003 to 2019, he has been an Associate Editor of *Radio Science*. Since 1998, he has been on the Standing Committee of the Electromagnetic Non-Destructive Evaluation Workshop Series and the International Steering Committee of the International Symposia on Applied Electromagnetics and Mechanics. He is a Fellow of the Institute of Physics and a Member of the Electromagnetics Academy and the International Union of Radio Science, Commission B.