Sparsity and Electromagnetic Imaging in Non-Linear Situations

Hidayet Zaimaga

université PARIS-SACLAY

NNT : 2017SACLE415

Thèse de doctorat

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir®

Sparsity and Electromagnetic Imaging in Non-Linear Situations

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-Sud

École doctorale n°580 Sciences et Technologies de l'Information et de la Communication (STIC)
Spécialité de doctorat: Traitement du Signal et des Images

Thèse présentée et soutenue à Gif-sur-Yvette, le 4 décembre 2017, par

Hidayet Zaimaga

Composition du Jury :

Oliver Dorn RCUK Fellow, Université de Manchester	Rapporteur
Amélie Litman Maître de Conférences, Université de Marseille	Rapportrice
Dominique Lesselier Directeur de Recherche CNRS, L2S UMR 8506	Président du jury
Mai Nguyen-Verger Professeur, Université de Cergy-Pontoise	Examinatrice
Andrea Massa Professeur, Université de Trente, DIGITEO Chair	Directeur de thèse
Marc Lambert Chargé de Recherche CNRS, GeePs UMR 8507	Co-Directeur de thèse
Aurélia Fraysse Maître de Conférences, Université Paris-Sud	Encadrante, Invitée

Acknowledgement

I would like to thank my supervisors Dr. Marc Lambert, chargé de recherche at Centre National de la Recherche Scientifique (CNRS), and Prof. Andrea Massa, director of the ELEDIA Research Center at the University of Trento, for their understanding and their guidance throughout the three years. Additionally, I would like to thank DIGITEO France for financing of this thesis through the SIRENA project (2014–2017) under the "Call for Chairs 2014".

I would like also to thank Dominique Lesselier, Directeur de Recherche at CNRS, and his lovely wife Laurence Lesselier who welcome to me Paris from the first day and made me feel like at home. I would like to thank Aurélia Fraysse, Maître de Conférences, for her advices and co-operation as well. I would like also thank all my friends who made my stay in Paris really pleasant and fruitful. Especially, I thank Zicheng Liu who shared my best and worst moments in the same office with patience and encouragement and also my friend Hüriye Gürsel who is always there for me.

I also wish to thank my two rapporteurs, Prof. Amélie Litman and Dr. Oliver Dorn along with all the other members Mai Nguyen-Verger, and Dominique Lesselier to have accepted to be part of the jury of my Ph.D. defense.

Special thanks and great appreciation go to my parents, Zühal and Caner, and my brother, Birkan, for their selfless support in all the stages of my life. Without their endless love and support I would not be where I am now.

Contents

Résumé en Français

Ce manuscrit de thèse décrit mes travaux de recherche dans le domaine de l'imagerie micro-onde, qui fait l'objet d'un intérêt de recherche important en raison de son potentiel en tant que technique pratique et efficace pour des systèmes médicaux [1, 2], des caractérisations de matériaux [3], la caractérisation du sous-sol, la télédétection, et les essais et évaluations non destructifs [4, 5]. L'objectif de l'imagerie micro-onde est notamment d'estimer l'emplacement de diffracteurs dans une région d'intérêt, en accédant en sus à la distribution de leurs propriétés électromagnétiques.

Ces dernières années, le besoin de méthodes et de techniques de reconstruction efficaces a fortement émergé pour résoudre les problèmes de diffraction inverse électromagnétique qui se posent dans l'imagerie micro-onde afin d'en réduire au mieux les difficultés théoriques et pratiques. La demande dans diverses applications impose le développement de méthodes efficaces et précises. Cependant, la mise en œuvre d'algorithmes de reconstruction qui soient stables, fiables et efficaces est un défi en raison de la non-linéarité des équations de la diffraction et du caractère mal-posé du problème auquel on se confronte [4, 6, 7].

En mathématique, il y a souvent deux problèmes qui font face l'un à l'autre, le problème *direct* et le problème *inverse*. Dans le problème direct, l'état observable d'un système est déterminé en utilisant tous les paramètres nécessaires. En comparaison, le problème inverse se préoccupe de la condition préalable qui crée une donnée observée. Mathématiquement, nous avons $y = K(\mathbf{x})$, où $\mathbf{x} \in X$ est l'inconnue et $\mathbf{y} \in Y$ est le vecteur des observations. Habituellement, $K()$ est un opérateur continu bien posé et la solution du problème direct (trouver $\mathbf{y}$ donné $\mathbf{x}$) ne rencontre aucun obstacle significatif. D'autre part, le mapping inverse de $\mathbf{y}$ à $\mathbf{x}$ dans les problèmes d'intérêt n'est pas aisé à traiter. Normalement, les problèmes inverses sont tels que les modèles mathématiques ne sont pas bien posés dans le sens de Hadamard, signifiant pas de solution unique, pas d'existence de solution et la non-dépendance des données de manière continue [6, 7]. Ces problèmes sont dits mal-posés et ils sont la source de nombreuses difficultés numériques car ils rendent la plupart des algorithmes numériques instables pour des perturbations de données [4, 8].

Les méthodes numériques qui peuvent traiter ces problèmes sont appelées méthodes de régularisation.

Les problèmes de diffraction électromagnétique inverse possèdent donc deux propriétés fondamentales qui font de leur résolution une tâche difficile, la première est la non-linéarité des équations de diffraction [5, 9], la seconde le caractère mal-posé du problème [5, 9]. Le champ diffracté est une fonction intégrale du champ total induit dans le domaine investigué, multiplié par le contraste (ici la différence entre les permittivités complexes du domaine d'investigation et du milieu environnant). Le champ total induit est également une fonction non linéaire de la permittivité [5, 9]. En conséquence, la reconstruction de cette permittivité à partir de champs diffractés est un problème non linéaire.

En outre, le fait que le champ diffracté soit une fonction intégrale du champ total et de la permittivité dans le domaine sous investigation conduit à un problème inverse mal-posé [5, 9]. L'opérateur intégral supprime la contribution des composantes à variation rapide de la permittivité au champ diffracté et rend difficile leur identification [5, 9]. Cette nature physique du phénomène de diffraction est combinée avec le fait que seul un ensemble fini d'échantillons, non nécessairement aussi de distribution optimale par rapport aux diffracteurs recherchés et insuffisamment indépendants les uns des autres, peut être collecté par les mesures, mesures qui ne fournissent pas suffisamment d'information, et que ces échantillons sont presque toujours corrompus par le bruit, de sorte qu'un modèle approximatif discrétisé doit remplacer la description mathématique exacte du problème. Ainsi, cela rend le problème inverse sévèrement mal posé [5, 9].

La force de la non-linéarité dans le problème de diffraction s'accroît avec l'amplitude des champs diffractés couplés avec le/les objet/s dans le domaine investigué. Par conséquent, une permittivité supérieure (cette notion étant aussi dépendante de la taille des diffracteurs) signifie une non-linéarité plus forte. La plupart des méthodes déterministes qui se révèlent efficaces pour résoudre des problèmes inverses électromagnétiques impliquant des faibles diffracteurs utilisent des approximations linéaires de premier ordre [10], telles que la tomographie par diffraction [11], l'approximation de Born de premier ordre [5, 12], l'approximation de Kirchhoff [5] et l'approximation de Rytov [5, 12]. Bien que ces méthodes soient moins exigeantes sur le plan informatique, elles ne fournissent pas de solutions précises lorsque de puissants diffracteurs sont présents dans le domaine étudié. Dans de tels cas, lorsque la non-linéarité est forte, la linéarisation ne produit pas une approximation précise du problème non linéaire et il faut utiliser des techniques plus rigoureuses telles que le solveur non linéaire de Newton combiné avec des formulations de champ de contraste ou de source de contraste [5, 13], la méthode de gradient con-

jugué non linéaire [14], l'algorithme de descente de plus grande pente non linéaire (par ex. Landweber non-linéaire) [15, 16], et le schéma de Levenberg-Marquardt [17], ainsi que les méthodes de Born distordu [18]. Ces méthodes sont de plus exigeantes en terme de calcul comparativement à celles qui utilisent les schémas de linéarisation du premier ordre. D'autres méthodes qui bénéficient du meilleur des deux mondes ont également été développées. Celles-ci utilisent des schémas de linéarisation d'ordre supérieur ou une application itérative de premier ordre. Les exemples incluent l'approximation de Born étendu [19] et les approximations de Born d'ordre supérieur [5, 20], la méthode itérative de Born (BIM) [21] et la méthode itérative variationnelle [22], respectivement. Les méthodes de ce dernier groupe sont plus efficaces en terme de calcul par rapport aux méthodes d'inversion non linéaires, et s'appliquent également à une plus grande gamme de force des diffracteurs que les méthodes de linéarisation de premier ordre.

Indépendamment de la méthode utilisée pour s'attaquer à la non-linéarité, le caractère mal-posé du problème inverse de diffraction doit être pris en compte [9, 23, 24]. À cette fin, des méthodes de régularisation linéaires et non linéaires, qui minimisent une fonction de coût pondérée entre mesure/inadéquation des données et un terme de pénalité, ont été développées. Le choix le plus populaire du terme de pénalité est la norme l_2 de la solution [23, 24]. Le problème de minimisation qui en résulte peut être résolu en utilisant le schéma bien connu de Tikhonov [23, 24]. En outre, l'utilisation d'itérations tronquées de Landweber ou de gradient conjugué conduit à un type similaire de régularisation [5, 23]. En programmation linéaire, toutes ces méthodes filtrent efficacement les valeurs singulières du problème de minimisation discrétisé, qui sont inférieures à un niveau de seuil spécifié, ou font appel à un paramètre de régularisation pour atténuer le caractère mal-posé.

La régularisation est utilisée pour résoudre des problèmes mal-posés en incorporant des connaissances *a-priori* sur l'objet de façon à stabiliser le problème et fournir des solutions raisonnables et utiles. Par exemple, si l'on sait que la solution doit être une discrétisation d'une fonction continue, cette connaissance nous permet de rejeter les candidats les plus variables et de réduire considérablement l'ensemble des solutions possibles. Le choix d'une régularisation appropriée dépend étroitement de la propriété que l'on souhaite mettre en avant, et cela dépend de l'application particulière.

Dans de nombreux problèmes inverses mathématiques, les choix *a-priori* sont des formes différentes de régularités, et les pénalités correspondantes sont les normes l_2 de $\mathbf{x}$ ou de ses dérivése. Des a priori de parcimonie sont utiles lorsque les signaux que nous recherchons doivent être parcimonieux. Nous définissons la parcimonie d'un vecteur par la présence d'un petit nombre d'éléments non nuls comme dans le travail de Mallat [25], celui de

Donoho [26], et celui de Rao [27], entre autre, à propos de l'approximation de fonctions et la sélection optimale de bases.

La régularisation par la parcimonie a été un domaine de recherche fort de ces dernières années. Le caractère bien posé et la convergence des méthodes ont été analysés pour des problèmes inverses linéaires dans [28] et pour des problèmes inverses non linéaires dans [29]. Il a été démontré que la régularisation par parcimonie est simple à utiliser et très efficace pour des problèmes inverses avec des solutions parcimonieuses. Cette méthode a été appliquée aux problèmes d'imagerie compressive et de tomographie par impédance électrique (EIT) [30–32]. Cette méthode de régularisation conduit à des minimiseurs parcimonieux lorsque $p = 1$ pour la norme l_p et favorise la parcimonie pour $1 < p < 2$ [28].

Dans ce travail, nous avons exclu $p = 0$ puisque l_0 est une pseudo-norme (la propriété de distance des normes n'est pas satisfaite), qui compte le nombre d'éléments non nuls. Même si la parcimonie est le plus favorisé lorsque $p = 0$, notre problème d'optimisation devient NP-hard [33]. De plus, ce problème est non-convexe, ce qui crée un problème de minima multiples et la recherche des minima globaux optimaux nécessite une recherche intensive très exigeante en calcul [24]. Pour éviter l'apparition de minima multiples, la norme l_1 (quand $p \leq 1$) est utilisée comme la meilleure approximation convexe du problème d'optimisation en norme l_0. Ce concept est connu sous le nom de relaxation convexe [24] en tant qu'alternative au problème NP-Hard. En outre, la norme l_1 propose en général une solution avec d'excellentes caractéristiques de parcimonie. Par conséquent, nous choisissons une pénalité l_1 afin de promouvoir la connaissance *a-priori* de parcimonie tout au long de ce travail de thèse.

L'utilisation de normes l_1 pour assurer la parcimonie est connue depuis près d'une décennie. L'opérateur LASSO ou Least Absolute Shrinkage and Selection Operator a été introduit dans la littérature statistique [34] et l'algorithme dit de Basis Pursuit [35] a été proposé dans la communauté du signal à peu près au même moment. Les avantages les plus importants des systèmes de pénalisation l_1 sont leur convexité, et la forte parcimonie des résultats. Différentes versions de problèmes de régularisation l_1 peuvent être reformulées en tant que programmation quadratique linéaire ou convexe, contrainte ou non, ou de second ordre (SOC), ce qui autorise des algorithmes efficaces et globalement convergents. Un autre avantage important de l'utilisation de la pénalisation l_1 est l'existence d'un certain nombre de résultats théoriques récents (par exemple [36]) montrant que, sous certaines conditions de parcimonie sur le signal inconnu sous-jacent, le signal peut être récupéré exactement.

En fait, la fonction inconnue (par exemple, la fonction de contraste ou le courant équiv-

alent dans les problèmes inverses électromagnétiques) peut être considérée comme parcimonieuse (ou compressible) dans une base appropriée, du fait qu'elle puisse n'occuper qu'une petite partie d'un scénario autrement connu ou qu'elle puisse représenter une anomalie étendue presque homogène contenue dans un milieu de fond. En d'autres termes, un diffracteur peut être parcimonieux par rapport à une base et non parcimonieux par rapport à une autre. En outre, dans de nombreux cas pratiques, seul un nombre limité de mesures sont permises, de sorte que des outils de traitement appropriés pour gérer ce type de complexité sont nécessaires.

La question de choisir une base appropriée pour une famille de signaux a fait l'objet de beaucoup d'attention cette dernière décennie, et de nombreuses nouvelles bases ont été introduites, telles les bases d'ondelettes, les ridgelets et curvelets, parmi beaucoup d'autres [35]. Bien que toute base d'extension minimale pour un espace à dimension finie puisse représenter parfaitement tout signal dans l'espace, lorsque seul un sous-ensemble de signaux possibles est intéressant, certaines bases possèdent de meilleures propriétés de représentation parcimonieuse que d'autres. Certaines applications qui bénéficient grandement de la parcimonie de la représentation sont la compression du signal, le débruitage et l'estimation des paramètres [37–39]. En compression pour la transmission de l'information, si la représentation du signal n'est pas parcimonieuse, nous devons transmettre le signal complet. Cependant, si, sous un changement de base, la représentation devient parcimonieuse, des économies substantielles sont possibles. La plupart des coefficients de la représentation sont très petits (par définition de la parcimonie) et si nous les mettons à zéro, la qualité perceptuelle du signal sera très peu affectée. Par conséquent, on ne transmet que les grands coefficients, qui sont peu nombreux.

Dans cette thèse, nous avons proposé des méthodes qui exploitent la parcimonie en tant que brique principale afin de résoudre le problème inverse non linéaire. Initialement, nous avons proposé une méthode où la contrainte de parcimonie est directement appliquée au problème de la reconstruction des propriétés diélectriques internes complexes d'un objet en fonction de la connaissance du champ diffracté extérieur qui est généré par l'interaction entre l'objet et un champ incident connu. Le problème d'optimisation non linéaire est résolu par un algorithme itératif de contraction douce afin d'appliquer la contrainte de parcimonie. Cette parcimonie est appliquée à chaque itération par une fonction de seuillage doux. En outre, nous utilisons une sélection de pas adaptative selon la règle de Barzilai et Borwein (BB) et une projection sur le contraste inconnu de telle sorte que les contraintes de positivité soient prises en compte.

La base adoptée joue un rôle clé, car celle-ci doit être tant précise qu'efficace. L'efficacité

nécessite un nombre réduit de coefficients dans la représentation, tandis que la précision implique une faible erreur de représentation, qui quantifie la désadaptation entre la fonction réelle et sa projection sur la base considérée. Étant donné que le problème inverse de diffraction électromagnétique que nous considérons n'est pas linéaire et est mal posé, la réduction du nombre de paramètres inconnus est importante pour réduire le caractère mal posé [25, 40, 41]. En fait, la décomposition en ondelettes permet de réduire le nombre de paramètres inconnus par rapport à la représentation en pixels habituellement exploitée afin d'accroître la fiabilité de l'inversion [42]. À cet égard, après notre première méthode proposée avec un seuillage doux, nous avons étendu l'application du paradigme d'imagerie micro-onde aux fonctions non-pixelisées, élargissant ainsi l'ensemble des distributions de diffracteurs pouvant être considérés aux cibles de forme et taille arbitraires, sous réserve qu'elles soient parcimonieuses par rapport à une base appropriée. Dans ce but, les ondelettes sont utilisées et nous avons étudié si la base d'ondelettes est un bon choix pour tenir compte ou non du compromis entre l'efficacité et la précision de la représentation dans notre cas.

La transformée en ondelettes permet d'avoir des informations tant sur la variation spatiale que sur le contenu fréquentiel de la fonction, en la représentant en tant que somme pondérée de la version dilatée et translatée de l'ondelette dite mère [25, 39]. En particulier, les fonctions de base à grande échelle codent le contenu à basse fréquence de la fonction traitée, tandis que celles à petites échelles prennent compte des détails de haute fréquence ou des détails fins, [25, 39]. Ceci réduit le nombre de paramètres inconnus par rapport à la représentation habituellement exploitée des pixels, améliorant ainsi la fiabilité de l'inversion [42, 43]. En conséquence, de telles bases sont intrinsèquement capables de conduire à aborder le problème de diffraction inverse dans un cadre multi-échelle, ce qui s'est révélé être un moyen efficace d'améliorer les résultats d'imagerie [44, 45].

Une grande variété d'études considère dorénavant l'information structurelle des solutions afin de faciliter leur estimation. Par exemple, la structure dite de "group sparsity" ou de parcimonie structurée [46] a été mise en œuvre dans de nombreuses applications, où les composantes des solutions sont susceptibles d'être soit toutes nulles soit non nulles dans un groupe. Ainsi, on vise à diminuer la dispersion afin d'améliorer la solution en tenant compte de ce regroupement préalable. Dans ce cadre, nous nous sommes concentrés sur la parcimonie conjointe, ce qui constitue un cas particulier de la parcimonie structurée. Nous avons proposé une approche d'inversion en deux étapes afin de résoudre un problème inverse non linéaire en appliquant une parcimonie commune pour obtenir les sources équivalentes puis le contraste inconnu. Plus précisément, la parcimonie conjointe signifie

que des vecteurs parcimonieux inconnus partagent le même support non nul [46, 47].

Pour conclure, deux algorithmes d'inversion sont proposés pour une imagerie micro-onde efficace et précise de domaines d'investigation parcimonieux dans ce manuscrit. Le premier dépend du seuil de contraction douce avec des contraintes différentes et le second exploite une parcimonie conjointe afin de retrouver les propriétés électromagnétiques inconnues des diffracteurs. Nous avons travaillé sur différents diffracteurs afin de montrer la fiabilité des méthodes et nous avons proposé des résultats numériques qui illustrent leur efficacité, tout en proposant des améliorations complémentaires dans certains cas.

List of Figures

List of Tables

1. Introduction

1.1. General Context

This thesis presents my research work in the area of microwave imaging which has been attracting momentous research interests due to its potential as a convenient and efficient technique for medical systems [1, 2, 49], material characterization [3], subsurface probing, remote sensing, and non-destructive testing and evaluation [4, 5]. The main goal of microwave imaging is to retrieve the distribution of the dielectric properties in a region of interest, with usual options of quantitative imaging (wherein the values of the electromagnetic parameters are sought within this region) and qualitative imaging (wherein what in fine matters is to identify certain zones in which such properties differ from those of the embedding medium).

It is well known that the electromagnetic inverse scattering problem is the basic formulation for microwave imaging methods. Although this problem has been suitably studied from a theoretical point of view, new methods are continuously developed to face new and challenging applications. To inspect dielectric targets, several techniques can be adopted [14, 44, 49–53]. There are diverse research studies in this area and we only mentioned few of them here.

The need of efficient reconstruction methods and techniques has widely emerged for solving inverse electromagnetic scattering problems arising in microwave imaging in order to mitigate theoretical and practical difficulties. High demand of such methods in various applications enforces the importance and the development of effective and accurate methods. However, implementation of stable, reliable, and efficient reconstruction algorithms is challenging because of the nonlinearity of the scattering equations and ill-posedness of the problem [4, 6, 7].

In mathematics, there are often two problems which are opposite to each other. One of them is called *direct* (*forward*) and the other one is called *inverse* problem. In the forward problem, the observable state of a system is determined by using all needed parameters. However, the inverse problem is about the precondition which creates observed data.

Mathematically, we have $\mathbf{y} = K(\mathbf{x})$, where $\mathbf{x} \in X$ is the unknown and $\mathbf{y} \in Y$ is the vector of observations. Usually, $K()$ is a well-behaved continuous operator, and the solution of the forward problem (find $\mathbf{y}$ given $\mathbf{x}$) meets no significant obstacles. On the other hand, the inverse mapping from $\mathbf{y}$ to $\mathbf{x}$ in the problems of interest is not easy to deal with. Normally, inverse problems cause mathematical models not to be well-posed in the sense of Hadamard such as no unique solution, no solution existence and non dependence on the data continuously [6, 7]. These problems are called ill-posed and they lead to many numerical difficulties as it makes most numerical algorithms unstable under data perturbations [4, 8]. Numerical methods that can deal with these problems are called regularization methods.

Regularization is used to solve ill-posed problems by incorporating *a-priori* knowledge about $\mathbf{x}$ to stabilize the problem and to provide reasonable and useful solutions. For example, if it is known that the solution should be a discretization of a continuous function, this knowledge allows us to discard the wildest looking candidates, and to considerably reduce the set of possible solutions. The task is to minimize some measure $J_1(\mathbf{x})$ of proximity of $\mathbf{y}$ (observation data), as well as to satisfy as much as possible the *a-priori* information about $\mathbf{x}$ (the unknown parameter), by minimizing some appropriate measure $J_2(\mathbf{x})$. The two objectives typically cannot be both reached at the same time, so we need a compromise, which can be simply obtained by taking a linear combination of the two:

$$J(\mathbf{x}) = J_1(\mathbf{x}) + \alpha J_2(\mathbf{x}) \tag{1.1}$$

Scalar α is the regularization parameter balancing the tradeoff between the fidelity to the data, $J_1(\mathbf{x})$, and the fidelity to the prior information, $J_2(\mathbf{x})$. There is a whole family of solutions indexed by α, with the non-regularized (least squares) solution if $\alpha = 0$, and a solution strongly favoring the *a-priori* information when α is large. In general, choosing an appropriate α is problem-dependent, and is a nontrivial task. With an appropriate choice for $J(\mathbf{x})$, regularization effectively deals with all the three aspects of ill-posedness. Also, proper choice of $J_2(\mathbf{x})$ deals with lack of uniqueness and can dramatically reduce sensitivity to noise (improve the Lipschitz constant of the inverse function), making it continuous enough for practical applications.

The selection of a proper regularizer intimately depends on the property of $(\mathbf{x})$ that one wishes to enforce, and that depends on the particular application. In many mathematical inverse problems, priors of choice are different forms of smoothness or constraints, and the corresponding regularizers are the l_2 norms of $(\mathbf{x})$ or its derivatives. Sparsity prior is useful

when signals $\mathbf{x}$ that we look for have to be sparse. Sparsity of a vector ($\mathbf{x}$) can be defined by the presence of a number of non-zero elements and zeros elsewhere, with respect to an appropriate basis, introducing the notion of L-sparse signal wherein this number is L. An appropriate numerical measure of sparsity is the count of non-zero elements. The base for this discussion is the work of Mallat [25], Donoho [26], Rao [27] and others on function approximation and optimal basis selection.

The problem of choosing an appropriate basis for a family of signals has received a great deal of attention over the past decade, and many new bases were introduced, such as wavelet bases, ridgelets, and curvelets, among many others [35]. Despite the fact that any minimal spanning basis for a finite-dimensional space can represent perfectly any signal in the space, when only a subset of possible signals is of interest, some bases have better representational properties than others. Some applications which greatly benefit from sparsity of representation are signal compression, denoising, and parameter estimation.

In compression for information transmission, if the representation of the signal is not sparse then we need to transmit the whole signal. However, if under a change of basis the representation becomes sparse, then substantial savings are possible. Most coefficients of the representation are very small and if we set them to zero the perceptual quality of the signal will be affected very little. Hence, we are left with transmitting only the large coefficients, which are few in number. This idea found use in commercial compression algorithms.

Another application where sparsity plays a key role is denoising. If the signal is sparse then separating it from the noise requires considerably less effort than when signal power is evenly distributed along the support of the signal. Therefore, for the purpose of facility of denoising of a class of signals, it is worthwhile to find a basis in which the representation of all signals belonging to this class is as sparse as possible.

An inverse problem is often formulated in order to compute an approximation to a solution of the operator equation such as

$$K(\mathbf{x}) = \mathbf{y}, \tag{1.2}$$

where K is an ill-posed operator and with the case of noisy data $\mathbf{y}^\delta$

$$\left\|\mathbf{y} - \mathbf{y}^\delta\right\| \leq \delta \tag{1.3}$$

is available. For the stable approximation of a solution (1.2), a sparsity regularization

method is used to minimize the functional

$$\Theta(\mathbf{x}) := \frac{1}{2}\left\|K(\mathbf{x}) - \mathbf{y}^{\delta}\right\|^2 + \alpha \sum_{k\in\Lambda} w_k |\langle \mathbf{x}, \varphi_k\rangle|^p \qquad (1 \le p \le 2) \tag{1.4}$$

or more generally

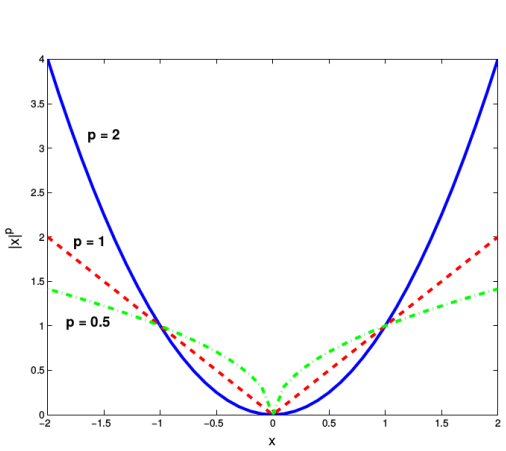
$$\Theta(\mathbf{x}) := F(K(\mathbf{x}), \mathbf{y}^{\delta}) + \alpha \sum_{k\in\Lambda} w_k |\langle \mathbf{x}, \varphi_k\rangle|^p \tag{1.5}$$

where $\alpha > 0$ is a regularization parameter, $\varphi_{k}{}_{k\in\Lambda}$ is a basis and $w_k \ge w_{min} > 0, \forall k$ is a weight parameter. The functional $F(K(\mathbf{x}), \mathbf{y}^{\delta})$ measures the error between $K(\mathbf{x})$ and $\mathbf{y}^{\delta}$. If we let $J_1(\mathbf{x}) = \frac{1}{2}\left\|K(\mathbf{x}) - \mathbf{y}^{\delta}\right\|^2$, and $J_2(\mathbf{x}) = \sum_{k\in\Lambda} w_k |\langle \mathbf{x}, \varphi_k\rangle|^p$, then we have nothing but a regularized inverse problem of the form in (1.1). When $p = 1$ this method is called basis pursuit [35] (or LASSO [34] in the statistical literature). The prior term, $J_2(\mathbf{x})$ has an effect of enforcing sparsity. Figure 1.1 gives some insight into why p-regularization with $p \le 1$ favors sparse $\mathbf{x}$. In Fig. 1.1b, we show the level sets of p norms to the p-th power ($\|\mathbf{x}\|_p$) for $p = 0.5$, $p = 1$, and $p = 2$ of a two-dimensional vector. For a fixed l_2-norm, i.e, for all vectors that lie on a circle with fixed radius, p norms with $p \le 1$ are minimized on the coordinate axes, i.e. preferring that some of the coefficients are set exactly to zero, while others are large. In other words, p norms with $p \le 1$ maintains sparse solutions. This argument can be generalized to vectors in higher dimensions.

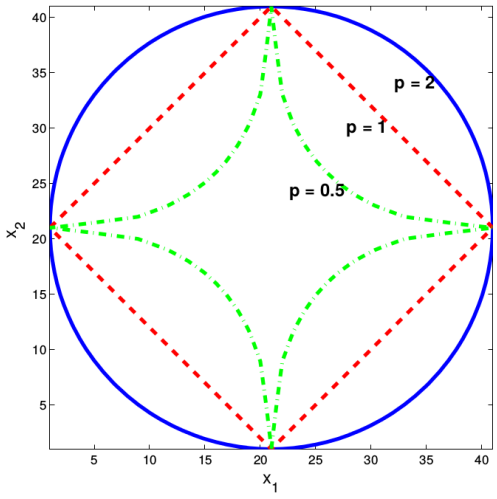
Figure 1.1a shows l_p norms for the same p's in one dimension. It shows that the penalty on large features (large x_i) is less for smaller p. Strong features are penalized much less severely in l_p penalization with $p \le 1$ than in l_2 penalization (Tikhonov regularization). This motivates the smoothing effect of l_2-penalization, and the feature-preserving behavior of p for $p \le 1$. It is well known that l_2-norm regularizers are known to promote the smoothness in the solution and hence do not efficiently produce accurate solutions when applied in domains with sharp variations, discontinuities, or sparse content (i.e., scatterers occupy much smaller volumes/areas in comparison to the whole investigation domain) [24]. Such domains exist in many practical applications, such as see-through-the-wall imaging, hydrocarbon reservoir detection, radar imaging, and crack detection.

Another observation from Figure 1.1 is that l_1-norm is convex, whereas when $p < 1$, l_p-norm is no longer convex. The computational complexity for the minimization of some non-convex cost functions (p in particular) can be improved by using the half-quadratic regularization method [54]. The key idea is to introduce a supplementary vector $\mathbf{s}$, and an extended cost function, $Q(\mathbf{x}, \mathbf{s})$, which is quadratic in $\mathbf{x}$ for a fixed $\mathbf{s}$, and $\operatorname{argmin}_{\mathbf{s}} Q(\mathbf{x}, \mathbf{s}) = J(\mathbf{x})$, for any $\mathbf{x}$. If $Q(\mathbf{x}, \mathbf{s})$ is also easy to minimize in $\mathbf{s}$ (or even better if there is a closed-

form solution), then the resulting extended cost function can be optimized with reasonable efficiency by iterative methods.



(a) 1-D plot



(b) 2-D plot

Figure 1.1.: (a) 1-D plot of $\|\mathbf{x}\|_p$, $p = 0.5, 1, 2$. (b) 2-D level sets of $\|\mathbf{x}\|_p$, same p.

Sparsity regularization has been the foremost research area for the last years. The well-posedness and convergence rates of the method have been analyzed for linear inverse problems in [28] and for nonlinear inverse problems in [29]. It has been shown that sparsity regularization is simple for use and very effective for inverse problems with sparse solutions. This method has been applied to Compressive Imaging and Electrical Impedance Tomography (EIT) problems [30–32]. This regularization method leads sparse minimizers of (1.4) for $p = 1$ and promotes sparsity for $1 < p < 2$ [28].

In this manuscript we have excluded $p = 0$ since l_0 is a pseudo-norm (the distance property of norms is not satisfied), which counts the number of non-zero elements. Even if sparsity is most favored when $p = 0$, the above optimisation problem will become NP-hard [33]. In other words, this creates a problem of multiple infima and solving it for the optimal global minima requires an intensive search that is computationally a burden [24]. To avoid the occurrence of multiple infima, the l_1-norm (when $p \leq 1$) is used as the best convex approximation to the l_0 -norm optimization problem. This concept is known as convex relaxation [24] as an alternative to the NP-hard problem. Furthermore, the l_1-norm has been verified to elect a solution with excellent sparsity features. Therefore, we choose l_1 penalty to promote *a-priori* knowledge of the sparse representation throughout the work in this thesis.

The use of l_1-norms to achieve sparsity has been known for almost a decade as already mentioned. The Least Absolute Shrinkage and Selection Operator (LASSO) has been introduced in the statistics literature [34], and Basis Pursuit algorithm [35] for choosing a sparse basis has been proposed in the signal representation community at around the same time. The most important advantages of l_1 penalization schemes are their convexity, and the strong sparsity of the results (most indices of the result are set exactly to zero). Different versions of l_1-regularization problems can be reformulated as linear, convex constrained or unconstrained quadratic, or second order cone (SOC) programming, all of which allow efficient and globally convergent algorithms. Another significant benefit of using l_1 penalization is a number of theoretical (e.g. [36]) results showing that under certain sparsity conditions on the underlying unknown signal, the signal can be recovered exactly. This is quite surprising since the direct combinatorial formulation of the problem requires comparing solutions with all possible permutations of non-zero indices, which is very hard.

As a matter of fact, the unknown function (e.g., the contrast function or the equivalent current) can be considered as sparse (or compressible) in some suitable basis, due to the fact it may occupy a small portion of an otherwise known scenario or that may represent an almost homogeneous extended anomaly hosted in a background medium. In other words, a scatterer can be sparse with respect to a basis and not sparse versus another one. Moreover, in many practical cases, only a limited number of measurements is allowed, so that proper processing tools to handle this kind of complexity are needed.

One of the aims of this thesis is to extend the application of the microwave imaging paradigm to nonpixel basis functions, thus enlarging the set of retrievable scatterer distributions to targets of arbitrary shape and size provided that they are sparse with respect to a suitable chosen basis. Toward this end, wavelet basis functions will be used.

The wavelet transform stores information on both the spatial variation and the frequency content of the processed function, by representing it as a weighted sum of dilated and translated version of the so-called mother wavelet [25, 39]. In particular, large scale basis functions encode the low frequency content of the processed function, while small scale ones account for high frequency or fine details. Such an encoding can be replicated by considering mother wavelets at different scales, thus decomposing the functions into a coarse approximation and detail coefficients at different levels [25, 39].

Moreover, differently from other representation bases, such as Fourier harmonics, the wavelet expansion enables a selective allocation of the unknown coefficients within the investigated domain, thus allowing to concentrate them only where the object is located

and to consider a finer texture only where needed. It allows to reduce the number of unknown parameters, as compared to the usually exploited pixel representation, thus improving inversion reliability [42, 43]. As such, they are intrinsically capable of approaching the inverse scattering problem within a multiscale framework, which has been shown to be an effective way to improve imaging results [44, 45].

After years of intensive research on l_1-problem solving, it would appear that most relevant algorithmic ideas have been either tried or, in many cases, re-discovered. Yet interestingly, additionally to the existing methods we also studied the classic idea of soft shrinkage in the application of nonlinear inverse problem (considering wavelet domain as well) where the inverse problem is solved without linearization and the application of the joint sparsity within microwave imaging.

1.2. Related works among others

As stated in above, inverse electromagnetic scattering problems have two fundamental properties that make their solution a challenging task, the first one is the nonlinearity of the scattering equations [5, 9] and the second one is the ill-posedness of the problem [5, 9]. The scattered field is an integral function of the total field induced inside the investigation domain multiplied by the contrast (the difference between the permittivities of the investigation domain and the background medium). The induced total field is also a nonlinear function of the permittivity [5, 9]. Consequently reconstruction of the permittivity from scattered fields is a nonlinear inverse problem.

Additionally, the fact that the scattered field is an integral function of the total field and the permittivity inside the investigation domain leads to an ill-posed inverse problem [5, 9]. The integral operator suppresses the contribution of the fast-varying components of the permittivity to the scattered field and makes them difficult to retrieve from the measurements [5, 9]. This physical nature of the scattering phenomena when combined with the fact that only a finite set of samples, possibly not at optimal location with respect to the sought scatterers and possibly as well not sufficiently independent from one another, can be collected by the measurements, which overall does not provide sufficient information. These samples are almost always corrupted by noise, and a discretized approximate model has to replace the exact mathematical description of the problem. Thus, this makes the inverse problem severely ill-posed [5, 9].

The strength of the nonlinearity in the scattering problem increases with the strength

of the scattered fields due to the object(s) in the investigation domain. Hence, higher permittivity (this notion being size-dependent as well) means stronger non-linearity. Most of the deterministic methods shown to be effective in solving electromagnetic inverse problems involving weak scatterers make use of linear first-order approximations [10], such as diffraction tomography [11], first-order Born approximation [5, 12], Kirchhoff approximation [5] and Rytov approximation [5, 12].

Even though these methods are computationally less demanding, they fail to provide accurate solutions when strong scatterers are present in the domain being investigated. In such cases, where the nonlinearity is strong, linearization does not produce an accurate approximation of the nonlinear problem and one needs to use more rigorous techniques such as inexact Newton nonlinear solver combined with either contrast-field or contrast-source formulations [5, 13], nonlinear conjugate gradient method [14], nonlinear steepest descent (e.g. nonlinear Landweber) algorithm [15, 16], and Levenberg-Marquardt scheme [17], and distorted Born methods [18]. These methods are computationally more demanding when compared to those that make use of the first order linearization schemes. Besides these deterministic methods, there are also nonlinear approaches based on stochastic (multiple/single agent) techniques such as simulated annealing [55], genetic algorithms [56], differential evolution [57], etc. Other methods that benefit from the best of the both worlds have also been developed. These make use of higher order linearization schemes or iterative application of the first-order ones. Examples include extended Born approximation [19] and the higher-order Born approximations [5, 20], Born iterative method (BIM) [21] and the variational Born iterative method [22], respectively. The methods in this last group are computationally more efficient when compared to nonlinear inversion methods but also applicable for a wider range of scatterer strength when compared to the first-order linearization methods.

Regardless of the method used for tackling the nonlinearity, the ill-posedness of the inverse scattering problem should be accounted for [9, 23, 24]. For this purpose, linear and nonlinear regularization methods, which minimize a cost function weighted between measurement/data misfit and a penalty term, have been developed. The most popular choice of the penalty term is the l_2-norm of the solution [23, 24]. The resulting minimization problem can be solved using the well-known Tikhonov scheme [23, 24]. Additionally, using truncated Landweber or conjugate gradient iterations leads to a similar type of regularization [5, 23]. In linear programming, all these methods effectively filter singular values of the discretized minimization problem, which are smaller than a specified threshold level or a regularization parameter to alleviate its ill-posedness.

As stated before, one common alternative is the convex relaxation of this combinatorial problem which consists in replacing the l_0-norm by the convex l_1-norm to take advantage of well-known algorithms in convex optimization. The basis pursuit (BP) principle or its denoising (BPDN) adaptation, proposed in [35], deals with the l_1 relaxation of problems using linear and quadratic programming algorithms (e.g., simplex, active set, or interior-point methods). Iterative thresholding algorithms (ISTA) [28] and its accelerated version FISTA [58], belonging to the family of forward-backward algorithms [59], are now well known to be very efficient for such l_1 relaxed problems. The fact that this convex relaxation provides sparse solutions comes from the singularity of the l_1-norm at zero [60].

On the other hand, the Compressive Sensing (CS) theory has been introduced for models relating the data to the unknowns that are linear [61]. As a consequence, a number of different CS toolboxes are available to tackle linear problems, whereas the development of such tools for the nonlinear case is ongoing and not yet fully assessed [62, 63]. For such a reason, the application of CS to inverse scattering problems has been first explored in those cases that can be handled through linear scattering approximations. These include the Born [4, 50, 64] and Rytov [65] approximations for weak scatterers, as well as the approximation based on the recently proposed virtual experiments framework [51, 66]. However, inverse scattering problems are non-linear, so that such an issue has to be addressed to gain a full advantage of CS in this framework. To this end, several authors are developing methods to take advantage of CS within inverse scattering approaches that are not limited to the range of validity of linear approximations.

For instance, the authors of [52, 65] have proposed an approach that is based on the solution of a number of linear inverse source problems. Such an approach does not consider approximations in modelling the non-linear scattering interactions and it can be applied to all cases in which the contrast source is sparse.

Another possible way to extend the applicability of CS for the solution of inverse scattering problems is to exploit distorted wave methods or distorted-iterated methods, in which the solution is iteratively achieved through a succession of linear inversion steps [49, 67–69]. Because of the linearity of the models considered at each step, these latter methods naturally lend themselves to be paired with CS. CS-based method based on l_1 norm minimization [61] was used to find the sparse solution in a sparse domain. In [52, 65], a method based on Bayesian framework [70] in compressive sensing [71] was utilized to solve an inverse scattering problem. The Bayesian framework has been suggested for some time for microwave medical imaging [70]. Compared with the traditional l_1 norm minimization [61], BCS searches for the sparse solution from a Bayesian probability

perspective.

Two popular non-quadratic cost functions are total variation and entropy [72]. Total variation puts a penalty on the sum of variations of the signal $J_2 = \|C\mathbf{x}\|_1$, where C is a discrete approximation to the gradient operator. Total variation is most frequently used in image processing applications, such as image restoration. In comparison to the Tikhonov regularization with $L = C$, the penalty on strong features is less severe, and the reconstruction can contain sharp edges. It works very well in practice with images that can be described as piecewise-smooth. Even if the regularizing function with l_1-penalty is similar with the total-variation we take the l_1 norm of the values of $\mathbf{x}$ instead of their derivatives. Total variation allows sparse jumps of the gradient of x, whereas the l_1 penalty favors sparse values of $\mathbf{x}$. To lower the penalty on strong features even further, several non-convex functions have also found use [73].

It is also worthwhile to mention that unlike inexact Newton [74] and Born iterative [67, 75] methods with sparsity constraints, the proposed scheme in Chapter 3 avoids generation of a sequence of linear sparse optimization problems and requires only one regularization parameter, which directly penalizes the nonlinear problem, to be set. Consequently, it simplifies the task of heuristic parameter tweaking, which is oftentimes very cumbersome for existing inversion algorithms.

1.3. Outline of thesis

While describing the contents of the thesis chapter by chapter, we briefly summarize our main contributions. The first major contribution is the development of a sparse signal reconstruction framework for microwave imaging. In this framework we formulate various optimization problems for microwave imaging scenarios. We adapt and use two paradigms for the numerical solution of the optimization problems. Finally, we carry out an extensive performance analysis of the proposed methods.

Chapter 2: Problem Statement

In this chapter, we presented the formulation of the direct problem modeling where incident wave and object are known by using a Method of Moments as discussed in [76]. We show numerical results in order to illustrate the sensitivity of the model to various noise levels and how the same level of noise has different randomization.

Chapter 3: Methods to solve Nonlinear Inverse Problem

We start by giving a brief overview of discrete ill-posed inverse problems, and motivate

the need for regularization. Then, we describe an important nonlinear inverse problem, sparse representation of signals using suitable bases. This problem serves a central role in the thesis: the basis of our work is the transformation of the microwave imaging problem into the problem of sparse signal representation. In this chapter we describe numerical optimization of the objective functions corresponding to l_1 regularization.

Next, we introduce our nonlinear inversion method in wavelet domain so that it takes the advantage of the adaptive multiresolution features of the wavelet basis to accommodate the trade off between spatial resolution and inversion stability. Specifically, we represent the unknown contrast sources in terms of the wavelet basis functions to reduce the number of nonzero coefficients whereas this proposed approach augments the work that has been done previously in this research area.

Chapter 4: Two-Step Inversion Method

This chapter is another contribution of our thesis. We address the analysis of the joint sparsity, a special case of group sparsity, which gives multiple sparse solutions that share a common nonzero support. We offer numerical results which demonstrate the potential of the proposed two-step inversion approach both in 2D and 3D.

Chapter 5: Summary, Conclusion and Future Work

This chapter summarizes the main ideas of the thesis and gives suggestions for further research in the area.

2. Formulation of the Direct Problem

2.1. Motivation

Modeling the associated direct problem in electromagnetic scattering is an initial yet essential step to describe how the parameters of the model are translated into observable effects. The direct problem consists in modeling the physics of the interaction between a known interrogating wave and the object which is supposed to be known as well. This interaction is described here by the Helmholtz wave equation. Applying the Green's theorem to this equation and taking into account the conditions of continuity of the fields and of radiation at infinity [77], we are led to an integral representation of the electric field consisting of two coupled integral equations, observation and coupling equation (or state), respectively. The solution of the direct problem that requires the discrete counterparts of these integral equations which are obtained in an algebraic framework using the "method of moments" [76, 78].

Validation of the direct model is a necessary step before switching to inversion. This is to verify that the direct model well describes the phenomena (in our case the scattered fields) to be observed in a controlled situation. Note that, in this thesis, we have no experimental data. This is why, when solving the inverse problem, we will generate synthetic data using a different direct model than the one used during the inversion. Therefore, we will avoid committing an 'inverse crime' in the sense of [9] which would consist in testing the inversion algorithm on the data obtained by means of a model closely related to the one used in the inversion. The validation of the direct model will thus consist in verifying that the differences between the scattered fields produced by these two models remain weak.

2.2. Formulation

In the following we consider homogeneous, non-magnetic and isotropic media and we place ourselves in a 2D imaging model in a transverse magnetic configuration (TM) where we consider that the object is of infinite extension and invariant according to one of its axes and is illuminated by a wave whose electric field ($\boldsymbol{E}^{\text{inc}}(\mathbf{r}) = E^{\text{inc}}\mathbf{z}$) is polarized parallel to this axis (perpendicular to the plane of the figure), which generates a scattered field perpendicular to the plane of the figure.

We now establish the equation of propagation of the electromagnetic wave on the basis of the Maxwell equations which constitute the basis of electromagnetism and which are written at an observation point $\mathbf{r}$ and at time t such as

$$\nabla \cdot \boldsymbol{D}(\mathbf{r}, t) = \rho(\mathbf{r}, t) \tag{2.1}$$

$$\nabla \times \boldsymbol{E}(\mathbf{r}, t) = -\frac{\partial \boldsymbol{B}(\mathbf{r}, t)}{\partial t} \tag{2.2}$$

$$\nabla \times \boldsymbol{H}(\mathbf{r}, t) = \boldsymbol{J}(\mathbf{r}, t) + \frac{\partial \boldsymbol{D}(\mathbf{r}, t)}{\partial t} \tag{2.3}$$

$$\nabla \cdot \boldsymbol{B}(\mathbf{r}, t) = 0, \tag{2.4}$$

where $\boldsymbol{D}(\mathbf{r}, t)$ is the electric displacement field in $[Cb/m^2]$, $\boldsymbol{E}(\mathbf{r}, t)$ is the electric field in V/m, $\boldsymbol{H}(\mathbf{r}, t)$ is the magnetic intensity or magnetic field in $[A/m]$, $\boldsymbol{J}(\mathbf{r}, t)$ is the current density in $[A/m^2]$ and $\boldsymbol{B}(\mathbf{r}, t)$ is the magnetic induction field in $[Wb/m^2]$, $t \in \mathbb{R}$, $\mathbf{r} = (x_1, x_2, x_3)$, $\nabla\times$ denotes curl operator and $\nabla\cdot$ denotes divergence operator. To solve the system of equations (2.2)-(2.4), one needs to introduce two additional equations, i.e., the so-called constitutive relations. Such equations enable to express the magnetic flux density $\boldsymbol{B}(\mathbf{r}, t)$ and the electric displacement $\boldsymbol{D}(\mathbf{r}, t)$ as a function of the electromagnetic parameters of the medium where Maxwell equations are set. The constitutive relations read as

$$\boldsymbol{B}(\mathbf{r}, t) = \mu \boldsymbol{H}(\mathbf{r}, t) \tag{2.5}$$

$$\boldsymbol{D}(\mathbf{r}, t) = \epsilon \boldsymbol{E}(\mathbf{r}, t), \tag{2.6}$$

where ϵ is the dielectric permittivity and μ is the magnetic permeability of the medium

at $\mathbf{r}$. We now find ourselves in the framework of a time-harmonic regime where a source emits a wave of pulsation ω and an implicit time dependence of the fields in $e^{-i\omega t}$ is chosen so that it will be omitted in the remainder of the discussion leading replacement of the derivations with respect to time appearing in the equations of Maxwell by a factor $-i\omega$. To establish the equation of propagation in electric field, we now place ourselves in a medium without sources or charges. From Maxwell's equations and taking into account constitutive relations, we arrive at

$$\nabla \times \nabla \times \boldsymbol{E} - \omega^2 \mu \epsilon \boldsymbol{E} = 0. \tag{2.7}$$

By developing the vector operator appearing in this equation, remembering that the electric field has only one component along the z-axis that we write E and introducing the propagation constant k of the medium considered ($k^2 = \omega^2 \epsilon \mu$), we obtain the equation of scalar Helmholtz waves verified by the electric field

$$\nabla^2 E(\mathbf{r}) + k^2 E(\mathbf{r}) = 0. \tag{2.8}$$

We shall now suppose that the object which we are interested in is immersed in a homogeneous medium D and is contained in a test domain C ($C \subset D$). The different media are non-magnetic (magnetic permeability equal to the one of vacuum $\mu_0 = 1,256 \times 10^{-6} Hm^{-1}$) and are characterized by their propagation constant such that $k(\mathbf{r})^2 = \omega^2 \varepsilon_0 \varepsilon_r(\mathbf{r}) \mu_0 + \mathrm{i}\omega\mu_0\sigma(\mathbf{r})$, where ε_0 and μ_0 are the permittivity and the permeability of air, respectively $\varepsilon_r(\mathbf{r})$ and $\sigma(\mathbf{r})$ are the relative permittivity and conductivity of the medium as $\mathbf{r} \in D$ is an observation point. The dielectric properties of D are described by the inhomogeneous contrast function defined as $\boldsymbol{\chi}(\mathbf{r}) = (k(\mathbf{r})^2 - k_B^2)$, where $k_B^2 = \omega^2 \varepsilon_0 \mu_0$ is the propagation constant of the embedding medium D. We assume that sources and receivers are located at the positions $\mathbf{r}_s$ and $\mathbf{r}_r$, respectively.

The scattered electric field $E^{\text{diff}}(\mathbf{r}_r, \mathbf{r}_s)$ measured via a receiver placed at $\mathbf{r}_r$ due to the incident wave emitted by a source placed at $\mathbf{r}_s$ adheres to the following domain integral equation [5]

$$E^{\text{diff}}(\mathbf{r}_r, \mathbf{r}_s) = \int_D G(\mathbf{r}_r, \mathbf{r}') \boldsymbol{\chi}(\mathbf{r}') E(\mathbf{r}', \mathbf{r}_s) d\mathbf{r}' \tag{2.9}$$

with $E(\mathbf{r}, \mathbf{r}_s)$ being the total electric field induced within the object by the incident wave, and $G(\mathbf{r}, \mathbf{r}')$ is a Green's function which represents the electromagnetic response to a line source radiating in free-space. In a homogeneous medium, in the case of two dimension

that concerns us, it is given by:

$$G(\mathbf{r},\mathbf{r}') = \frac{-i\omega\mu_0}{4}\mathrm{H}_0^{(1)}\left(k_B\|\mathbf{r}-\mathbf{r}'\|\right) \tag{2.10}$$

and $\mathrm{H}_0^{(1)}$ is the zero-th order Hankel function of the 1st kind for the 2D case. Furthermore, $E(\mathbf{r},\mathbf{r}_s)$ is obtained as

$$E(\mathbf{r},\mathbf{r}_s) = E^{\mathrm{inc}}(\mathbf{r},\mathbf{r}_s) + \int_D G(\mathbf{r},\mathbf{r}')\boldsymbol{\chi}(\mathbf{r}')E(\mathbf{r}',\mathbf{r}_s)d\mathbf{r}' \quad \forall \mathbf{r}\in D. \tag{2.11}$$

The direct problem is defined as the calculation of $E^{\mathrm{diff}}(\mathbf{r}_r,\mathbf{r}_s)$ by solving (2.9) and (2.11) when $\boldsymbol{\chi}(\mathbf{r})$, $G(\mathbf{r},\mathbf{r}')$ and $E^{\mathrm{inc}}(\mathbf{r},\mathbf{r}_s)$ are known whereas the inverse problem (or imaging problem) is defined as the determination of $\boldsymbol{\chi}(\mathbf{r})$ within a prescribed domain D from the knowledge of $E^{\mathrm{diff}}(\mathbf{r}_r,\mathbf{r}_s)$, $G(\mathbf{r},\mathbf{r}')$ and $E^{\mathrm{inc}}(\mathbf{r},\mathbf{r}_s)$ for N_s sources and N_r receivers.

Formulation of Contrast Source

We now rewrite the observation and coupling equations for the contrast source which is induced inside the object by the incident wave such as

$$E^{\mathrm{diff}}(\mathbf{r}_r,\mathbf{r}_s) = \int_D G(\mathbf{r}_r,\mathbf{r}')J(\mathbf{r}',\mathbf{r}_s)d\mathbf{r}' \tag{2.12}$$

and

$$J(\mathbf{r},\mathbf{r}_s) = J^{\mathrm{inc}}(\mathbf{r},\mathbf{r}_s) + \boldsymbol{\chi}(\mathbf{r}')\int_D G(\mathbf{r},\mathbf{r}')J(\mathbf{r}',\mathbf{r}_s)d\mathbf{r}' \quad \forall \mathbf{r}\in D. \tag{2.13}$$

the contrast source being defined as

$$J(\mathbf{r},\mathbf{r}_s) = \boldsymbol{\chi}(\mathbf{r})E(\mathbf{r},\mathbf{r}_s). \tag{2.14}$$

Solving equations (2.12) and (2.13) is done from their discrete counterparts obtained by using the method of moments. The domain D containing the unknown object is discretized in $N = N_x \times N_y$ small square pixels so that the electric field and the contrast can be considered as constants in each of them (Fig. 2.1).

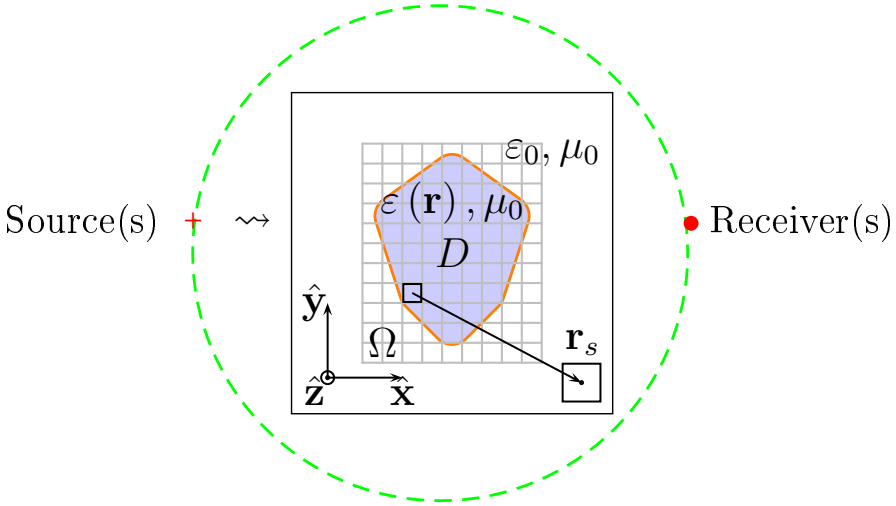


Figure 2.1.: Geometry of scattering experiment.

2.3. Discretization of the problem

The method of moments is a method commonly used to transform functional equations into matrix equations by projection on sets of basis functions and of test functions [76]. The problem being reduced under the form $\boldsymbol{AX} = \boldsymbol{B}$, where $\boldsymbol{A}$ is a matrix, $\boldsymbol{X}$ is a vector of unknown and $\boldsymbol{B}$ is a known vector, intuitively, when the matrix $\boldsymbol{A}$ is invertible, this allows us to get the desired solutions.

Let L be a linear operator $L : G \longrightarrow H$, and two functions $g \in G$ and $h \in H$ such as $h = L(g)$. One seeks to determine g knowing L and h. The function g can be written in the form $g = \sum_{i=1,\cdots,I} g_i u_i$, where $u_1, u_2, \cdots, u_I$ are constant coefficients which are a set of basis functions u_i. The property of linearity of the operator L gives us

$$h = L(g) = L\left(\sum_{i=1}^{I} g_i u_i\right) = \sum_{i=1}^{I} g_i L u_i \tag{2.15}$$

The goal is to approach equation (2.15) by a discrete linear system. It defines a set of test functions $t_1, t_2, \cdots, t_N$ on which projected the two members of this equation:

$$\langle t_n, h\rangle_H = \sum_{i=1}^{I} g_i \langle t_n, L u_i\rangle_H, \quad \forall n = 1, \cdots, N, \tag{2.16}$$

where $\langle ., .\rangle_H$ is a scalar product on H such as $\langle t_n, h\rangle_H = \int_H t_n(x) h^*(x) dx$.

Thus, by omitting the index H, one obtains the algebraic writing $h = Lg$ with

$$L = \begin{pmatrix} \langle t_1, Lu_1 \rangle & \langle t_1, Lu_2 \rangle & \cdots & \langle t_1, Lu_I \rangle \\ \langle t_2, Lu_1 \rangle & \langle t_2, Lu_2 \rangle & \cdots & \langle t_2, Lu_I \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle t_N, Lu_1 \rangle & \langle t_N, Lu_2 \rangle & \cdots & \langle t_N, Lu_I \rangle \end{pmatrix}, \quad g = \begin{pmatrix} g_1 \\ \vdots \\ g_I \end{pmatrix}, \quad h = \begin{pmatrix} \langle t_1, h \rangle \\ \vdots \\ \langle t_N, h \rangle \end{pmatrix},$$

The choice of basic functions and test functions depends essentially upon the nature of the physical problem to be treated. Note that, if we choose test functions identical to the basis functions, we come across with the special case of the Galerkin method [79].

The two coupled equations (2.9) and (2.11) have a bilinearity property with respect to the two variables $\boldsymbol{\chi}$ and E. Then, the method of moments can be applied to these equations. Indeed, if one considers the equations of observation and of state like two equations linear with respect to the variable E, equations (2.9) and (2.11) can be written as follows

$$E^{\text{diff}}(\mathbf{r}_r, \mathbf{r}_s) = \int_D G(\mathbf{r}_r, \mathbf{r}')\boldsymbol{\chi}(\mathbf{r}')E(\mathbf{r}', \mathbf{r}_s)d\mathbf{r}' \Rightarrow E^{\text{diff}} = L_1 E \tag{2.17}$$

$$E^{\text{inc}}(\mathbf{r}, \mathbf{r}_s) = E(\mathbf{r}, \mathbf{r}_s) - \int_D G(\mathbf{r}, \mathbf{r}')\boldsymbol{\chi}(\mathbf{r}')E(\mathbf{r}', \mathbf{r}_s)d\mathbf{r}' \Rightarrow E^{\text{inc}} = L_2 E, \tag{2.18}$$

where L_1 and L_2 are linear operators.

Then, in order to apply the principle of the method of moments, we must choose the basis functions u_{1i} and u_{2i} and the test functions t_{1i} and t_{1i} in order to construct scalar products

$$\left\langle \int_D G(\mathbf{r}_r, \mathbf{r}')\boldsymbol{\chi}(\mathbf{r})E(\mathbf{r}, \mathbf{r}_s)d\mathbf{r}', t_{1i} \right\rangle = L_{1ij} \tag{2.19}$$

$$E^{\text{inc}}(\mathbf{r}, \mathbf{r}_s) = E(\mathbf{r}, \mathbf{r}_s) - \int_D G(\mathbf{r}, \mathbf{r}')\boldsymbol{\chi}(\mathbf{r})E(\mathbf{r}, \mathbf{r}_s)d\mathbf{r}' \Rightarrow E^{\text{inc}} = L_2 E, \tag{2.20}$$

Here, we will take basis functions as the characteristic functions of the pixels partitioning the domain D and test functions as distributions of Dirac δ located at the centers of each of the elementary pixels for the coupling equation and at the points of measurement for the observation equation.

Thanks to a classical method of moments using pulse basis/point matching method [80] where the domain under test D is discretized into $N = N_x \times N_y$ pixels, a discretized version of the previous equations is obtained. In so doing $\boldsymbol{\chi}(\mathbf{r})$ is approximated by using

pulse basis functions defined over square cells:

$$\boldsymbol{\chi}(\mathbf{r}) = \sum_{n=1}^{N} \boldsymbol{\chi}_n C_n(\mathbf{r}), \tag{2.21}$$

where $C_n(\mathbf{r})$ is the basis function defined as

$$C_n(\mathbf{r}) = \begin{cases} 1, & \mathbf{r} \in D_n \\ 0, & \mathbf{r} \notin D_n, \end{cases} \tag{2.22}$$

in a 'pixel' representation. Here, $\boldsymbol{\chi}$ is a $N \times 1$ vector storing the samples of $\boldsymbol{\chi}(\mathbf{r})$. Additionally, the discretized version of (2.9) stands as

$$\mathbf{E}_i^{\text{diff}} = \mathbf{G_{or}} \mathbf{J}_i, \quad i = 1, \ldots, N_s, \tag{2.23}$$

where $\mathbf{E}_i^{\text{diff}}$ and $\mathbf{J}_i$ are complex vectors of size N_r and N respectively and $\mathbf{G_{or}}$ a complex matrix of size $N_r \times N$. The discretized version of (2.14) is then

$$\mathbf{J}_i = \text{diag}\,(\boldsymbol{\chi}) \mathbf{E}_i, \quad i = 1, \ldots, N_s, \tag{2.24}$$

where $\mathbf{E}_i$ and $\boldsymbol{\chi}$ are complex vectors of size N and $\text{diag}\,(\boldsymbol{\chi})$ is a diagonal matrix of size $N \times N$ obtained from $\boldsymbol{\chi}$. Finally the discretized version of (2.11) is

$$\mathbf{E}_i = \mathbf{E}_i^{\text{inc}} + \mathbf{G_{oo}} \mathbf{J}_i, \quad i = 1, \ldots, N_s \tag{2.25}$$

where $\mathbf{E}_i^{\text{inc}}$ is a complex vector of size N and $\mathbf{G_{oo}}$ a matrix of size $N \times N$.

Thus, if we denote the center of the elementary pixel μ_i (permeability) by r_i and assuming that the contrast $\boldsymbol{\chi}$ and the electric field E are constant in each pixel μ_i and equal to $\boldsymbol{\chi}(\mathbf{r}_i)$ and $E(\mathbf{r}_i)$ respectively, the elements of the coupling matrix are written as

$$L_{2ij} = \delta_{ij} - \boldsymbol{\chi}(\mathbf{r}_i) \int_{\mu_j} G(\mathbf{r}_i, \mathbf{r}') d\mathbf{r}', \tag{2.26}$$

where δ_{ij} is Kronecker delta. The numeric integration of the Green's function on the square cell Δ_i is analytical following [76] where the square cell is approximated by a circular cell

of the same surface leading to:

$$\int_{\Delta_i} G(\mathbf{r}_j, \mathbf{r}')\mathrm{d}\mathbf{r}' = \begin{cases} \dfrac{1}{k_B^2}\left[\dfrac{\mathrm{i}}{2}\pi k_B R \mathrm{H}_1^{(1)}\left(k_B R\right) - 1\right], & \text{if } i = j, \\ \dfrac{\mathrm{i}}{2k_B}\pi R \mathrm{H}_0^{(1)}\left(k_B r_{ij}\right)\mathrm{J}_1\left(k_B R\right), & \text{if } i \neq j, \end{cases} \tag{2.27}$$

with $r_{ij} = |x_j - x_i|$ and $R = \sqrt{\frac{\Delta_x \Delta_y}{\pi}}$; $\mathrm{H}_1^{(1)}$ being the Hankel function of first order and first kind and J_1 the Bessel function of the first kind.

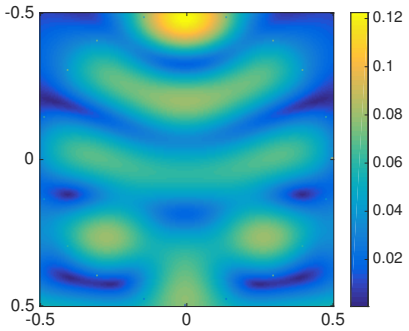
2.4. Validation of the model

Firstly our numerical code has been validated by comparison with numerical results provided by an analytical solution available for the case of a cylindrical obstacle illuminated by a line source for a single frequency and a single incidence.

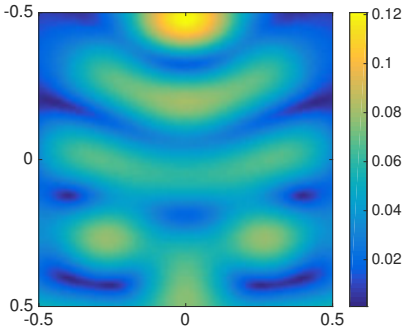
The validation in Fig. 2.2 has been done with the frequency being 3 GHz and the radius of the cylinder equal to $\lambda_0/2$ whereas $\lambda_0 = 1$. The relative permittivity of the object is 2 whereas it is 1 for the embedding medium. The source is located at $(0, 2\lambda_0)$ while there is 36 receivers.

Generally the difference between the model and the data is low but increases with frequency. This is explained by the fact that the discretization, i.e. pixel size, is constant whatever the frequency is. Thus, we have seen that a good convergence of the calculations carried out using the method of moments requires to index this size to the wavelength. Fig. 2.2 validates that with a high discretization ($N = 125$) our model is well-fitted with the data. On the other hand, Fig. 2.3 shows the evolution of the normalized error on to the total field and the scattered field as a function of the discretization number N.

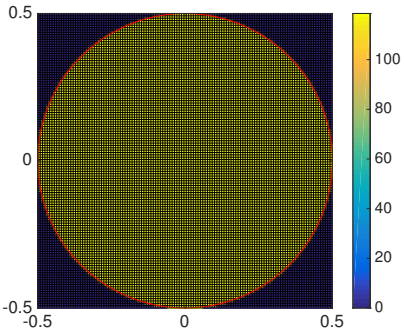
It is also worthwhile to mention that we are not in Born approximation case as it can be seen in Fig. 2.6. The scattered fields are calculated for the case of without Born approximation throughout this manuscript.



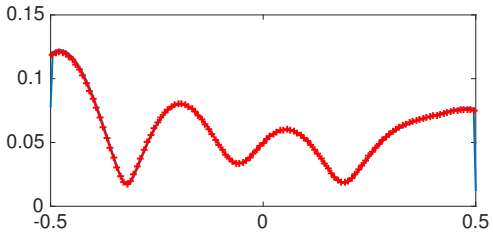
(a) Modulus of the exact total field



(b) Modulus of the total field obtained via MoM

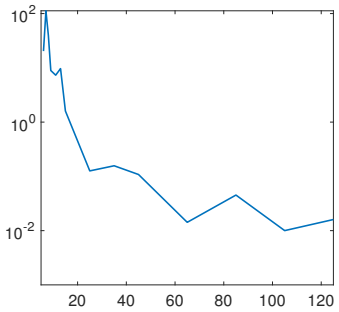


(c) Discretization of the obstacle 125×125

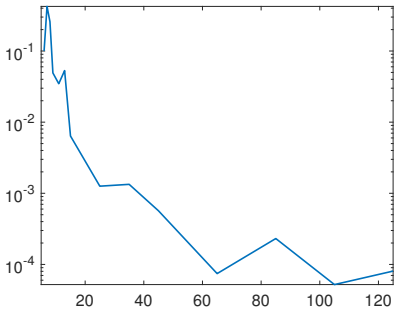


(d) Comparison of the modulus of the exact and the computed total field at the center of the cylinder

Figure 2.2.: Validation of the model when $N = 125 \times 125$.



(a) Total field



(b) Scattered field

Figure 2.3.: Evolution of the normalized error of the total field (left) and on the scattered field (right) as a function of the discretization number N.

2.5. Comparison of the data with different models

In this section we studied an example where has the true value of the relative permittivity of the object is 2 whereas it is 1 for embedding medium. The scattering object which is a dielectric square sided λ (1 m) under test is contained in a $l = 3 \times \lambda$-sided square investigation area D centered at the origin, and the discretization size is $n \times n = 36 \times 36$ for the forward problem. The number of transmitters and receivers located around the investigation area is 29. The frequency of the transmitters is 300 MHz. The measured field samples are generated by adding either 10 dB or 20 dB noise with a zero mean additive Gaussian noise with unknown variance.

From Fig. 2.4 we notice that the outputs of the models coincide relatively well with the data, both in modulus and in phase. However, as we increase the level of noise, the discrepancy between the model and the data increases. On the other hand, Fig. 2.5 shows us that whenever we add noise of the same level to our model we get different randomization of the same noise level with respect to each other and also to the data. This specified issue led us to be careful while choosing the model to solve the inverse problem due to the sensitivity of the data to noise.

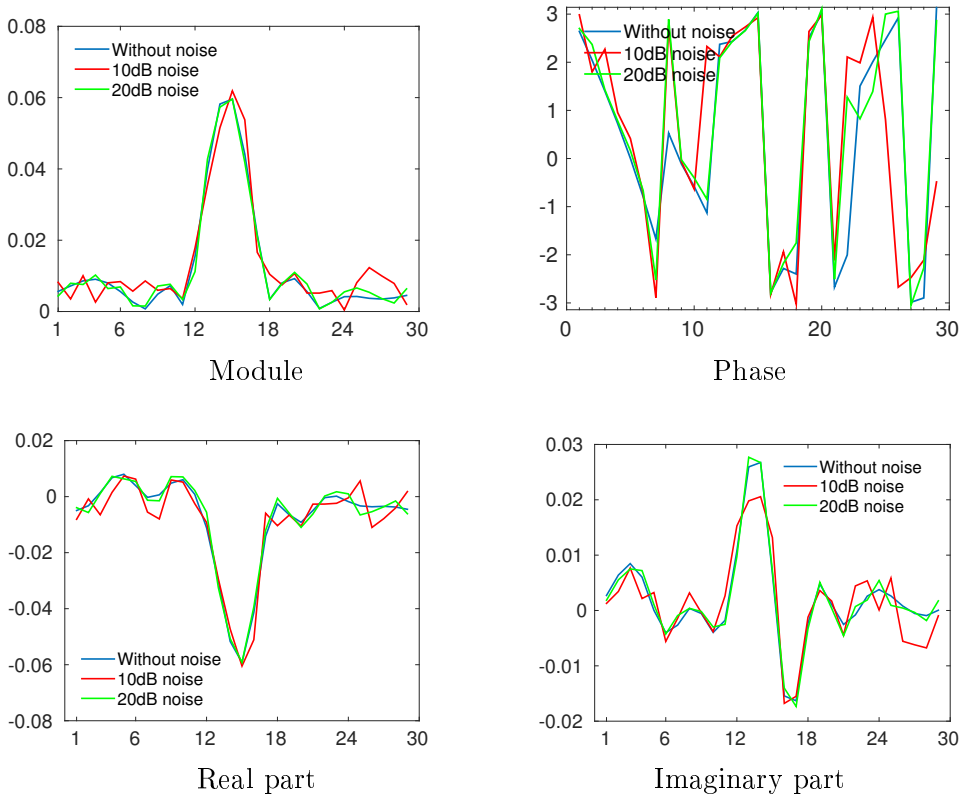


Figure 2.4.: Example for 1 source and $N = 29$ receivers for various SNR.

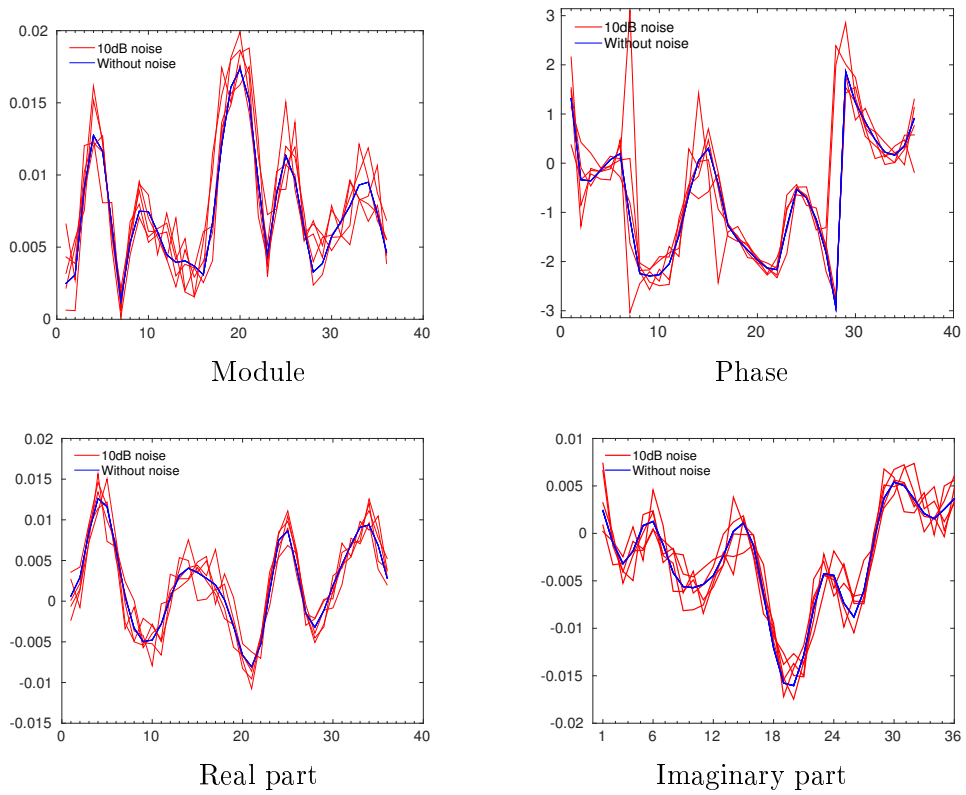


Figure 2.5.: Example for 1 source and $N = 29$ receivers for various 10 dB noise level.

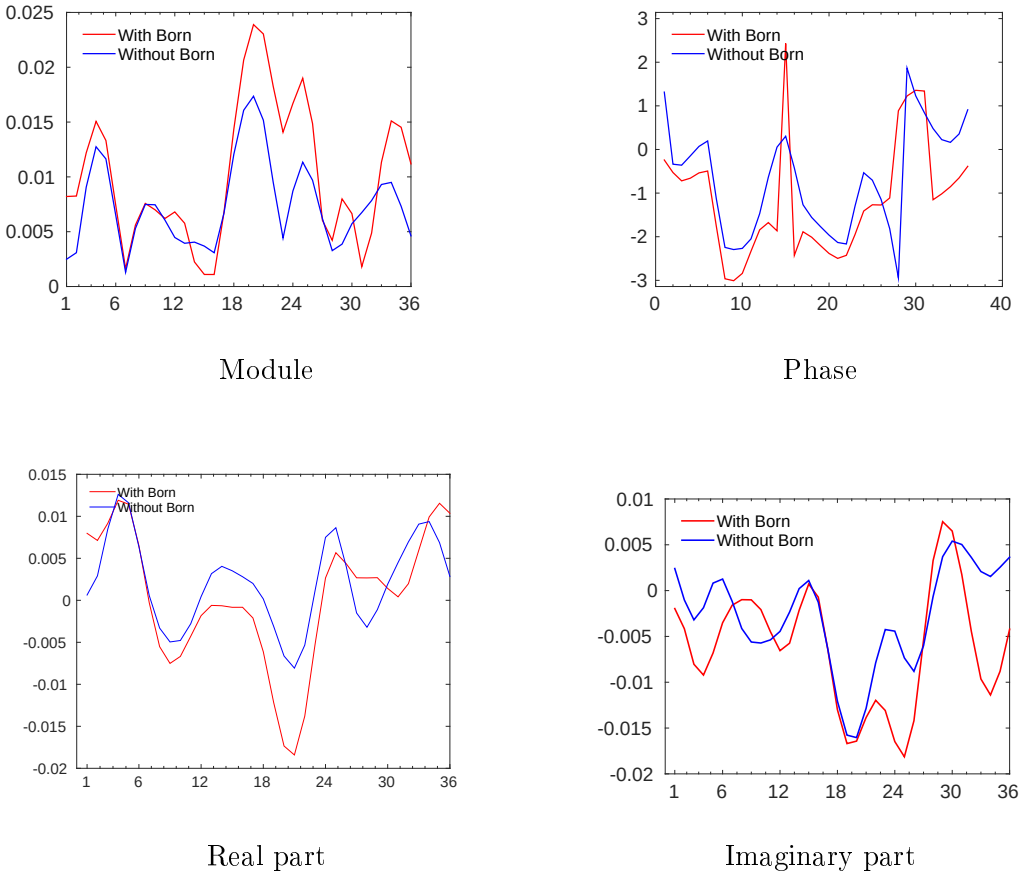


Figure 2.6.: Difference of scattered field obtained with and without Born approximation.

2.6. Conclusion

In this chapter, we presented the modeling of the direct problem where incident wave and object are known. Primarily, we are interested in the calculation of the scattered field resulting from their interaction. The resolution of the latter corresponds to a first necessary step to solve the inverse problem in which it is a matter of reconstructing the unknown object, the incident wave and the scattered field then being known. In the absence of experimental data, we have studied the electromagnetic field calculations carried out using this forward model by comparison with synthetic results obtained using the forward model added with noise.

3. Methods to Solve Nonlinear Inverse Problem

3.1. Motivation

In this chapter we propose two methods in order to solve the nonlinear inverse problem. Initially, we propose a method where sparsity constraint is directly applied to the problem of reconstructing the complex internal dielectric properties of an object based on knowledge of the external scattered field which is generated by the interaction between the object and a known incident field. The nonlinear optimization problem is solved by the iterative algorithm of soft shrinkage in order to enforce the sparsity constraint. Sparsity is applied at each iteration by a soft thresholding function.

Since the electromagnetic inverse scattering problem which we consider is nonlinear and ill-posed, the reduction of the number of unknown parameters is important to alleviate the ill-posedness [25, 40, 41]. As a matter of fact, wavelet expansion (as considered hereafter) should allow (this is in effect a topic of discussion in the present work, noticing that optimality of a basis is strongly linked to the way the scatterers at hand fit it) to reduce the number of unknown parameters with respect to the usually exploited pixel representation so that it advances the inversion reliability [42].

Moreover, the adopted basis plays a key role as it has to be accurate and efficient. Efficiency requires a reduced number of coefficients in the representation, while accuracy involves a low representation error, which quantifies the mismatch between the actual function and its projection onto the considered basis. In this respect, this chapter argues a second approach which studies whether the wavelet basis is a good choice to accommodate for the trade off between efficiency and accuracy of the representation or not in microwave imaging with simulation results provided.

3.2. Soft Shrinkage Method for Nonlinear equations

Soft shrinkage is an approach which minimizes a nonlinear Tikhonov functional with sparsity promoting penalty term. The algorithm is based on the iterated soft shrinkage approach originated for linear operators in the work [28]. A generalization to nonlinear inverse problems has been studied in [7, 81].

The algorithm performs a gradient descent step which involves the adjoint gradient of the cost function with a step size τ, like Landweber method, and then a shrinkage step. The latter enforces the sparsity of the reconstruction by setting the small coefficients to zero. Following this, Barzilai and Borwein (BB) method is suggested in order to choose the step size to overcome the slow convergence of the iterative soft shrinkage algorithm with fixed step size. The solution of the inverse problem can be obtained by minimizing the cost function which is the difference, in other words called an error, between the measured quantity and the solution obtained by a forward problem. The cost function to be minimized is of the form

$$F(\boldsymbol{\chi}) = \underbrace{\frac{1}{2}\left\|\zeta(\boldsymbol{\chi}) - E^{\text{diff}}\right\|_2}_{K(\boldsymbol{\chi})} + \alpha\|\boldsymbol{\chi}\|_1. \tag{3.1}$$

In other words, by combining (2.23) and (2.25) and using $\mathbf{J}_i^{\text{inc}} = \text{diag}\,(\boldsymbol{\chi})\mathbf{E}_i^{\text{inc}}$ $i = 1, \ldots, N_s$ the inverse problem can be rewritten as

$$\boldsymbol{\chi}^\star = \underset{\boldsymbol{\chi}}{\text{argmin}} \left\|\boldsymbol{\zeta}_i - \mathbf{G}_{\mathbf{or}}\,\text{diag}\,(\boldsymbol{\chi})\left[\mathbf{I} - \text{diag}\,(\boldsymbol{\chi})\mathbf{G}_{\mathbf{oo}}\right]^{-1}\mathbf{J}_i^{\text{inc}}\right\|_2 + \alpha\|\boldsymbol{\chi}\|_1, \tag{3.2}$$

where $\boldsymbol{\zeta}_i$ is a vector of size N_r which gathered the signal due to the source $\#i$ measured by the N_r receivers. It can be seen from (3.2) that the inverse problem is nonlinear in $\boldsymbol{\chi}$ and ill-posed. The dimensions of these matrices may become quickly important as soon as the domain D is more than a few wavelengths. Indeed, the convergence of the calculations carried out using the method of moments imposes an upper limit on the size of the pixels to be considered [76, 82]. Pixel sides $a \leq \lambda/10$ generally ensure a good convergence.

Iterative soft shrinkage is a standard approach for minimizing a functional involving the l_1-penalty and it has the form as in (3.1) where $\zeta : X \mapsto Y$ is a bounded and nonlinear operator. The l_1 penalty can promote *a-priori* knowledge of the sparse representation. At first, the algorithm is started by choosing an initial guess χ^1, and the iteration continues

as

$$\boldsymbol{\chi}^{k+1} = S_\alpha\Big(\boldsymbol{\chi}^k - \tau\zeta'^*(\boldsymbol{\chi}^k)\Big[\zeta(\boldsymbol{\chi}^k) - E^{\text{diff}}\Big]\Big), \tag{3.3}$$

where τ is the step size, $\zeta'(\boldsymbol{\chi})$ is the gradient of the nonlinear function $\zeta(\boldsymbol{\chi})$ with respect to $\boldsymbol{\chi}$, and $\zeta'^*(\boldsymbol{\chi})$ is the adjoint of the operator $\zeta'(\boldsymbol{\chi})$. S_α is the soft shrinkage operator defined componentwise by

$$(S_\alpha(\boldsymbol{\chi}))_i = \begin{cases} (|\boldsymbol{\chi}_i| - \alpha)\text{sign}(\boldsymbol{\chi}_i), & \text{if} \quad |\boldsymbol{\chi}_i| > \alpha \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

The term $\zeta'^*(\boldsymbol{\chi}^k)\Big[\zeta(\boldsymbol{\chi}^k) - E^{\text{diff}}\Big]$ is the gradient of the discrepancy $\frac{1}{2}\big\|\zeta(\boldsymbol{\chi}) - E^{\text{diff}}\big\|^2$.

3.2.1. Gradient K'

The gradient is obtained by using the adjoint method. The main idea is to obtain the gradient (Fréchet derivative) of the whole discrepancy term and avoid calculating $\zeta'^*(\boldsymbol{\chi})$ in the iteration rule (3.3). For the calculation of $K(\boldsymbol{\chi})$ we use the adjoint method which is presented in [54] to reduce the calculation costs. This has been achieved by solving the adjoint problem as in A.1. For simplicity, at first it can be again stated with simpler notations that the least squared cost function describing the goodness of fit of a hypothesized case to a measured case can be written as

$$K(\boldsymbol{\chi}) = \frac{1}{2}\big\|\zeta\left(\boldsymbol{\chi}\right) - E^{\text{diff}}\big\|^2 = \frac{1}{2}\langle\zeta\left(\boldsymbol{\chi}\right) - E^{\text{diff}}, \zeta\left(\boldsymbol{\chi}\right) - E^{\text{diff}}\rangle \tag{3.5}$$

If $\boldsymbol{\chi}$ is perturbed by a small amount $\delta\boldsymbol{\chi}$, then the $K(\boldsymbol{\chi})$ changes according to

$$K\left(\boldsymbol{\chi} + \delta\boldsymbol{\chi}\right) = K\left(\boldsymbol{\chi}\right) + \langle\delta\boldsymbol{\chi}, \zeta\left(\boldsymbol{\chi}\right)^*\left(\zeta\left(\boldsymbol{\chi}\right) - E^{\text{diff}}\right)\rangle + \overline{\langle\delta\boldsymbol{\chi}, \zeta\left(\boldsymbol{\chi}\right)^*\left(\zeta\left(\boldsymbol{\chi}\right) - E^{\text{diff}}\right)\rangle} + O\left(\delta\boldsymbol{\chi}\right). \tag{3.6}$$

By definition of the Fréchet derivative,

$$\boldsymbol{\nabla}\boldsymbol{\chi}K = \overline{\zeta\left(\boldsymbol{\chi}\right)^*\left(\zeta\left(\boldsymbol{\chi}\right) - E^{\text{diff}}\right)}, \tag{3.7}$$

where the overlined terms are the complex conjugate ones and gradient $K'\left(\boldsymbol{\chi}\right) = \nabla_{\boldsymbol{\chi}}K\left(\boldsymbol{\chi}\right)\big|_{\boldsymbol{\chi}}$. The detailed work can be found in A.1 as stated above.

3.2.2. Smooth Gradient K'_s

It is emerged in practice that the gradient $K'(\boldsymbol{\chi})$ has unnatural oscillating properties which can be avoided by using the smoother gradient. This process is also called denoising [54]. Therefore, we look for a Sobolev smoothed gradient $K'_s(\boldsymbol{\chi})$. For instance, $K'(\boldsymbol{\chi})\zeta$ is the image of $\zeta \in D$ under application of the adjoint operator considered as an operator mapping from D into $P = L_2(\Omega)$. Its image denoted by $K'_s(\boldsymbol{\chi})$ under the adjoint operator with respect to the newly defined weighted inner product mapping into the smaller space $\hat{P}$ such as

$$\langle K'(\boldsymbol{\chi})x, \zeta\rangle_D = \langle x, K'(\boldsymbol{\chi})\zeta\rangle_P = \langle x, K'_s(\boldsymbol{\chi})\zeta\rangle_{\hat{P}} \tag{3.8}$$

following

$$K'_s(\boldsymbol{\chi})\zeta = (\delta I - \beta\Delta)^{-1}K'(\boldsymbol{\chi})\zeta, \tag{3.9}$$

where I refers to the identity, and Δ refers to the Laplacian operator. A proper choice of the weighting parameters δ and β will allow us to drive the regularization properties of our algorithm in an efficient and predictable way. We should choose β close to one in order not to lose the differentiability for the next iteration. For theoretical justification we refer to [83]. Applying Green's formula to the right hand side of equation (3.8) yields equation (3.9). Following this, we can rewrite the iteration formula with this property following as

$$\boldsymbol{\chi}^{k+1} = S_\alpha(\boldsymbol{\chi}^k - \tau K'_s(\boldsymbol{\chi}^k)). \tag{3.10}$$

3.2.3. The Step Size

The step size τ can be determined in order to fasten the algorithm. The motivation for increasing the rate of convergence is the comparison with the classical Landweber iteration whose slow convergence results from using a constant step size which is very small. Therefore, we select the step size in a way to increase the convergence speed where we consider only the steepest descent operation $\boldsymbol{\chi}^k - \tau K'_s(\boldsymbol{\chi}^k)$ of the algorithm. The selection is done by the two-point rule of Barzilai and Borwein (BB) which calculates the step size as

$$\tau_k = \arg\min_\tau \left\|\tau(\boldsymbol{\chi}^k - \boldsymbol{\chi}^{k-1}) - (K'_s(\boldsymbol{\chi}^k) - K'_s(\boldsymbol{\chi}^{k-1}))\right\|_2. \tag{3.11}$$

The choice of this functional is motivated by the secant equation $K'(\boldsymbol{\chi}^k) = K'\boldsymbol{\chi}^{k-1} + B(\boldsymbol{\chi}^k - \boldsymbol{\chi}^{k-1})$ with τ times the identity operator as the approximation of the Hessian B and describes the approximation performance of the last iteration step. This equation does

not necessarily have a solution, so it is solved in a least-squared sense. In one-dimensional real case of this procedure implies the secant method. By minimizing the functional we get the following formula for the step size

$$\tau_k = \frac{\left\langle \boldsymbol{\chi}^k - \boldsymbol{\chi}^{k-1}, \boldsymbol{\chi}^k - \boldsymbol{\chi}^{k-1} \right\rangle}{\left\langle \boldsymbol{\chi}^k - \boldsymbol{\chi}^{k-1}, K_s'\left(\boldsymbol{\chi}^k\right) - K_s'\left(\boldsymbol{\chi}^{k-1}\right) \right\rangle}. \tag{3.12}$$

Another approach, also from BB, is that τ times the identity operator imitates the inverse of the Hessian B over the last step. This results in

$$\tau_k = \arg\min_{\tau} \left\| (\boldsymbol{\chi}^k - \boldsymbol{\chi}^{k-1}) - \tau (K_s'(\boldsymbol{\chi}^k) - K_s'(\boldsymbol{\chi}^{k-1}) \right\| \tag{3.13}$$

and therefore we get [30, 83]

$$\tau_k = \frac{\left\langle \boldsymbol{\chi}^k - \boldsymbol{\chi}^{k-1}, K_s'\left(\boldsymbol{\chi}^k\right) - K_s'\left(\boldsymbol{\chi}^{k-1}\right) \right\rangle}{\left\langle K_s'\left(\boldsymbol{\chi}^k\right) - K_s'\left(\boldsymbol{\chi}^{k-1}\right), K_s'\left(\boldsymbol{\chi}^k\right) - K_s'\left(\boldsymbol{\chi}^{k-1}\right) \right\rangle} \tag{3.14}$$

with

$$K_s'\left(\boldsymbol{\chi}^k\right) = (\delta I - \beta \Delta)^{-1} \zeta'^{*}(\boldsymbol{\chi}^k) \left[\zeta(\boldsymbol{\chi}^k) - E^{\text{diff}} \right]. \tag{3.15}$$

In the implementation we use this step size as an initial guess and it is decreased geometrically until the Armijo condition [84] is satisfied. In general, descent methods determine a descent direction d_k with a step size $\alpha_k \in (0, 1]$ by an inexact line search such as Armijo, Wolfe or Goldstein backtracking [84] schemes forming $\mathbf{x}_{k+1} = \mathbf{x}_k + \tau_k d_k$. The scheme is repeated until a stopping criteria is reached. Firstly, choosing a suitable inexact line search maintains that the sequence of function values is monotonically decreasing ($f_{k+1} \leq f_k$). Secondly, the sequence x_k is converging globally. In other words, the method is convergent even if the initial point is far away from the minimizer. The first case implies the minimization of the objective function and the second case provides that the method is not depending on the initial point. Particularly, Armijo's line search satisfies

$$f\left(\mathbf{x}_k + \tau_k d_k\right) \leq f(\mathbf{x}_k) + c_1 \tau_k \nabla f(\mathbf{x}_k) d_k, \tag{3.16}$$

where $c_1 \in (0, \frac{1}{2})$ and τ_k is the largest $\tau \in \{s, \rho s, \ldots\}$ with $s > 0$ and $\rho \in (0, 1)$ such that (3.16) is satisfied. This concludes that function values satisfy the condition $f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k)$ imposing the monotonicity to the sequence of functions generated by this scheme and this scheme is globally convergent [84, 85].

However, in [85] a modified version of Armijo condition is proposed such as

$$F\left(\boldsymbol{\chi}^1 + S_\alpha\left(\boldsymbol{\chi}^k - \tau K_s'\left(\boldsymbol{\chi}^k\right)\right)\right) \leq \max_{k-M+1\leq n\leq k} F\left(\boldsymbol{\chi}^n\right) - \tau s\left\|S_\alpha\left(\boldsymbol{\chi}^k - \tau K_s'\left(\boldsymbol{\chi}^k\right)\right) - \boldsymbol{\chi}^k\right\|, \tag{3.17}$$

where s is a small number and M is an integer. The right hand side of the new Armijo type line search is greater than the original Armijo's rule implying that the new method can take bigger step sizes compared to the descent methods using original Armijo condition. Thus, we can get faster convergence. This is also called weaker monotonicity. In original Armijo condition if no step size can be obtained to satisfy the condition the algorithm usually stops by rounding errors and preventing further progress. Further details can be found in [30, 85]. The equation in (3.17) is motivated by the Taylor approximation of $F\left(\boldsymbol{\chi}^{k+1}\right)$ as following

$$F\left(\boldsymbol{\chi}^{k+1}\right) = F\left(\boldsymbol{\chi}^k\right) + F'\left(\boldsymbol{\chi}^k + \xi\left(\boldsymbol{\chi}^{k+1} - \boldsymbol{\chi}^k\right)\right)\left[\boldsymbol{\chi}^{k+1} - \boldsymbol{\chi}^k\right], \tag{3.18}$$

where $\xi \in (0,1)$. We can now restrict the gradient to a negative direction by choosing τ such that $F\left(\boldsymbol{\chi}^{k+1}\right) \leq F\left(\boldsymbol{\chi}^k\right)$. Hence,

$$F\left(\boldsymbol{\chi}^{k+1}\right) = F\left(\boldsymbol{\chi}^k\right) - \left|F'\left(\boldsymbol{\chi}^k + \xi\left(\boldsymbol{\chi}^{k+1} - \boldsymbol{\chi}^k\right)\right)\left[\boldsymbol{\chi}^{k+1} - \boldsymbol{\chi}^k\right]\right|. \tag{3.19}$$

Furthermore, we take a sufficiently small and fixed $s > 0$ such as

$$s\tau\left|\langle\boldsymbol{\chi}^{k+1} - \boldsymbol{\chi}^k, \boldsymbol{\chi}^{k+1} - \boldsymbol{\chi}^k\rangle\right| \leq \left|F'\left(\boldsymbol{\chi}^k + \xi\left(\boldsymbol{\chi}^{k+1} - \boldsymbol{\chi}^k\right)\right)\left[\boldsymbol{\chi}^{k+1} - \boldsymbol{\chi}^k\right]\right|. \tag{3.20}$$

The aim is to get a negative direction, therefore we can estimate it as

$$F\left(\boldsymbol{\chi}^{k+1}\right) \leq F\left(\boldsymbol{\chi}^k\right) - s\tau\left|\langle\boldsymbol{\chi}^{k+1} - \boldsymbol{\chi}^k, \boldsymbol{\chi}^{k+1} - \boldsymbol{\chi}^k\rangle\right|. \tag{3.21}$$

However, this estimation causes slow convergence because of strong monotonicity. Therefore, we change $F\left(\boldsymbol{\chi}^k\right)$ to $\max_{k-M+1\leq n\leq k} F\left(\boldsymbol{\chi}^n\right)$.

Adding $\boldsymbol{\chi}^{k+1} = S_\alpha\left(\boldsymbol{\chi}^k - \tau K_s'\left(\boldsymbol{\chi}^k\right)\right)$ leads us back to equation (3.17).

We should point out that we cannot use the rule developed by Brazilai and Borwein and the monotonicity criterion in the first iteration of the algorithm. Therefore, we use an initial step size obtained by 'fminsearch' which finds the minimum of a scalar function of several variables, starting at an initial estimate zero. This is generally referred to as

unconstrained nonlinear optimization [86]. In our simulations, we follow the BB method in

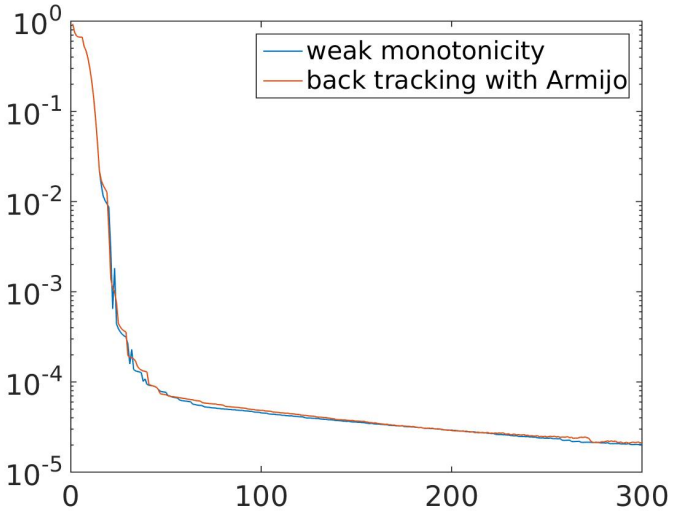


Figure 3.1.: Comparison between weak monotonicity and back-tracking with Armijo condition, error on cost function (vertical axis) w.r.t iteration number (horizontal axis).

order to obtain the required step size τ. We have also discovered that weak monotonicity does not improve the convergence rate of the cost function in our formulations compared with the back tracking line search with Armijo condition as can be seen in Fig. 3.1 where iteration number which is 300 is our stopping criteria for the scatterer Fig. 3.4f. A standard monotonicity condition holds for $M = 1$. We have used the value $M = 2$ in our reconstructions.

Algorithm 1 The Barzilai-Borwein (BB) approach for choosing τ in line 6 of Algorithm 1.

1: **Input** The iteration counter k, $\tau_0 = 0$, $h^k = \boldsymbol{\chi}^k - \boldsymbol{\chi}^{k-1}$ and $g^k = K'(\boldsymbol{\chi}^k) - K'(\boldsymbol{\chi}^{k-1})$
2: **if** $k = 0$ **then**
3: $\quad \tau = \text{fminsearch}(K(\boldsymbol{\chi}^1), \tau_0)$
4: **else**
5: $\quad$ **output** $\tau_k = \langle h^k, h^k \rangle / \langle h^k, g^k \rangle$
6: **end for**

3.2.4. Soft Thresholding

Recently, algorithms known as iterative soft-thresholding (or forward-backward splitting) for optimization with sparse regularizers, including [28, 59, 87, 88] have been used in many

applications. These methods address problems of the form

$$\min_{\mathbf{x}} f(\mathbf{x}) = J_1(\mathbf{x}) + J_2(\mathbf{x}) \tag{3.22}$$

wherein $J_2(\mathbf{x})$ is convex and possibly non-differentiable while $J_1(\mathbf{x})$ is assumed to be differentiable and convex with a Lipschitz-continous gradient as stated before. These algorithms solve the non-smooth optimization problem directly with a projection-like operator instead of converting this problem into a constrained optimization problem. Particularly, these methods take the form of

$$\mathbf{x}_{k+1} \leftarrow S_{J_2}\left(\mathbf{x}_k - \alpha \triangledown J_1(\mathbf{x}_k), \alpha\right), \tag{3.23}$$

where $S_{J_2}(\mathbf{x}, \alpha)$ is the solution of a 'soft-threshold' problem at $\mathbf{x}$ with step size α and regularizer $J_2(\mathbf{x})$. Moreover, the soft-threshold operator is given by the solution to the soft-threshold problem

$$S_{J_2}(\mathbf{x}, \alpha) = \operatorname*{argmin}_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 + \alpha J_2(\mathbf{x}). \tag{3.24}$$

In our case, $J_2(\mathbf{x}) = \sum \lambda \|\mathbf{x}\|_1$ so the soft-threshold step for problem

$$\min_{\mathbf{x}} f(\mathbf{x}) = J_1(\mathbf{x}) + \sum \lambda \|\mathbf{x}\|_1 \tag{3.25}$$

would be

$$\operatorname*{argmin}_{\mathbf{x}} \frac{1}{2}\left\|\mathbf{y} - \left(\mathbf{x}^k - \alpha \boldsymbol{\nabla} J_1(\mathbf{x}^k)\right)\right\|_2^2 + \alpha \sum \lambda \|\mathbf{x}\|_1. \tag{3.26}$$

The soft-threshold rule is the nonlinear function defined as

$$(S_\alpha(t)) = \begin{cases} t - \alpha, & t > \alpha \\ 0, & |t| \leq \alpha \\ t + \alpha & t < -\alpha \end{cases} \tag{3.27}$$

or more compactly the shrinkage function defined as [28, 30, 32, 83]

$$(S_\alpha(\boldsymbol{\chi}))_i = \begin{cases} (|\boldsymbol{\chi}_i| - \alpha)\,\mathrm{sign}(\boldsymbol{\chi}_i), & \text{if} \quad |\boldsymbol{\chi}_i| > \alpha \\ 0, & \text{otherwise.} \end{cases} \tag{3.28}$$

It truncates small values to zero and shrinks large values as in Fig. 3.2. If we consider $J_2(\mathbf{x}) = \sum \lambda \|\mathbf{x}\|_0$, $S_\alpha(t)$ is termed hard thresholding function [89]. However, we apply the sparsity constraint through the soft-thresholding function because of the reasons mentioned in the Introduction section recalling that the l_0 norm counts the number of non-zero elements in the solution, which is directly related to the solution's sparseness. However, this makes the minimization problem non-convex such that the problem might contain more than a single infimum [58]. Moreover, a global minimization of the cost function under l_0-norm penalty is NP-hard problem (extensively burden computational cost is required to achieve the solution) [29].

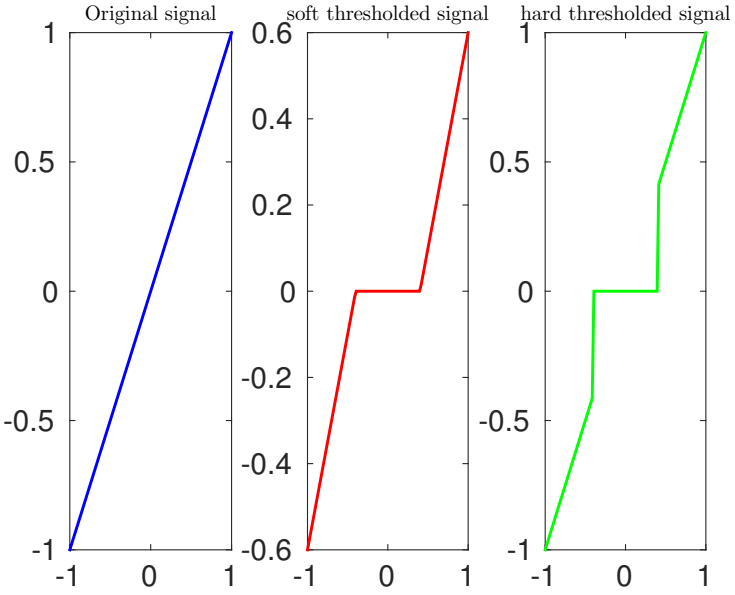


Figure 3.2.: Performance of soft and hard thresholding.

We first take a step along the negative gradient of the function, and then compute this projection-like soft-threshold operator to take into account the effect of the regularizer. The latter step effectively sparsifies the result of the (generally dense) gradient step. As discussed by [59], the soft-threshold operator is a generalization of the projection operator, and we recognize the iterative soft-thresholding algorithm as the classic gradient-projection algorithm but with projection replaced by soft-thresholding. Similar with the classic gradient projection algorithm, this algorithm may converge very slowly. However, analogous to the SPG algorithm which is a non-monotone projected gradient algorithm that combines the classical projected gradient method with the spectral gradient choice of step length and a non=monotone line-search strategy, [89] proposes to use Barzilai-Borwein steps and a non-monotonic line search which we have discussed in previous section

to speed up the convergence of the method.

It is important to underline that the regularization parameter determines the relative weight of the regularization term with respect to the other terms of the cost functional. Therefore, if it is too small the sparsity regularization does not play a significant role in the minimization process, whereas a large value entails that less importance is given to the fitting of the data with respect to the sparsity enforcement. As we are dealing with a nonlinear problem, an optimal choice of this parameter is not straightforward and in any case depends on the available *a-priori* information.

3.2.5. Positivity Constraints

Imposing a priori constraints can improve the quality of solutions to the inverse problems in a great portion [54]. Non-negativity is important in applications like imaging [28, 54]. We know that in order to have a physical solution with a convergence, there should not be a negative conductivity arising after a few iterations. However, the real and imaginary parts of the contrast function are not necessarily non-negative. On the other hand, there are possible ways which enforce the positivity constraints.

Projection Operator: One possible way is to define a closed set such as $C = \{\boldsymbol{\chi} \in \mathrm{Re}^n \,|\boldsymbol{\chi} \geq 0\}$ so that this condition can be imposed to our concerned problem. Thus, we can demonstrate the projection of $\boldsymbol{\chi}$ onto a set C as

$$P_C(\boldsymbol{\chi}) = \arg\min_{v \in C} \|\boldsymbol{\chi} - v\|. \tag{3.29}$$

In other words, $P_C(\boldsymbol{\chi})$ is the closest point to $\boldsymbol{\chi}$ in C [54]. We drop the subscript C from the projection operator for simplicity and define the ith component of $P(\boldsymbol{\chi})$ as

$$[P(\boldsymbol{\chi})_i] = \max(\boldsymbol{\chi}_i, 0) = \begin{cases} \boldsymbol{\chi}_i & \text{if} \quad \boldsymbol{\chi}_i \geq 0 \\ 0 & \text{if} \quad \boldsymbol{\chi}_i < 0. \end{cases} \tag{3.30}$$

Reformulation of Parameters: The other possibility can be to choose two new real-valued unknowns κ and η such that they can keep *a-priori* information. Let $\varepsilon_r(\mathbf{r}) = 1 + \kappa^2$ and $\sigma(\mathbf{r}) = \eta^2$ with $\varepsilon_r(\mathbf{r}) \geq 1$ and $\sigma(\mathbf{r}) \geq 0$. Thus, we can write the gradient of contrast

function with respect to these unknowns such as

$$\frac{\mathrm{d}\boldsymbol{\chi}(\mathbf{r})}{\mathrm{d}\kappa(\mathbf{r})} = 2k_0^2\kappa(\mathbf{r}), \qquad \text{and} \qquad \frac{\mathrm{d}\boldsymbol{\chi}(\mathbf{r})}{\mathrm{d}\eta^2(\mathbf{r})} = \frac{2k_0^2\eta(\mathbf{r})}{\omega\varepsilon_0(\mathbf{r})}. \tag{3.31}$$

Hence, by applying a chain rule, the gradient of the cost functional is obtained as

$$\frac{\mathrm{d}K(\boldsymbol{\chi}(\eta,\kappa))}{\mathrm{d}\kappa(\mathbf{r})} = \frac{\mathrm{d}K(\boldsymbol{\chi}(\eta,\kappa))}{\mathrm{d}\boldsymbol{\chi}(\eta,\kappa)} \times \frac{\mathrm{d}\boldsymbol{\chi}(\eta,\kappa)}{\mathrm{d}\kappa(\mathbf{r})}. \tag{3.32}$$

In our simulations we have applied positivity constraint by projection rather than the reformulation of parameters. The reason for this is that positivity constraint by reformulation of parameters has a computationally burden and it slows down the convergence of cost function.

3.2.6. Stopping Criteria

The algorithm is terminated if the following holds

$$\max_{k-\varsigma+1\leqslant n\leqslant k} \frac{\|K(\boldsymbol{\chi}^{n+1}) - K(\boldsymbol{\chi}^{n})\|}{\|K(\boldsymbol{\chi}^{n+1})\|} < s. \tag{3.33}$$

This means the algorithm is terminated if the last ς steps have sufficiently small changes.

Another possible stopping criterion is to check τ when falling below a small positive constant. It can be interpreted as when maximum absolute value of step size times search direction goes below a small positive constant.

One other termination criteria can be the number of iterations. Having enough number of iterations can lead us to have a convergence in the cost function. In our simulations we have used 100 iterations unless it is stated otherwise.

The algorithm which has been put forth, is as follows:

Algorithm 2 Steepest descent reconstruction algorithm (+ sparsity constraint)

1: Initialization $\boldsymbol{\chi}^1$ and α
2: **for** $k = 1, \cdots, T$ **do**
3: Solve the direct problem $E^{\text{diff}}\left(\boldsymbol{\chi}^k\right)$
4: Compute the gradient $K'\left(\boldsymbol{\chi}^k\right) = \nabla_{\boldsymbol{\chi}} K\left(\boldsymbol{\chi}\right)|_{\boldsymbol{\chi}^k}$ ▷ Adjoint method
5: Smooth the gradient $K'_s\left(\boldsymbol{\chi}^k\right)$ ▷ by solving (3.9)
6: Determine the step size τ_j
7: Update inhomogeneity by $\boldsymbol{\chi}^{k+1} = \boldsymbol{\chi}^k - \tau_k K'_s\left(\boldsymbol{\chi}^k\right)$
8: Threshold $\boldsymbol{\chi}^{k+1}$ by $S_\alpha\left(\boldsymbol{\chi}^{k+1}\right)$ (3.4) ▷ Sparsity constraint
9: Imposing prior constraint via projection ▷ Positivity constraint
10: check stopping criterion.
11: **end for**
12: **output** approximate the minimizer of (3.1)

We have used this algorithm throughout our simulations with addition of projection constraint when we call projection and the algorithm without projection constraint when we call pixel basis approach.

3.3. Soft Shrinkage Method in Wavelet Transform

The intrinsic multiresolution feature of the wavelet transform [25] allows an accurate representation of the unknown function with a reduced number of coefficients. In particular, the wavelet transform decomposes a given profile into two sets of coefficients: coarse and detail. The coarse coefficients account for the profile's low frequency content acting as a low-pass filter of the original function, while detail coefficients account for high frequency content and allow representing the finer details of the function [25]. As the decomposition level increases, the number of nonzero coefficients is reduced. These properties motivate an approach that starts from a high level of decomposition, which is gradually reduced to retrieve finer details in the image. In practice, this is achieved by starting from a high order coarse representation and progressively moving to lower order coarse images to improve resolution.

In this section we will give a brief flavor about the wavelet transform and reformulate the forward and the inverse problem of interest in terms of wavelet decomposition.

3.3.1. Review of wavelet representation and the wavelet transform

Any continuous function in $\mathbb{R}^2$ can be approximated in terms of scaling and wavelet functions up to some decomposition level J according to the following formula [41]

$$y(\mathbf{r}) = \sum_{k,l} a_{J_0,k,l}\Phi_{J_0,k,l}(\mathbf{r}) + \sum_{s=1}^{3}\sum_{j=J_0}^{J}\sum_{k,l=0}^{2^J-1} w^s_{j,k,l}\Psi^s_{j,k,l}(\mathbf{r}), \tag{3.34}$$

where $\mathbf{r} = (x, y)$ represents the spatial coordinates in the Cartesian system and J_0 is the initial decomposition level. Scaling functions $\Phi_{J_0,k,l}$ and wavelet function $\Psi^s_{j,k,l}$ are defined as tensor products of the scaled and translated scaling and wavelet functions in $\mathbb{R}$ such as

$$\Phi_{j,k,l}(\mathbf{r}) = \phi_{j,k}(x)\phi_{j,l}(y), \tag{3.35}$$
$$\Psi^V_{j,k,l}(\mathbf{r}) = \phi_{j,k}(x)\Psi_{j,l}(y), \tag{3.36}$$
$$\Psi^H_{j,k,l}(\mathbf{r}) = \Psi_{j,k}(x)\phi_{j,l}(y), \tag{3.37}$$
$$\Psi^D_{j,k,l}(\mathbf{r}) = \Psi_{j,k}(x)\Psi_{j,l}(y), \tag{3.38}$$

where $\Psi_{j,k}(x)$ is derived from the mother wavelet $\Psi(x)$ as follows:

$$\Psi_{j,k}(x) = 2^{-j/2}\Psi(2^j x - k), \tag{3.39}$$

and $\phi_{j,k}(x)$ is the corresponding scaling function. This means that there is a unique scale function to compute the low frequency components in the previous decomposition level and three wavelet functions to compute the detail coefficients along horizontal, vertical and diagonal directions (H,V,D).

The coefficients $a_{J_0,k,l}$ and $w^s_{j,k,l}$ are obtained by inner products in $\mathbb{R}^2$,

$$a_{j,k,l} = \langle y, \Phi_{j,k,l}\rangle \quad w_{j,k,l} = \langle y, \Psi_{j,k,l}\rangle. \tag{3.40}$$

For the sake of simplicity, we rewrite the equation (3.34) as

$$y(\mathbf{r}) = \sum_j c_j C_j(\mathbf{r}), \tag{3.41}$$

where c_j and $C_j(\mathbf{r})$ are wavelet coefficients and scaling basis functions for $j = 1, \cdots, 2^{2J_0}$

and for $j = 2^{2J_0} + 1, \cdots, 2^{2J}$. We denote the calculations in (3.40) in an operator form following as

$$c = \mathcal{W}y, \tag{3.42}$$

where $\mathcal{W}$ is a wavelet transform operator which maps y from $\mathbb{R}$ to a sparse set of coefficients, c [25]. In our application, we are dealing with complex-valued functions. For instance, by using the linear property of wavelet transforms we obtain

$$c = \mathcal{W}y^{\mathrm{re}} + i\mathcal{W}y^{\mathrm{im}} = c^{\mathrm{re}} + ic^{\mathrm{im}}. \tag{3.43}$$

As in the case of 1D wavelet transform (see appendix B), the 2D wavelet could be implemented using Low pass L and High pass H filters, respectively, as shown in Fig. 3.3. By this way, the two-dimensional signal, generally images, is divided into four subbands

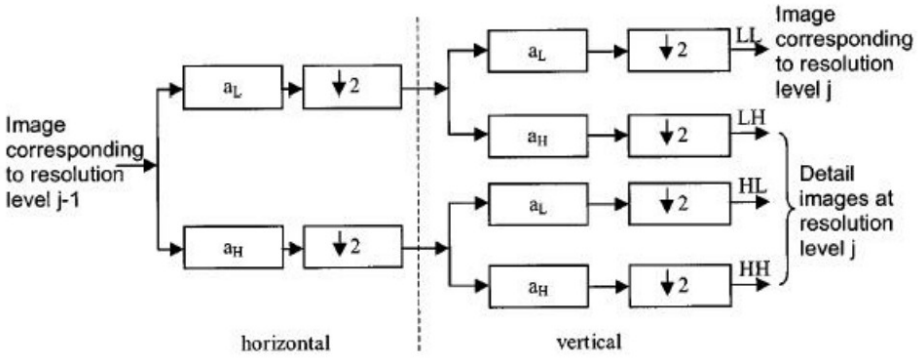


Figure 3.3.: Filter bank of the 2D wavelet transform [48].

LL, HL, LH and HH. The band HL indicates the variation of the original signal along the horizontal direction x and the subband LH for the vertical variation y.

3.3.2. Forward Problem in the Wavelet Domain

First, we explore the applicability of the wavelet representation for the contrast sources by solving the forward scattering problem. We start by rewriting the domain integral equation which is in the form of contrast source such as

$$J^{\mathrm{inc}}(\mathbf{r}, \mathbf{r}_s) = \int_\Omega \underbrace{[I - \chi(\mathbf{r})G(\mathbf{r}, \mathbf{r}')]}_{A(\mathbf{r},\mathbf{r}')} J(\mathbf{r}', \mathbf{r}_s)d\mathbf{r}'. \tag{3.44}$$

Using the same basis decomposition as in (2.21) along $\mathbf{r}'$ we can write

$$J^{\text{inc}}(\mathbf{r}, \mathbf{r}_s) = \sum_{i=1}^{\infty} \mathcal{A}_i(\mathbf{r}, \Omega)\mathcal{J}_i(\mathbf{r}_s) \tag{3.45}$$

with $\mathcal{A}_i(\mathbf{r}, \Omega) = \int_\Omega A(\mathbf{r}, \mathbf{r}')\psi_i(\mathbf{r}')d\mathbf{r}'$, $\psi_i(\mathbf{r}')$ being wavelet basis and $\mathbf{r}_s$ represents the position of sources. Applying basis decomposition along $\mathbf{r}$ on (3.45) leads to

$$\mathcal{J}_j^{\text{inc}}(\Omega, \mathbf{r}_s) = \sum_{i=1}^{\infty} \mathcal{A}_{ij}(\Omega, \Omega)\mathcal{J}_i(\mathbf{r}_s), \quad j = 1, \cdots, \infty \tag{3.46}$$

with $\mathcal{A}_{ij}(\Omega, \Omega) = \int_\Omega \mathcal{A}_i(\mathbf{r}, \Omega)\psi_i(\mathbf{r})d\mathbf{r}$. Moreover, N pixels leads to N basis functions so that

$$J^{\text{inc}}(\mathbf{r}, \mathbf{r}_s) = \sum_{i=1}^{N} \mathcal{A}_i(\mathbf{r}, \Omega)\mathcal{J}_i(\mathbf{r}_s. \tag{3.47}$$

Applying the basis decomposition along $\mathbf{r}$ on (3.47) leads to

$$\mathcal{J}_j^{\text{inc}}(\Omega, \mathbf{r}_s) = \sum_{i=1}^{N} \mathcal{A}_{ij}(\Omega, \Omega)\mathcal{J}_i(\mathbf{r}_s), j = 1, \cdots, N \tag{3.48}$$

with $\mathcal{A}_{ji}(\Omega, \Omega) = \int_\Omega \mathcal{A}_i(\mathbf{r}, \Omega)\psi_i(\mathbf{r})d\mathbf{r}$. Solving the state equation in the wavelet domain can be obtained as

$$\begin{pmatrix} \mathcal{J}_1(\mathbf{r}_s) \\ \mathcal{J}_2(\mathbf{r}_s) \\ \vdots \\ \mathcal{J}_{j=N}(\mathbf{r}_s) \end{pmatrix} = \begin{pmatrix} \mathcal{A}_{1,1}(\Omega, \Omega) & A_{1,2}(\Omega, \Omega) & \cdots & A_{1,N}(\Omega, \Omega) \\ \mathcal{A}_{2,1}(\Omega, \Omega) & \mathcal{A}_{2,2}(\Omega, \Omega) & \cdots & \mathcal{A}_{2,N}(\Omega, \Omega) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{A}_{N,1}(\Omega, \Omega) & \mathcal{A}_{N,2}(\Omega, \Omega) & \cdots & \mathcal{A}_{N,N}(\Omega, \Omega) \end{pmatrix}^{-1} \begin{pmatrix} \mathcal{J}_1^{\text{inc}}(\Omega, \mathbf{r}_s) \\ \mathcal{J}_2^{\text{inc}}(\Omega, \mathbf{r}_s) \\ \vdots \\ \mathcal{J}_N^{\text{inc}}(\Omega, \mathbf{r}_s) \end{pmatrix} \tag{3.49}$$

Combining (3.48) and (3.49) shows that the direct problem can be worked in the wavelet basis domain.

3.3.3. Inverse Algorithm in Wavelet Domain

Since we would like to represent the contrast function in the wavelet domain, we first write the data equation as follows:

$$E^{\text{diff}}(\mathbf{r}_r, \mathbf{r}_s) = \int_\Omega G(\mathbf{r}_r, \mathbf{r}')\mathcal{W}^*(\mathcal{W}(\boldsymbol{\chi}(\mathbf{r}')E(\mathbf{r}', \mathbf{r}_s)))d\mathbf{r}', \tag{3.50}$$

where $\mathcal{W}^*$ is the inverse wavelet transform operator which maps $\boldsymbol{\chi}$ back to a spatial domain. Similarly, the domain equation is written in terms of the contrast sources in the wavelet domain

$$\mathcal{W}\left((\chi(\mathbf{r}))E(\mathbf{r}, \mathbf{r}_s)\right) = \mathcal{W}\left(\chi(\mathbf{r})E^{\text{inc}}(\mathbf{r})\right) - \mathcal{W}\left(\chi(\mathbf{r})\int_D G(\mathbf{r}, \mathbf{r}')\mathcal{W}^*\left(\mathcal{W}\left(\chi(\mathbf{r}')E(\mathbf{r}', \mathbf{r}_s)\right)\right)\right). \tag{3.51}$$

By using the above equations, we solve the nonlinear problem which minimizes the following cost functional:

$$F(\boldsymbol{\chi}) = \underbrace{\|\zeta(\boldsymbol{\chi}) - G(\mathbf{r}_r, \mathbf{r}')\mathcal{W}^*\left(W(\boldsymbol{\chi}(\mathbf{r}')E(\mathbf{r}', \mathbf{r}_s)\right)\|_2}_{K(\boldsymbol{\chi})} + \|\mathcal{W}(\boldsymbol{\chi})\|_1. \tag{3.52}$$

We update the unknown wavelet domain contrast function with the formula

$$\mathcal{W}(\boldsymbol{\chi}_{k+1}) = \mathcal{W}(\boldsymbol{\chi}_k) - \tau_k \mathcal{W}(K'\left(\boldsymbol{\chi}_k\right)), \tag{3.53}$$

where

$$\tau_k = \frac{\left\langle \mathcal{W}(\boldsymbol{\chi}^k) - \mathcal{W}(\boldsymbol{\chi}^{k-1}), \mathcal{W}(K'\left(\boldsymbol{\chi}^k\right)) - \mathcal{W}(K'\left(\boldsymbol{\chi}^{k-1}\right))\right\rangle}{\left\langle \mathcal{W}(K'\left(\boldsymbol{\chi}^k\right)) - \mathcal{W}(K'\left(\boldsymbol{\chi}^{k-1}\right)), \mathcal{W}(K'\left(\boldsymbol{\chi}^k\right)) - \mathcal{W}(K'\left(\boldsymbol{\chi}^{k-1}\right))\right\rangle}. \tag{3.54}$$

In other words, to retrieve the unknown contrast function by solving (2.9), a wavelet expansion of the problem unknown in discretized form is considered as

$$E^{\text{diff}}(\mathbf{r}_r, \mathbf{r}_s) = \sum_{i=1}^{N} c_i \int_D G_{\text{or}}(\mathbf{r}_r, \mathbf{r}')C_i(\mathbf{r}')E(\mathbf{r}', \mathbf{r}_s)d\mathbf{r}' \tag{3.55}$$

The algorithm used to solve inverse scattering problem is stated below:

Algorithm 3 Steepest descent reconstruction algorithm with sparsity constraint in wavelet domain

Initialize $\boldsymbol{\chi}^1$ and α
for $k = 1, \cdots, T$ **do**
  Solve the direct problem $E^{\text{diff}}\left(\boldsymbol{\chi}^k\right)$
  Compute the gradient $\mathcal{W}(K'\left(\boldsymbol{\chi}^k\right)) = \nabla_{\boldsymbol{\chi}} K\left(\boldsymbol{\chi}\right)|_{\boldsymbol{\chi}^k}$
  Determine the step size τ_k
  Update inhomogeneity by $\mathcal{W}(\boldsymbol{\chi}_{k+1}) = \mathcal{W}(\boldsymbol{\chi}^k) - \tau_k \mathcal{W}(K'\left(\boldsymbol{\chi}^k\right))$
  Threshold $\mathcal{W}(\boldsymbol{\chi}_{k+1})$ by $S_\alpha\left(\mathcal{W}(\boldsymbol{\chi}^{k+1})\right)$
  Back to spatial domain $\mathcal{W}^*(\boldsymbol{\chi}_{k+1})$
  check stopping criterion
end for

3.4. Simulations

In this chapter, we present a number of numerical results to justify the effectiveness and accuracy of the proposed methods. We have tried different scatterers having different permittivities and conductivities in order to identify and compare the performance and efficiency of each method with different constraints (i.e, addition of smoothness, projection, etc.). The description of physical characteristics scatterers and simulation configuration for each scatterer has been given in Tabs. 3.1 and 3.2 and in Fig. 3.4. The relative error norms as a comparison criterion are expressed as

$$\varepsilon_r^{\text{err}} = \frac{\|\varepsilon_r^{\text{rec}} - \varepsilon_r\|_2}{\|\varepsilon_r\|_2}, \quad \sigma^{\text{err}} = \frac{\|\sigma^{\text{rec}} - \sigma\|_2}{\|\sigma\|_2}, \tag{3.56}$$

where $\varepsilon_r^{\text{rec}}$ and σ^{rec} are the reconstructed permittivity and conductivity, respectively, and ε_r and σ the exact ones. The minimum of the cost function is the one reached at the end of the process for each regularization parameter α. Additionally, the relative error norm on $\boldsymbol{\chi}$ is defined as

$$\boldsymbol{\chi}^{\text{err}} = \frac{\text{Tr}\left(\|\text{diag}(\boldsymbol{\chi}) - \text{diag}(\boldsymbol{\chi}^{\text{true}})\|_2\right)}{\text{Tr}\left(\|\text{diag}(\boldsymbol{\chi}^{\text{true}})\|_2\right)}, \tag{3.57}$$

where $\boldsymbol{\chi}^{\text{true}}$ is the true contrast to be found and $\text{Tr}\left(\mathbf{A}\right)$ stands for the trace of the matrix $\mathbf{A}$. The discrepancy between the model and the measurement is as

$$\mathbf{E}^{\text{diff}} = \frac{\sum_{i=1}^{N_i} \left\|\boldsymbol{\zeta}_i - \mathbf{E}_i^{\text{diff}}(\boldsymbol{\chi})\right\|_2}{\sum_{i=1}^{N_i} \|\boldsymbol{\zeta}_i\|_2} \tag{3.58}$$

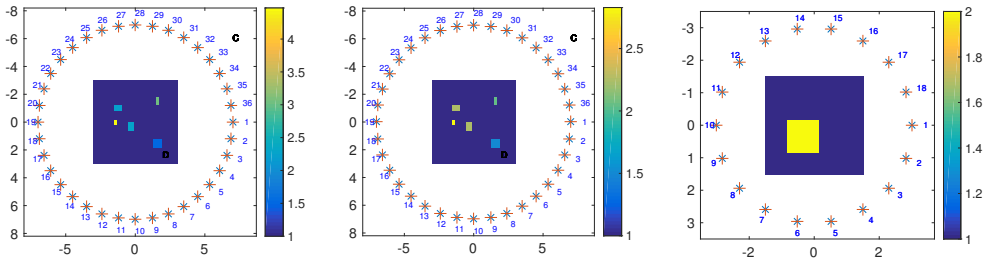
Scatterer	#	x	y	L_x	L_y	ε_r	σ
3.4a	1	-1.5	−1.5	0.5	0.5	1.5	0.002 20
	2	-0.33	0.44	0.5	0.5	2.25	0
	3	1.5	−1.5	0.5	0.33	3	0.004 45
	4	0	1.5	0.5	0.33	4	0.022 25
	5	1	1.33	0.5	0.5	1	0.022 25
3.4b,3.4g 3.4j,3.4k	1	-1.5	−1.5	0.5	0.5	1.5	0.002 20
	2	-0.33	0.44	0.5	0.5	2.25	0
	3	1.5	−1.5	0.5	0.33	2	0.004 45
	4	0	1.5	0.5	0.33	2	0.022 25
	5	1	1.33	0.5	0.5	1	0.022 25
3.4c,3.4d,3.4f 3.4h,3.4l	1	-0.3	−0.3	1	1	2	0
3.4e	1	0	0	1	1	2	0
	2	-0.68	−0.68	0.33	0.33	1.5	0
3.4i	1	0.5	0.5	1	1	2	0.008 30
	2	-0.5	−0.5	1	1	2	0.004 45

Table 3.1.: Description of the scatterers, x, y being the coordinate of the center of the obstacle (in m), L_x, L_y its lengths (in m) and ε_r and σ its relative permittivity and conductivity (the latter in $\mathrm{S\,m^{-1}}$)

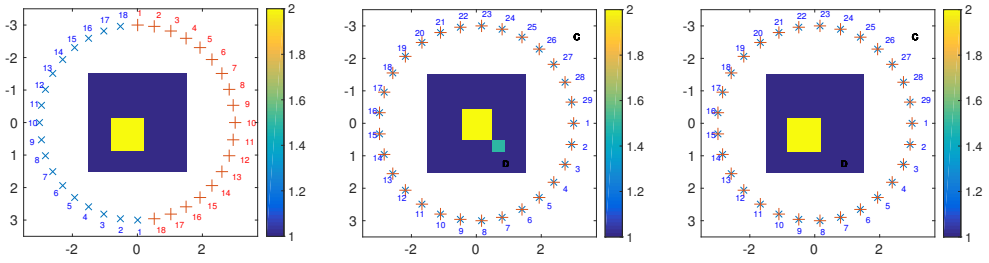
where $\mathbf{E}^{\mathrm{diff}}(\boldsymbol{\chi})$ is computed by solving the direct problem with the reconstructed $\boldsymbol{\chi}$. The laptop that has been used in order to run the simulations has a processor such as: Intel Core i7−4600U CPU@2.10 GHz×4

3.4.1. Choice of Regularization Parameters

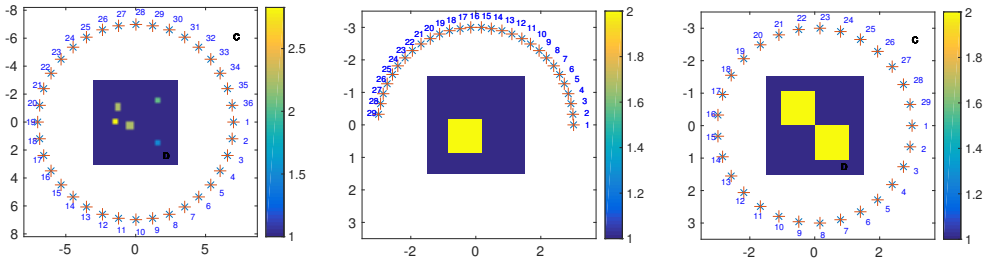
One of the key points of such an inversion either in spatial domain (pixel basis decomposition) or wavelet domain is the choice of the regularization parameters α in (3.4) and β in (3.9) (δ being kept constant and equal to 1). Different tests have been performed in order to evaluate the sensitivity of the choice of those parameters on the solution in spatial domain. The weighting parameters α and β should be chosen suitably so that we can drive the regularization properties of our algorithm in an efficient and predictable way. In order to achieve a reconstruction which is close to the real case we make modifications that affect the parameters.



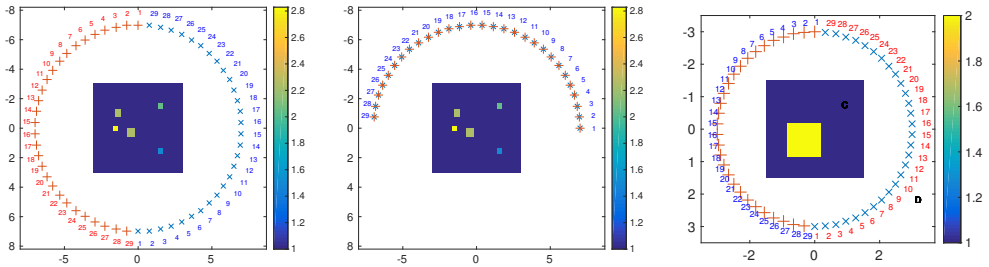
(a) Scatterer 1 (b) Scatterer 2 (c) Scatterer 3



(d) Scatterer 4 (e) Scatterer 5 (f) Scatterer 6



(g) Scatterer 7 (h) Scatterer 8 (i) Scatterer 9



(j) Scatterer 10 (k) Scatterer 11 (l) Scatterer 12

Figure 3.4.: Measured configuration of actual permittivity profiles for different scatterers and source-receiver locations.

Scatterer	ROI	frequency (MHz)	Discretization	Nsources	Nreceivers
3.4a	$6 \times \lambda$	200MHz	Forward P.= 80×80 Inverse P.= 30×30	36 r= $7m$ $\theta = [0, 2\pi]$	36 r= $7m$ $\theta = [0, 2\pi]$
3.4b	$6 \times \lambda$	300MHz	Forward P.= 80×80 Inverse P.= 30×30	36 r= $7m$ $\theta = [0, 2\pi]$	36 r= $7m$ $\theta = [0, 2\pi]$
3.4c	$3 \times \lambda$	300MHz	Forward P.= 36×36 Inverse P.= 18×18	36 r= $7m$ $\theta = [0, 2\pi]$	36 r= $7m$ $\theta = [0, 2\pi]$
3.4d	$3 \times \lambda$	300MHz	Forward P.= 36×36 Inverse P.= 18×18	36 r= $7m$ $\theta = [\pi/2, -\pi/2]$	36 r= $7m$ $\theta = [3\pi/2, \pi/2]$
3.4e,3.4f 3.4g,3.4i	$3 \times \lambda$	300MHz	Forward P.= 64×64 Inverse P.= 32×32	29 r= $7m$ $\theta = [0, 2\pi]$	29 r= $7m$ $\theta = [0, 2\pi]$
3.4h,3.4k	$3 \times \lambda$	300MHz	Forward P.= 64×64 Inverse P.= 32×32	29 r= $7m$ $\theta = [0, -\pi]$	29 r= $7m$ $\theta = [0, -\pi]$
3.4j,3.4l	$3 \times \lambda$	300MHz	Forward P.= 64×64 Inverse P.= 32×32	29 r= $7m$ $\theta = [\pi/2, -\pi/2]$	29 r= $7m$ $\theta = [3\pi/2, \pi/2]$
3.4g,	$3 \times \lambda$	300MHz	Forward P.= 64×64 Inverse P.= 32×32	36 r= $7m$ $\theta = [0, -\pi]$	36 r= $7m$ $\theta = [0, -\pi]$

Table 3.2.: Properties of the geometry of each system: Description of the region of interest (ROI), frequency, discretization, and number of sources (Nsources) and receivers (Nreceivers).

From a practical point of view, supported by an extensive numerical analysis, it has been observed that a convenient choice for α can be $1/\#ofpixels$. This choice is suggested by the fact that the penalty functions acts on the amplitude of the unknowns and it is so normalized by their number. On the other hand, this value is not the optimal one in our case. We have followed trial and error approach to be able to choose the most appropriate regularization parameter α in this application.

At first we consider the reconstruction as a function of α with a constant β as in Fig. 3.5 for the scatterer in Fig. 3.4f. Secondly, we examine the same reconstruction with a constant α as a function of β as in Fig. 3.6.

In the first case we observe a better reconstruction of the exact case by increasing the parameter α such as in Figs. 3.8e and 3.8g or in Figs. 3.9e and 3.9g. The higher the value we choose for α the sharper the reconstructions are. However, the choice of α is not arbitrary as can be seen in Fig. 3.5. The error in permittivity gets larger for larger α. We get a minimum error when α is 2.5×10^{-2}.

In the other case, we would expect the inclusion to get smoother when we increase β. However, in Fig. 3.6 it can be observed that there are no crucial differences when β is 1, 1.5 and 2.5 differently than 2. Fig. 3.7 shows that even if we do not have smoothness, small error can be obtained by a properly chosen regularization parameter. However, if we cannot choose the appropriate regularization parameter, addition of smoothness can help us to improve the quality of reconstruction. Nevertheless, we should choose β carefully and close to 1 in order not to loose the differentiability for the next iteration. Using β as 1.5 shows that the reconstruction becomes smoother as can be seen in Figs. 3.8i, Figs. 3.8j, 3.9i, and 3.9j compared to the figures without any smoothness.

3.4.2. Addition of Projection

Addition of projection to the sparsity reconstruction method improves the quality of the reconstruction even more and the error range for both permittivity and conductivity decreases until a value of α such as 1.5×10^{-2} for permittivity as in Figs. 3.10b and 3.11b in spatial domain. In here, different randomization of 10 dB noise levels have been tested to see how the addition of noise data can be discretized. Figs. 3.10b, 3.10d, 3.11b, and 3.11d show us that we can choose the regularization parameter in a wider range when we add projection to our algorithm. On the other hand, we cannot base our solutions on the minimum cost function as it does not give us the best solution even if we add projection as in Fig. 3.14. This shows us that our problem is ill-posed and we need *a-priori* information.

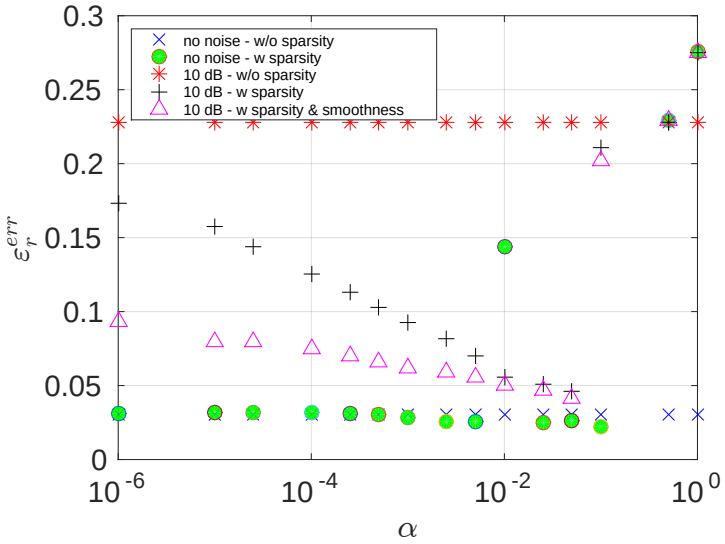


Figure 3.5.: Error of permittivity w.r.t different values of α when β is equal to 0 for the cases with (w) sparsity, without (w/o) sparsity and equal to 1.5 for sparsity and smoothness (smoothing the gradient (Sec. 3.2.2))

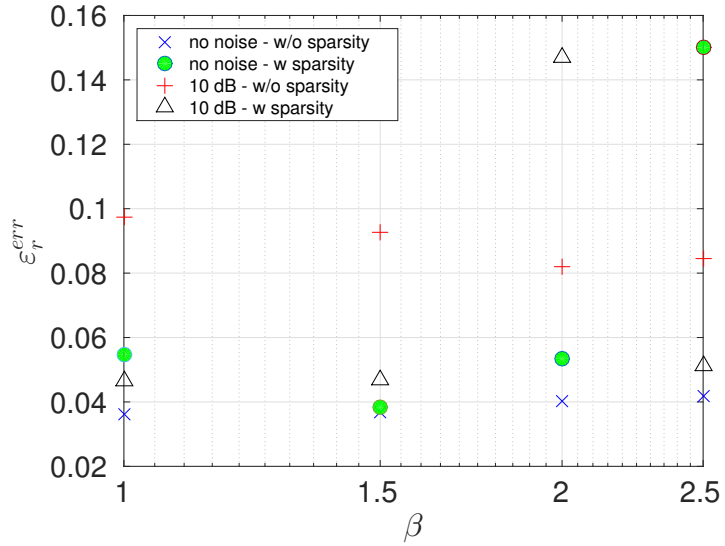


Figure 3.6.: Error of permittivity w.r.t different values of β when α is equal to 2.5×10^{-2} with (w) sparsity and α is equal to 1×10^{-6} without (w/o) sparsity.

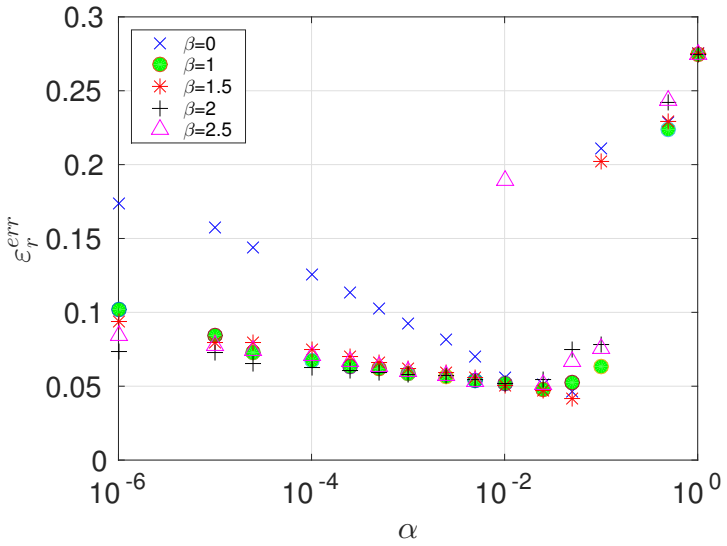
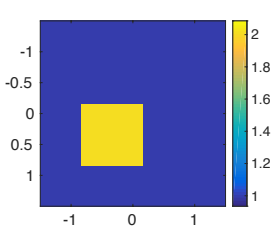


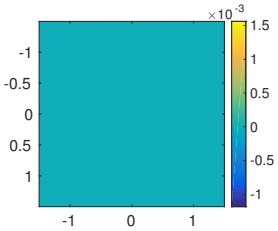
Figure 3.7.: Error of permittivity w.r.t different values of β and α with 10 dB noise data.

Figs. 3.12 and 3.13 present a comparison of the maps of the permittivity and conductivity, respectively for various inversion parameters where Tab.3.3 exhibits the error in permittivity in each case. The influence of the α parameter without any projection constraint onto the permittivity and conductivity is important as we get better reconstruction when α is 5×10^{-3} than when α is 1×10^{-6}.

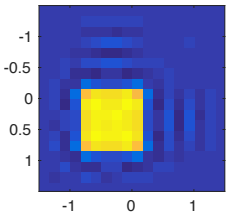
On the other hand, a clearer image is obtained by addition of projection and choice of α has wider range. Comparing Figs. 3.12 and 3.13 shows us that when the permittivity is low, the reconstruction quality is better. Moreover, Tab.3.3 highlights that we can get the minimum error no matter the regularization parameter is, if we have addition of projection. Even if we have less sparsity enforcement (smaller regularization parameter), addition of projection to our inverse problem yields a better quality of reconstruction.



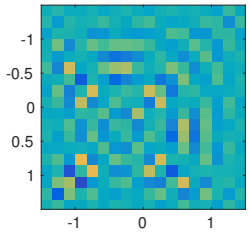
(a) Exact



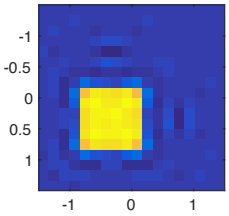
(b) Exact



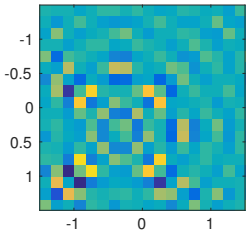
(c) Without sparsity



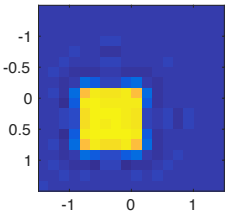
(d) Without sparsity



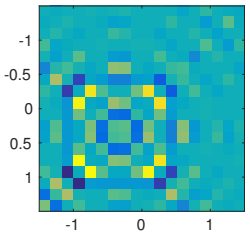
(e) $\alpha = 1 \times 10^{-3}$



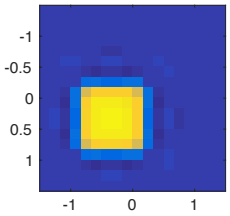
(f) $\alpha = 1 \times 10^{-3}$



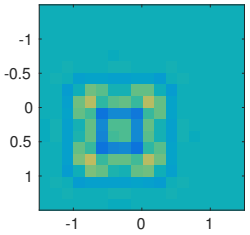
(g) $\alpha = 2.5 \times 10^{-2}$



(h) $\alpha = 2.5 \times 10^{-2}$

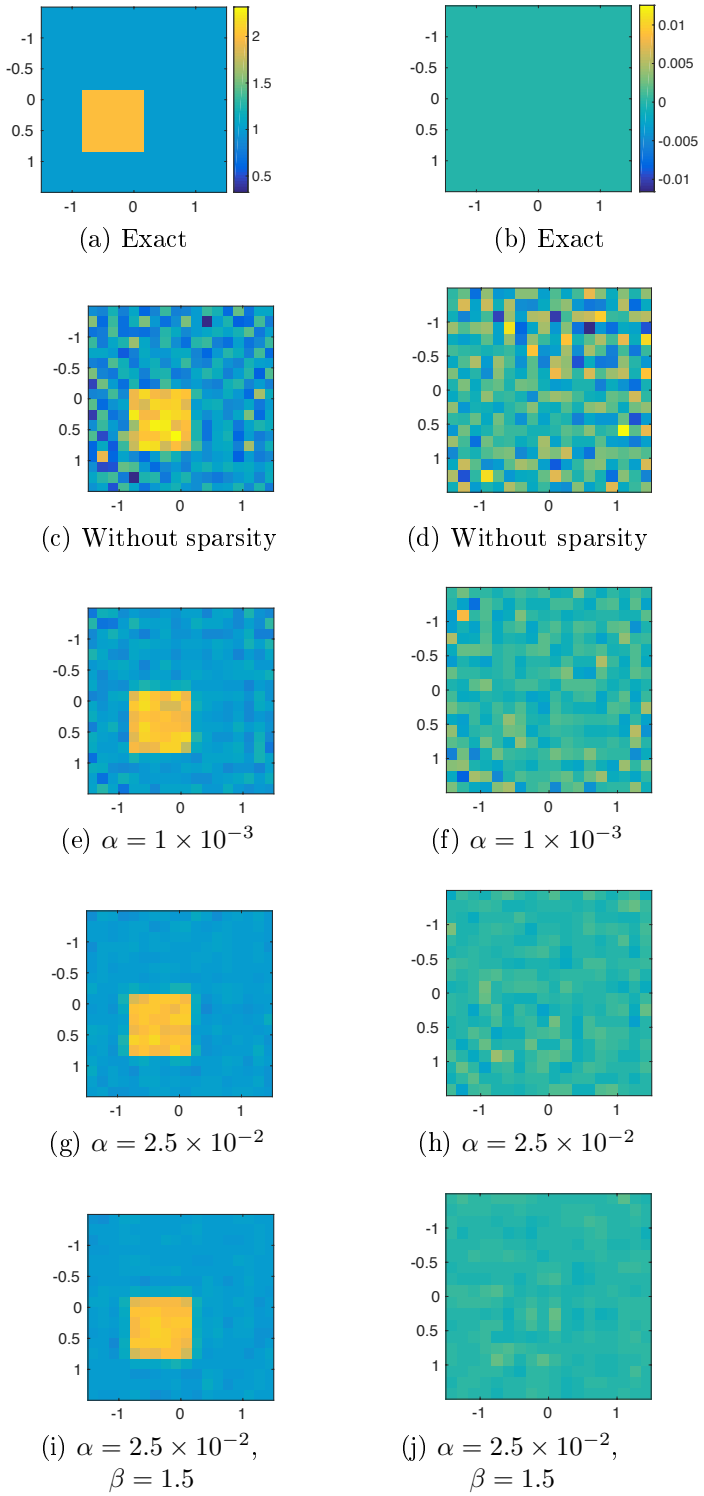


(i) $\alpha = 2.5 \times 10^{-2}$, $\beta = 1.5$



(j) $\alpha = 2.5 \times 10^{-2}$, $\beta = 1.5$

Figure 3.8.: Scatterer 6 (Fig. 3.4f): Retrieval of permittivity ε_r (left) and conductivity σ (right) using sparsity and smoothness without noise.



(a) Exact (b) Exact

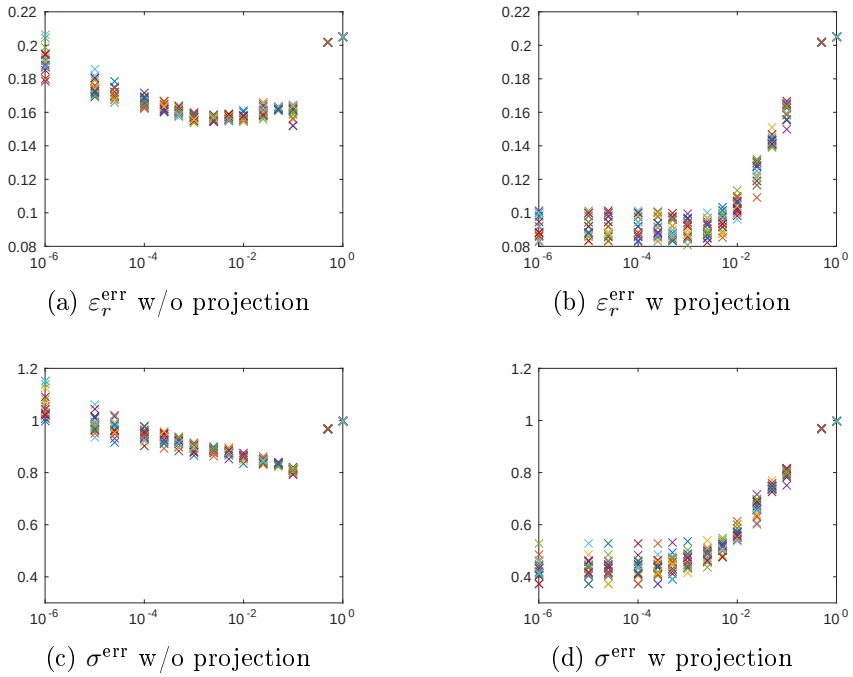
(c) Without sparsity (d) Without sparsity

(e) $\alpha = 1 \times 10^{-3}$ (f) $\alpha = 1 \times 10^{-3}$

(g) $\alpha = 2.5 \times 10^{-2}$ (h) $\alpha = 2.5 \times 10^{-2}$

(i) $\alpha = 2.5 \times 10^{-2}$, $\beta = 1.5$ (j) $\alpha = 2.5 \times 10^{-2}$, $\beta = 1.5$

Figure 3.9.: Scatterer 6 (Fig. 3.4f): Retrieval of permittivity ε_r (left) and conductivity σ (right) using sparsity and smoothness with 10 dB noise data.



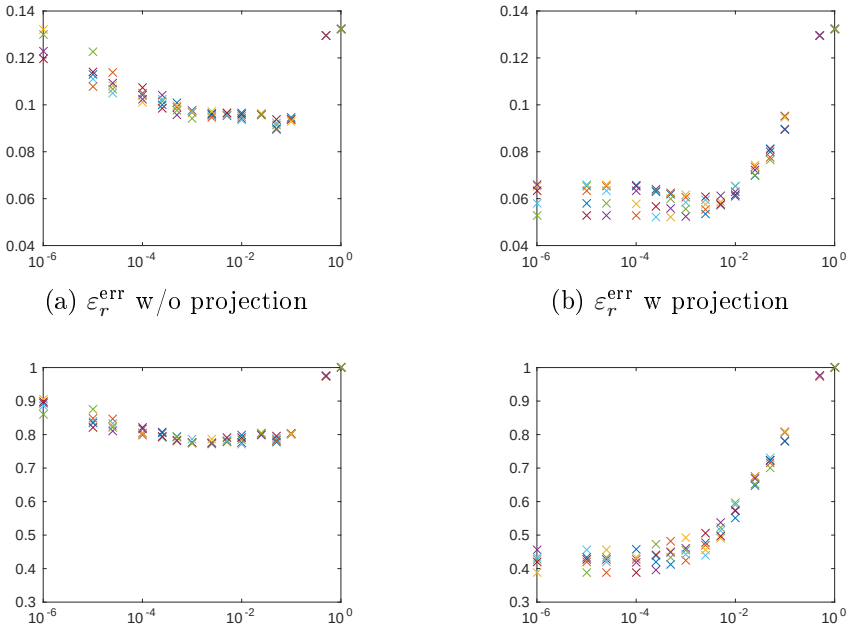
(a) $\varepsilon_r^{\text{err}}$ w/o projection (b) $\varepsilon_r^{\text{err}}$ w projection

(c) σ^{err} w/o projection (d) σ^{err} w projection

Figure 3.10.: Scatterer 1 (Fig. 3.4a): Error in ε_r and error in σ as a function of α by using sparsity with (w) and without (w/o) projection with 20 noise levels of 10 dB (vertical axis corresponds to an error and horizontal axis corresponds to a regularization parameter α).

Scatterer	Scenario	α	$\varepsilon_r^{\text{err}}$	σ^{err}
Fig. 3.4a	Without projection	$\alpha = 1 \times 10^{-5}$	0.1714	0.9072
	Without projection	$\alpha = 5 \times 10^{-3}$	0.1546	0.8673
	With projection	$\alpha = 1 \times 10^{-6}$	0.0976	0.4493
Fig. 3.4b	Without projection	$\alpha = 1 \times 10^{-5}$	0.21226	0.8760
	Without projection	$\alpha = 5 \times 10^{-3}$	0.0952	0.7767
	With projection	$\alpha = 1 \times 10^{-6}$	0.0654	0.3879
Fig. 3.4c	Without projection	$\alpha = 1 \times 10^{-5}$	0.1460	
	Without projection	$\alpha = 2.5 \times 10^{-2}$	0.0690	
	With projection	$\alpha = 2.5 \times 10^{-2}$	0.0494	
Fig. 3.4d	Without projection	$\alpha = 1 \times 10^{-5}$	0.2719	
	Without projection	$\alpha = 2.5 \times 10^{-2}$	0.1079	
	With projection	$\alpha = 2.5 \times 10^{-2}$	0.0947	

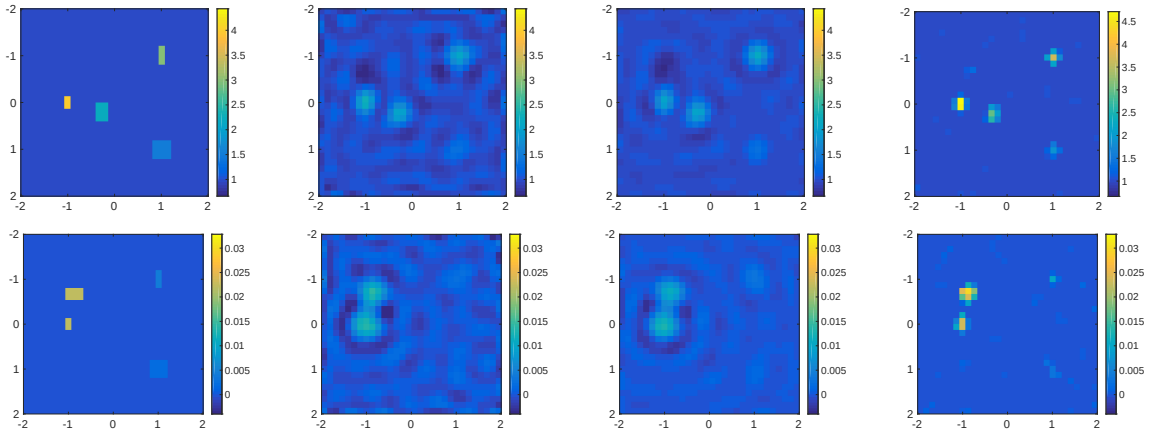
Table 3.3.: Error on permittivity with or without addition of projection



(a) $\varepsilon_r^{\text{err}}$ w/o projection (b) $\varepsilon_r^{\text{err}}$ w projection

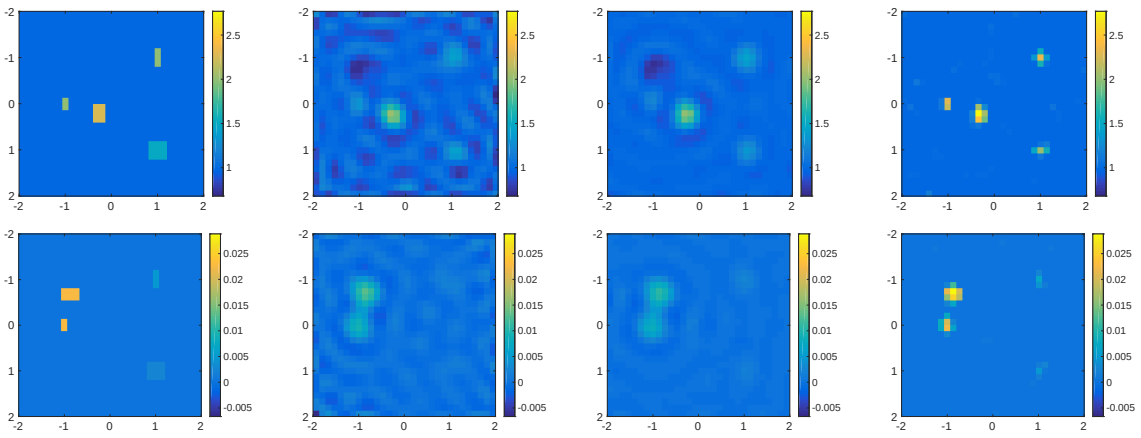
(c) σ^{err} w/o projection (d) σ^{err} w projection

Figure 3.11.: Scatterer 2 (Fig. 3.4b): Error in ε_r and error in σ as a function of α by using sparsity with (w) and without (w/o) projection with 5 noise levels of 10 dB (vertical axis corresponds to an error and horizontal axis corresponds to a regularization parameter α).



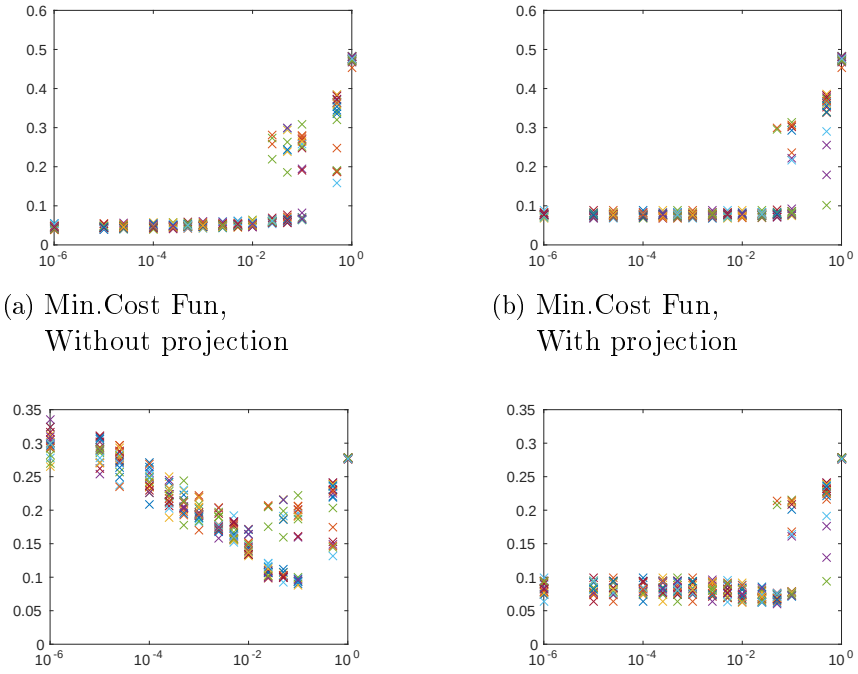
Exact $\alpha = 1 \times 10^{-5}$ w/o projection $\alpha = 5 \times 10^{-3}$ w/o projection $\alpha = 1 \times 10^{-6}$ w projection

Figure 3.12.: Scatterer 1 (Fig. 3.4a): Retrieval of permittivity ε_r (top) and retrieval of conductivity σ (bottom) by using sparsity and projection with 10 dB noise data.



Exact | $\alpha = 1 \times 10^{-5}$ w/o projection | $\alpha = 5 \times 10^{-3}$ w/o projection | $\alpha = 1 \times 10^{-6}$ w projection

Figure 3.13.: Scatterer 2 (Fig. 3.4b): Retrieval of permittivity ε_r (top) and retrieval of conductivity σ (bottom) by using sparsity and projection with 10 dB noise data.



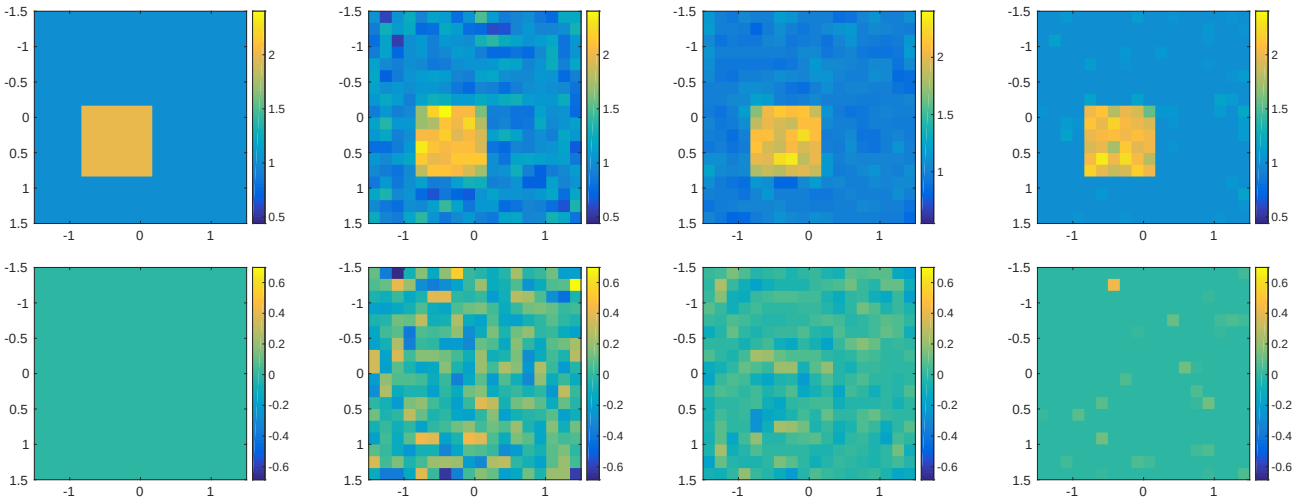
(a) Min.Cost Fun, Without projection

(b) Min.Cost Fun, With projection

(c) ε_r, Without projection

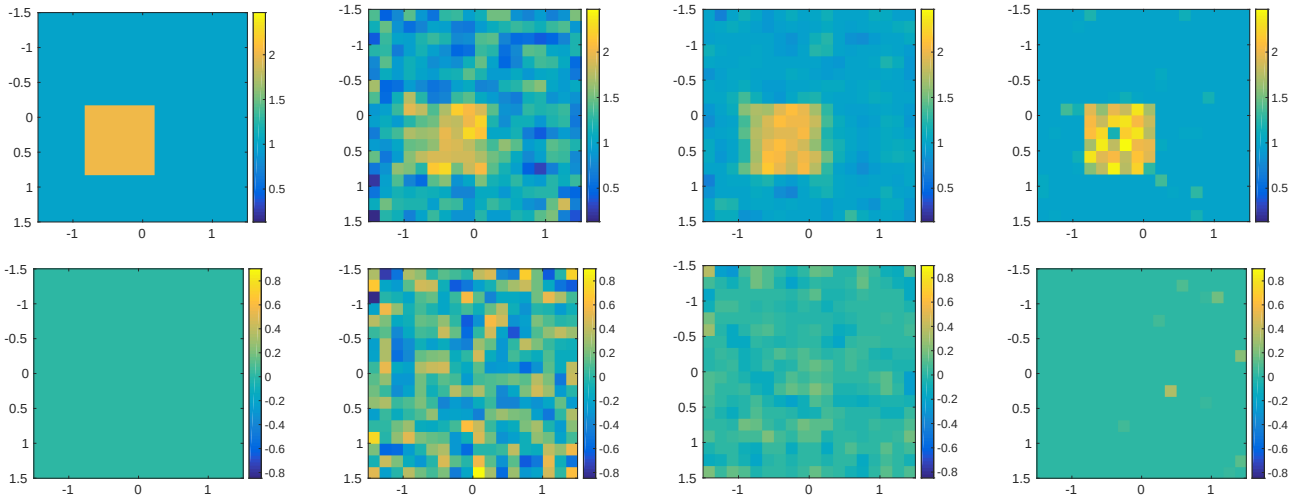
(d) ε_r, With projection

Figure 3.14.: Scatterer 4 (Fig. 3.4d): Error in ε_r and minimum cost function obtained by using sparsity and positivity with 20 noise levels of 10 dB data.



Exact | $\alpha = 1 \times 10^{-5}$ Without Projection | $\alpha = 2.5 \times 10^{-2}$ Without Projection | $\alpha = 2.5 \times 10^{-2}$ With Projection

Figure 3.15.: Scatterer 3 (Fig. 3.4c): Retrieval of permittivity (ε_r, top) and conductivity (σ, bottom) by using sparsity and projection with 10 dB noise data.



Exact | $\alpha = 1 \times 10^{-5}$ Without Projection | $\alpha = 2.5 \times 10^{-2}$ Without Projection | $\alpha = 2.5 \times 10^{-2}$ With Projection

Figure 3.16.: Scatterer 4 (Fig. 3.4d): Retrieval of permittivity (ε_r, top) and conductivity (σ, bottom) by using sparsity and projection with 10 dB noise data.

3.4.3. Analysis of Proposed Approaches

In this section, the effectiveness of the proposed approaches has been assessed and compared against different scatterers as in Fig.3.4. In particular, we study the Stationary Wavelet Transform (SWT) in terms of Haar basis of level two and three, and the Discrete Wavelet transform (DWT).

The SWT is a wavelet transform achieving translation invariance, a property that is missing in the DWT [38]. 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 SWT 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 [38]. In other words, such a representation is not univocal, as the considered functions are not orthonormal. 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 [38].

On the other hand, 'Wavelab' which is a library of Matlab routines for wavelet analysis, wavelet-packet analysis, cosinepacket analysis and matching pursuit is used in Matlab to study our problem in Discrete Wavelet Transform. DWT is any wavelet transform where the wavelets are discretely sampled. In practice, we only have a finite number of values of the contrast function in a rectangular domain D, which is discretized into $2^J \times 2^J$ grid cells. The values of the contrast function on these cells represent the scaling coefficients at the highest scale J (decomposition level). In our case, we have used 32×32 ($J = 5$) discretization for each geometry while solving our inverse problem in the wavelet domain.

As mentioned before the choice of the wavelet family plays an important role since the priority is to choose a wavelet basis that produces as many zero coefficients as possible. This choice is based on the analysis of three main criteria which are the vanishing moments, support size and the regularity. High-amplitude coefficients occur when the support of the wavelet overlap with transitions such as edge and they are proportional to the width of the wavelet support, which needs preferably to be as small as possible. However, the number of vanishing moments is proportional to the support size so that the choice of the optimal wavelet is a trade-off between the number of vanishing moment and support size. Furthermore, the wavelet regularity can reduce the visibility of artifacts.

Considering all these features and the Tab.3.4 it can be observed that the Daubechies and Symmlet families represent good options to be employed in the numerical analysis. However, as it was pointed out before, support size with respect to our unknown scatterers plays an important role. Taking account the sparsity of our unknown scatterers within these wavelet basis, we will firstly exploit the 'Haar' (db1) wavelet and do the comparison with Daubechies (db4) wavelet basis.

Secondly, we will do the optimization on detail coefficients of the contrast function, which are its finer details by exploiting Haar basis for SWT (swt-db1-opt.details) and Wavelab case (WL-db1-opt.details). Before passing to the comparisons of the methods with different characteristic properties of wavelet functions, we would like to provide some notations for simplicity:

- Haar wavelet: db1
- Daubechies 4 wavelet: db4
- Optimization on both detail and approximation coefficients by wavelet decomposition using Wavelab: WL
- Optimization on both detail and approximation coefficients by wavelet decomposition using Stationary Wavelet Transform: SWT
- Optimization on detail coefficients by exploiting Haar basis through SWT: swt-db1-opt.details
- Optimization on detail coefficients by exploiting Haar basis through Wavelab: WL-db1-opt.details
- level two: $l = 2$ and level three: $l = 3$.

We have used 20 dB and 10 dB noise data set in order to illustrate our proposed approaches. Moreover, we exclude the smoothing the gradient constraint in soft-shrinkage algorithm both in spatial and wavelet domains for the rest of the simulations since addition of smoothness does not improve the solution accuracy with the proper choice of regularization parameter.

20 dB noise data: For instance, the plots of the real part and the imaginary one in Fig. 3.17 of the contrast retrieved from blurred scattering data with a SNR on the scattered field amplitude equal to 20 dB prove a good reconstruction of the actual contrast despite its nonsparse nature with respect to a standard single-pixel basis. Comparing the performance

Property	Haar	Daubechies N	Symmlet N	Coiflet N	Bi-orthogonal $N_r.N_d$	Bi-orthogonal Reverse $N_r.N_d$
Support width	1	$2N-1$	$2N-1$	$6N-1$	$2N_r+1$	$2N_d+1$
Number of vanishing moments	1	N	N	$2N$	N_r	N_d
Regularity	≈ 0.2	$\approx 0.2N$	$2N$	Not defined	N_r-1	N_d-1

Table 3.4.: Features of the main wavelet families

of the approach in the wavelet domain with the one in pixel basis, wavelet based soft shrinkage reconstruction scheme is faster. However, as it can be noticed in Tab. 3.5, the wavelet based approach improvement is of about 0.3% which is quite low. Moreover, addition of projection to the pixel based reconstruction algorithm gives us better result. On the other hand, choosing Daubechies (db4) basis decomposition does not give us a good quality of reconstruction while its performance is slower when compared to others.

For the case of Scatterer 9 (Fig. 3.4i), we can see from Tab. 3.5 when we increase the decomposition level from level 2 to 3 we get a slightly better solution. Even if there is not a big difference in the reconstruction error of permittivity between Haar and Daubechies (db4) basis decompositions, the reconstruction obtained by Wavelab with Haar basis is smoother than the one obtained with db4.

10 dB noise data: When we increase the noise such as in Tab. 3.6, the reconstruction error increases directly. However, interestingly, the error difference between WL-db1 and pixel reconstruction schemes is around 0.0015, which is not remarkable while the scatterer is not exactly sparse in the domain of the pixel basis function. On the contrary, the performance of WL-db1 is better than the one with pixel basis decomposition.

When we consider scatterer 7 (Fig. 3.4g), which counts as sparse with respect to the pixel basis, we can see in Tab. 3.5, 3.6 that the soft shrinkage algorithm gives a slightly better reconstruction in the spatial domain. On the other hand, it is worthwhile to mention that the algorithm in wavelet decomposition also achieves a good resolution of the image whereas wavelet decomposition in terms of swt is not preferable because of its low performance and lower reconstruction quality.

We also consider the applicability of the proposed methods with the scatterers which have partial views (Fig. 3.4d, 3.4j) and a case where the sources and receivers are located on the top half of the region of interest as in Fig. 3.4h and Fig. 3.4k. Compared to the full view scatterers, these configurations have worse reconstruction quality. However, scatterer

4 (Fig. 3.4d) has lower error in permittivity compared to its full view case (Fig. 3.4f) as can be seen in Tab. 3.7 and Tab. 3.8 especially when we have the wavelet decomposition with swt. For the scatterer 11 (Fig. 3.4k) we have better results by using wavelet basis functions than the pixel functions albeit the scatterer is sparse with respect to pixel basis.

When we have both reflection and transmission configurations the effectiveness of the proposed approaches reduces as the *a-priori* information is limited due to the information obtained at the receivers. However, we can still obtain the best approximation to the actual profiles by enforcing sparsity through projection as it can be seen in Tabs. 3.7 and 3.8. Like for all other results, application of projection gives us the better minimum error for the permittivity and conductivity for both 20 dB and 10 dB noise data sets while keeping its slow performance compared to performances of other approaches.

It is also worthwhile to mention that in [90] the authors studied wavelet decomposition in breast cancer imaging and according to their results the wavelet decomposition recovers the unknown profile around 15% more than the pixel basis case based on the fact that there is no frequency hopping for wavelet decomposition. However, in studies [41, 91, 92] wavelet decomposition gives better results than pixel basis decomposition taking into consideration scatterers which are sparse in wavelet domain and not in spatial one.

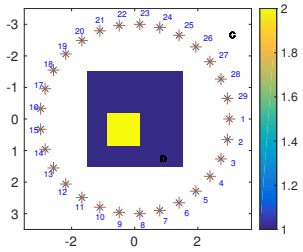
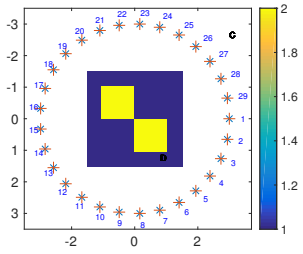
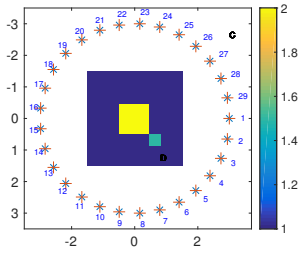
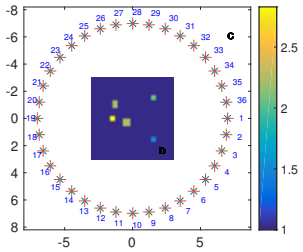
Scatterer	Test	Time (s)	χ^{err}	ε^{err}	σ^{err}
	swt-db1 (l=2) full	133.3	0.0928	0.0847	
	swt-db1 (l=3) full	137.7	0.0927	0.0853	
	WL-db1 (l=2) full	117.2	0.0913	0.0831	
	WL-db1 (l=3) full	111.9	0.0911	0.0832	
	projection	165.7	0.0770	0.0768	
	pixel	185.6	0.0914	0.0831	
	swt-db1-opt. details	117.6	0.0919	0.0837	
	WL-db1-opt. details	136.3	0.0917	0.0833	
	WL-Daubechies	135.7	0.2354	0.2266	
	swt-db1 (l=2) full	150.1	0.1538	0.1322	0.4914
	swt-db1 (l=3) full	144.3	0.1534	0.1313	0.4953
	WL-db1 (l=2) full	194.2	0.1554	0.1348	0.4850
	WL-db1 (l=3) full	160.5	0.1516	0.1317	0.4705
	projection	169.6	0.1221	0.1129	0.3016
	pixel	198.7	0.1549	0.1345	0.4824
	swt-db1-opt. details	125.5	0.1535	0.1333	0.4780
	WL-db1-opt. details	180.2	0.1520	0.1320	0.4723
	WL-Daubechies	153.3	0.1531	0.1330	0.4760
	swt-db1 (l=2) full	107.6	0.1109	0.0971	
	swt-db1 (l=3) full	114.3	0.1110	0.0969	
	WL-db1 (l=2) full	104.2	0.1106	0.0976	
	WL-db1 (l=3) full	105.3	0.1104	0.0975	
	projection	183.4	0.0920	0.0918	
	pixel	162.7	0.1108	0.0970	
	swt-db1-opt. details	119.4	0.1109	0.0972	
	WL-db1-opt. details	133.7	0.1105	0.0978	
	WL-Daubechies	104.0	0.1105	0.0976	
	swt-db1 (l=2) full	135.0	0.1486	0.0914	0.6237
	swt-db1 (l=3) full	128.6	0.1495	0.0915	0.6289
	WL-db1 (l=2) full	147.3	0.1467	0.0903	0.6151
	WL-db1 (l=3) full	126.1	0.1478	0.0905	0.6218
	projection	170.5	0.1113	0.0677	0.4825
	pixel	142.1	0.1466	0.0901	0.6152
	swt-db1-opt. details	126.9	0.1473	0.0904	0.6190
	WL-db1-opt. details	124.5	0.1471	0.0905	0.6175

Table 3.5.: Error and average simulation time in seconds with 20 dB

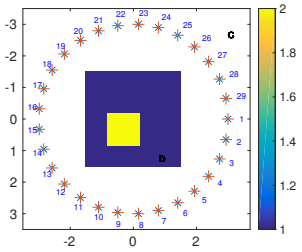
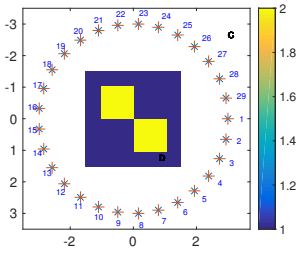
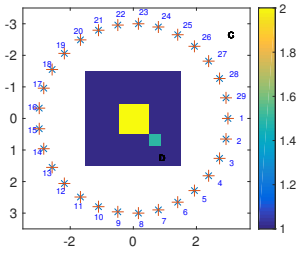
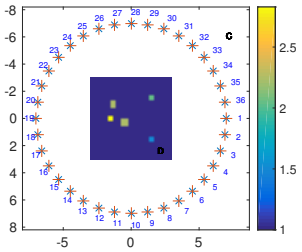
Scatterer	Test	Time (s)	χ^{err}	ε^{err}	σ^{err}
	swt-db1 (l=2) full	143.5	0.1422	0.1128	
	swt-db1 (l=3) full	147.8	0.1404	0.1108	
	WL-db1 (l=2) full	132.4	0.1358	0.1104	
	WL-db1 (l=3) full	118.4	0.1340	0.1093	
	projection	166.1	0.0960	0.0919	
	pixel	197.0	0.1304	0.1068	
	swt-db1-opt. details	125.0	0.1395	0.1127	
	WL-db1-opt. details	140.6	0.1314	0.1121	
	WL-Daubechies	148.3	0.1339	0.1092	
	swt-db1 (l=2) full	123.0	0.2275	0.1785	0.8686
	swt-db1 (l=3) full	128.1	0.2292	0.1780	0.8922
	WL-db1 (l=2) full	119.7	0.2322	0.1852	0.8636
	WL-db1 (l=3) full	126.2	0.2281	0.1829	0.8406
	projection	181.2	0.1783	0.1575	0.5272
	pixel	177.2	0.2296	0.1839	0.8484
	swt-db1-opt. details	141.2	0.2330	0.1857	0.8680
	WL-db1-opt. details	152.2	0.2289	0.1833	0.8453
	WL-Daubechies	166.4	0.2296	0.1838	0.8488
	swt-db1 (l=2) full	126.6	0.1447	0.1148	
	swt-db1 (l=3) full	127.9	0.1456	0.1237	
	WL-db1 (l=2) full	122.9	0.1644	0.1277	
	WL-db1 (l=3) full	146.5	0.1654	0.1285	
	projection	207.8	0.1043	0.1018	
	pixel	169.0	0.1632	0.1265	
	swt-db1-opt. details	135.1	0.1674	0.1291	
	WL-db1-opt. details	120.8	0.1725	0.1328	
	WL-Daubechies	130.9	0.1641	0.1275	
	swt-db1 (l=2) full	146.7	0.1579	0.1010	0.6467
	swt-db1 (l=3) full	141.5	0.1614	0.0964	0.6883
	WL-db1 (l=2) full	147.9	0.1564	0.0959	0.6570
	WL-db1 (l=3) full	120.1	0.1566	0.0956	0.6596
	projection	179.3	0.1220	0.0719	0.5237
	pixel	186.8	0.1561	0.0954	0.6575
	swt-db1-opt. details	113.6	0.1563	0.0959	0.6564
	WL-db1-opt. details	130.8	0.1564	0.0960	0.6570

Table 3.6.: Error and average simulation time in seconds with 10 dB

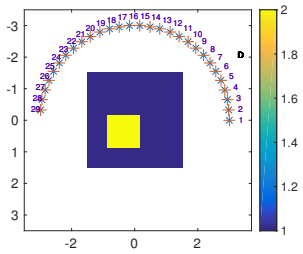
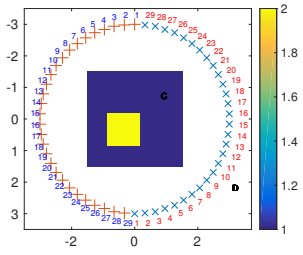
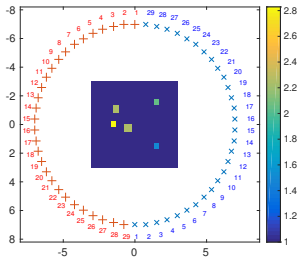
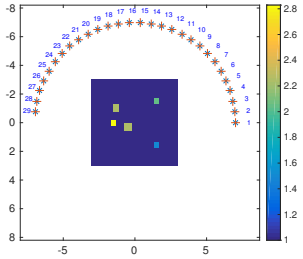
Scatterer	Test	Time (s)	χ^{err}	ε^{err}	σ^{err}
	swt-db1 (l=2) full	149.6	0.2816	0.2734	
	swt-db1 (l=3) full	121.6	0.2805	0.2720	
	WL-db1 (l=2) full	134.1	0.2806	0.2722	
	WL-db1 (l=3) full	167.6	0.2807	0.2726	
	projection	184.5	0.2718	0.2284	
	pixel	159.0	0.2816	0.2735	
	swt-db1-opt. details	133.2	0.2809	0.2725	
	WL-db1-opt. details	121.4	0.2812	0.2731	
	WL-Daubechies	122.0	0.2810	0.2728	
	swt-db1 (l=2) full	134.3	0.1166	0.1082	
	swt-db1 (l=3) full	127.7	0.1211	0.1127	
	WL-db1 (l=2) full	131.0	0.1219	0.1146	
	WL-db1 (l=3) full	141.2	0.1211	0.1142	
	projection	184.0	0.0791	0.0784	
	pixel	188.5	0.1217	0.1143	
	swt-db1-opt. details	124.5	0.1234	0.1152	
	WL-db1-opt. details	133.5	0.1233	0.1152	
	WL-Daubechies	142.7	0.1224	0.1150	
	swt-db1 (l=2) full	158.7	0.2263	0.1414	0.9407
	swt-db1 (l=3) full	134.2	0.2258	0.1411	0.9387
	WL-db1 (l=2) full	127.3	0.2268	0.1407	0.9468
	WL-db1 (l=3) full	114.9	0.2234	0.1398	0.9464
	projection	203.2	0.2260	0.1410	0.9402
	pixel	162.5	0.2259	0.1413	0.9381
	swt-db1-opt. details	139.2	0.2263	0.1411	0.9416
	WL-db1-opt. details	118.3	0.2268	0.1407	0.9469
	swt-db1 (l=2) full	148.7	0.2546	0.1607	1.0521
	swt-db1 (l=3) full	151.1	0.2526	0.1588	1.0464
	WL-db1 (l=2) full	132.4	0.2547	0.1607	1.0523
	WL-db1 (l=3) full	137.3	0.2546	0.1606	1.0522
	projection	392.4	0.2387	0.1457	1.0056
	pixel	458.5	0.2898	0.1884	1.1746
	swt-db1-opt. details	139.4	0.2549	0.1609	1.0528
	WL-db1-opt. details	134.8	0.2547	0.1607	1.0526

Table 3.7.: Error and average simulation time in seconds with 20db

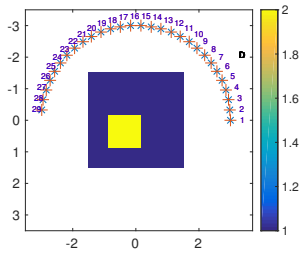
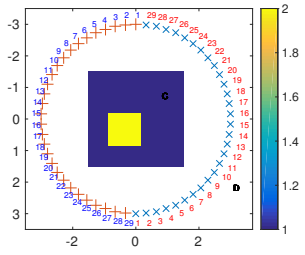
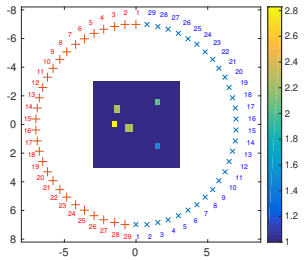
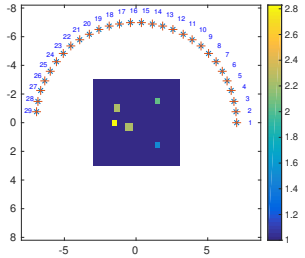
Scatterer	Test	Time (s)	χ^{err}	ε^{err}	σ^{err}
	swt-db1 (l=2) full	158.7	0.2807	0.2720	
	swt-db1 (l=3) full	131.8	0.2817	0.2729	
	WL-db1 (l=2) full	125.0	0.2820	0.2735	
	WL-db1 (l=3) full	127.0	0.2819	0.2737	
	projection	188.7	0.2728	0.2284	
	pixel	176.0	0.2825	0.2739	
	swt-db1-opt. details	133.3	0.2820	0.2732	
	WL-db1-opt. details	126.1	0.2818	0.2731	
	WL-Daubechies	119.4	0.2817	0.2730	
	swt-db1 (l=2) full	133.4	0.1612	0.1344	
	swt-db1 (l=3) full	141.5	0.1749	0.1438	
	WL-db1 (l=2) full	136.3	0.1675	0.1414	
	WL-db1 (l=3) full	125.9	0.1699	0.1434	
	projection	193.0	0.1098	0.1055	
	pixel	168.0	0.1591	0.1353	
	swt-db1-opt. details	146.1	0.1815	0.1515	
	WL-db1-opt. details	122.0	0.1785	0.1494	
	WL-Daubechies	120.1	0.1790	0.1499	
	swt-db1 (l=2) full	131.0	0.2275	0.1416	0.9481
	swt-db1 (l=3) full	146.0	0.2275	0.1423	0.9453
	WL-db1 (l=2) full	114.3	0.2279	0.1413	0.9520
	WL-db1 (l=3) full	112.1	0.2280	0.1414	0.9521
	projection	160.0	0.2035	0.1244	0.8564
	pixel	164.0	0.2273	0.1417	0.9462
	swt-db1-opt. details	124.3	0.2275	0.1417	0.9474
	WL-db1-opt. details	137.9	0.2279	0.1413	0.9521
	swt-db1 (l=2) full	148.3	0.2536	0.1600	1.0481
	swt-db1 (l=3) full	150.7	0.2525	0.1589	1.0453
	WL-db1 (l=2) full	137.2	0.2539	0.1602	1.0490
	WL-db1 (l=3) full	117.0	0.2544	0.1607	1.0503
	projection	406.3	0.2389	0.1456	1.0075
	pixel	440.5	0.2901	0.1884	1.1764
	swt-db1-opt. details	129.1	0.2535	0.1598	1.0478
	WL-db1-opt. details	148.8	0.2542	0.1605	1.0498

Table 3.8.: Error and average simulation time in seconds with 10dB

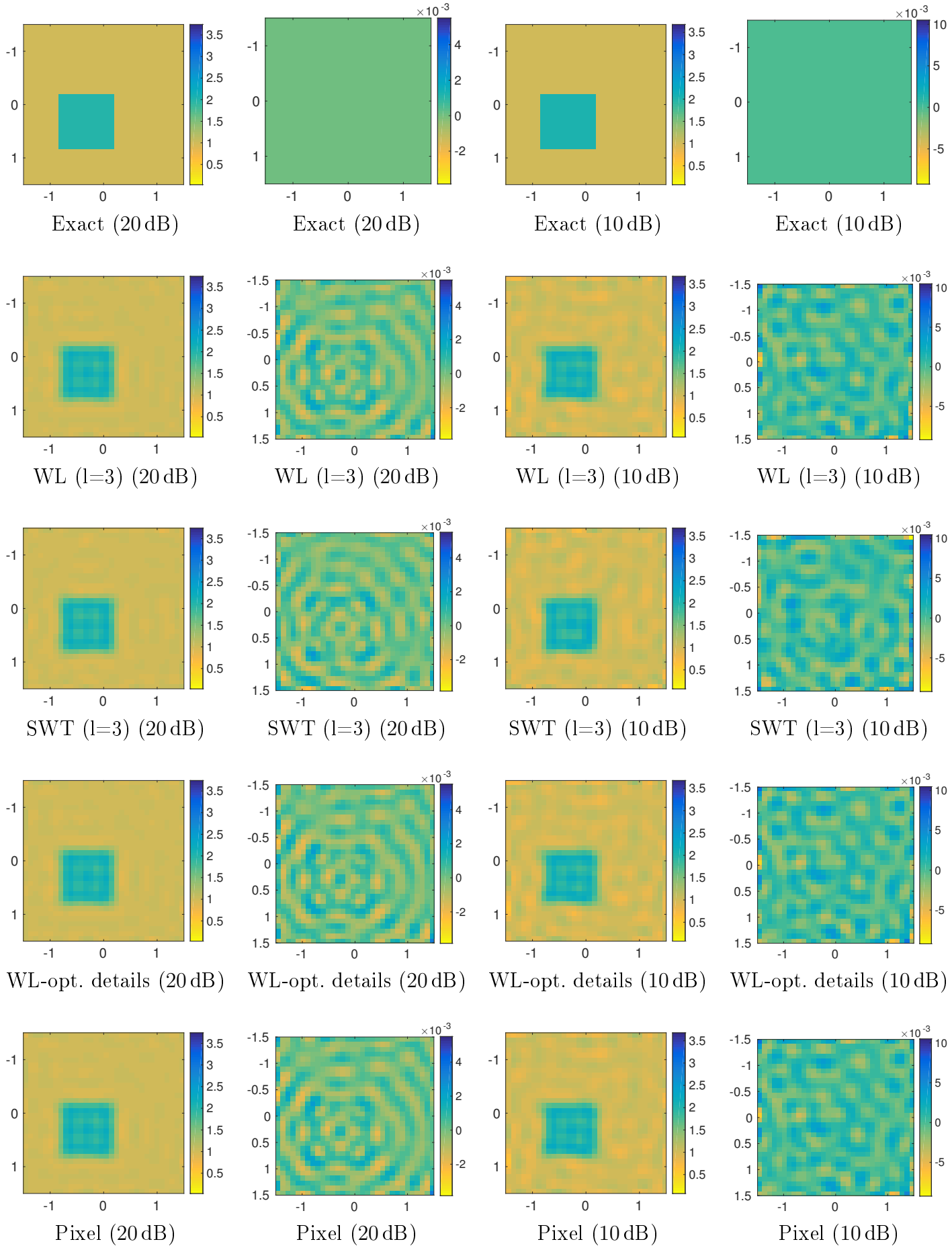


Figure 3.17.: Retrieval of permittivity (ε_r, 1st and 3rd column) and conductivity (σ, 2nd and 4th column) with 20 dB and 10 dB noise data ($\alpha = 2.5 \times 10^{-2}$ without projection and smoothness).

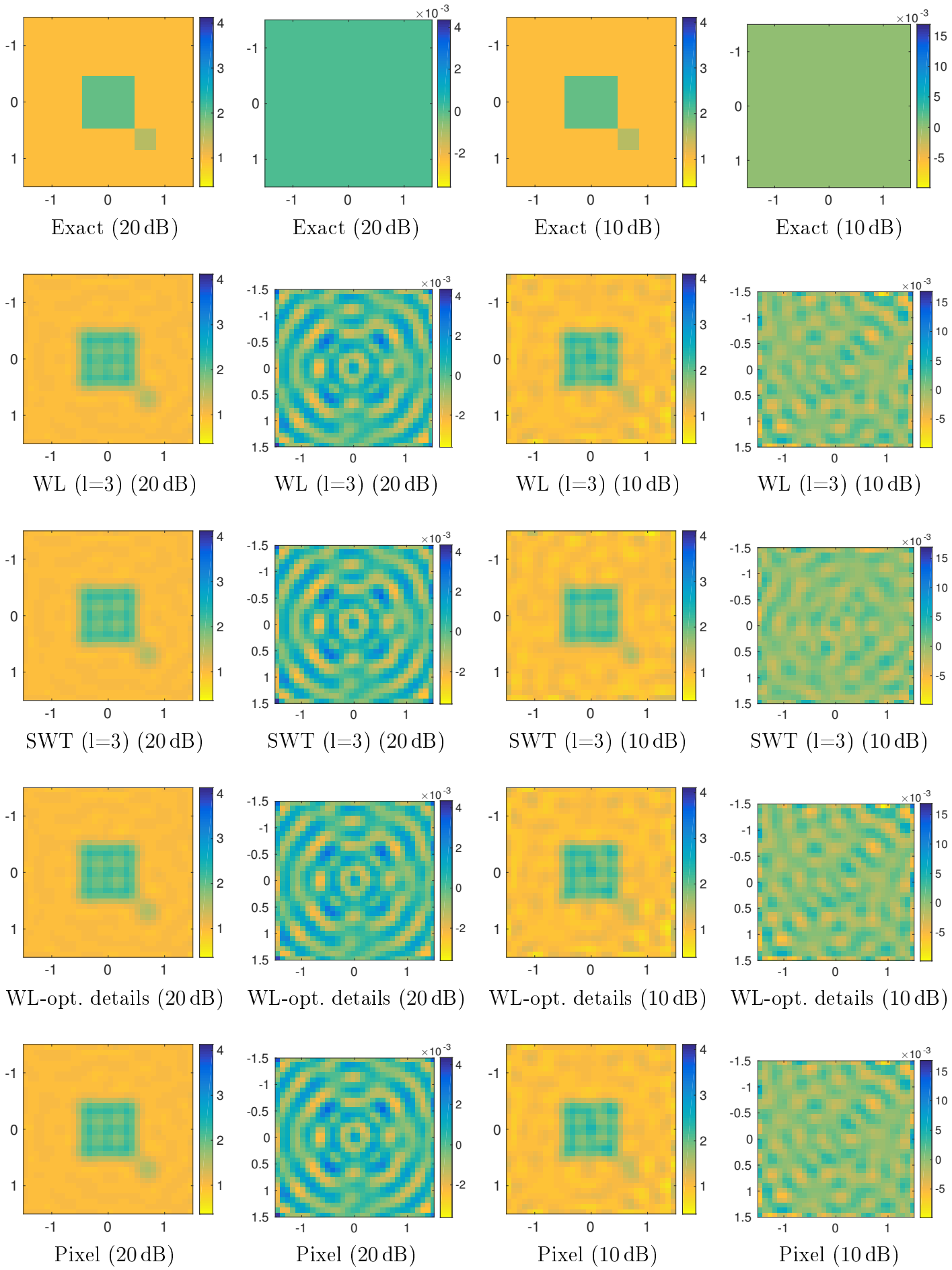


Figure 3.18.: Retrieval of permittivity (ε_r, 1st and 3rd column) and conductivity (σ, 2nd and 4th column) with 20 dB and 10 dB noise data ($\alpha = 2.5 \times 10^{-2}$ without projection and smoothness).

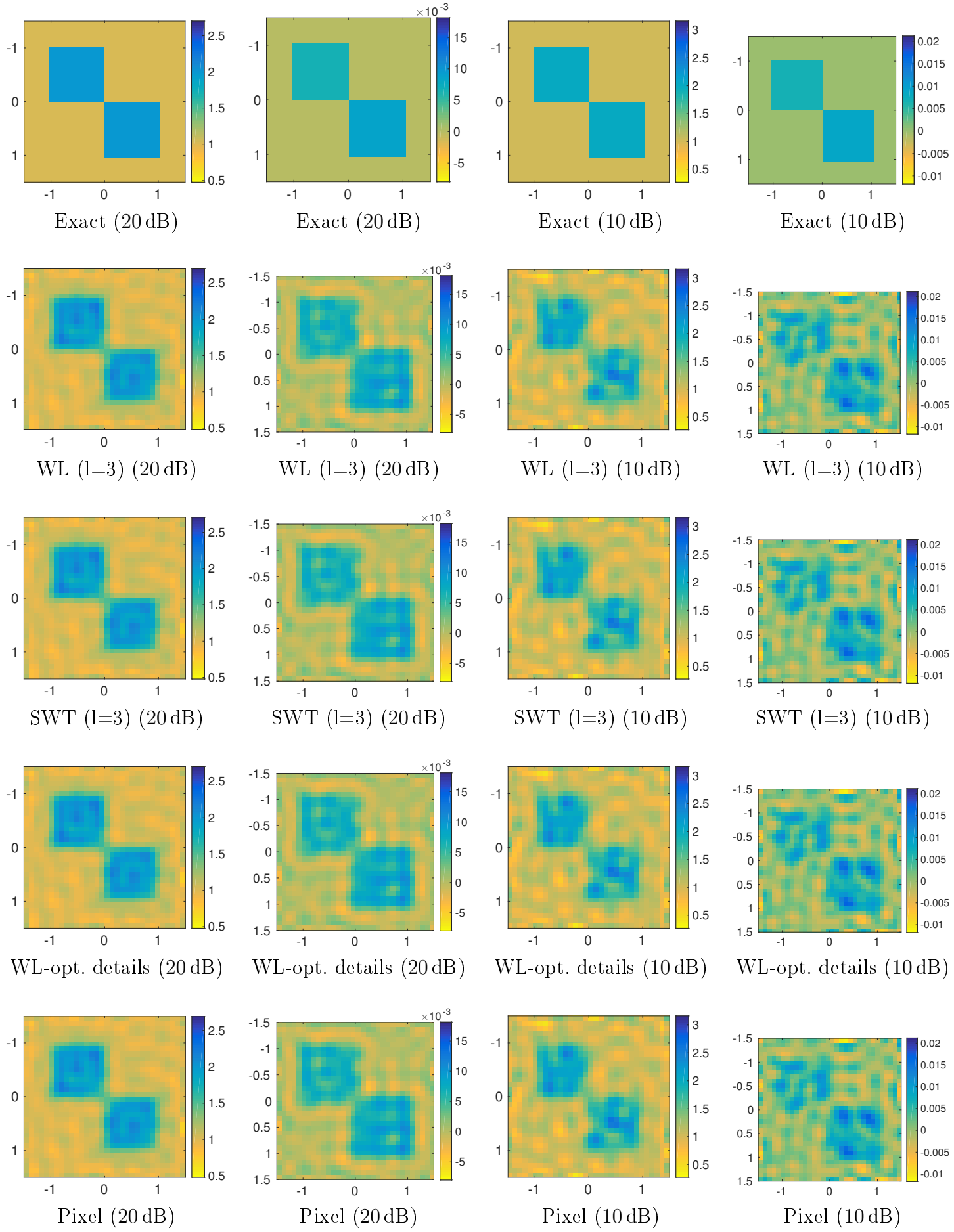


Figure 3.19.: Retrieval of permittivity (ε_r, 1st and 3rd column) and conductivity (σ, 2nd and 4th column) with 20 dB and 10 dB noise data ($\alpha = 2.5 \times 10^{-2}$ without projection and smoothness).

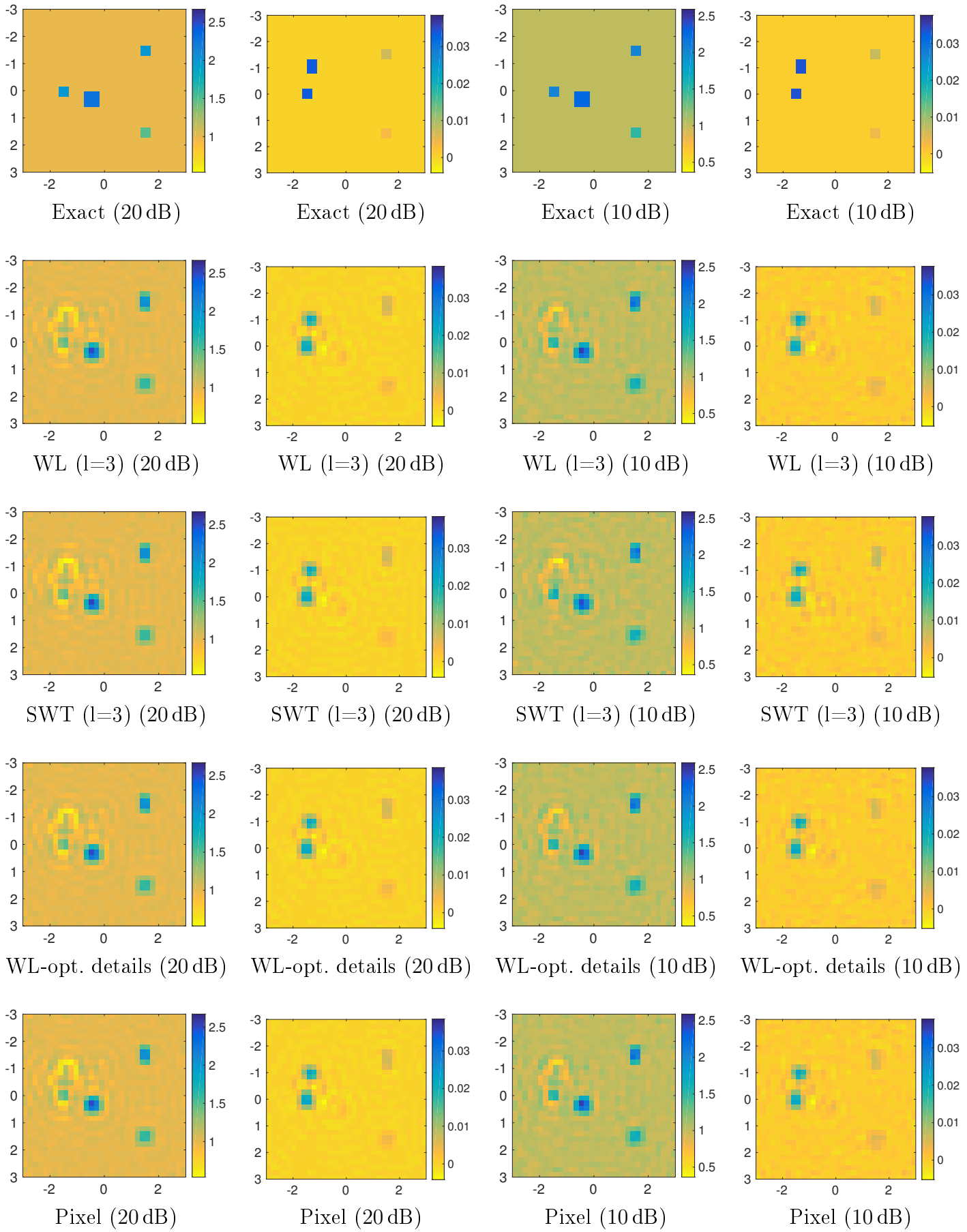


Figure 3.20.: Retrieval of permittivity (ε_r, 1st and 3rd column) and conductivity (σ, 2nd and 4th column) with 20 dB and 10 dB noise data ($\alpha = 5 \times 10^{-3}$ without projection and smoothness).

3.5. Conclusion

Two imaging techniques have been introduced in this Chapter in order to solve the nonlinear inverse scattering problem. The first technique is based on the iterative algorithm of soft shrinkage type which enforces the sparsity constraint at each nonlinear iteration. We have introduced constraints such as Barzilai and Borwein (BB) step size selection criteria and projection constraint on contrast in order to favor the performance and quality of the proposed algorithm.

We extend our proposed inversion algorithm based on soft shrinkage enforcement over the reconstruction of small isolated scatterers. By adopting wavelet basis functions we discretized the formulation of the 2D inverse scattering problem where the imaging problem has been solved with a soft shrinkage reconstruction algorithm as a second technique. Selected results from numerical experiments have been presented and discussed to give some insights about the robustness, the flexibility, and the accuracy of the proposed approaches as well as to illustrate their advantages and limitations with respect to differently constrained inversion methods.

Since sparsity is a relative concept, the effectiveness of each approach depends on the actual scatterer and the adopted representation basis, i.e, on the available *a-priori* information on the scattering scenario to be used for choosing the basis. We have studied different scatterers within this respect. The results showed us that whenever the proper wavelet basis is selected for the scenario at hand (Fig. 3.4), the proposed method based on soft shrinkage enforcement gives good results in terms of computational efficiency and accuracy (Tabs. 3.5, 3.6, 3.7, 3.8). The wavelet based method can be reliably applied to a wide set of scattering configurations by handling a reduced set of scattering data.

Overall, even if the soft shrinkage algorithm which has been studied both in spatial domain and wavelet domain gives us satisfactory results, an inversion based on projection constraint gives us slightly better results yet its performance is more time consuming.

4. Two-Step Inversion Method

4.1. Motivation

Additional to the interest of exploiting sparsity which has been mentioned in previous chapters, the other key point of the sparse estimation problem is to reveal the identification of the support, which denotes the indices of the nonzeros. If the support is known, the estimation of the sparse vectors reduces to a standard overdetermined linear inverse problem [93].

A wide variety of the studies starts to contemplate the structure information of the solutions in order to facilitate a better estimation. For instance, 'group sparsity' structure [46] has been presented in many applications, where the components of solutions are likely to be either all zero or all nonzero in a group. Thus, one aims to decrease the dispersion to enhance the solution by taking account the grouping prior. In this chapter, we focus on joint sparsity, which is a special case of the group sparsity. Specifically, joint sparsity means that multiple unknown sparse vectors ($\mathbf{x}_j \in \mathbb{R}^n, j = 1, \cdots, s$) share a common unknown nonzero support set [46, 47].

We propose a two-step inversion approach in order to solve a nonlinear inverse problem as ours by applying joint sparsity to get the equivalent current, then the unknown contrast. We show the efficiency, accuracy and the limitations of our proposed method with the results obtained in 2D and 3D.

4.2. Two-Step Inversion

In compressive sensing, joint sparsity desires to reconstruct unknown signals from m measurement vectors based on a common measurement matrix. This is also called the multiple measurement vectors (MMV) problem [94]. Given the vectors $\mathbf{y}_j, \in \mathbb{R}^m$, and a measurement matrix $A \in \mathbb{R}^{m \times n}$, we want to recover the $\mathbf{x}_j$ from the noisy underdetermined systems $\mathbf{y}_j = A\mathbf{x}_j + \mathbf{n}_j$ $(j = 1, \cdots, s)$, where $\mathbf{n}_j$ is the noise vector. The vectors $\mathbf{x}_1, \cdots, \mathbf{x}_s$

share the sparsity pattern M, i.e., the nonzero entries of $\mathbf{x}_1, \cdots, \mathbf{x}_s$ appear at the same positions. A common recovery model is

$$\min_X |M| \quad \text{s.t} \quad \mathbf{y}_j = A\mathbf{x}_j + \mathbf{n}_j, \quad j = 1, \cdots, s, \tag{4.1}$$

where $|M|$ is the cardinality of M [95]. Since (4.1) is NP-hard [33, 96], this problem is usually relaxed with a convex alternative which is computationally efficient at the expense of more being required measurements. Like l_1-norm being the convex relaxation of l_0-norm [24], the (weighted) $l_{2,1}$-norm is widely used as the convex replacement of $|M|$ as stated below [46, 95]:

$$\min_X \|X\|_{w,2,1} := \sum_{i=1}^{n} w_i \left\|\mathbf{x}^i\right\|_2 \quad \text{s.t} \quad AX = B + N, \tag{4.2}$$

where $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times l}$ is an available measurement matrix consisting of l measurement vectors, and $\mathbf{x}_j \in \mathbb{R}^n$ denote the i-th row and the j-th column of X, whereas $X = [\mathbf{x}_1, \cdots, \mathbf{x}_s] \in \mathbb{R}^{n\times l}$ denotes a collection of l jointly sparse solutions (unknown source matrix) while $w_i \geq 0$. $N \in \mathbb{R}^{m\times l}$ is an unknown noise matrix.

A key assumption in the MMV model is that the support of every column of X is identical. Similarly with the constraint in the single measurement vector model, the number of nonzero rows in X has to be below a threshold to maintain a unique global solution [97]. This leads to the fact that X has a small number of nonzero rows.

Following all the above properties, the other study of the thesis is coming from adopting a two-step method which firstly consists of finding the equivalent current $\mathbf{J}_i$ for $i = 1, \ldots, N_s$ (N_s represents the number of sources) using (2.9) and then looking for $\boldsymbol{\chi}$ by combining (2.14) and (2.11) instead of solving directly the nonlinear inverse problem given by (3.2) as we studied in Chapter 3. This approach leads us to have two linear minimization problems solved by exploiting the jointly-sparse aspect of the sought equivalent currents solution of the first step and a classical l_2-minimization of a linear problem for the second step.

4.2.1. First step: Reconstruction of the equivalent currents

The following optimization problem is solved

$$\mathbf{J}_i^\star = \underset{\mathbf{J}_i}{\operatorname{argmin}} \left[\frac{1}{2}\|\boldsymbol{\zeta}_i - \mathbf{G}_{\mathbf{or}}\mathbf{J}_i\|_2\right] \quad i = 1, \cdots, N_s. \tag{4.3}$$

where $\boldsymbol{\zeta}_i$ is a vector of size N_r which gathered the signal due to the source $\#i$ measured by the N_r receivers.

The main idea of the approach is to take into account that, as shown by (2.24), $\mathbf{J}_i$ and $\boldsymbol{\chi}$ share the same support which means that when $\chi^j = 0$ then $J_i^j = 0, \forall i = 1, \ldots, N_s$ where χ^j and J_i^j are the j^{th} element of $\boldsymbol{\chi}$ and $\mathbf{J}_i$ respectively. Taking into account this hypothesis, (4.3) based on the original joint sparsity model (4.2) can be recast as a minimization problem under the constraints that $\mathbf{J}_i$ has the same sparse support for each source i leading to the use of a (weighted) $l_{2,1}$-regularization to enforce joint sparsity such as

$$\min_{\mathbf{J}} \|\mathbf{J}\|_{w,2,1} := \sum_{j=1}^{N} w_j \|\mathbf{J}^j\|_2 \quad \text{s. t.} \quad \mathbf{G}_{\mathbf{or}}\mathbf{J}_i = \boldsymbol{\zeta}_i \tag{4.4}$$

while w_j is the corresponding weight [46].

In [98], Yong *et al.* applied the alternating direction method (ADM) technique to solve the l_1 problem in compressed sensing and developed the corresponding MATLAB package termed Your ALgorithms for L1 (YALL1). Furthermore, Deng *et al.* extended the YALL1 to the group version for solving the group sparse optimization with $l_{2,1}$-norm regularization in [46]. In this respect, we use YALL1 Group [46] package which encodes the joint sparsity model in order to solve (4.4).

Alternatively, we use T-MSBL which is a block sparse Bayesian learning to solve (4.4). T-MBSL identifies the MMV model in order to exploit the correlation that exists in each nonzero row of X while automatically choosing the optimal regularization value. More details related to this algorithm can be found in [99]. The MATLAB code for both algorithms can be obtained online as well.

In [98], it has been also proved that YALL1 gives better performance and solution accuracy compared to state of art algorithms such as spectral projection gradient method (SPGL1) [35], SpaRSA (a sparse reconstruction algorithm for more general regularizers) [89], FISTA (a fast iterative shrinkage thresholding algorithm that attains an optimal convergence rate in function values) [58]. Therefore, we have used YALL1 toolbox and T-MSBL which applies joint sparsity approach by adopting a probabilistic approach to incorporate correlation structure in each nonzero row of the solution matrix differently than the existing algorithms to apply our proposed method.

4.2.2. Second step: Reconstruction of the contrast function

Once $\mathbf{J}_i$ is known, the contrast function is obtained by solving the following minimization problem [100]

$$\boldsymbol{\chi}^\star = \operatorname{argmin} \sum_{i=1}^{N_s} \|\mathbf{J}_i - \operatorname{diag}(\boldsymbol{\chi})\mathbf{E}_i\|_2 = \frac{\sum_{i=1}^{N_s} \mathbf{J}_i \cdot \bar{\mathbf{E}}_i}{\sum_{i=1}^{N_s} \mathbf{E}_i \cdot \bar{\mathbf{E}}_i}, \tag{4.5}$$

where $\mathbf{E}_i$ has been obtained using (2.25) and $\bar{\mathbf{E}}_i$ being its conjugate.

4.2.3. Two-Step Inversion Method Results in 2D

It is worthwhile to restate that sparseness is a relative concept with respect to a basis [4]. In the case of a pixel basis a fast way to estimate the sparsity of our problem is to define it as the ratio of the obstacle's areas to the investigation domain D area. By this way, sparseness of the first example (Fig. 3.4b) is around 3 % whereas the one of the second example (Fig. 3.4f) is around 11 %. One of the key points of our examples is that in the first one the scatterers are sparse with respect to their pixel expansion basis while in the second one the scatterer is less sparse and our proposed approach in this chapter performs differently for these two cases.

We exploit the joint sparsity by using YALL1 and T-MSBL algorithms in our two-step inversion method. Thereafter, we compare these approaches with SpaRSA and Born Approximation (B.A). SpaRSA is a method based on iterative shrinkage/thresholding (IST) [89]. It computes the steepest descent direction on the l_2 norm in (4.3) and uses a very simple soft thresholding function related to the regularization term.

As for the parameters of YALL1 algorithm we choose the primal-based solver where the linear system is exactly solved. The weights have been chosen as $w_i = 1$ $(i = 1, \cdots, N)$ and the initialization $\mathbf{J} \equiv 0$ (no prior information). We have also not added projection while keeping the number of iterations as $10 \times$ # of measurements.

It can be observed that when we have sparse scatterers (as in Fig. 4.1) the two-step inversion approach gives a better reconstruction in terms of quality of $\boldsymbol{\chi}$ and the processing time compared to the soft shrinkage algorithm as shown in Tab. 4.1. However, application of joint sparsity through T-MSBL is more favorable than the YALL1 when we have sparse scatterers with respect to their basis function. Error on contrast and permittivity is calculated as in equations 3.56 and 3.57

On the other hand, even if the two-step inversion approach is faster in the reconstruction

of the scatterer in Fig. 4.2, the soft shrinkage algorithm yields a better reconstruction quality with an appropriate regularization parameter. Furthermore, one can also observe that when we decrease the number of discretization (comparing Fig. 3.4b to Fig. 3.4g), the difference in reconstruction quality is quite small such as 1.21%. Also, Born approximation and SpaRSA approaches give us poor reconstruction even if they are less time consuming as it can be seen in Tab. 4.1.

When we transform the two-step inversion approach from spatial to wavelet domain, our minimization problem transforms into

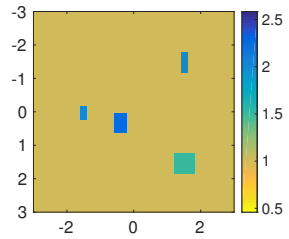
$$\min_{\tilde{\mathbf{J}}} \|\tilde{\mathbf{J}}\|_{w,2,1} := \sum_{j=1}^{N} w_j \|\tilde{\mathbf{J}}^j\|_2 \quad \text{s. t.} \quad \mathcal{W}(\mathbf{G_{or}}\mathbf{J}_i) = \mathcal{W}(\boldsymbol{\zeta}_i), \tag{4.6}$$

where $\tilde{\mathbf{J}} = \sum_{j=1}^{N} J_j C_j$ implying $\tilde{\mathbf{J}} = \mathcal{W}(\mathbf{J})$ as in (3.42). $\{C_j\}_{j=1}^{N}$ represent the wavelet functions and $\{J_j\}$ are the decomposition coefficients.

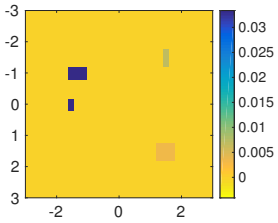
When we compare both scatterers (Fig. 3.4f and Fig. 3.4g) in the wavelet domain (WD) and the spatial domain, soft shrinkage algorithm provides approximately the same error. However, the soft shrinkage algorithm in wavelet domain achieves slightly better results than the two-step inversion method applied through YALL1 and T-MSBL as shown in Tab. 4.2. On the other hand, when we compare the approaches studied in the spatial and the wavelet domain, we can see in Tab. 4.1 and Tab. 4.2 for the case of Fig. 3.4g where we have 5 small scatterers, the two-step inversion method through T-MSBL gives better results in the spatial domain than is the wavelet domain.

Furthermore, as it can be seen in Fig. 4.3 when we reduce the noise level on our synthetic data, the error on the contrast function shows us that the effectiveness of the approximation gets better. For instance, when the scatterer is as in Fig. 3.4f, the error obtained through the soft shrinkage algorithm gives us better estimate than the other methods applied.

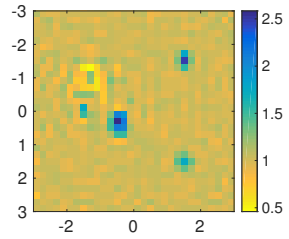
Contrarily, the two-step inversion method applied by T-MSBL in the spatial domain reduces the error more than the other methods are doing when the scatterer is small as in Fig. 3.4g. On the other hand, we can observe that the soft shrinkage algorithm provides a good approximation both in the spatial and the wavelet domains even if we have our data disturbed with a high level of noise.



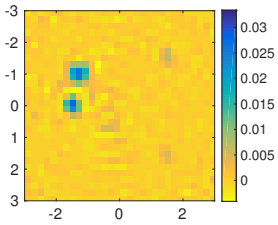
(a) Exact



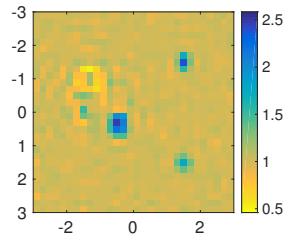
(b) Exact



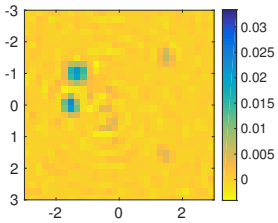
(c) Soft Shrinkage Alg. $\alpha = 1 \times 10^{-6}$



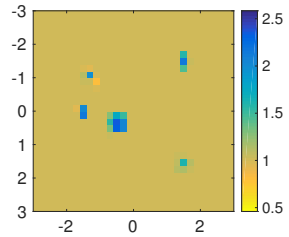
(d) Soft Shrinkage Alg. $\alpha = 1 \times 10^{-6}$



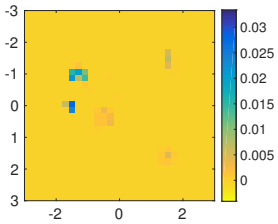
(e) Soft Shrinkage Alg. $\alpha = 5 \times 10^{-3}$



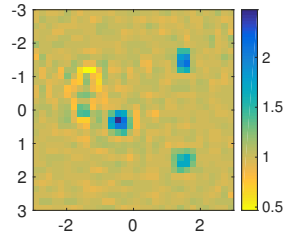
(f) Soft Shrinkage Alg. $\alpha = 5 \times 10^{-3}$



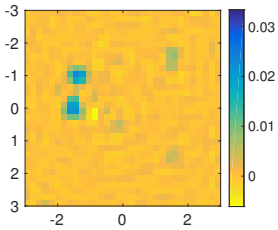
(g) Jointly-sparse model-T-MSBL



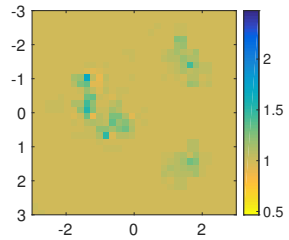
(h) Jointly-sparse model-T-MSBL



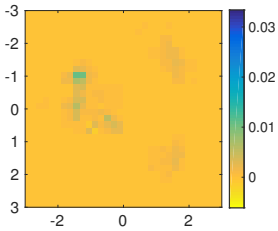
(i) Soft Shrinkage Alg. $\alpha = 5 \times 10^{-3}$-WD



(j) Soft Shrinkage Alg. $\alpha = 5 \times 10^{-3}$-WD

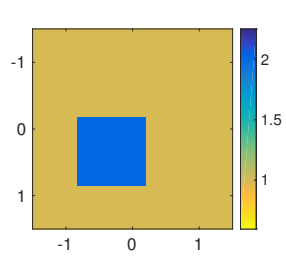


(k) Jointly-sparse model-T-MSBL-WD

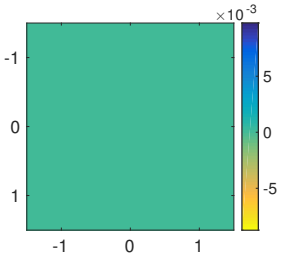


(l) Jointly-sparse model-T-MSBL-WD

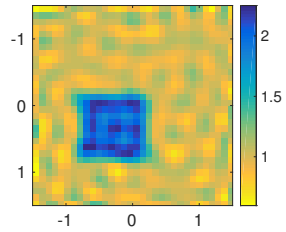
Figure 4.1.: Scatterer 2 (Fig. 3.4b): Retrieval of permittivity (ε_r, left) and conductivity (σ, right) both in the spatial domain and the wavelet domain (WD) by using sparsity with 10 dB noise data.



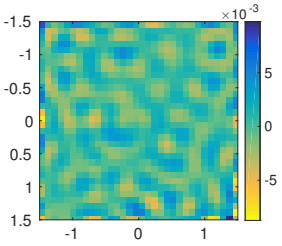
(a) Exact



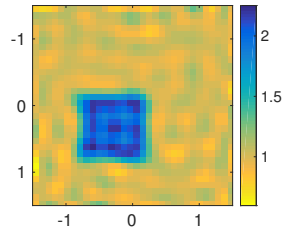
(b) Exact



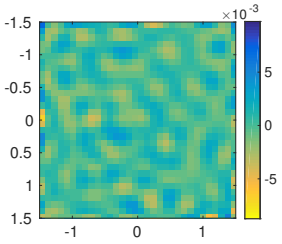
(c) Soft Shrinkage Alg. $\alpha = 1 \times 10^{-6}$



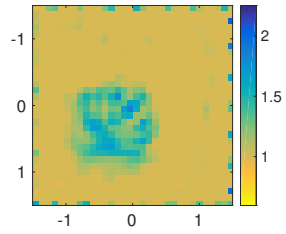
(d) Soft Shrinkage Alg. $\alpha = 1 \times 10^{-6}$



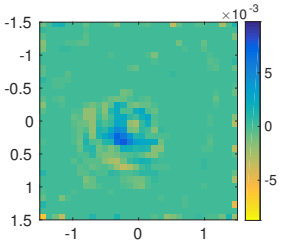
(e) Soft Shrinkage Alg. $\alpha = 2.5 \times 10^{-2}$



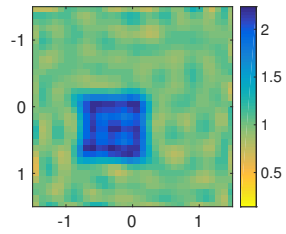
(f) Soft Shrinkage Alg. $\alpha = 2.5 \times 10^{-2}$



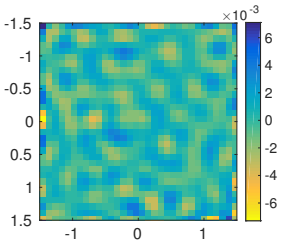
(g) Jointly-sparse model-YALL1



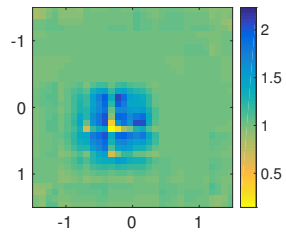
(h) Jointly-sparse model-YALL1



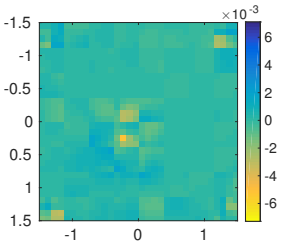
(i) Soft Shrinkage Alg. $\alpha = 2.5 \times 10^{-2}$-WD



(j) Soft Shrinkage Alg. $\alpha = 2.5 \times 10^{-2}$ WD



(k) Jointly-sparse model-YALL1-WD



(l) Jointly-sparse model-YALL1-WD

Figure 4.2.: Scatterer 6 (Fig.3.4f): Retrieval of permittivity (ε_r, left) and conductivity (σ, right) both in the spatial domain and the wavelet domain (WD) by using sparsity with 10 dB noise data.

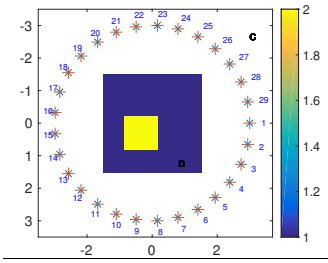
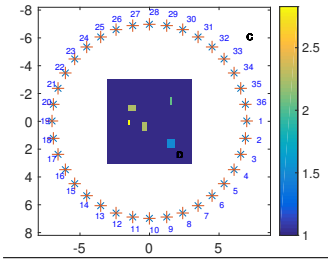
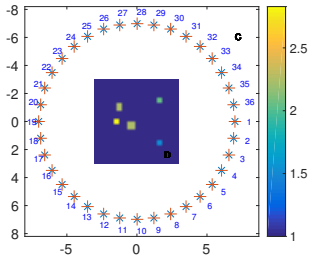
Scatterer	Test	Time (s)	χ^{err}	ε^{err}	σ^{err}
	SpaRSA	5.4	0.2952	0.2951	
	T-MSBL	117.7	0.3190	0.3151	
	YALL1	24.1	0.2092	0.2047	
	Soft Shrinkage Alg.	197.0	**0.1304**	**0.1068**	
	YALL1 (B.A)	32.1	0.2633	0.2570	
	SpaRSA	6.3	0.1975	0.1056	0.9194
	T-MSBL	**13.45**	**0.1169**	**0.0686**	**0.5224**
	YALL1	25.3	0.1385	0.0733	0.6472
	Soft Shrinkage Alg.	145.7	0.1432	0.0872	0.6277
	T-MSBL (B.A)	18.15	0.3902	0.1290	2.0191
	SpaRSA	4.5	0.2124	0.1246	0.9143
	T-MSBL	18.48	**0.1571**	**0.0900**	0.6837
	YALL1	24.73	0.1715	0.0980	0.7474
	Soft Shrinkage Alg.	179.8	0.1590	0.0990	**0.6624**
	T-MSBL (B.A)	23.3	0.4001	0.1389	2.3454

Table 4.1.: Error and average simulation time in seconds with 10 dB

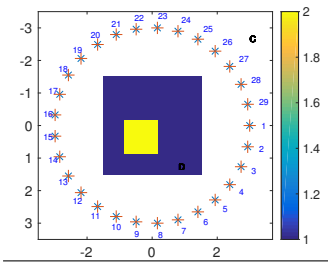
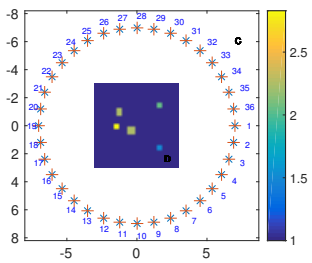
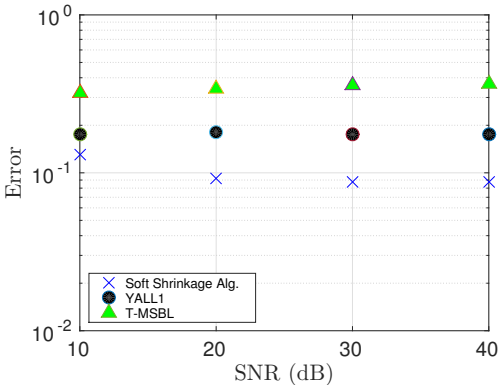
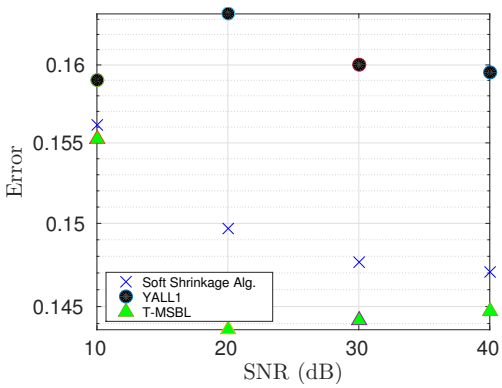
Scatterer	Test	Time (s)	χ^{err}	ε^{err}	σ^{err}
	SpaRSA	7.4	0.2513	0.2495	
	T-MSBL	140.2	0.5160	0.3643	
	YALL1	23.3	0.2014	0.1994	
	Soft Shrinkage Alg.	132.2	**0.1358**	**0.1104**	
	SpaRSA	6.3	0.2260	0.1345	0.9654
	T-MSBL	34.19	0.2051	0.1251	0.8646
	YALL1	24.8	0.2118	0.1220	0.9192
	Soft Shrinkage Alg.	123.7	**0.1595**	**0.0997**	**0.6630**

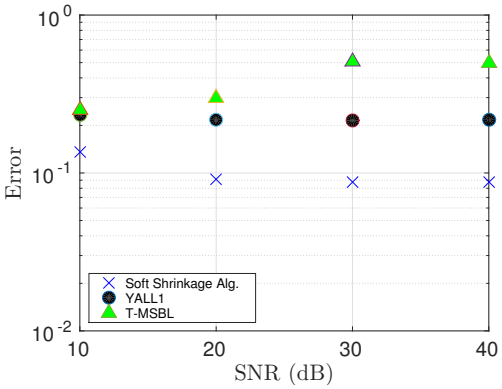
Table 4.2.: Error and average simulation time in seconds with 10 dB-Wavelet Domain



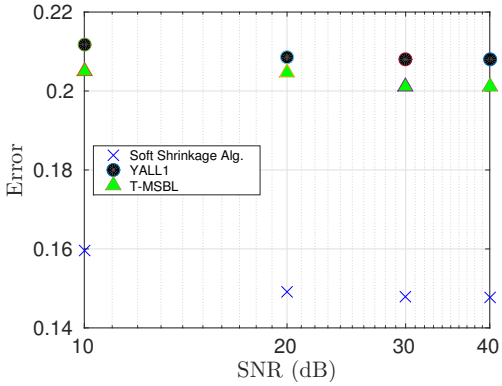
(a) Scatterer 6 (Fig. 3.4f)



(b) Scatterer 7 (Fig. 3.4g)



(c) Scatterer 6 (Fig. 3.4f)- Wavelet Domain



(d) Scatterer 7 (Fig. 3.4g)- Wavelet Domain

Figure 4.3.: Error in $\boldsymbol{\chi}$ ($\boldsymbol{\chi}^{\mathrm{err}}$) against varying SNR both in the spatial and the wavelet domain

4.2.4. Two-Step Inversion Method: Preliminary Results in 3D

Let us consider a complex, multi-layer structure, where each of the n layers is assumed to be non-magnetic ($\mu = \mu_0$) homogeneous uniaxial. (This type of structure usually results from a large-scale (homogenized) view of fibered planar laminates as in aeronautics.) In the local coordinate system (material frame), each layer is characterized by a diagonal complex permittivity tensor

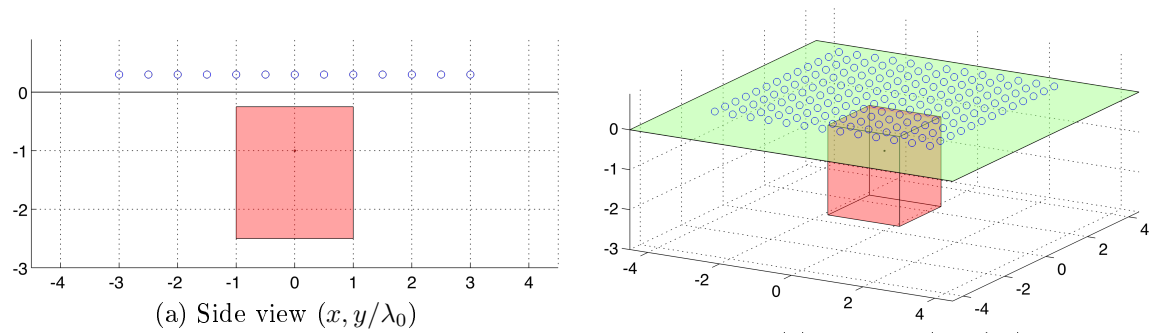
$$\bar{\bar{\epsilon}}_e^{(n)} = \begin{bmatrix} \epsilon_{11}^{(n)} & 0 & 0 \\ 0 & \epsilon_{22}^{(n)} & 0 \\ 0 & 0 & \epsilon_{22}^{(n)} \end{bmatrix} \qquad \bar{\bar{\Xi}}(\theta_n) = \begin{bmatrix} \cos\theta_n & \sin\theta_n & 0 \\ -\sin\theta_n & \cos\theta_n & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.7}$$

One is able to carry the local coordinate system to the global Cartesian one where θ_n is the rotation Euler angle. The electromagnetic response of such a structure which satisfies the radiation condition at infinity (the Sommerfeld condition) and in a complex multi-layer continuity conditions at each interface is computed as

$$\mathbf{E}_j^{\text{diff}}(\mathbf{r}) = \mathrm{i}\omega\mu_0 \int_V \bar{\bar{G}}^{\text{ee}}(\mathbf{r},\mathbf{r}')\cdot\bar{\bar{\chi}}(\mathbf{r}')\cdot\mathbf{E}_j^{\text{tot}}(\mathbf{r}')\,\mathrm{d}\mathbf{r}', \text{ with } \bar{\bar{\chi}}(\mathbf{r}) = -\mathrm{i}\omega\epsilon_0\bar{\bar{\Xi}}^{-1}(\theta_n)\cdot\left(\bar{\bar{\epsilon}}_i - \bar{\bar{\epsilon}}_e^{(n)}\right)\cdot\bar{\bar{\Xi}}(\theta_n) \tag{4.8}$$

$$\bar{\bar{\chi}}(\mathbf{r})\cdot\mathbf{E}_j^{\text{inc}}(\mathbf{r}) = \frac{\mathbf{J}_j(\mathbf{r})}{-i\omega\epsilon_0} - \bar{\bar{\chi}}(\mathbf{r})\cdot i\omega\mu_0\int_V \bar{\bar{G}}^{\text{ee}}(\mathbf{r},\mathbf{r}')\cdot\mathbf{J}_j(\mathbf{r}')\,\mathrm{d}\,\mathbf{r}' \tag{4.9}$$

where $\bar{\bar{G}}^{\text{ee}}(\mathbf{r},\mathbf{r}')$ is the electric-electric dyadic Green's function, $\bar{\bar{\epsilon}}_i$ the background permittivity tensor, $\bar{\bar{\epsilon}}_e$ the permittivity tensor of an inclusion of volume V within the backgroundand volume V. $\mathbf{E}_j^{\text{diff}}(\mathbf{r})$, $\mathbf{E}_j^{\text{tot}}(\mathbf{r})$ and $\mathbf{E}_j^{\text{inc}}(\mathbf{r})$ are the scattered, total and incident fields respectively due to the j^{th} source. For theoretical and numerical details one should refer to [80], [101] and [102].



(a) Side view $(x, y/\lambda_0)$

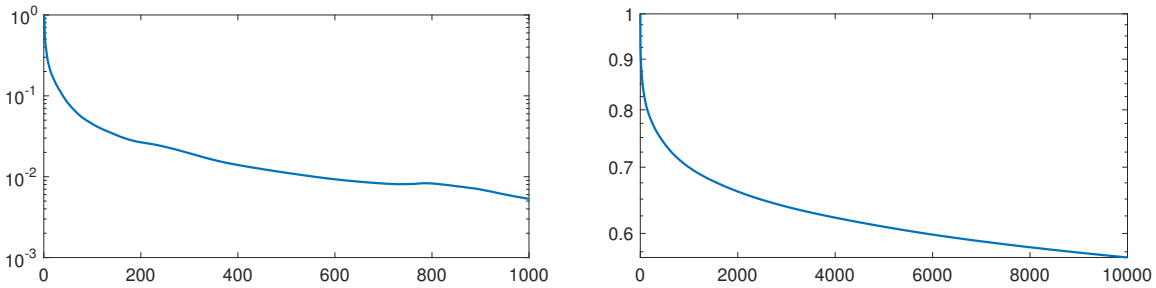
(b) 3D view $(x, y/\lambda_0)$

Figure 4.4.: Configuration for inverse imaging testing.

Here is the description of the configuration on which the two-step inversion is applied:

- Frequency $f = 6\,\text{GHz}$; $\lambda_0 = 5\,\text{cm}$; $\lambda_1 = 2.1\,\text{cm}$
- Acquisition:
  - Surface area of $L_x \times L_y$ at height $L_{\text{acquisition}}$

    $L_x = L_y = 6\lambda_1$, $L_{\text{acquisition}} = 0.3\lambda_0$
  - Number of antennas: $N_x \times N_y$; $N_x = N_y = 13$
  - Each antenna: 3 orthogonal unit dipoles along x, y and z



(a) Without noise (b) With 30 dB noise data

Figure 4.5.: Convergence of the cost function, vertical axis corresponds to number of iterations and horizontal axis corresponds to error on cost function (two-step method applied via YALL1).

- Fiber-glass composite $\bar{\bar{\epsilon}}_1 = \epsilon_0 \text{diag}\left[5.46 + \text{i}2.29, 5.21 + \text{i}2.08, 5.21 + \text{i}2.08\right]$ $\theta_1 = 0°$
- Two inclusions of complex permittivity $\epsilon_\Omega = (6, 0)$, $\bar{\bar{\epsilon}}_{\text{inclusion}} = \bar{\bar{I}}\epsilon_0\epsilon_\Omega$
  - Size $(0.2 \times 0.2 \times 0.2)\,\lambda_1$, centered at $(1.4, 0.7, 0.7) \times \lambda_1$
  - Size $(0.2 \times 0.2 \times 0.2)\,\lambda_1$, centered at $(0.5, 1.4, 1.4) \times \lambda_1$
- Region Of Interest:
  - Size: $l_x \times l_y \times l_z$ with $l_x = l_y = 2\lambda_1$ and $l_z = 2\lambda_1$
  - Discretization: $n_x \times n_y \times n_z$, $n_x = n_y = n_z = 10$
  - Depth of top of ROI $z = 0.25\lambda_1$

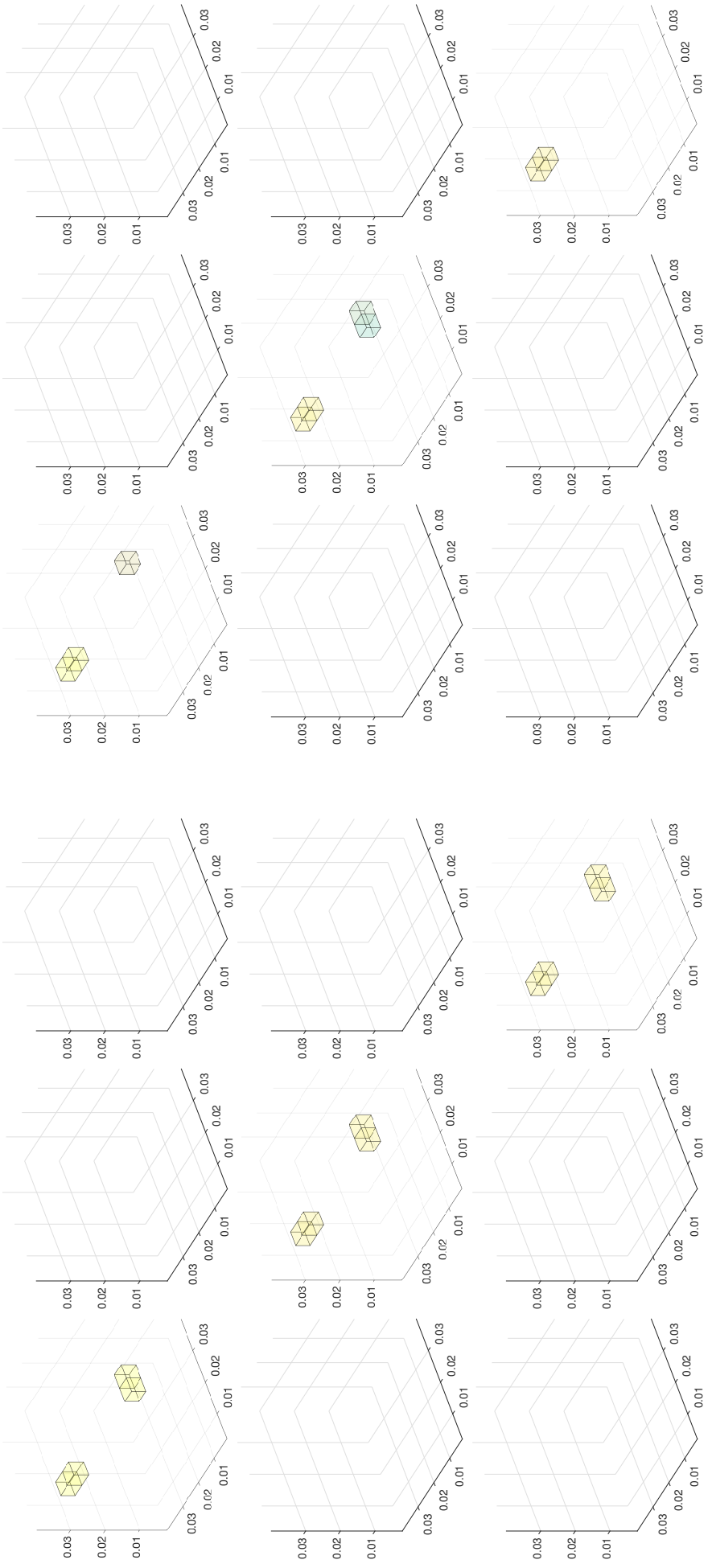
In this section, the unknown scatterer is not too large and/or its contrast not too high. The preliminary results obtained using two-step inversion method already show some good results. In [103] a Bayesian Compressive Sensing solver, is used whereas, in this work, an algorithm exploiting a joint sparsity regularization is proposed. Let us notice that results using the one-shot MUSIC retrieval method and the iterative Subspace Optimization Method in the 3D anisotropic case are put in perspective in [104], this being behind the scope of our discussion here however.

From Fig. 4.6 we can see that the two-step inversion method applied through YALL1 gives us a good localization and estimation of the contrast when there is no noise and it maintains a good convergence as shown in Fig. 4.5. However, the method does not give us good results when we have noise in our data as in Fig. 4.7.

Similarly, the two-step inversion method applied through T-MSBL gives us a good reconstruction for the real and imaginary parts of the contrast (see Fig. 4.8 and Fig. 4.9)

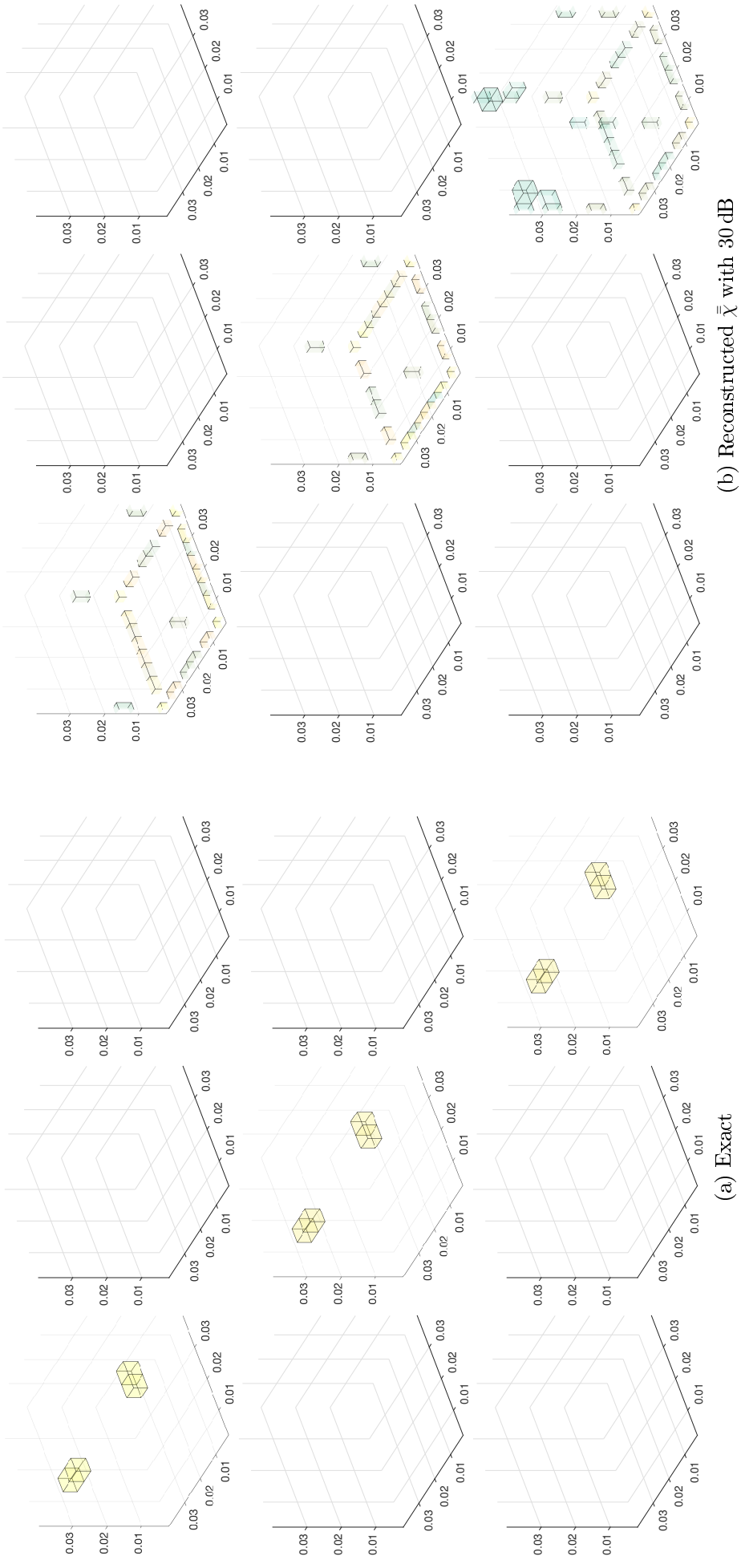
when there is no noise. However, it can still keep its good reconstruction quality with the addition of the noise to the data as can be seen Fig. 4.10 and Fig. 4.11. We have used 3000 iterations while running our simulations. On the other hand, this method is sensitive to initialization.

In 3D, nonlinearity and ill-posedness become even more severe than 2D applications. This reminds us of the importance of the enforcement of the sparsity constraint. As it can be observed, the two-step inversion gives better results when it is applied through T-MSBL than the ones of YALL1. This might be due to the stronger correlation that applied through T-MSBL to the coefficients of each row of the unknown contrast. This point as well as other open questions will be considered in the Conclusion chapter.



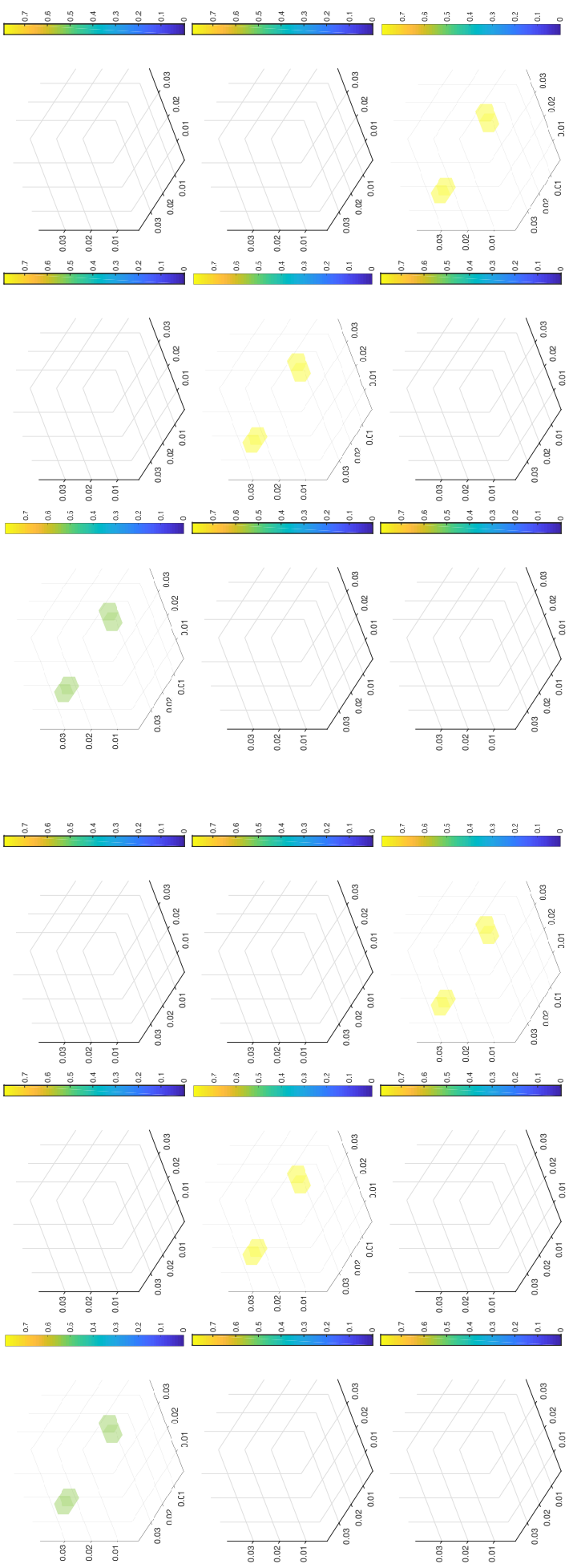
(a) Exact (b) Reconstructed $\bar{\bar{\chi}}$ without noise

Figure 4.6.: Reconstruction of contrast ($\bar{\bar{\chi}}$) without noise through YALL1



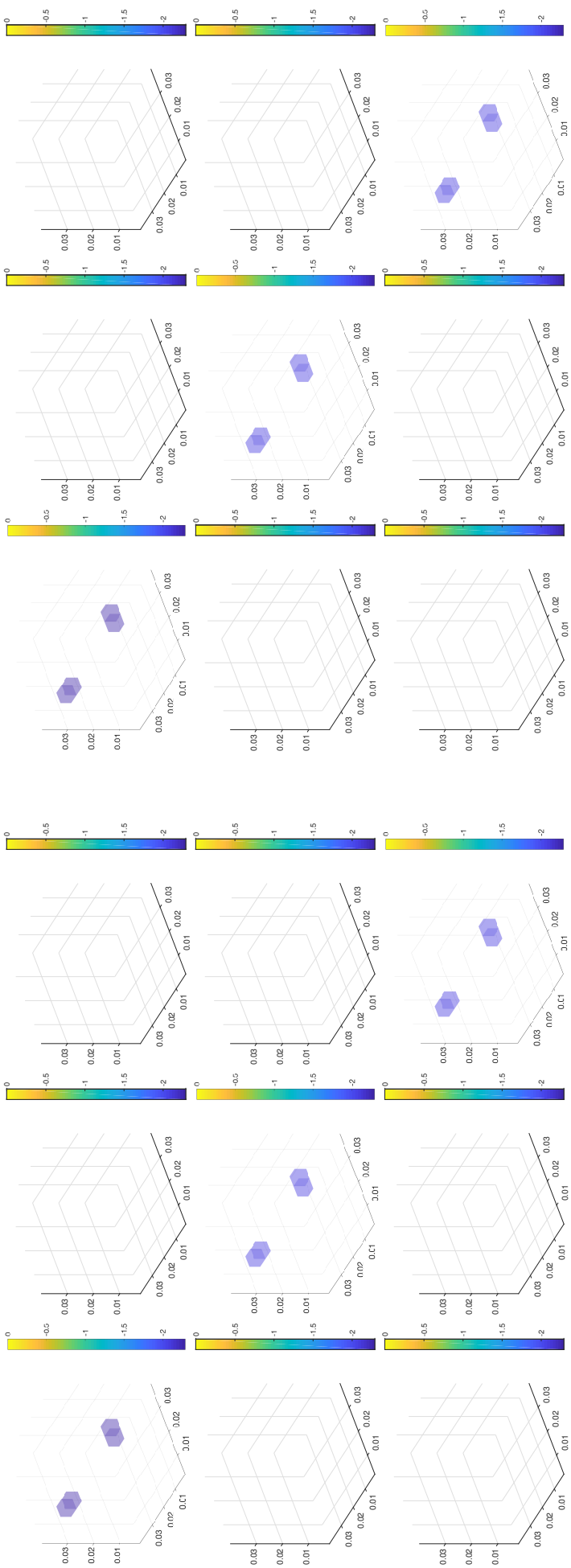
(a) Exact (b) Reconstructed $\bar{\bar{\chi}}$ with 30 dB

Figure 4.7.: Reconstruction of contrast ($\bar{\bar{\chi}}$) with 30 dB noise through YALL1.



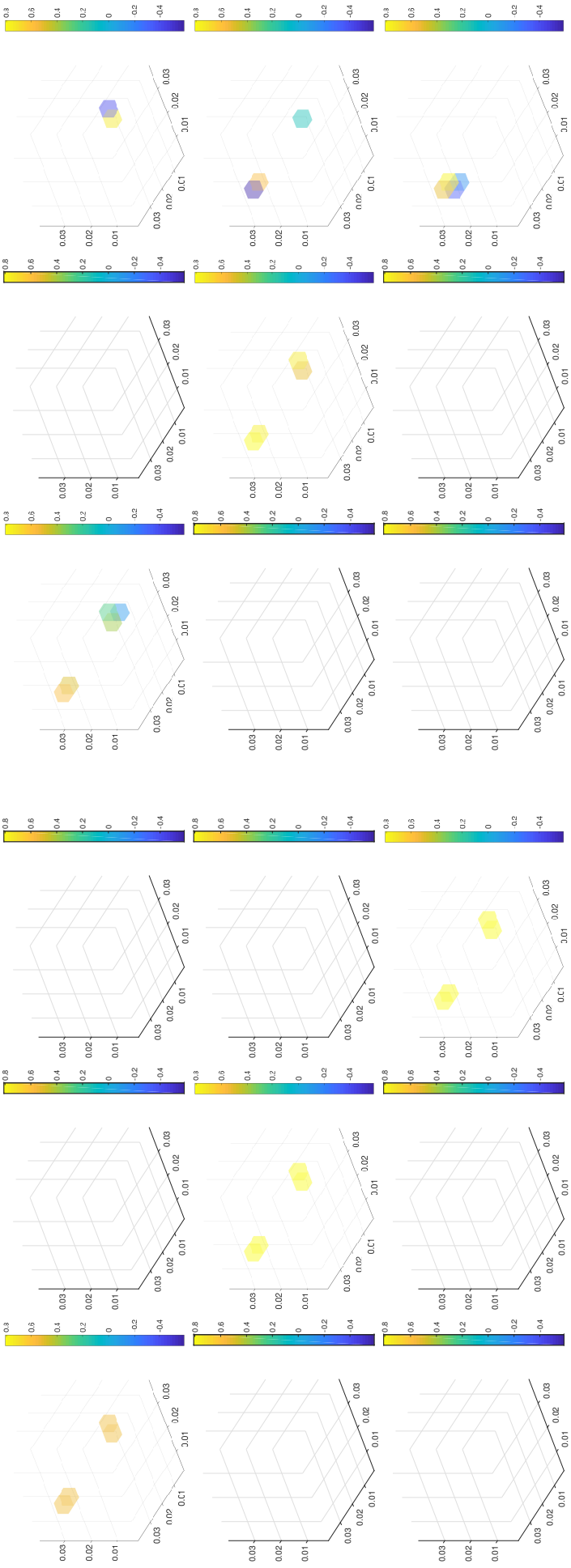
(a) Exact-Re$\{\bar{\bar{\chi}}\}$ (b) Reconstructed Re$\{\bar{\bar{\chi}}\}$ without noise

Figure 4.8.: Reconstruction of real part of contrast (Re$\{\bar{\bar{\chi}}\}$) without noise through T-MSBL.



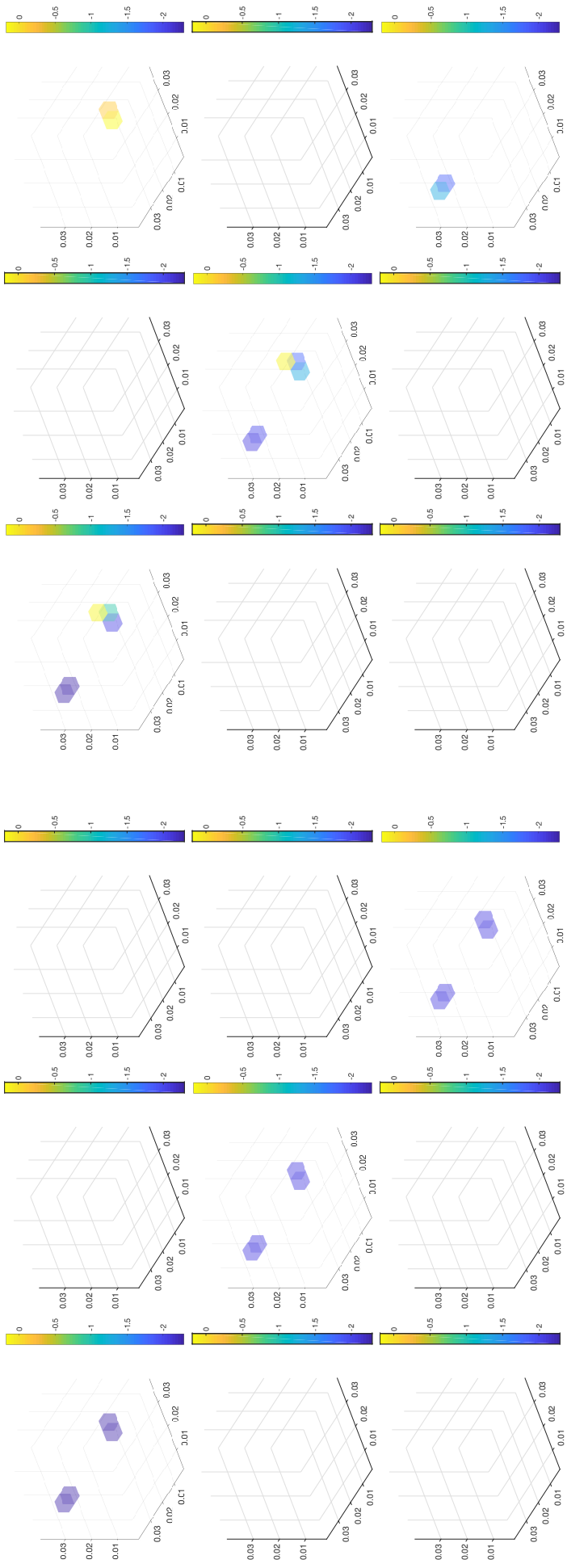
(a) Exact-($\mathrm{Im}\{\bar{\bar{\chi}}\}$) (b) Reconstructed ($\mathrm{Im}\{\bar{\bar{\chi}}\}$) without noise

Figure 4.9.: Reconstruction of imaginary part of contrast ($\mathrm{Im}\{\bar{\bar{\chi}}\}$) without noise through T-MSBL.



(a) Exact-Re$\{\bar{\bar{\chi}}\}$ (b) Reconstructed Re$\{\bar{\bar{\chi}}\}$ with 10 dB noise

Figure 4.10.: Reconstruction of real part of contrast (Re$\{\bar{\bar{\chi}}\}$) with 10 dB noise through T-MSBL.



(a) Exact-Im$\{\bar{\bar{\chi}}\}$ (b) Reconstructed Im$\{\bar{\bar{\chi}}\}$ with 10 dB noise

Figure 4.11.: Reconstruction of imaginary part of contrast (Im$\{\bar{\bar{\chi}}\}$) with 10 dB noise through T-MSBL.

4.3. Conclusion

In this chapter, we presented a two-step inverse process which allows sparse recovery of the unknown (complex) dielectric profiles of scatterers. The proposed approach is correlated with joint sparsity which gives multiple sparse solutions that share a common nonzero support. The principal interest of such a method is that it accurately reconstructs the unknown scatterers without linear approximation and presents an efficient recovery algorithm of sparse scatterers by reducing possible ambiguities on the scatterer sparsity deduced from the null values of the equivalent current [4, 105].

Thereafter, we compared the method proposed in this chapter with the approach that directly obtains the contrast through enforcement of sparsity by soft shrinkage thresholding. Both approaches produce sharp and good reconstruction of dielectric profiles in sparse domains and keep their convergence during the reconstruction.

On the other hand, when we have a scatterer which is sparser with respect to pixel basis then two-step inversion method through T-MSBL gives better results in the spatial domain than the wavelet domain. This might be due to the correlation of the nonzero rows of the unknown contrast that we are interested in. When we have a scatterer which is not that sparse in spatial domain then there is not a big difference for both two-step inversion and soft-shrinkage approaches either in spatial or wavelet domains.

Moreover, a three-dimensional inversion strategy is introduced for the detection of scatterer in uniaxially anisotropic layered media having principal axes with arbitrary orientation. The two-step method is adopted for the solution of the imaging problem. A set of preliminary numerical results is reported to assess the accuracy of the proposed method even if more studies have to be done in order to evaluate the limitations of the proposed method.

Recently, several computational advances have been made in the nonconvex sparse regularization since its performance is better than the one of the convex sparse regularization. While there exist many algorithms for solving the nonconvex sparse regularized models, it is still a challenging problem to obtain the global optimal solution efficiently. In addition, the behavior of a local solution is hard to analyze and, more importantly, structural information of the solution is also hard to be incorporated into these algorithms [47].

5. Summary, Conclusions and Future Work

In this last chapter, the work that has been done and presented in the previous chapters is summarized. The main points, contributions, results and inferences are summarized to provide a global view and highlight the goal of the project. The conclusions are subsequently followed by a list of suggestions and ideas for future work. These perspectives aspire to complement the work carried out and presented here and, if possible, to bring it closer to real life applications.

5.1. Summary and Conclusions

In this dissertation, we have considered the non-linear microwave imaging problem by transforming it into a problem of sparse signal representation using dictionaries. This is a very attractive way of looking at retrieving profiles of unknown scatterers in a region of interest because when the scatterers can be well-modeled with an appropriate basis, and the number of their unknown coefficients is small, then the true underlying spatial spectrum is sparse. The problem of signal representation in bases is an ill-posed non-linear inverse problem, and as such, it requires regularization to have unique well-behaved solutions. We are interested in sparse signal representations, so the regularization has to enforce sparsity.

To enforce sparsity we utilized l_p penalties with $p \leq 1$. There is an important distinction between l_1 penalties ($p = 1$), and penalties with $p < 1$. For the l_1 case, the penalty leads to convex optimization problems, whereas for $p < 1$, the associated optimization is nonconvex. For the sake of simplicity in terms of computational complexity we focused on optimization involving l_1 penalties for which we used an algorithm that is a special case of gradient descent. The algorithm has the important benefit of allowing efficient global solutions by adapting different properties.

In this work, thresholded nonlinear Landweber iterations (special case of gradient descent) are used to solve the sparse minimization problem constructed directly from the nonlinear scattering equations. It is adapted from the classical iterative soft shrinkage algorithm, and its main ingredients include a Sobolev smoothing of the estimated gradients, a soft shrinkage iteration, and an adaptive step size selection based on the Barzilai and Borwein rule. The resulting scheme requires the user to set only one simulation parameter before the execution, which significantly simplifies the application of the method to different problems. Additionally, since the regularization is applied directly to the nonlinear problem without any linearization approximation, the scheme can be applied to higher contrast level of scatterers.

Furthermore, a projected steepest descent algorithm, which increases convergence rate of the nonlinear Landweber iterations, is used for reducing potential high computational cost. The projection operator replaces the thresholding function and enforces the sparsity constraint. Indeed, numerical results demonstrate that the resulting projection scheme is very efficient in recovering permittivity profiles.

We adopted as already said the Barzilai-Borwein direction as a step size selection, and the numerical results show the significant development in efficiency of the gradient method. Being computationally efficient and needing low memory requirement makes this scheme interesting to solve large-scale optimization problems.

On the other hand, the decoding process requires finding a sparse solution of an underdetermined linear system. What makes such a scheme work is sparsity; i.e., the original signal must have a sparse or nearly sparse representation under some known basis. Following this fact, we used the wavelet transform to retrieve the unknown scatterer profile by solving the sparse Tikhonov minimization problem constructed directly from the nonlinear scattering equations. This inversion is usually more robust than simultaneously inverting all the pixels in the spatial domain because the number of unknowns is less and the non-uniqueness is thereby reduced. We explained how to extract the difference of information between successive resolutions and thus define a new (complete) representation called the wavelet representation. This representation is computed by decomposing the original signal using a wavelet orthonormal basis.

Let us remind that the scatterers are sparse with respect to their expansion basis. In this work, we have studied different profiles of scatterers which are sparse in pixel basis and/or sparse in Haar or in Daubechies basis. The integration of wavelet bases within sparseness-regularized formulations for microwave imaging has been done. Then, the introduction of a generalized regularized imaging strategy applicable to a very wide

set of scattering scenarios depending on the chosen wavelet family (i.e., the available *a-priori* information about the scattering scenario) has been worked out while keeping the advantages of pixel-basis optimization techniques.

The wavelet transform is localized in space, and we can first invert only the lower level wavelet coefficients to get a rough image. This inversion is usually more robust than simultaneously inverting all the pixels in the spatial domain because the number of unknowns is less and the non-uniqueness is thereby reduced. After this, we can improve the inversion results by adding wavelet coefficients at higher levels. On the other hand, when we increase the decomposition level in order to have the optimization on sparser coefficients, there is no such large difference both in accuracy and in performance for the examples that we have studied.

The analysis in Chapter 3 shows that a soft shrinkage algorithm is sufficiently robust to work even when the *a-priori* knowledge about the scatterer is only approximate, that is, the scatterer is not exactly sparse within the domain of the considered basis functions. Moreover, the reconstruction accuracy turns out to be acceptable also when higher contrast and lower SNRs are.

It has been analysed that the discrete wavelet transform shows a better performance and effectiveness in the reconstructions compared to the stationary wavelet transform in some cases. However, the error on retrieving various profiles is quite small between each transformation. One can also conclude that the accuracy, the convergence rate, and the robustness vs. noise of the proposed approaches are not dependent only on the problem setup, but also on the choice of the wavelet basis for the contrast expansion, which is of course strongly related to the available *a-priori* information.

The third direction of the work carried out in the thesis is exploiting the joint-sparsity benefits by applying a two-step method to solve our non-linear problem of concern. The two-step inversion method uses the contrast-source formulation of the scattering equations, which allows for straightforward application of the sparsity constraint to the solution. The nonlinear system of equations is solved using an iterative scheme, which calls for the solution of a linear system at every iteration.

The two-step inversion method is applied through available YALL1 and T-MSBL algorithms both in spatial and wavelet domains. Both methods exploit joint-sparsity which reduces possible ambiguities on the scatterer sparsity deduced from the null values of the equivalent current. This approach consumes less time and ensures better imaging quality compared to an iterative method with soft thresholding in the reconstruction of sparse scatterers. However, even if the two-step method is faster than the usual soft shrinkage

algorithm, the reconstruction of the scatterers is better with soft shrinkage algorithm in wavelet domain. Above of all this, addition of projection to our proposed methods always gives us better approximation for the unknown.

5.2. Future Work

A very important issue in the framework of this work is the choice of the regularization parameters, α in the l_1 formulations and β and δ in smooth gradient formulations. It is worthwhile to continue the investigation of methods for regularization parameter selection from other fields since being able to identify the most accurate parameter is challenging. In our application we have not be able to use 'L-curve'. There are other methods such cross-validation, and universal and min-max regularization parameter selection rules as well. The viability of these methods for our problem is on discussion for our specific case of application. However, much more work has to be made to get insights into how to select the regularization parameter for our problem, or to dismiss these methods as inappropriate for our problem.

In the thesis we exclusively used l_1 regularization for enforcing sparsity. However, regularization that favors sparsity is not limited to this. Many other forms exist, such as Huber regularization, entropy-based regularization [72], and the other ones stated in Introduction section. Also, an analysis of the specific features that are necessary for the regularizing term to favor sparsity would provide much insight into the selection of a particular functional. Such analysis has been previously done at some level [26, 96], but deeper understanding can be gained by putting the analysis on firm theoretical grounds and considering much wider sets of regularizing functionals.

All of the schemes proposed in this work, namely soft shrinkage algorithm in spatial and wavelet domain, and two-step inversion method, can be extended to imaging of domains residing in layered media. This extension calls for replacing the free-space Green function with the planary or cylindrically layered medium Green functions. The resulting inversion tools have applications in subsurface and borehole imaging, as an example.

The efficiency of the sparsity-constrained regularization can be increased by making use of the fact that scatterers are represented as block sparse vectors after discretization [53, 89]. This *a-priori* information about the distribution of non-zero elements can be incorporated within the regularization scheme to increase the robustness of the solution and the quality of the recovered images.

According to some studies such as [41, 91, 92], wavelet decomposition is an efficient transformation while retrieving scatterers which are sparse with respect to their expansion functions. In our specific cases, the proposed methods do not exhibit a remarkable difference between wavelet and pixel basis functions. This might be due to the fact that the scattering matrix, which is obtained by sampling the Green function between investigation domain and receiver locations, does not satisfy the restricted isometry property [106, 107]. Following the work done in [107], we have adapted our scattered field by adding a constant parameter to its diagonal entries. However, we have not been able to get better results so that it can be a good idea to investigate the problem within this perspective in the future.

Although Haar and Daubechies wavelet basis functions beside pixel basis function are studied in this work to represent the unknown model in Sec. 3.3, the optimal choice of the basis set is an application-dependent open problem. The recovery problem solution is the sparsest one with respect to a basis function and the sparsest reconstruction (fewest non-zero coefficients) has to be picked up. Therefore, the available *a-priori* (physical) solution information plays a fundamental role in the selection of the most suitable basis for a given application. Within this respect, more research can be carried out in order to deduce which basis function can be the most appropriate to specific applicative domains in a more practical and faster manner. Other options such as Curvelets [40] are of interest to a future research also. Future works can be aimed at extending the validation of the proposed inversion scheme to other wavelet families more suitable for dealing with different applicative domains such as nondestructive testing/evaluation.

Furthermore, more investigations can be performed for 3D electromagnetic imaging, where the nonlinearity and ill-posedness become even more severe. For instance, we have already started looking into this direction of study in which reconstruction of 3D defects affecting an anisotropic laminate is dealt with the two-step inversion method as stated in Sec. 4.2.4. We can obtain good localization and estimation of the contrast by applying joint sparsity through T-MSBL. However, this method is applicable for small scatterers so far and more work needs to be performed for larger scatterers. This method can also be studied for the localization and estimation of the contrast of a small defect in a complex anisotropic structure. Improvements with respect to the convergence speed and radius of the convergence of the method should be investigated as well.

Bibliography

[1] S. Mustafa, B. Mohammed, and A. Abbosh, "Novel preprocessing techniques for accurate microwave imaging of human brain," *IEEE Antennas Wireless Propagat. Lett.*, vol. 12, pp. 460–463, 2013.

[2] T. Henriksson, N. Joachimowicz, C. Conessa, and J.-C. Bolomey, "Quantitative microwave imaging for breast cancer detection using a planar 2.45 GHz system," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 10, pp. 2691–2699, 2010.

[3] I. T. Rekanos and T. D. Tsiboukis, "A finite element based technique for microwave imaging of two-dimensional objects," in *Proceedings of the 16th Instrumentation and Measurement Technology Conference, 1999. IMTC/99.*, vol. 3, pp. 1576–1581, 1999.

[4] L. Poli, G. Oliveri, and A. Massa, "Microwave imaging within the first-order Born approximation by means of the contrast-field Bayesian compressive sensing," *IEEE Trans. Antennas Propagat.*, vol. 60, no. 6, pp. 2865–2879, 2012.

[5] M. Pastorino, *Microwave Imaging*, vol. 208. John Wiley & Sons, 2010.

[6] F. Di Benedetto, C. Estatico, J. G. Nagy, and M. Pastorino, "Numerical linear algebra for nonlinear microwave imaging," *Electron. Trans. Numer. Anal.*, vol. 33, pp. 105–125, 2009.

[7] R. Ramlau and G. Teschke, "A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints," *Numer. Math.*, vol. 104, no. 2, pp. 177–203, 2006.

[8] M. Hanke, A. Neubauer, and O. Scherzer, "A convergence analysis of the Landweber iteration for nonlinear ill-posed problems," *Numer. Math.*, vol. 72, no. 1, pp. 21–37, 1995.

[9] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, vol. 93. Springer Science & Business Media, 2012.

[10] A. J. Devaney, *Mathematical Foundations of Imaging, Tomography and Wavefield inversion*. Cambridge University Press, 2012.

[11] R. Zorgati, B. Duchane, D. Lesselier, and F. Pons, "Eddy current testing of anomalies in conductive materials. i. qualitative imaging via diffraction tomography techniques," *IEEE Trans. Magn.*, vol. 27, no. 6, pp. 4416–4437, 1991.

[12] S. D. Rajan and G. V. Frisk, "A comparison between the Born and Rytov approximations for the inverse backscattering problem," *Geophysics*, vol. 54, no. 7, pp. 864–871, 1989.

[13] G. Bozza and M. Pastorino, "An inexact Newton-based approach to microwave imaging within the contrast source formulation," *IEEE Trans. Antennas Propagat.*, vol. 57, no. 4, pp. 1122–1132, 2009.

[14] P. Abubakar, "Contrast source inversion method: State of art," *Prog. Electromagn. Res.*, vol. 34, pp. 189–218, 2001.

[15] Y. Li and W. Yang, "Image reconstruction by nonlinear Landweber iteration for complicated distributions," *Meas. Sci. Technol.*, vol. 19, no. 9, p. 094014, 2008.

[16] F. Hettlich, "The Landweber iteration applied to inverse conductive scattering problems," *Inverse Prob.*, vol. 14, no. 4, p. 931, 1998.

[17] A. Franchois and C. Pichot, "Microwave imaging-complex permittivity reconstruction with a Levenberg-Marquardt method," *IEEE Trans. Antennas Propagat.*, vol. 45, no. 2, pp. 203–215, 1997.

[18] 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, 1990.

[19] Z. Q. Zhang and Q. H. Liu, "Two nonlinear inverse methods for electromagnetic induction measurements," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 6, pp. 1331–1339, 2001.

[20] C. Estatico, M. Pastorino, and A. Randazzo, "An inexact-Newton method for short-range microwave imaging within the second-order Born approximation," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 11, pp. 2593–2605, 2005.

[21] Y. Wang and W. C. Chew, "An iterative solution of the two-dimensional electromagnetic inverse scattering problem," *Int. J. Imaging Syst. Technol.*, vol. 1, no. 1, pp. 100–108, 1989.

[22] N. Zaiping, Y. Feng, Z. Yanwen, and Z. Yerong, "Variational Born iteration method and its applications to hybrid inversion," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 4, pp. 1709–1715, 2000.

[23] A. Tveito, H. P. Langtangen, B. F. Nielsen, and X. Cai, "Parameter estimation and inverse problems," in *Elements of Scientific Computing*, pp. 411–421, Springer, 2010.

[24] M. Fornasier, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, vol. 9. Walter de Gruyter, 2010.

[25] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[26] D. L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern, "Maximum entropy and the nearly black object," *J. Royal Statistical Soc. Ser. B (Methodological)*, pp. 41–81, 1992.

[27] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.

[28] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on pure and applied mathematics*, vol. 57, no. 11, pp. 1413–1457, 2003.

[29] M. Grasmair, M. Haltmeier, and O. Scherzer, "Sparse regularization with lp penalty term," *Inverse Prob.*, vol. 24, no. 5, p. 055020, 2008.

[30] B. Jin and P. Maass, "Sparsity regularization for parameter identification problems," *Inverse Prob.*, vol. 28, no. 12, p. 123001, 2012.

[31] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "An architecture for compressive imaging," in *International Conference in Image Processing, 2006 IEEE*, pp. 1273–1276, IEEE, 2006.

[32] M. Gehre, T. Kluth, A. Lipponen, B. Jin, A. Seppänen, J. P. Kaipio, and P. Maass, "Sparsity reconstruction in electrical impedance tomography: an experimental evaluation," *J. Comput. Appl. Math.*, vol. 236, no. 8, pp. 2126–2136, 2012.

[33] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.

[34] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[35] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.

[36] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[37] I. Daubechies, *Ten Lectures on Wavelets*. SIAM, 1992.

[38] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 1, pp. 59–73, 2011.

[39] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 7, pp. 674–693, 1989.

[40] E. J. Candes and D. L. Donoho, "Curvelets: A surprisingly effective nonadaptive representation for objects with edges," tech. rep., DTIC Document, 2000.

[41] M. Li, O. Semerci, and A. Abubakar, "A contrast source inversion method in the wavelet domain," *Inverse Prob.*, vol. 29, no. 2, p. 025015, 2013.

[42] T. Isernia, V. Pascazio, and R. Pierri, "On the local minima in a tomographic imaging technique," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 7, pp. 1596–1607, 2001.

[43] O. Bucci and T. Isernia, "Electromagnetic inverse scattering: Retrievable information and measurement strategies," *Radio Sci.*, vol. 32, no. 6, pp. 2123–2137, 1997.

[44] S. Caorsi, M. Donelli, D. Franceschini, and A. Massa, "A new methodology based on an iterative multiscaling for microwave imaging," *IEEE Trans. Microwave Theory Tech.*, vol. 51, no. 4, pp. 1162–1173, 2003.

[45] O. M. Bucci, L. Crocco, T. Isernia, and V. Pascazio, "Subsurface inverse scattering problems: quantifying, qualifying, and achieving the available information," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 11, pp. 2527–2538, 2001.

[46] W. Deng, W. Yin, and Y. Zhang, "Group sparse optimization by alternating direction method," in *SPIE Optical Engineering+ Applications*, pp. 88580R–88580R-15, International Society for Optics and Photonics, 2013.

[47] Y. Fan, Y. Wang, and T. Huang, "Enhanced joint sparsity via iterative support detection," *arXiv preprint arXiv:1412.2675*, 2014.

[48] R. C. Gonzalez and R. E. Woods, "Processing," 2002.

[49] 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, 2015.

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

[51] M. Bevacqua, L. Crocco, L. Di Donato, and T. Isernia, "Microwave imaging of nonweak targets via compressive sensing and virtual experiments," *IEEE Antennas Wireless Propagat. Lett.*, vol. 14, pp. 1035–1038, 2015.

[52] L. Poli, G. Oliveri, F. Viani, and A. Massa, "MT–BCS-based microwave imaging approach through minimum-norm current expansion," *IEEE Trans. Antennas Propagat.*, vol. 61, no. 9, pp. 4722–4732, 2013.

[53] O. Dorn and D. Lesselier, "Level set methods for inverse scattering," *Inverse Prob.*, vol. 22, no. 4, p. R67, 2006.

[54] C. R. Vogel, *Computational Methods for Inverse Problems*, vol. 23. SIAM, 2002.

[55] P. J. Van Laarhoven and E. H. Aarts, "Simulated annealing," in *Simulated Annealing: Theory and Applications*, pp. 7–15, Springer, 1987.

[56] C.-C. Chiu and P.-T. Liu, "Image reconstruction of a perfectly conducting cylinder by the genetic algorithm," *IEE Proceedings-Microwaves, Antennas and Propagation*, vol. 143, no. 3, pp. 249–253, 1996.

[57] A. K. Qin, V. L. Huang, and P. N. Suganthan, "Differential evolution algorithm with strategy adaptation for global numerical optimization," *IEEE Trans. on Evolution. Comput.*, vol. 13, no. 2, pp. 398–417, 2009.

[58] 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.

[59] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[60] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Am. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, 2001.

[61] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[62] H. Ohlsson, A. Y. Yang, R. Dong, and S. S. Sastry, "Nonlinear basis pursuit," in *2013 Asilomar Conference on Signals, Systems and Computers*, pp. 115–119, IEEE, 2013.

[63] T. Blumensath and M. E. Davies, "Gradient pursuit for non-linear sparse signal modelling," in *16th European Signal Processing Conference, 2008*, pp. 1–5, IEEE, 2008.

[64] M. T. Bevacqua and R. Scapaticci, "A compressive sensing approach for 3D breast cancer microwave imaging with magnetic nanoparticles as contrast agent," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 665–673, 2016.

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

[66] M. Bevacqua, L. Crocco, L. D. Donato, T. Isernia, and R. Palmeri, "Exploiting sparsity and field conditioning in subsurface microwave imaging of nonweak buried targets," *Radio Sci.*, vol. 51, no. 4, pp. 301–310, 2016.

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

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

[69] P. Shah, U. K. Khankhoje, and M. Moghaddam, "Inverse scattering using a joint $l1-l2$ norm-based regularization," *IEEE Trans. Antennas Propagat.*, vol. 64, no. 4, pp. 1373–1384, 2016.

[70] L. Gharsalli, H. Ayasso, B. Duchêne, and A. Mohammad-Djafari, "Inverse scattering in a Bayesian framework: application to microwave imaging for breast cancer detection," *Inverse Prob.*, vol. 30, no. 11, p. 114011, 2014.

[71] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.

[72] W. C. Karl, "Regularization in image restoration and reconstruction-3.6," 2000.

[73] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Lett.*, vol. 14, no. 10, pp. 707–710, 2007.

[74] A. Desmal and H. Bagci, "A preconditioned inexact Newton method for nonlinear sparse electromagnetic imaging," *IEEE Geosci. Remote. S.*, vol. 12, no. 3, pp. 532–536, 2015.

[75] H. Bagci, R. Raich, A. Hero, and E. Michielssen, "Sparsity-regularized born iterations for electromagnetic inverse scattering," in *Antennas and Propagation Society International Symposium, 2008. AP-S 2008. IEEE*, pp. 1–4, IEEE, 2008.

[76] J. H. Richmond, "Scattering by a dielectric cylinder of abritrary cross-section shape," *IEEE Trans. Antennas Propagat.*, vol. 13, no. 3, pp. 334–341, 1965.

[77] W. C. Chew, *Waves and Fields in Inhomogeneous Media*, vol. 522. IEEE Press New York, 1995.

[78] W. C. Gibson, *The Method of Moments in Electromagnetics*, vol. 1. Chapman & Hall/CRC London, UK, 2008.

[79] R. F. Harrington and J. L. Harrington, *Field Computation by Moment Methods*. Oxford University Press, 1996.

[80] J. Richmond, "Scattering by a dielectric cylinder of arbitrary cross section shape," *IEEE Trans. Antennas Propagat.*, vol. 13, no. 3, pp. 334–341, 1965.

[81] T. Bonesky, K. Bredies, D. A. Lorenz, and P. Maass, "A generalized conditional gradient method for nonlinear operator equations with sparsity constraints," *Inverse Prob.*, vol. 23, no. 5, p. 2041, 2007.

[82] M. Hagmann, O. Gandhi, and C. Durney, "Upper bound on cell size for moment-method solutions," *IEEE Trans. Microwave Theory Tech.*, vol. 25, no. 10, pp. 831–832, 1977.

[83] B. Jin, T. Khan, and P. Maass, "A reconstruction algorithm for electrical impedance tomography based on sparsity regularization," *Int. J. Numer. Methods Eng.*, vol. 89, no. 3, pp. 337–353, 2012.

[84] L. Armijo *et al.*, "Minimization of functions having Lipschitz continuous first partial derivatives," *Pacific J. Mathematics*, vol. 16, no. 1, pp. 1–3, 1966.

[85] L. Grippo, F. Lampariello, and S. Lucidi, "A nonmonotone line search technique for Newton's method," *SIAM J. Numer. Anal.*, vol. 23, no. 4, pp. 707–716, 1986.

[86] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder–mead simplex method in low dimensions," *SIAM Journal on optimization*, vol. 9, no. 1, pp. 112–147, 1998.

[87] E. T. Hale, W. Yin, and Y. Zhang, "A fixed-point continuation method for l1-regularized minimization with applications to compressed sensing," *CAAM TR07-07, Rice University*, vol. 43, p. 44, 2007.

[88] Y. Nesterov, "Gradient methods for minimizing composite functions," 2013.

[89] S. J. Wright, R. D. Nowak, and M. A. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Processing*, vol. 57, no. 7, pp. 2479–2493, 2009.

[90] M. Ambrosanio, P. Kosmas, and V. Pascazio, "Exploiting wavelet decomposition to enhance sparse recovery in microwave imaging," in *11th European Conference on Antennas and Propagation (EUCAP), 2017*, pp. 1607–1610, IEEE, 2017.

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

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

[93] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *IEEE Trans. Inform. Theory*, vol. 58, no. 6, pp. 3613–3641, 2012.

[94] S. Sarvotham, D. Baron, M. Wakin, M. F. Duarte, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Asilomar Conference on Signals, Systems, and Computers*, pp. 1537–1541, 2005.

[95] R. Heckel and H. Bolcskei, "Joint sparsity with different measurement matrices," in *50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2012*, pp. 698–702, IEEE, 2012.

[96] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing*, vol. 6, no. 2, pp. 298–311, 1997.

[97] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

[98] J. Yang and Y. Zhang, "Alternating direction algorithms for l_1-problems in compressive sensing," *SIAM J. Sci. Comput.*, vol. 33, no. 1, pp. 250–278, 2011.

[99] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," vol. 5, no. 5, pp. 912–926, 2011.

[100] A. Abubakar and P. M. van den Berg, "Iterative forward and inverse algorithms based on domain integral equations for three-dimensional electric and magnetic objects," *J. Comput. Phys.*, vol. 195, no. 1, pp. 236–262, 2004.

[101] Y. Zhong, M. Lambert, D. Lesselier, and X. Chen, "Electromagnetic response of anisotropic laminates to distributed sources," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 1, pp. 247–256, 2014.

[102] Y. Zhong, P.-P. Ding, M. Lambert, D. Lesselier, and X. Chen, "Fast calculation of scattering by 3-d inhomogeneities in uniaxial anisotropic multilayers," *IEEE Trans. Antennas Propagat.*, vol. 62, no. 12, pp. 6365–6374, 2014.

[103] G. Oliveri, P. P. Ding, and L. Poli, "3-D crack detection in anisotropic layered media through a sparseness-regularized solver," *IEEE Antennas and Wireless Propagation Letters*, vol. 14, pp. 1031–1034, 2015.

[104] D. Lesselier, P.-P. Ding, G. Rodeghiero, M. Lambert, and Y. Zhong, "On inverse scattering and imaging solutions for objects buried within uniaxially anisotropic media," in *Microwave Symposium (MMS), 2015 IEEE 15th Mediterranean*, pp. 1–4, IEEE, 2015.

[105] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[106] R. V. Şenyuva, Ö. Özdemir, G. K. Kurt, and E. Anarım, "Electromagnetic imaging of closely spaced objects using matching pursuit based approaches," *IEEE Antennas Wireless Propagat. Lett.*, vol. 15, pp. 1179–1182, 2016.

[107] A. I. Sandhu and H. Bagci, "A modified CoSaMP algorithm for electromagnetic imaging of two dimensional domains," in *2017 International Applied Computational Electromagnetics Society Symposium-Italy (ACES)*, pp. 1–2, IEEE, 2017.

[108] D. Brandwood, "A complex gradient operator and its application in adaptive array theory," in *IEEE Proc.s F (Communicat., Radar and Signal Process.)*, vol. 130, pp. 11–16, IET, 1983.

[109] P. Busch, T. Heinonen, and P. Lahti, "Heisenberg's uncertainty principle," *Phys. Rep.*, vol. 452, no. 6, pp. 155–176, 2007.

[110] A. Graps, "An introduction to wavelets," *IEEE Computational Science and Engineering*, vol. 2, no. 2, pp. 50–61, 1995.

A. Adjoint Gradient of Cost Function

A.1. Adjoint Gradient of Cost Function

The discrepancy D under consideration is the standard least squared cost function such as

$$\begin{aligned} K &= \frac{1}{2}\left\|\zeta(\mathbf{r}) - E^{\text{diff}}(\mathbf{r})\right\|^2_{L\times S} \\ &= \frac{1}{2}\|\Theta(\mathbf{r})\|^2_{L\times S} \end{aligned} \tag{A.1}$$

where $\zeta(\mathbf{r})$ is the measured scattered field at the receiver position L while E^{diff} is the scattered field obtained analytically and S is the source position. By multiplying both sides of the state equation by $\chi(\mathbf{r})$, the following equation

$$J(\mathbf{r}) = \chi(\mathbf{r})E^{\text{inc}} + \chi(\mathbf{r})\int_\Omega G(\mathbf{r},\mathbf{r}')\chi(\mathbf{r})E^{\text{tot}}(\mathbf{r},\mathbf{r}')d\mathbf{r}' \tag{A.2}$$

is obtained where $J(\mathbf{r}) = \chi(\mathbf{r})E^{\text{tot}}(\mathbf{r})$ and $\chi(\mathbf{r}) = \chi_{\text{real}}(\mathbf{r}) + \chi_{\text{imag}}(\mathbf{r})$. Taking the first-order development of the kind $K(\chi(\mathbf{r}) + \delta\chi_{\text{real}}(\mathbf{r})$:

$$\begin{aligned} K + \delta K &= \frac{1}{2}\left\|\zeta(\mathbf{r}) - (E^{\text{diff}}(\mathbf{r}) + \delta E^{\text{diff}}(\mathbf{r}))\right\|^2_L \\ &= \frac{1}{2}\langle\zeta - (E^{\text{diff}}(\mathbf{r}) + \delta E^{\text{diff}}(\mathbf{r})), \zeta - (E^{\text{diff}}(\mathbf{r}) + \delta E^{\text{diff}}(\mathbf{r}))\rangle \\ &= \frac{1}{2}\int_S\int_L [\zeta - (E^{\text{diff}}(\mathbf{r}) + \delta E^{\text{diff}}(\mathbf{r}))][\overline{\zeta(\mathbf{r}) - E^{\text{diff}}(\mathbf{r}) + \delta E^{\text{diff}}(\mathbf{r})}] \\ &= \frac{1}{2}\int_S \Big\{\int_L \Theta(\mathbf{r})\overline{\Theta(\mathbf{r})}d(\mathbf{r}) - \int_L \delta E^{\text{diff}}(\mathbf{r})\overline{\Theta(\mathbf{r})} - \int_L \overline{\delta E_{\text{diff}}(\mathbf{r})}\Theta(\mathbf{r})dr + \int_L \delta E^{\text{diff}}(\mathbf{r})\overline{\delta E_{\text{diff}}(\mathbf{r})}\Big\} \end{aligned} \tag{A.3}$$

so that a small change in the cost function w.r.t the real part of contrast is described as

$$\delta K \simeq \mathrm{Re} \int_S \int_L \overline{\Theta(\mathbf{r},\mathbf{r}')} \delta E^{\mathrm{diff}}(\mathbf{r},\mathbf{r}') d\mathbf{r} d\mathbf{r}', \tag{A.4}$$

where overlined terms are denoted as complex conjugate ones [108].

Considering the terms forming the δK, δE^{diff} can be expressed as

$$\delta E^{\mathrm{diff}}(\mathbf{r}) = \int_D G^{\mathrm{or}}(\mathbf{r},\mathbf{r}') \delta J(\mathbf{r}) d\mathbf{r}'. \tag{A.5}$$

where $\delta J(\mathbf{r})$ can be also defined as

$$\begin{aligned} \delta J(\mathbf{r}) &= \delta\chi(\mathbf{r}) E^{\mathrm{inc}} + \delta\chi(\mathbf{r}) \int_D G_{\mathrm{oo}}(r,r)\chi(\mathbf{r}) E^{\mathrm{tot}}(\mathbf{r}) dr + \chi(\mathbf{r}) \int_D G_{\mathrm{oo}}(r,r)\delta J(\mathbf{r}) dr \\ &= \delta\chi(\mathbf{r}) E^{\mathrm{tot}}(\mathbf{r}) + \chi(\mathbf{r}) \int_D G_{\mathrm{oo}}(r,r)\delta(\chi(\mathbf{r}) E^{\mathrm{tot}}(\mathbf{r})) dr \\ &= \delta\chi(\mathbf{r}) E^{\mathrm{tot}} + \chi(\mathbf{r}) \int_D G_{\mathrm{oo}}(r,r)\delta J(\mathbf{r}) dr \end{aligned} \tag{A.6}$$

Considering

$$G_{\mathrm{oo}}(\mathbf{r},\mathbf{r}') = G_{\mathrm{oo}}(\mathbf{r}',\mathbf{r})^T$$

$$G_{\mathrm{ro}}(\mathbf{r},\mathbf{r}') = G_{\mathrm{or}}(\mathbf{r}',\mathbf{r})^T$$

and substituting A.5 into A.4 we can write δK as follows

$$\begin{aligned} \delta K &= \mathrm{Re} \int_S \Big\{ \int_L \overline{\Theta(\mathbf{r})} . \int_D G_{\mathrm{or}}(\mathbf{r},\mathbf{r}') \delta J(\mathbf{r}') d\mathbf{r}' d\mathbf{r} \Big\} \\ &= \mathrm{Re} \int_S \Big\{ \int_L \overline{\Theta(\mathbf{r})} d\mathbf{r} . \int_D G_{\mathrm{or}}(\mathbf{r},\mathbf{r}') \delta J(\mathbf{r}) d\mathbf{r} d\mathbf{r}' \Big\} \\ &= \mathrm{Re} \int_S \Big\{ \int_D \delta J(\mathbf{r}) . \int_L G_{\mathrm{or}}(\mathbf{r},\mathbf{r}') \overline{\Theta(\mathbf{r})} d\mathbf{r} d\mathbf{r}' \Big\} \end{aligned} \tag{A.7}$$

By definition as in [108], the adjoint is

$$P(\mathbf{r}) = \int_L G_{ro}(\mathbf{r},\mathbf{r}') \overline{\Theta(\mathbf{r})} dr + \int_D G_{oo}(\mathbf{r},\mathbf{r}') \chi(\mathbf{r}) P(\mathbf{r}') d\mathbf{r}' \tag{A.8}$$

so that substituting A.8 into A.7, δK becomes

$$\begin{aligned}
\delta K &= \mathrm{Re}\int_S \Big\{ \int_D \delta J(\mathbf{r}).\Big[P(\mathbf{r}) - \int_D G_{\mathrm{oo}}(\mathbf{r},\mathbf{r}')\chi(\mathbf{r}')P(\mathbf{r}')d\mathbf{r}'\Big]d\mathbf{r}\Big\} \\
&= \mathrm{Re}\int_S \Big\{ \int_D \Big[P(\mathbf{r}).\delta J(\mathbf{r}) - \delta J(\mathbf{r}) \int_D G_{\mathrm{oo}}(\mathbf{r},\mathbf{r}')\chi(\mathbf{r})P(\mathbf{r}')d\mathbf{r}'\Big]d\mathbf{r}\Big\} \\
&= \mathrm{Re}\int_S \Big\{ \int_D P(\mathbf{r})\Big[\delta J(\mathbf{r}) - \chi(\mathbf{r}) \int_D G_{\mathrm{oo}}(\mathbf{r},\mathbf{r}')\delta J(\mathbf{r}')d\mathbf{r}'\Big]d\mathbf{r}\Big\} \\
&= \mathrm{Re}\int_S \Big\{ \int_D P(\mathbf{r},\mathbf{r}_s)\delta\chi(\mathbf{r})E^{\mathrm{tot}}(\mathbf{r},\mathbf{r}_s)d\mathbf{r}\Big\}.
\end{aligned} \tag{A.9}$$

Therefore, the gradient of the discrepancy K w.r.t $\chi_{\mathrm{real}}(\mathbf{r})$ is given by

$$K'(\chi_{\mathrm{real}}(\mathbf{r})) = 2 \times \mathrm{Re}(P(\mathbf{r})E^{\mathrm{tot}}(\mathbf{r}_s,\mathbf{r})) \tag{A.10}$$

The same process can be followed for the gradient w.r.t $\chi_{\mathrm{imag}}(\mathbf{r})$ as well. It is known that $\frac{\mathrm{d}K(\chi(\mathbf{r}))}{\mathrm{d}\chi(\mathbf{r})} = \frac{\mathrm{d}K(\chi(\mathbf{r}))}{\mathrm{d}\chi_{\mathrm{real}}(\mathbf{r})} - i\frac{\mathrm{d}K(\chi(\mathbf{r}))}{\mathrm{d}\chi_{\mathrm{imag}}(\mathbf{r})}$. Therefore,

$$K'(\chi(\mathbf{r})) = 2 \times P(\mathbf{r})E^{\mathrm{tot}}(\mathbf{r}_s,\mathbf{r}) \tag{A.11}$$

It can be noted that the equation in (A.8) has the same form as equation (7). Therefore, it can be solved through MoM in the same way. By discretization, we can obtain

$$P_i = \sum_{r=1}^{N_R} G_{ir}\Theta_k^* + \sum_{j=1}^{N} G_{ij}\chi_j P_j \quad j = 1,2,\ldots,N \tag{A.12}$$

where N_R is the number of receivers, G_{ir} is the Green's function between the position of the ith point in the region of interest and the position of the kth receiver. The equation above can also be written as the following linear equations such as

$$\begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \end{pmatrix} = \begin{pmatrix} \sum_{r=1}^{N_R} G_{1r}\Theta_k^* \\ \sum_{r=2}^{N_R} G_{2r}\Theta_k^* \\ \vdots \\ \sum_{r=N}^{N_R} G_{Nr}\Theta_k^* \end{pmatrix} + \begin{pmatrix} G_{11}\chi_1 & G_{12}\chi_2 & \cdots & G_{1N}\chi_N \\ G_{21}\chi_1 & G_{22}\chi_2 & \cdots & G_{1N}\chi_N \\ \vdots & \vdots & \ddots & \vdots \\ G_{N1}\chi_1 & G_{N2}\chi_2 & \cdots & G_{NN}\chi_N \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \end{pmatrix} \tag{A.13}$$

Solving for P gives

$$\begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \end{pmatrix} = \begin{pmatrix} 1-G_{11}\chi_1 & -G_{12}\chi_2 & \cdots & -G_{1N}\chi_N \\ -G_{21}\chi_1 & 1-G_{22}\chi_2 & \cdots & -G_{1N}\chi_N \\ \vdots & \vdots & \ddots & \vdots \\ -G_{N1}\chi_1 & -G_{N2}\chi_2 & \cdots & 1-G_{NN}\chi_N \end{pmatrix}^{-1} \begin{pmatrix} \sum_{r=1}^{N_R} G_{1r}\Theta_k^* \\ \sum_{r=2}^{N_R} G_{2r}\Theta_k^* \\ \vdots \\ \sum_{r=N}^{N_R} G_{Nr}\Theta_k^* \end{pmatrix} \tag{A.14}$$

whereas

$$\begin{pmatrix} \sum_{r=1}^{N_R} G_{1r}\Theta_k^* \\ \sum_{r=2}^{N_R} G_{2r}\Theta_k^* \\ \vdots \\ \sum_{r=N}^{N_R} G_{Nr}\Theta_k^* \end{pmatrix} = \begin{pmatrix} G_{11} & G_{12} & \cdots & G_{1N_R} \\ G_{21} & G_{22} & \cdots & G_{2N_R} \\ \vdots & \vdots & \ddots & \vdots \\ G_{N_R1} & G_{N_R2} & \cdots & G_{NN_R} \end{pmatrix} \left(\begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_{N_R} \end{pmatrix} - \begin{pmatrix} E_1^{\text{diff}} \\ E_2^{\text{diff}} \\ \vdots \\ E_{N_R}^{\text{diff}} \end{pmatrix} \right)^* \tag{A.15}$$

Furthermore,

$$\begin{pmatrix} E_1^{\text{diff}} \\ E_2^{\text{diff}} \\ \vdots \\ E_{N_R}^{\text{diff}} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N} G_{i1}\chi_i E_i^{\text{tot}} \\ \sum_{i=1}^{N} G_{i2}\chi_i E_i^{\text{tot}} \\ \vdots \\ \sum_{i=1}^{N} G_{iN_R}\chi_i E_i^{\text{tot}} \end{pmatrix} \tag{A.16}$$

$$\begin{pmatrix} E_1^{\text{diff}} \\ E_2^{\text{diff}} \\ \vdots \\ E_{N_R}^{\text{diff}} \end{pmatrix} = \begin{pmatrix} G_{11} & G_{21} & \cdots & G_{N_R1} \\ G_{12} & G_{22} & \cdots & G_{N_R2} \\ \vdots & \vdots & \ddots & \vdots \\ G_{1N_R} & G_{2N_R} & \cdots & G_{NN_R} \end{pmatrix} \begin{pmatrix} \chi_1 & 0 & \cdots & 0 \\ 0 & \chi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \chi_N \end{pmatrix} \begin{pmatrix} E_1^{\text{tot}} \\ E_2^{\text{tot}} \\ \vdots \\ E_{N_R}^{\text{tot}} \end{pmatrix}$$

$$\begin{pmatrix} E_1^{\text{diff}} \\ E_2^{\text{diff}} \\ \vdots \\ E_{N_R}^{\text{diff}} \end{pmatrix} = \begin{pmatrix} G_{11} & G_{12} & \cdots & G_{1N_R} \\ G_{21} & G_{22} & \cdots & G_{2N_R} \\ \vdots & \vdots & \ddots & \vdots \\ G_{N_R1} & G_{N_R2} & \cdots & G_{NN_R} \end{pmatrix}^{\dagger} \begin{pmatrix} \chi_1 & 0 & \cdots & 0 \\ 0 & \chi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \chi_N \end{pmatrix} \begin{pmatrix} E_1^{\text{tot}} \\ E_2^{\text{tot}} \\ \vdots \\ E_{N_R}^{\text{tot}} \end{pmatrix} \tag{A.17}$$

B. The Wavelet Theory

B.1. The Wavelet Transform

The wavelet transform has been proposed to overcome the limits of the Fourier transform. However, if we apply the Fourier transform on a given signal, we cannot determine the moments corresponding to high frequencies. The Short Fourier Transform (SFT) uses a windowing technique to calculate the spectrogram, which provides information in the time-frequency domain.

The problem of the SFT is the Heisenberg uncertainty principle [109]. This principle states that it is impossible to know the exact time frequency correspondence of a signal, i.e., we cannot know which spectral component corresponds to a signal at given instant. Although the temporal and frequency resolution problems come from physical phenomenon and they are independent of the used transform, it is possible to analyze any signal using an alternative approach called wavelet transform (WT). This transform allows to analyze the signals at different frequencies resolutions unlike the Fourier transform that perform a unique frequency resolution analysis. The wavelet transform is designed to give a poor temporal resolution and good frequency resolution at high frequencies and vice-versa at low frequencies. This compromise makes sense especially when the analyzed signal has short duration high frequency components and long duration low frequencies components. This is the case of most biological and natural signals [37]. So the major difference between the Fourier and wavelet transforms is that the window length varies with the transform computed for each spectral component.

B.1.1. Approximation theory and multiresolution analysis

The theory of multiresolution analysis in wavelet domain is founded by Mallat in 1989 [39]. Its principle consists to assume that there is a series of closed subsets $V_{j\,j\in\mathbb{Z}}$ that approximate the integrable $L^2(\mathbb{R})$ functions. The subsets V_j are designed in such a way that $V_j \subset V_{j+1}$. Let set W_j the orthogonal complement of V_j in V_{j+1}, i.e., $V_{j+1} = V_j \oplus W_j$

. Then, for a given function $f \in L^2(\mathbb{R})$, there are $f_N \in V_N$ approximating f. If we set $g_i \in W_i$ and $f_i \in V_i$ then we can write:

$$f_N = f_{N-1} + g_{N-1} = \sum_{i=1}^{M} g_{N-i} + f_{N-M}. \tag{B.1}$$

This is the wavelet decomposition of the function f. In the signal processing context, this decomposition is established under a pyramidal coding algorithm. The time frequency representation of a signal in wavelet domain is obtained using filtering techniques. The series $h(k)$, and $g(k)$, $k \in \mathbb{Z}$ are quadrature mirror filters and the relationship between h and g is given by:

$$g(k) = (-1)^n h(1-n) \tag{B.2}$$

The series $h(k)$ is a family of low pass filters and $g(k)$ are high pass filters. These two filters families belong to the Finite Impulse Response (FIR) filters. By using the Fourier transform and the orthogonality property between h and g, we can prove that [110]:

$$\sum_k h(k) = \sqrt{2} \quad \text{and} \quad \sum_k g(k) = 0 \tag{B.3}$$

The wavelet decomposition process begins by passing the signal through the low frequency band $h(n)$. The filtering operation of the signal consists to calculate its convolution with the impulse response of the filter. Note that a half band low pass filter removes all frequencies that are higher than half of the maximum frequency in the signal. So after applying the low pass filter, half of the samples will be eliminated and the filtered signal will have only half of samples in the next decomposition level. On other hand, the low pass filtering divides the resolution of the signal by two since this resolution depends only on the quantity of information in the signal. Then, the signal is undersampled by two since redundancy of the half of the samples and this operation doubles the scale.

B.1.2. Continuous Wavelet Transform

The continuous wavelet transform coefficients applied to a given signal f are defined as follows [110]:

$$C(a,b) = \int_{-\infty}^{+\infty} f(t)\Theta_{a,b}(t)d(t) \tag{B.4}$$

where

$$\Theta_{a,b}(t) = a^{\frac{-1}{2}}\Theta\left(\frac{t-b}{a}\right) \tag{B.5}$$

is the window function of the WT, a is the scale parameter and b is the translation parameter.

The translation coefficient b is related to the window location, it corresponds to the temporal information in the wavelet domain. On the other hand, we have not the same parameter for the frequency information, like in the Fourier transform, but the scale parameter a. This latter is used to delay or compress the transformed signal. The small scale values correspond to a dilation and the large values correspond to a compression. Note that the relationship between the scale and the frequency is reverse, i.e., low scale values correspond to high frequencies and vice-versa.

B.1.3. Multiband coding in wavelet domain

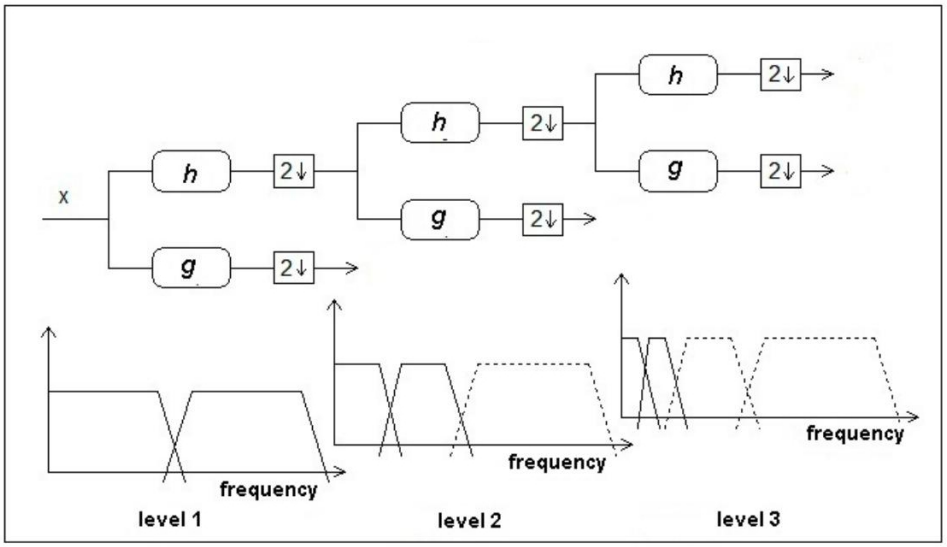


Figure B.1.: Filter bank of the 2D wavelet transform.

To express the multiband coding process in the wavelet domain, let us set the series $f = f_n$ that represents the discrete signal to be decomposed and the operators H and G defined as follows:

$$(Hf)_k = \sum_n h(n-2k)f(n)(Gf)_k = \sum_n h(n-2k)f(n) \tag{B.6}$$

These two equations refer to the filtering operation made by filters $h(k)$ and $g(k)$ using the convolution products of f with their impulse responses respectively, and the factor $2k$ refers to the undersampling. Therefore, the discrete wavelet transform (DWT) can be summarized as follows:

$$f \longrightarrow (Gf, GHf, GH^2f, \cdots, GH^{j-1}f, H^jf) = (d^{(j-1)}, d^{(j-2)}, \cdots, d^{(0)}, c^{(0)}) \tag{B.7}$$

where $d^{(j-1)}, d^(j-2), \cdots, d^{(0)}$ are the details and $a^(0)$ is the approximation coefficient defined as:

$$c^{(j-1)} = Hc^{(j)} \quad \text{and} \quad d^{(j-1)} = Gd^{(j)} \tag{B.8}$$

The reconstruction process is similar with the decomposition. For each decomposition level, the signal is oversampled by two and passed by the synthetic high pass and low pass filters $\bar{G}$ and $\bar{H}$ that will be added later. Therefore, we define the operators $\bar{G}$ and $\bar{H}$ as follows:

$$(\bar{H}f)_k = \sum_n h(n-2k)f(n)(\bar{G}f)_k = \sum_n h(n-2k)f(n). \tag{B.9}$$

The recursive calculation gives:

$$f = \sum_{j=0}^{n-1} (\bar{H})^j \bar{G} d^j + (\bar{H}f)^n c^{(0)} \tag{B.10}$$

or in the temporal domain:

$$D_j = (\bar{H})^j \bar{G} d^j \quad \text{and} \quad A = (\bar{H})^n a^{(0)} \tag{B.11}$$

where D_j and A are the details and the approximation, respectively.

Publications

JOURNAL PAPERS

Published

H. Zaimaga and M. Lambert, "Soft Shrinkage Thresholding Algorithm for Nonlinear Microwave Imaging", *Journal of Physics: Conference Series*, IOP Publishing, 2016, 756 (1), pp.012011. <http://stacks.iop.org/1742-6596/756/i=1/a=012011>.<10.1088/1742-6596/756/1/012011>. <hal-01390779>

CONFERENCE PAPERS

Published and accepted

H. Zaimaga, A. Fraysse, and M. Lambert, "Sparse Reconstruction Algorithms for Nonlinear Microwave Imaging", *25th European Signal Processing Conference (EUSIPCO 2017)*, Aug.-Sept 2017, Kos island, Greece. pp.743-747, European Signal Processing Conference (EUSIPCO 2017), accepted. <http://www.eusipco2017.org/>.<hal-01587676>

H. Zaimaga, A. Fraysse, and M.Lambert, "Nonlinear Imaging of 3D Defect in Anisotropic Laminate Using Joint Sparsity Constraints", *18th International Symposium on Applied Electromagnetics and Mechanics (ISEM 2017)*, Sept. 2017, Chamonix Mont-Blanc, France, accepted. <hal-01587782>

H. Zaimaga and M. Lambert, "Sparsity-Enforced Microwave Inverse Scattering using Soft Shrinkage Thresholding", *European Signal Processing Conference (EUSIPCO)*, Aug 2016, Budapest, Hungary. pp.350-354, Proceedings of the European Signal Processing Conference (EUSIPCO 2016). <http://www.eurasip.org/Proceedings/Eusipco/Eusipco2016/papers/1570256250.pdf>.<hal-01367566>

H. Zaimaga and M. Lambert, "Sparsity Reconstruction Algorithm for Nonlinear Microwave Problems", *URSI Asia-Pasific Radio Science Conference, URSI AP-RASC*, Aug 2016, Séoul, South Korea. S-B13a-5 (4 p.), Proceedings of the URSI Asia-Pacific Radio Science Conference (URSI RASC 2016). <hal-01367564>

H. Zaimaga and M. Lambert, "Sparsity-Enforced Microwave Inverse Scattering", Assemblée Générale 2015, GDR ONDES, Lyon, France, October 2015. <hal-01265732>

Poster Presentations

H. Zaimaga and M. Lambert, "Sparsity-Enforced Microwave Inverse Scattering: Preliminary Results", Compressive Sensing and Sparsity:Theory and Applications in Tomography, Manchester, UK, November 2015.

H. Zaimaga and M. Lambert, "Sparsity-Enforced Microwave Inverse Scattering", Assemblée Générale 2015, GDR ONDES, Lyon, France, October 2015.

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre : Parcimonie et imagerie électromagnétique dans des situations non-linéaires

Mots clefs : imagerie électromagnétique, non-linéarité, parcimonie, décomposition en ondelettes

Résumé : L'imagerie électromagnétique est le problème de la détermination de la distribution de matériaux à partir de champs diffractés mesurés venant du domaine les contenant et sous investigation. Résoudre ce problème inverse est une tâche difficile car il est mal posé en raison de la présence d'opérateurs intégraux (de lissage) utilisés dans la représentation des champs diffractés en terme de propriétés des matériaux, et ces champs sont obtenus à un ensemble fini et non nécessairement optimal de points via des mesures bruitées. En outre, le problème inverse est non linéaire simplement en raison du fait que les champs diffractés sont des fonctions non linéaires des propriétés des matériaux. Le travail décrit traite du caractère mal posé de ce problème d'imagerie électromagnétique en utilisant des techniques de régularisation basées sur la parcimonie, qui supposent que le(s)diffracteurs(s) ne capture(nt) de fait qu'une petite fraction du domaine d'investigation. L'objectif principal est d'étudier de manière approfondie la régularisation de parcimonie pour les problèmes inverses non linéaires. Par conséquent, nous nous concentrons sur la méthode de Tikhonov non linéaire normalisée qui résout directement le problème de minimisation non linéaire en utilisant les itérations de Landweber, où une fonction de seuillage est appliquée à chaque étape pour promouvoir la contrainte de parcimonie. Ce schéma est accéléré à l'aide d'une méthode de descente de plus grande pente projetée et remplace l'opération de seuillage pour faire respecter cette contrainte. Cette approche a également été implémentée dans un domaine d'ondelettes qui permet une représentation précise de la fonction inconnue avec un nombre réduit de coefficients. En outre, nous étudions une méthode corrélée à la parcimonie qui offre de multiples solutions parcimonieuses qui partagent un support commun non nul afin de résoudre le problème non linéaire concerné.

Title : Sparsity and electromagnetic imaging in non-linear situations

Keywords : electromagnetic imaging, non-linear, sparsity, wavelet decomposition

Abstract : So-called quantitative electromagnetic imaging focused onto here is the problem of determining material properties from scattered fields measured away from the domain under investigation. Solving this inverse problem is a challenging task because it is ill-posed due to the presence of (smoothing) integral operators used in the representation of scattered fields in terms of material properties, and scattered fields are obtained at a finite set of points through noisy measurements. Moreover, the inverse problem is nonlinear simply due the fact that scattered fields are nonlinear functions of the material properties. The work described in this thesis deals with the ill-posedness of the electromagnetic imaging problem using sparsity-based regularization techniques, which assume that the scatterer(s) capture only a small fraction of the investigation domain and/or can be described in sparse fashion on a certain basis. The primary aim of the thesis is to intensively investigate sparsity regularization for nonlinear inverse problems. Therefore, we focus on sparsity-regularized nonlinear Tikhonov method which directly solves the nonlinear minimization problem using Landweber iterations, where a thresholding function is applied at every iteration step to promote the sparsity constraint. This scheme is accelerated using a projected steepest descent method and replaces the thresholding operation to enforce the sparsity constraint. This approach has also been implemented in wavelet domain which allows an accurate representation of the unknown function with a reduced number of coefficients. Additionally, we investigate a method correlated with the joint sparsity which gives multiple sparse solutions that share a common nonzero support in order to solve concerned nonlinear problem.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France