Approches bayésiennes en tomographie micro-ondes : applications à l'imagerie du cancer du sein

Leila Gharsalli

▶ **To cite this version:**

Leila Gharsalli. Approches bayésiennes en tomographie micro-ondes : applications à l'imagerie du cancer du sein. Traitement du signal et de l'image [eess.SP]. Université Paris Sud - Paris XI, 2015. Français. NNT : 2015PA112048 . tel-01159537

HAL Id: tel-01159537

https://theses.hal.science/tel-01159537

Submitted on 3 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE PARIS-SUD

ÉCOLE DOCTORALE : STITS

Laboratoire des Signaux et Systèmes

DISCIPLINE PHYSIQUE

THÈSE DE DOCTORAT

soutenue le 10/04/2015

par

Leila GHARSALLI

Approches bayésiennes en tomographie micro-onde. Application à l'imagerie du cancer du sein.

Directeur de thèse :	Bernard Duchêne	Chargé de recherche CNRS (L2S)
Co-directeur de thèse :	Ali Mohammad-Djafari	Directeur de recherche CNRS (L2S)
Encadrant :	Hacheme Ayasso	Maître de conférence (GIPSA-lab)

Composition du jury :

Président du jury :	Christian Pichot du Mezeray	Directeur de recherche CNRS (LEAT)
Rapporteurs :	Amélie Litman	Maître de conférence (Institut Fresnel)
	Andréa Massa	Professeur (ELEDIA)
Examinateurs :	Laure Blanc-Féraud	Directrice de recherche CNRS (I3S)

Résumé

Ce travail concerne l'imagerie micro-onde en vue d'application à l'imagerie biomédicale. Cette technique d'imagerie a pour objectif de retrouver la distribution des propriétés diélectriques internes (permittivité diélectrique et conductivité) d'un objet inconnu illuminé par une onde interrogatrice connue à partir des mesures du champ électrique dit diffracté résultant de leur interaction.

Un tel problème constitue un problème dit inverse par opposition au problème direct associé qui consiste à calculer le champ diffracté, l'onde interrogatrice et l'objet étant alors connus. La résolution du problème inverse nécessite la construction préalable du modèle direct associé. Celui-ci est ici basé sur une représentation intégrale de domaine des champs électriques donnant naissance à deux équations intégrales couplées dont les contreparties discrètes sont obtenues à l'aide de la méthode des moments.

En ce qui concerne le problème inverse, hormis le fait que les équations physiques qui interviennent dans sa modélisation directe le rendent non linéaire, il est également mathématiquement mal posé au sens de Hadamard, ce qui signifie que les conditions d'existence, d'unicité et de stabilité de la solution ne sont pas simultanément garanties. La résolution d'un tel problème nécessite sa régularisation préalable qui consiste généralement en l'introduction d'information *a priori* sur la solution recherchée. Cette résolution est effectuée, ici, dans un cadre probabiliste bayésien où l'on introduit une connaissance *a priori* adaptée à l'objet sous test et qui consiste à considérer ce dernier comme étant composé d'un nombre fini de matériaux homogènes distribués dans des régions compactes. Cette information est introduite par le biais d'un modèle de « Gauss-Markov-Potts ». Le calcul bayésien nous donne la loi *a posteriori* de toutes les inconnues à partir de laquelle on peut définir les estimateurs ponctuels. On s'attache ensuite à déterminer les estimateurs *a posteriori* via des méthodes d'approximation variationnelles et à reconstruire ainsi l'image de l'objet recherché.

Les principales contributions de ce travail sont d'ordre méthodologique et algorithmique. Elles sont illustrées par une application de l'imagerie micro-onde à l'imagerie du cancer du sein. Cette dernière constitue en soi un point très important et original de la thèse. En effet, l'imagerie du cancer du sein par la technique micro-onde est une alternative très intéressante à la mammographie par rayons X, mais n'en est encore qu'à un stade exploratoire.

Mots-clefs : Imagerie de diffraction, imagerie micro-onde, modèle direct, problème inverse, information *a priori*, modèle de Gauss-Markov-Potts, approches variationnelles, imagerie du cancer du sein.

Bayesian approaches to microwave tomography. Application to breast cancer imaging.

Abstract

This work concerns microwave tomography for application to biomedical imaging. The aim is to retreive both permittivity and conductivity of an unknown object from measurements of the scattered field that results from its interaction with a known interrogating wave. Such a problem is denoted as “the inverse problem” as opposed to the associated forward problem that consists of calculating the scattered field while the interrogating wave and the object are known.

The resolution of the inverse problem requires the prior construction of the associated forward model. The latter is based upon an integral representation of the electric field resulting in two coupled integral equations whose discrete counterparts are obtained by means of the method of moments.

Regarding the inverse problem, in addition to the fact that the physical equations involved in the forward modeling make it nonlinear, it is also mathematically ill-posed in the sense of Hadamard, which means that the conditions of existence, uniqueness and stability of its solution are not simultaneously guaranteed. Hence, solving this problem requires its prior regularization which usually involves the introduction of *a priori* information on the sought solution. This resolution is done here in a Bayesian probabilistic framework where we introduce prior knowledge appropriate to the sought object by considering that the latter is composed of a finite number of homogeneous materials distributed into compact and homogeneous regions. This information is introduced by means of a "Gauss-Markov-Potts" model. In addition, Bayesian computations yield the posterior distribution of all the unknowns from which we can define the point estimators. We proceed then to identify the posterior estimators via variational approximation methods and thereby to reconstruct images of the thought object.

The main contributions of this work are methodological and algorithmic. They are illustrated by an application of microwave tomography to breast cancer imaging. The latter is in itself a very important and original aspect of the thesis. Indeed, imaging of breast cancer using microwaves is a very interesting alternative to X-ray mammography, but it is still at an exploratory stage.

Keywords : Diffraction imaging, microwave imaging, forward problem, inverse problem, prior information, Gauss-Markov-Potts model, variational approaches, breast cancer imaging.

Remerciements

Quoi que je vais écrire, je ne remercierai jamais assez les gens sans qui cette thèse ne serait pas ce qu'elle est (on ne se rend compte de la chance qu'on a eu qu'une fois les années écoulées...).

Je tiens tout d'abord à remercier mes co-directeurs de thèse, M. Bernard Duchêne et M. Ali Mohammad-Djafari pour m'avoir fait confiance malgré que je détenais à la base un profil de mathématicienne et malgré les connaissances plutôt légères que j'avais en octobre 2011 sur l'électromagnétisme. Ils ont su me guider, m'encourager, me conseiller et m'ont fait beaucoup voyager durant cette thèse en me laissant une grande liberté et en me faisant l'honneur de me déléguer plusieurs responsabilités dont j'espère avoir été à la hauteur.

Mes profonds remerciements vont également à mon encadrant, M. Hacheme Ayasso pour sa gentillesse, ses conseils, ses réponses minutieuses à mes nombreuses questions et aussi pour l'hospitalité dont il a fait preuve envers moi lors des trois séjours que j'ai effectués dans son groupe.

Je remercie énormément les membres de jury de s'être arrangé pour venir malgré la grève du transport aérien et le report d'un jour de ma soutenance où j'ai du subir un stress énorme. Je remercie sincèrement les deux rapporteurs, Mme. Amélie Litman et M. Andréa Massa pour leur lecture, leurs remarques et suggestions précieuses qui m'ont permis d'avoir une idée plus claire sur comment améliorer ce document et d'autres idées pour des travaux futurs.

Je remercie sincèrement le Groupe Problèmes Inverses (GPI) pour m'avoir accueilli d'abord en tant que stagiaire où j'ai travaillé avec mes tuteurs M. Thomas Rodet et Mme. Aurélia Fraysse qui m'ont d'ailleurs permis de franchir mes premiers pas au sein du groupe et bien m'intégrer. Je remercie par ailleurs Nicolas Gac pour sa sympathie et les discussions qu'a pu souvent avoir ainsi que pour ses divers conseils.

Certes, l'ambiance avec mes collègues (Yuling, Long, Mircea, Li, Thomas...) qui sont devenus à ce jour de bons amis était l'un des facteurs importants contribuant à ma réussite ; en déjeunant, autour d'un café ou même en étant chacun scotché à son poste, on a toujours su créer des moments pour rigoler, détendre l'atmosphère du travail et déstresser ensemble. J'espère que les nouveaux stagiaires et doctorants du GPI ainsi que mes collègues des autres unités (Safa, Boutheina, Mohamed, Mouaadh, Philippes, Seif, Soufiane, Djawad...) effectueront leurs travaux dans d'aussi bonnes conditions.

Je n'oublie pas de remercier du fond du cœur mes amis en dehors du laboratoire, avec qui j'ai eu une vie sociale très active et sans qui je n'aurais jamais pu dépasser les moments de stress et de pression vécus durant la thèse ; Slim, Azza, Haythem, Sarah, Saliha, Safae, Wafae, Mariem, Monaim et enfin Hassane qui a toujours été là pour moi et qui ma toujours soutenu spécialement lors des procédures administratives que je devais accomplir.

Je ne trouverai jamais les mots pour exprimer ma profonde et sincère reconnaissance à mes chers parents qui m'ont encouragé dès mon jeune âge à exceller dans mes études et qui m'ont offert les meilleures conditions possibles afin d'avancer aisément dans mon cursus. Ma grande gratitude va également à mon grand frère et mon idole Ghazi qui

m'a beaucoup soutenu dès mon entrée en France et a énormément facilité mon intégration. Une forte pensée va à mon petit frère Kamel pour qui je souhaite la réussite au baccalauréat d'abord ensuite un avenir radieux.

Enfin, je présente mes excuses aux personnes que j'ai peut être oubliées de citer, et qui pourtant ont été présentes lors de cette thèse, et j'adresse encore une fois mes remerciements à tous ceux qui, de loin ou de près, ont contribué à la réussite de ce présent travail.

Pour tout
ceux qui m'ont soutenue durant ces trois
courtes années, je
vous aime, je ne
vous oublie
jamais !
♡

Table des matières

1 Introduction générale **13**

1.1 Avant propos 13

1.2 Contexte général d'étude 14

1.3 Imagerie micro-onde 15

1.4 L'application visée 16

1.4.1 Approches de l'imagerie micro-onde 18

1.5 Méthodes de résolution du problème d'imagerie micro-onde 19

1.5.1 Les approches linéarisées 19

1.5.2 Les approches non linéaires 20

1.6 Cadre du travail 23

1.7 Plan de l'étude et contributions 23

2 Le problème direct **27**

2.1 Introduction 27

2.2 Représentation intégrale de domaine du champ électrique 28

2.3 Modèle discret 31

2.3.1 Méthode des Moments 31

2.3.2 Application de la méthode des moments au domaine intégral . . . 32

2.4 Génération des données 34

2.5 Configuration de mesure 36

2.6 Validation des modèles 39

2.7 Conclusions 39

3 Inférence bayésienne en imagerie micro-onde **43**

3.1 Introduction 43

3.2 Mise en équation probabiliste du modèle direct 44

3.3 Modèle *a priori* hiérarchique . . . 45
3.3.1 Modèle de Mélange de Gaussiennes conditionnellement Indépendantes (MGI) . . . 48
3.3.2 Modèle de Mélange de Gauss-Markov (MGM) . . . 48
3.4 Cadre quasi non-supervisé . . . 50
3.4.1 Estimation conjointe . . . 52
3.4.2 Estimateur $\mathcal{J}^{MAP}$ (optimisation alternée par point fixe) . . . 52
3.4.3 Discussion . . . 56
3.5 Conclusions . . . 57

4 Méthodes bayésiennes variationnelles 59
4.1 Introduction . . . 59
4.2 Principe de la méthode bayésienne variationnelle . . . 60
4.3 Approche bayésienne variationnelle en imagerie micro-onde . . . 62
4.3.1 Équations de mise à jour . . . 63
4.3.2 Déroulement de l'algorithme . . . 66
4.4 Méthode du gradient pour l'approche bayésienne variationnelle . . . 67
4.4.1 Principe de la méthode . . . 68
4.4.2 Équations de mise à jour . . . 70
4.4.3 Pas de descente . . . 71
4.4.4 Déroulement de l'algorithme . . . 72
4.5 Comparaison théorique entre BV, G-BV et $\mathcal{J}^{MAP}$ . . . 73
4.6 Conclusions . . . 74

5 Résultats d'inversion 75
5.1 Introduction . . . 75
5.2 Résultats de reconstruction . . . 76
5.2.1 Qualité de reconstruction . . . 77
5.2.2 Complexité algorithmique . . . 87
5.3 Données bruitées . . . 88
5.4 Paramètre de corrélation . . . 91
5.5 MGI vs MGM . . . 92
5.6 Conclusions . . . 92

6 Bilan et perspectives **95**

6.1 Bilan de l'étude . . . 95

6.2 Perspectives du travail . . . 97

6.2.1 Autre modèle *a priori* . . . 97

6.2.2 Autre modélisation du bruit de mesure . . . 98

6.2.3 Mises à jour non simultanées des paramètres de forme . . . 98

6.2.4 Approche totalement non supervisée . . . 99

6.2.5 Prise en compte du caractère dispersif des tissus . . . 99

6.2.6 Taux de fausse alarme . . . 100

A Solution analytique de l'approche bayésienne variationnelle **101**

B Mise à jour des paramètres de forme du G-BV **103**

B.1 Les sources induites $\boldsymbol{w}$ . . . 103

B.2 Le contraste $\boldsymbol{\chi}$ . . . 105

B.3 Le champ caché $\boldsymbol{z}$ . . . 106

B.4 La variance du bruit d'observation v_ϵ . . . 107

B.5 La variance du bruit de couplage v_ξ . . . 108

B.6 La variance des classes v_k . . . 108

B.7 La moyennes des classes m_k . . . 109

C Détermination du pas optimal de descente du G-BV **111**

C.1 L'énergie libre négative . . . 111

C.2 Le calcul du pas optimal de descente . . . 112

D Publications **115**

Références . . . 143

1 Introduction générale

Sommaire

1.1 Avant propos 13
1.2 Contexte général d'étude 14
1.3 Imagerie micro-onde 15
1.4 L'application visée 16
1.4.1 Approches de l'imagerie micro-onde 18
1.5 Méthodes de résolution du problème d'imagerie micro-onde 19
1.5.1 Les approches linéarisées 19
1.5.2 Les approches non linéaires 20
1.6 Cadre du travail 23
1.7 Plan de l'étude et contributions 23

1.1 Avant propos

Ce chapitre introductif expose le problème d'imagerie micro-onde, qui est un problème inverse non linéaire et mal posé. Il présente les motivations applicatives qui ont fait que cette technique d'imagerie a connu un véritable essor dans la dernière décennie en applications biomédicales, et plus précisément en imagerie du cancer du sein. Par ailleurs, il détaille l'état des recherches dans le domaine de résolution de ce type de problème où les techniques les plus utilisées sont rappelées.

Ainsi, cette introduction générale nous place dans le contexte de la thèse tout en mettant l'accent sur les principales difficultés et limitations rencontrées ainsi que les contributions majeures réalisées durant ces travaux.

1.2 Contexte général d'étude

Contrairement aux ondes acoustiques ou élastiques qui nécessitent un milieu matériel pour se propager, les ondes électromagnétiques, qui sont le résultat de la propagation simultanée d'un champ électrique et d'un champ magnétique (figure (1.1)), ont la particularité de pouvoir se propager sans support matériel, c'est-à-dire qu'elles peuvent se propager dans le vide.

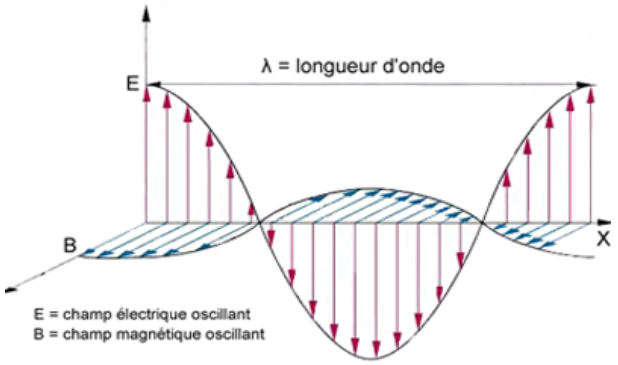


FIGURE 1.1 – Propagation d'une onde éléctromagnétique.

Ces ondes couvrent un spectre de fréquence extrêmement large dont la partie centrale (allant des ondes radio aux rayons gamma) est représentée sur la figure (1.2) avec les longueurs d'onde correspondantes. La partie du spectre qui nous intéresse ici est celle des micro-ondes (300 MHz - 300 GHz) ou plus précisément sa partie basse (300 MHz - 20 GHz) dans laquelle se situe la plupart des dispositifs d'imagerie micro-onde pour application biomédicale.

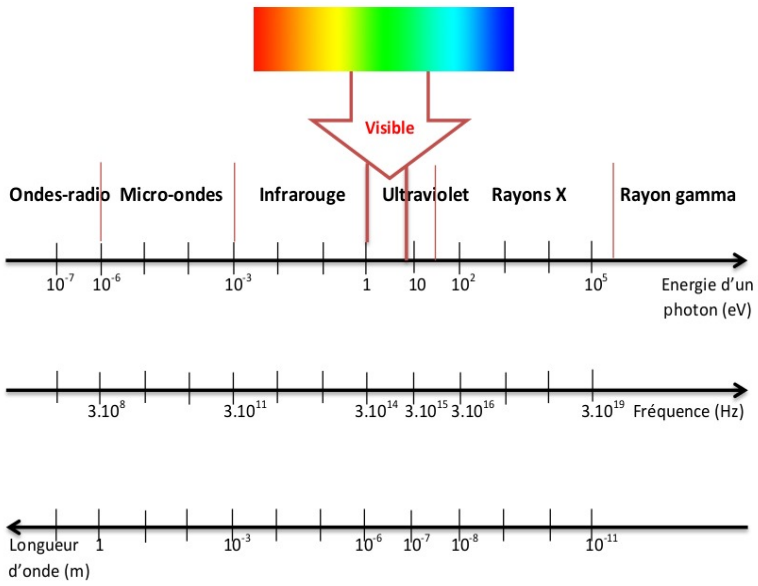


FIGURE 1.2 – Spectre éléctromagnétique.

Lorsqu'une onde électromagnétique rencontre un objet, le résultat de leur interaction dépend essentiellement des propriétés électromagnétiques de l'objet (caractérisées par sa

permittivité diélectrique ϵ, sa perméabilité magnétique μ et sa conductivité σ) ainsi que des dimensions respectives de ce dernier par rapport à la longueur d'onde λ ($\lambda = c/\nu$ où $c = 299\,792\,458\ m/s$ est la vitesse de la lumière dans le vide et ν est la fréquence de l'onde). En effet, lorsque la longueur d'onde est très grande par rapport aux dimensions caractéristiques de l'objet ou de ses inhomogénéités, il n'y a quasiment pas d'interaction. En revanche, lorsque dimensions caractéristiques et longueur d'onde sont du même ordre de grandeur (domaine dit de résonance), l'interaction onde - objet donne lieu à des phénomènes de diffraction importants caractérisés par une dispersion de l'énergie dans tout l'espace, tandis que lorsque la longueur d'onde est très petite devant les dimensions caractéristiques, l'onde subira réflexion, réfraction et atténuation sans dispersion notable de son énergie.

Lorsqu'il y a interaction entre l'onde et l'objet, la mesure du champ résultant de celle-ci peut être exploitée pour caractériser ou imager l'objet. Le dernier mode d'interaction est, par exemple, exploité depuis les années 1970 au travers du scanner X dans le domaine de l'imagerie biomédicale et celui du contrôle non destructif. L'imagerie à rayon X vise ici à reconstruire une image du contraste de l'objet qui dépend à la fois de l'épaisseur et du coefficient d'atténuation des structures traversées. L'interaction onde-objet est ici décrite à l'aide de l'optique géométrique [1] en ce qui concerne la propagation de l'onde et de la loi de Beer-Lambert [2] en ce qui concerne son atténuation. Dans le cadre de la thèse, nous nous intéressons au premier mode d'interaction, dont la physique est décrite à l'aide des équations de Maxwell, et, plus précisément, à l'imagerie micro-onde qui constitue le corps de nos travaux. Il s'agit d'imager un objet inconnu, c'est-à-dire de dresser une cartographie de ses paramètres physiques (électromagnétiques), à partir de la mesure du champ dit "diffracté" résultant de son interaction avec une onde interrogatrice connue, ce qui constitue un problème inverse dit de "diffraction".

Dans la suite de ce chapitre, nous présentons l'application visée dans le cadre de ce travail. Ensuite nous discuterons les techniques de résolution de ce problème en survolant l'état de l'art. Et enfin, nous listerons les contributions majeures de nos travaux ainsi que le plan de la thèse.

1.3 Imagerie micro-onde

L'imagerie micro-onde a suscité un intérêt considérable au cours de la dernière décennie en vue d'applications dans différents domaines tels que l'exploration géophysique [3], le contrôle de la teneur en eau du sol [4], le radar à pénétration de sol et l'imagerie d'objets enfouis [5, 6, 7], le radar de détection à travers les murs [8, 9, 10], l'évaluation ou le contrôle non destructif [11, 12, 13] et l'imagerie biomédicale [14, 15, 16, 17, 18, 19, 20]. En ce qui concerne l'imagerie biomédicale, une part importante de ce succès provient de l'innocuité et de la sensibilité à l'égard de facteurs physiques ou physiologiques d'intérêt clinique (comme la teneur en eau, le flux sanguin et la température) des micro-ondes comparées aux techniques classiques comme l'imagerie à rayons X qui émet des rayons ionisants nocifs pour le corps humain. En revanche l'imagerie micro-onde présente également quelques inconvénient et notamment l'atténuation importante des micro-ondes dans les tissus à forte teneur en eau, qui peut compromettre la propagation des ondes à travers des volumes importants et donc l'imagerie d'organes enfouis profondément dans le corps, et une résolution relativement médiocre du fait des longueurs d'onde re-

lativement importantes (de l'ordre du centimètre) contrairement à celles rencontrées en tomographie à rayon X (de l'ordre du nanomètre). Ceci peut provoquer une dégradation de la résolution de l'image obtenue par rapport à l'imagerie X, mais cela n'empêche pas que la tomographie micro-onde permet d'obtenir une résolution suffisante pour le diagnostic précoce de tumeurs du sein [21, 22], l'application visée dans le cadre de cette thèse.

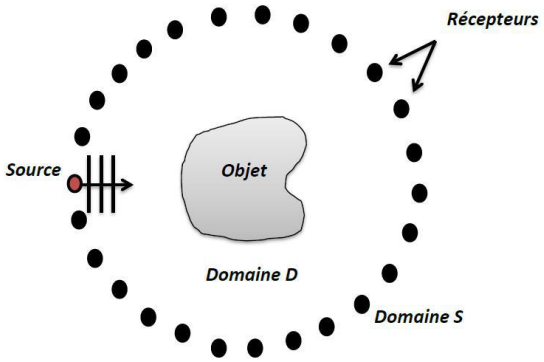


FIGURE 1.3 – Configuration géométrique en imagerie micro-onde bi-dimensionnelle (2D)

1.4 L'application visée

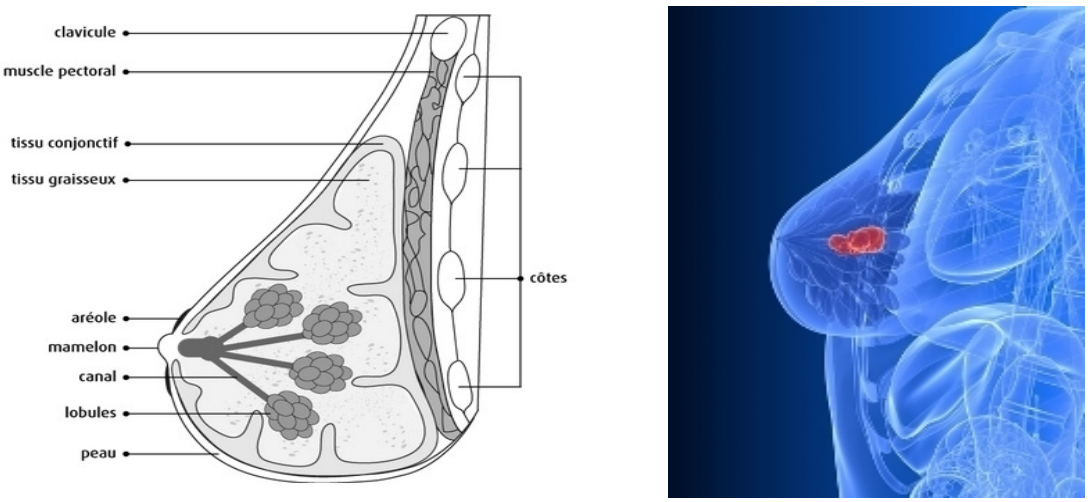
L'imagerie médicale est le premier geste intuitif pour confirmer ou infirmer un diagnostic supposé. Elle englobe différentes techniques comme la radiographie X, la résonance magnétique nucléaire, l'échographie ultrasonore ou la tomographie par émission de positons auxquelles on associe parfois l'endoscopie, l'imagerie optique et l'imagerie micro-onde. Ces techniques trouvent aussi des applications de plus en plus nombreuses dans différents domaines telles que la détection des cancers. Ici, nous nous intéressons plus particulièrement à la tomographie micro-onde qui, au cours des dernières décennies, a reçu un intérêt croissant pour des applications médicales telles que l'imagerie du cerveau ou la détection du cancer du sein [20, 23, 24, 25, 26, 27, 28, 29].

Le cancer du sein est un problème de santé majeur, car il est l'une des formes de cancer les plus courantes chez la femme. Chaque année, un million de nouveaux cas sont enregistrés dans le monde, ce qui fait de ce cancer le plus diagnostiqué au monde. En France également, il constitue le premier type de cancer chez la femme. Par exemple, selon un sondage réalisé en 2010, 52500 nouveaux cas de cancer du sein ont été estimés, soit plus d'un tiers (34%) de l'ensemble des nouveaux cas de cancers féminins. Cette tumeur maligne prend naissance dans les cellules du sein. Elle se développe le plus souvent dans les cellules qui couvrent les canaux, ou tubes, qui transportent le lait des glandes au mamelon (figure (1.4)). Elle peut se former aussi dans les lobules. Ces deux types de cancer peuvent rester localisés, ou bien ils peuvent se propager dans les tissus voisins du sein et endommager d'autres organes.

Toutes ces raisons font que le diagnostic de cette maladie à un stade précoce est

un véritable défi. Actuellement, la méthode de dépistage la plus utilisée est la mammographie par rayons X. Cependant, cette technique présente plusieurs inconvénients dont les principaux sont le caractère ionisant des rayons X et le coût élevé des dispositifs d'imagerie. En outre, selon l'âge de la patiente et la densité des tissus rencontrés, elle conduit souvent à un taux de fausses alarmes non négligeable. En revanche, l'imagerie micro-onde, qui donne des résultats encourageants dans d'autres domaines d'application, fournit une alternative à la mammographie par rayons X moins coûteuse et moins risquée que cette dernière.

L'imagerie micro-onde est exploitée à des fins biomédicales depuis plusieurs décennies. C'est à la fin des années 1970 que Larsen et Jacobi ont réussi à obtenir des images de la structure interne de reins canins à partir des coefficients de transmission entre deux antennes déplacées en parallèle, en garantissant ainsi la première pénétration suffisante de micro-ondes dans un objet biologique [30]. L'imagerie micro-onde est utilisée essentiellement au travers de trois approches, i.e. passive, duale et active. Avec les méthodes passives [31, 32] et duales (micro-ondes - acoustiques ou encore thermo-acoustiques [33]) on cherche à bénéficier d'un échauffement sélectif des tumeurs par micro-ondes, échauffement facilité par la conductivité plus élevée des tissus malins, comparés au tissus sains ; les premières visent à mesurer, à l'aide de radiomètres, l'augmentation de la température des tumeurs par rapport aux tissus normaux, tandis que les secondes détectent, à l'aide de transducteurs ultrasonores, les ondes de pression émises par les tumeurs, suite à leur échauffement. Enfin, l'imagerie micro-onde active, à laquelle nous nous intéressons ici, cherche à construire une image de l'organe sous test à partir de la mesure du champ diffracté résultant de son interaction avec une onde interrogatrice (dite onde incidente) connue, cette image étant, en fait, une cartographie des paramètres électromagnétiques (permittivité diélectrique et conductivité) de l'organe. Notons que ces derniers sont très variables selon les tissus et la fréquence d'illumination ([34], table (1.1)) et que les propriétés diélectriques des tumeurs présentent un contraste non négligeable par rapport à celles des tissus biologiques sains dans la bande de fréquence micro-onde [14].



(a) (b)

FIGURE 1.4 – a) Schéma anatomique d'un sein (source : Société Canadienne du Cancer) et b) représentation 3D d'un sein atteint par une tumeur (source : Vivant Santé, Médecine et Sciences du Vivant).

TABLE 1.1 – Permittivité diélectrique et conductivité de divers organes humains à la fréquence de 3 GHz.

Tissu	ϵ_r	σ (S/m)
Sang	$56-60$	$2,5$
Os	12	$0,4$
Graisse	$4-5$	$0,07-0,1$
Muscle	$50-55$	$1,8-2,2$
Peau	38	$1,5$
Cerveau	45	2
Rein	55	$2,5$
Foie	42	$1,8$
Cœur	55	$2,3$

1.4.1 Approches de l'imagerie micro-onde

Dans la suite, on se penche sur les principales méthodes trouvées dans la littérature permettant de résoudre le problème d'imagerie micro-onde. Mais avant cela, il faut mentionner que la formulation de ce problème peut se faire dans deux domaines différents selon la nature des données collectées par le système d'imagerie, ces dernières pouvant être collectées dans le domaine temporel ou dans le domaine fréquentiel. Dans le domaine temporel l'objet sous test est éclairé par une onde implusionnelle et le champ diffracté est mesuré à différents temps d'échantillonnage en différents points de mesure. Le processus d'inversion fait alors appel à méthodes numériques temporelles dont la plus fréquemment rencontrée est certainement l'imagerie confocale [25, 35, 36] dont le but est de retrouver la position et la taille de possibles inhomogénéités de l'objet en faisant intervenir des algorithmes de reconstruction simples et rapides couramment utilisés dans le domaine du radar. Mais d'autres méthodes plus sophistiquées permettant de reconstruire une cartographie des paramètres diélectriques de l'objet sous test ont également été développées telles que *forward-backward time stepping method* [37] ou encore *Chirp-Pulse Microwave Computed Tomography* [38].

Dans le domaine fréquentiel, l'objet est éclairé de façon séquentielle par un nombre N_ϑ d'émetteurs à une ou plusieurs fréquences fixes et, pour chacune des illuminations, l'amplitude et la phase du champ électrique diffracté sont mesurés par N_S récepteurs. Nous nous placerons ici dans le domaine fréquentiel car c'est le mode de fonctionnement d'un dispositif expérimental d'imagerie micro-onde développé depuis de nombreuses années au laboratoire [28].

1.5 Méthodes de résolution du problème d'imagerie micro-onde

Le problème d'imagerie micro-onde est pris ici au sens d'un problème inverse de diffraction où il s'agit de reconstruire une cartographie des paramètres électromagnétique

d'un objet inconnu (ici le sein) à partir de la mesure du champ diffracté qui résulte de son interaction avec une onde interrogatrice connue. La résolution d'un tel problème nécessite la construction préalable d'un modèle direct associé où, l'onde interrogatrice et l'objet étant connus, il s'agit de calculer le champ diffracté résultant de leur interaction. Le modèle direct est ici basé sur une représentation intégrale des champs électriques obtenue en appliquant le théorème de Green à l'équation des ondes de Helmholtz satisfaite par ceux-ci et en tenant compte des conditions de continuité des champs et de rayonnement à l'infini [39]. Nous sommes alors conduits à l'obtention de deux équations intégrales couplées dont l'une, dite équation d'observation, relie le champ diffracté observé au contraste, une fonction qui dépend des paramètres diélectriques de l'objet, et au champ électrique total induit à l'intérieur de l'objet par l'onde incidente et l'autre, dite équation de couplage (ou d'état), relie le champ total dans l'objet au contraste et à lui-même. Des contreparties discrètes de ces équations sont obtenues à l'aide de la méthode des moments [40, 41], ce qui permet d'obtenir un modèle algébrique plus facile à manipuler.

En ce qui concerne le problème inverse, où il s'agit de reconstruire une cartographie du contraste, il est non linéaire du fait que dans l'équation d'observation, reliant le champ diffracté observé au contraste, il apparaît également le champ total à l'intérieur de l'objet (inconnu) qui dépend lui-même du contraste. Il est de plus connu pour être mal posé au sens de Hadamard, ce qui signifie que l'existence, l'unicité et la stabilité de sa solution ne sont pas simultanément garanties [42].

1.5.1 Les approches linéarisées

En ce qui concerne la non-linéarité, une première réflexion pour gérer ce problème est de le ramener à un problème linéaire et d'utiliser les techniques de résolution de systèmes linéaires classiques pour trouver la solution. C'est ce que permet l'approximation de Born au premier ordre, née en théorie de la diffusion des ondes et en mécanique quantique [43], qui consiste à ne tenir compte que du champ incident au lieu du champ total dans la relation qui lie le contraste au champ diffracté sur les points de mesure. Cependant, cette approximation n'est valable que pour des objets dont la taille est petite devant la longueur d'onde et le contraste faible et elle présente également d'autres limitations [44, 45]. Une autre approximation, celle de Rytov, consiste à rechercher une approximation de la phase complexe du champ diffracté dans le domaine de l'objet ; elle est généralement considérée comme moins contraignante que l'approximation de Born en ce qui concerne son domaine de validité, les limites portant plus ici sur les variations du contraste à l'intérieur de l'objet que sur ses dimensions [46]. Ces approximations linéarisantes, ainsi que leurs versions "distordues" [47, 48] permettant de considérer des milieux plus fortement contrastés, sont à la base des techniques dites de *tomographie par diffraction* [49, 50, 51], terme qui, selon [52], désigne « des applications qui utilisent des ondes diffractées dans le processus de reconstruction tomographique », mais qui sous-entendait alors généralement des processus de reconstruction basés sur le *théorème coupe-projection généralisé*, une extension au cas de la diffraction du *théorème coupe-projection* de la tomographie à rayons X classique dont le modèle direct est décrit par une transformée de Radon. Le terme *tomographie par diffraction* était alors paradoxalement utilisé pour désigner des techniques de reconstruction adaptées à des milieux faiblement diffringents et qui ne permettent pas d'obtenir des informations quantitatives sur les objets de fort contraste

diélectrique tels que ceux rencontrés dans les applications biomédicales, où la diffraction multiple ne peut être négligée. Il faut souligner que les techniques de tomographie par diffraction ne sont pas les seules méthodes d'inversion basées sur des approximations linéaires. En particulier, un champ de recherche tout-à-fait nouveau concerne les méthodes de régularisation des problèmes inverses en imagerie micro-onde [53, 54, 55, 56] basées sur la parcimonie au travers de l'acquisition comprimée (en anglais *compressed sampling* ou *compressive sensing* [57, 58]). Notons cependant que, en plus du fait qu'elles supposent que le problème inverse soit linéaire, ces méthodes nécessitent également que l'objet recherché soit parcimonieux, c'est-à-dire qu'il puisse être représenté par un petit nombre de coefficients lorsqu'il est projeté sur une base judicieusement choisie.

1.5.2 Les approches non linéaires

Dès le début des années 1990, de nombreux chercheurs se sont intéressés au développement d'algorithmes d'inversion capables de traiter le problème sous sa forme non linéaire. Le problème se transforme alors en un problème d'optimisation où la solution est recherchée de façon itérative jusqu'à satisfaction d'un critère d'arrêt ou en minimisant une fonctionnelle coût qui exprime l'écart entre les données et les champs diffractés calculés au moyen du modèle direct et de la solution courante. Notons que l'introduction d'informations *a priori* sur la solution recherchée dans la fonctionnelle coût à minimiser peut permettre de surmonter le caractère mal posé du problème inverse en restreignant l'espace des solutions admissibles. Les méthodes itératives peuvent se classer en deux grandes familles : les méthodes déterministes et les méthodes stochastiques.

Les approches déterministes

Avec les méthodes déterministes, l'optimisation d'un critère adéquat au problème est la clé importante pour trouver la solution recherchée. Parmi ces méthodes on peut citer de façon non exhaustive celles faisant partie de la gamme des méthodes dites de *linéarisations successives* telles que "les méthodes de Newton-Kantorovich" [59], de "Born itérative distordue" (DBIM) [60, 61, 62] et "de Newton inexacte" [63, 64, 65] ou encore "la méthode de Levenberg-Marquardt" [66]. On peut également citer les méthodes d'inversion basées sur une minimisation non linéaire de la fonctionnelle coût par une méthode de gradient et qui ne nécessitent pas la résolution du problème direct, telles que "la méthode du gradient modifiée" [67] ou celle de "l'inversion contraste-source" (CSI) [68]. Notons ici que cette dernière nous servira par la suite de méthode de référence pour évaluer les méthodes développées dans cette thèse. Elle consiste à rechercher alternativement à chaque itération le contraste et les sources induites (produit du contraste par le champ total à l'intérieur de l'objet) en minimisant une fonctionnelle coût tenant compte à la fois de l'équation d'observation et de l'équation de couplage à l'aide d'un algorithme de gradient conjugué. Cette méthode a déjà fourni des résultats intéressants en imagerie du cancer du sein [69, 70]. Elle a donné lieu à de nombreuses variantes permettant de rendre l'algorithme plus efficace notamment au travers de l'introduction de termes régularisants additifs ou multiplicatif dans la fonctionnelle à minimiser [71, 72, 73] et, finalement, elle a pu être interprétée d'une manière probabiliste comme une maximisation d'une loi *a posteriori* [74, 75]. Les méthodes déterministes sont généralement assez efficaces en ce sens qu'elles convergent assez rapidement vers une solution. Cepen-

dant, comme elles sont basées sur des techniques d'optimisation locales, elles présentent l'inconvénient de pouvoir se retrouver bloquées sur des solutions sous-optimales correspondant à des minima locaux de la fonction coût et elles nécessitent d'être initialisées dans des états proches de la solution recherchée. Tel n'est pas le cas des méthodes stochastiques qui permettent généralement de trouver le minimum global de la fonction coût car elles peuvent explorer tout l'espace des solutions admissibles.

Les approches stochastiques

Une des premières méthodes stochastiques à avoir été appliquées à l'imagerie micro-onde est certainement "la méthode du recuit simulé" [76, 77]. Puis sont venus les méthodes dites "évolutionnistes" [78, 79] censées reproduire l'évolution de systèmes observés dans la nature et qui consistent à travailler sur un ensemble de solutions possibles améliorées progressivement, plutôt que de n'en considérer qu'une seule. Ces dernières se répartissent en deux classes dont la première concerne des méthodes, telles que "les algorithmes génétiques" [80, 81] ou "les algorithmes d'évolution différentielle" [82, 83], qui reposent sur une analogie avec la théorie de l'évolution naturelle de Darwin et sont basées sur l'idée que les individus d'une population les mieux adaptés à leur environnement ont une grande probabilité de survivre et de se reproduire en donnant des descendants encore mieux adaptés, le processus de sélection naturelle causant l'évolution des espèces. La seconde classe d'algorithmes évolutionnistes regroupe les méthodes dites "coopératives", telles que "l'optimisation par essaims de particules" [84, 85] ou par "colonies de fourmis" [86, 87], qui sont basées sur une analogie avec de systèmes naturels dont le comportement est guidé par l'intelligence collective et l'échange d'informations au sein d'un groupe social.

Enfin une dernière classe de méthodes stochastiques concerne les méthodes probabilistes développées dans un cadre bayésien. C'est cette classe qui nous intéresse ici, c'est pourquoi nous allons nous y attarder un peu plus longuement. Pour pouvoir appliquer les méthodes de reconstruction probabilistes, il faut une mise en équations probabiliste du modèle direct afin de permettre la modélisation de tous les paramètres sous la bonne forme. Ainsi, il est nécessaire de tenir compte des bruits de mesure, de modélisation et de discrétisation que l'on introduit au niveau des équations d'observation et de couplage. Ensuite, nous définissons les distributions de probabilité des bruits et nous faisons appel à la règle de "Bayes" pour extraire la loi *a posteriori* des inconnues connaissant les données et le modèle. Notons également que la modélisation des informations *a priori* traduisant la connaissance de certaines caractéristiques de l'objet à reconstruire est une étape primordiale dans la modélisation bayésienne du problème inverse. En effet, son choix est décisif pour assurer une meilleure qualité de reconstruction tenant compte des informations disponibles sur l'objet recherché. Plusieurs travaux ont abordé cette question. Par exemple, dans [88], un modèle *a priori* markovien a été proposé. Ce même choix *a priori* est discuté dans [74]. Un modèle de Gauss-Markov-Potts avec un estimateur de l'espérance *a posteriori* basé sur l'échantillonnage stochastique est étudié dans [89], avec application en imagerie micro-onde 2D. Enfin, un modèle de Gauss-Markov-Potts, utilisant soit l'échantillonnage de Gibbs soit l'approche bayésienne variationnelle, a été proposé dans [90] avec application en tomographie optique de diffraction et en imagerie micro-onde 2D et 3D.

C'est ce dernier modèle qui est utilisé dans cette thèse. Il traduit l'information *a*

priori que l'objet recherché est composé d'un nombre fini de matériaux homogènes répartis en régions compactes. Cette information *a priori* peut paraître surprenante dans l'application considérée ici, car le sein est généralement constitué de tissus très hétérogènes qui montrent une grande variabilité entre les individus. Il a cependant été montré que ces tissus peuvent être divisés en quatre groupes différents : un pour les tumeurs [91] et trois pour les tissus normaux qui se distinguent par leur teneur en matières grasses [92], et les propriétés diélectriques moyennes de chacun de ces groupes peuvent être décrites par des modèles de Debye valables sur une très large bande de fréquence [93]. Ce que nous chercherons donc ici c'est non seulement à dresser une cartographie du contraste mais également à segmenter le sein en distinguant ces différents types de tissus. Le modèle *a priori* est donc constitué d'un mélange de gaussiennes, où chaque gaussienne représente un type de tissu, et d'un champ de Markov caché qui représente la segmentation du sein en différents tissus. Notons ici que si le nombre de tissus différents constituant le sein est supposé connu *a priori*, en revanche le contraste correspondant à chaque tissu lui n'est pas supposé connu ; il peut prendre des valeurs continues qui satisfont une distribution gaussienne dont la moyenne correspond à la valeur moyenne du contraste pour le tissu correspondant et dont l'écart type correspond aux fluctuations du contraste autour de cette moyenne. La situation est donc très différente de celle considérée dans [94] où l'objet recherché est homogène par morceaux et de contraste connu *a priori*, les différentes régions le constituant étant alors recherchées à l'aide d'une "méthode déterministe d'évolution contrôlée d'ensembles de niveaux".

Une fois le modèle *a priori* choisi, intuitivement l'estimation du contraste et des autres quantités inconnues du modèle peut se faire à travers le maximum a posteriori (MAP) ou l'espérance *a posteriori* (EAP). Cependant, l'expression analytique de ces estimateurs peut être difficilement calculable, comme c'est le cas en imagerie micro-onde. C'est pourquoi la loi *a posteriori* doit être approchée pour obtenir la solution recherchée avec un coût de calcul acceptable. Deux familles d'approximations sont connues :

1. "Les méthodes d'échantillonnage stochastique de Monte-Carlo par Chaîne de Markov" (MCMC)
   L'échantillonnage stochastique consiste à approcher une loi de probabilité (loi *a posteriori*) par des échantillons numériques et à les utiliser par la suite pour calculer des moments de la loi considérée ou estimer les valeurs correspondant aux maxima à l'aide d'un histogramme. Plusieurs approches d'échantillonnage stochastique existent dans la littérature [95], telles que "l'échantillonnage d'importance", "l'échantillonnage de rejet", "l'échantillonnage imbriqué", "l'échantillonnage de Metropolis-Hastings" et "l'échantillonnage de Gibbs". Chaque technique a ses propres limitations ; par exemple pour les deux premières méthodes, une difficulté réside dans le choix de la distribution qui approche bien la loi d'intérêt pour les problèmes de grandes dimensions. La troisième méthode est complexe et coûteuse en temps de calcul car elle nécessite l'échantillonnage de la loi *a priori* sous contrainte de vraisemblance, en plus de l'échantillonnage de la loi *a posteriori*. Les deux dernières méthodes semblent les plus adaptées pour la résolution du problème inverse d'imagerie micro-onde. L'échantillonneur de Gibbs a été déjà exploité pour ce problème ainsi qu'en imagerie optique de diffraction [89, 90]. Des résultats intéressants ont été obtenus mais le temps de calcul et la complexité de l'algorithme sont importants.
2. "Les méthodes bayésiennes variationnelles" (BV)

La méthode bayésienne variationnelle (apprentissage dans un ensemble) utilisée ici, née en physique statistique pour une application en réseaux de neurones [96], a été introduite comme une alternative aux méthodes d'échantillonnage stochastiques MCMC. L'idée ici est d'approcher la distribution *a posteriori* jointe des inconnues par une densité séparable minimisant la divergence de Kullback-Leibler entre ces deux lois. La solution du problème peut ensuite être obtenue analytiquement, de sorte que l'on s'attend à ce que la vitesse de convergence de l'algorithme BV soit bien meilleure que celle des approches stochastiques MCMC.

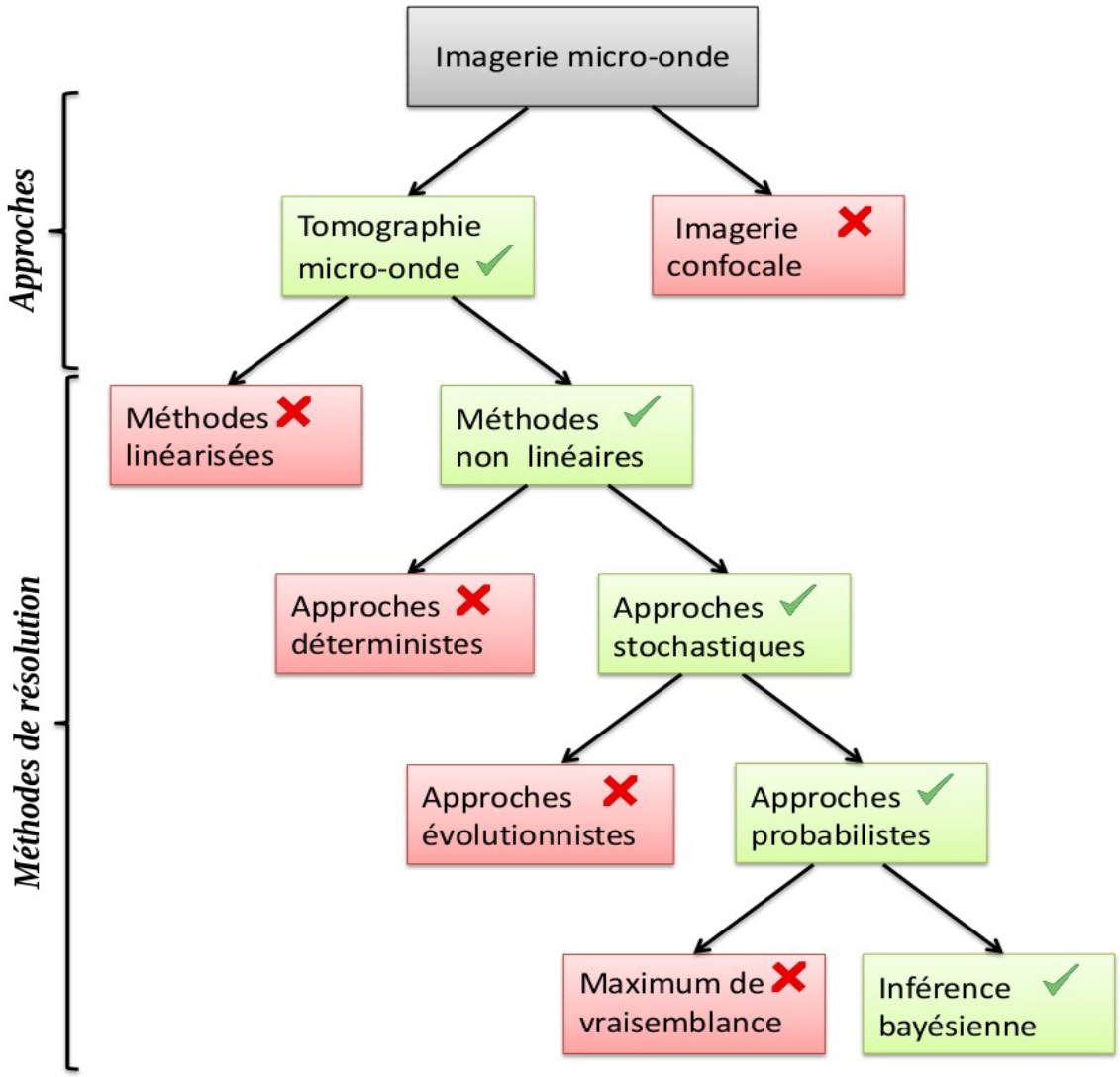


FIGURE 1.5 – Schéma de la réflexion sur les approches de l'imagerie micro-onde et les méthodes de résolution du problème inverse.

1.6 Cadre du travail

Les travaux présentés dans cette thèse étaient l'occasion d'un rapprochement, au sein du Laboratoire des Signaux et Systèmes (L2S), de deux communautés qui, même si

elles s'intéressent toutes deux aux "problèmes inverses", les abordent cependant avec des points de vue, des outils et des langages bien différents. En effet, d'un côté, le Groupe Problèmes Inverses (GPI) de la Division Signaux aborde les problèmes inverses dans un cadre stochastique avec des outils bayésiens et un point de vue de "traitement de signal", et de l'autre, le Département de Recherche en Électromagnétisme (DRÉ) privilégie un cadre déterministe avec une vision plus proche de la "physique des ondes".

Sur le plan technique, les fantômes numériques de sein utilisés dans cette thèse pour tester les approches proposées sont issus, pour une partie d'entre eux, d'une collaboration informelle avec les universités de Bristol (UK) et de Mälardalen (Suède) et l'Université Polytechnique de Catalogne et, pour l'autre partie, de la base de données de l'Université du Wisconsin-Madison [97].

1.7 Plan de l'étude et contributions

Dans le chapitre 2, nous exposons le problème direct qui consiste en la détermination du champ diffracté en supposant l'objet et le champ incident connus. Nous présentons alors les équations qui modélisent ce problème d'abord dans un cadre continu, puis dans un cadre discret. Nous présentons ensuite la configuration de mesure considérée durant la thèse ainsi que les différents modèles étudiés, partant du plus simple jusqu'au plus sophistiqué. Enfin, nous discuterons la validation du modèle direct en présentant une comparaison des résultats obtenus à l'aide de deux modèles différents, dont l'un est utilisé pour la génération des données du problème inverse et l'autre est utilisé lors de l'inversion.

Le chapitre 3 est, pour sa part, consacré à la modélisation du problème inverse en se plaçant dans un cadre probabiliste bayésien. Nous mettons l'accent sur la prise en compte de l'information *a priori* qui conduit à une segmentation par régions *via* un modèle de Gauss-Markov-Potts. Nous montrons ensuite que cette information *a priori* est introduite de façon aisée dans le processus d'inversion, ce qui confert à l'approche bayésienne un avantage certain par rapport a beaucoup d'autres méthodes d'inversion. Cependant, les expressions analytiques des estimateurs classiques souvent utilisés pour retrouver la loi jointe *a posteriori* des inconnues (maximum *a posteriori* et moyenne *a posteriori*) ne sont pas toujours exploitables et la loi jointe *a posteriori* doit être approchée pour obtenir la solution désirée avec un coût de calcul raisonnable.

Le chapitre 4 est dédié à la présentation des approches bayésiennes étudiées et développées durant cette thèse pour la résolution du problème inverse. Dans un premier temps, nous présentons la technique bayésienne variationnelle déjà développée en imagerie optique de diffraction et adaptée ici au cas de l'imagerie micro-onde. La nouveauté par rapport aux travaux précédents [90] est que le contraste recherché est à valeurs complexes. Ainsi, il s'agit de reconstruire ici non seulement une carte de permittivité de l'objet mais également une carte de conductivité et il faut adapter l'information *a priori* à la fois à la partie réelle et à la partie imaginaire du contraste.

Le chapitre 5 présente les contributions majeures, aussi bien méthodologiques qu'algorithmiques, apportées durant cette thèse : l'application des approches bayésiennes, discutées dans le chapitre précédent, aux modèles de seins supposés atteints par des tumeurs. Les résultats de reconstruction sont exposés et discutés en les comparant es-

sentiellement aux résultats donnés par la méthode déterministe d'inversion contraste-source.

Enfin, dans le chapitre 6 nous concluons sur les techniques bayésiennes utilisées et sur l'apport de nos travaux et nous donnons quelques perspectives pour la poursuite de ceux-ci.

2

Le problème direct

Sommaire

2.1 Introduction 27
2.2 Représentation intégrale de domaine du champ électrique . 28
2.3 Modèle discret 31
2.3.1 Méthode des Moments 31
2.3.2 Application de la méthode des moments au domaine intégral 32
2.4 Génération des données 34
2.5 Configuration de mesure 36
2.6 Validation des modèles 39
2.7 Conclusions 39

2.1 Introduction

Lors de la résolution d'un problème inverse, la modélisation du problème direct associé est une étape initiale et primordiale qui permet de décrire comment les paramètres du modèle se traduisent en effets observables expérimentalement. Le problème direct consiste ici à modéliser la physique de l'interaction entre une onde interrogatrice connue et l'objet supposé également connu. Cette interaction est décrite par l'équation des ondes de Helmholtz. En appliquant le théorème de Green à cette équation et en tenant compte des conditions de continuité des champs et de rayonnement à l'infini [39], nous sommes conduits à une représentation intégrale du champ électrique constituée de deux équations intégrales couplées dites équation d'observation et équation de couplage (ou d'état), respectivement. La résolution du problème directe nécessite les contreparties discrètes de ces équations intégrales qui sont obtenues dans un cadre algébrique à l'aide de "la méthode des moments" [41]. Une validation expérimentale du modèle direct est une étape nécessaire avant de passer à l'inversion. Il s'agit en effet de vérifier

que le modèle direct décrit bien les phénomènes (dans notre cas les champs diffractés) observés expérimentalement en situation contrôlée. Notons qu'ici nous ne disposons pas de données expérimentales. C'est pourquoi, lors de la résolution du problème inverse, nous générerons des données synthétiques à l'aide d'un modèle direct différent de celui utilisé lors de l'inversion, ce qui nous permettra d'éviter de commettre un "crime inverse" au sens de [98] qui consisterait à tester l'algorithme d'inversion sur les données obtenues au moyen d'un modèle étroitement apparenté à celui utilisé dans l'inversion. La validation du modèle direct consistera donc ici à vérifier que les différences entre les champs diffractés produits par ces deux modèles restent faibles.

2.2 Représentation intégrale de domaine du champ électrique

Équation de propagation de l'onde électromagnétique

Dans la suite nous considérons des milieux homogènes, non magnétiques, linéaires et isotropes et nous nous plaçons dans un mode d'imagerie 2D en configuration transverse magnétique (TM) où l'on considère que l'objet est d'extension infinie et invariant selon l'un de ses axes (symétrie cylindrique) et qu'il est éclairé par une onde dont le champ électrique ($\vec{\boldsymbol{E}}^{inc} = \boldsymbol{E}_z^{inc}\vec{z}$) est polarisé parallèlement à cet axe (perpendiculaire au plan de la figure), ce qui engendre un champ diffracté perpendiculaire également au plan de la figure, d'où une écriture scalaire du modèle direct.

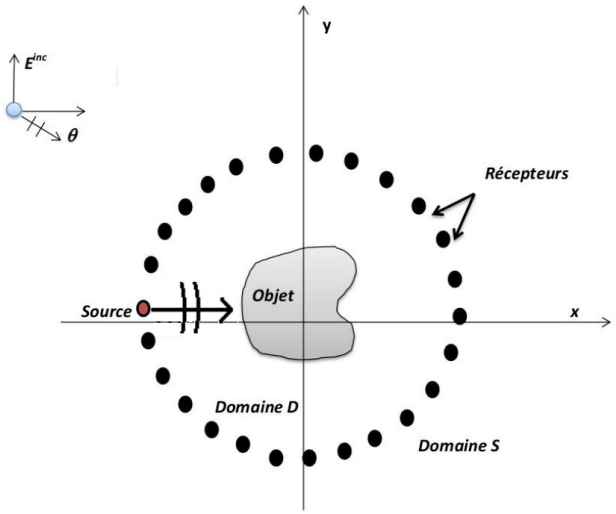


FIGURE 2.1 – La configuration en imagerie micro-onde bi-dimensionnelle en polarisation transverse magnétique (2D-TM).

Nous établissons maintenant l'équation de propagation de l'onde électromagnétique à partir des équations de Maxwell qui constituent la base de l'électromagnétisme et qui s'écrivent en un point d'observation $\boldsymbol{r}$ et au temps t implicites :

$$\begin{cases} \nabla . \boldsymbol{D} &= \rho \ : \text{ équation de Maxwell-Gauss,} \\ \nabla . \boldsymbol{B} &= 0 \ : \text{ équation de conservation du flux,} \\ \nabla \times \boldsymbol{E} &= -\dfrac{\partial \boldsymbol{B}}{\partial t} \ : \text{ équation de Maxwell-Faraday,} \\ \nabla \times \boldsymbol{H} &= \boldsymbol{J} + \dfrac{\partial \boldsymbol{D}}{\partial t} \ : \text{ équation de Maxwell-Ampère,} \end{cases} \tag{2.1}$$

où ∇ est l'opérateur de dérivation par rapport aux coordonnées d'espace, ρ la densité volumique de charge électrique, $\boldsymbol{J}$ la densité de courant, $\boldsymbol{D}$ le déplacement (induction) électrique, $\boldsymbol{B}$ l'induction magnétique, $\boldsymbol{E}$ le champ électrique et $\boldsymbol{H}$ le champ magnétique. A ceci s'ajoutent les équations constitutives reliant champs et inductions :

$$\begin{cases} \vec{D} &= \epsilon \boldsymbol{E}, \\ \vec{B} &= \mu \boldsymbol{H}, \end{cases} \tag{2.2}$$

où ϵ est la permittivité diélectrique et μ la perméabilité magnétique du milieu de propagation.

On se place maintenant dans le cadre d'un régime harmonique où une source émet une onde de pulsation ω et l'on choisit une dépendance temporelle implicite des champs en $\exp\{-\mathrm{i}\omega\mathrm{t}\}$ que l'on omettra dans la suite de l'exposé et qui conduit à remplacer les dérivations par rapport au temps apparaissant dans les équations de Maxwell par un facteur $-\mathrm{i}\omega$. Pour établir l'équation de propagation en champ électrique on se place maintenant dans un milieu sans sources ni charges. A partir des équations de Maxwell et en tenant compte des relations constitutives, on arrive alors à :

$$\nabla \times \nabla \times \boldsymbol{E} - \omega^2 \epsilon \mu \boldsymbol{E} = 0. \tag{2.3}$$

En développant l'opérateur vectoriel apparaissant dans cette équation, en se souvenant que le champ électrique n'a qu'une composante selon l'axe Oz que nous écrirons E et en introduisant la constante de propagation k du milieu considéré ($k = \omega\sqrt{\epsilon\mu}$), nous obtenons l'équation des ondes de Helmholtz scalaire vérifiée par le champ électrique :

$$\nabla^2 E(\boldsymbol{r}) + k^2 E(\boldsymbol{r}) = 0, \tag{2.4}$$

où $\boldsymbol{r}$ est le point d'observation ($\boldsymbol{r} = (x, y)$).

Nous supposerons maintenant que l'objet auquel nous nous intéressons est plongé dans un milieu homogène $\mathcal{D}_1$ et est contenu dans un domaine test $\mathcal{D}$ ($\mathcal{D} \subset \mathcal{D}_1$). Les différents milieux sont non magnétiques (de perméabilité magnétique égale à celle du vide $\mu_0 = 1,256 \times 10^{-6}\,\mathrm{H\,m^{-1}}$) et sont caractérisés par leur constantes de propagation k_m ($m = 1$ ou $\mathcal{D}$) telle que $k(\boldsymbol{r})^2 = \omega^2 \epsilon_0 \epsilon_r(\boldsymbol{r}) \mu_0 + i\omega\mu_0\sigma(\boldsymbol{r})$, où ϵ_0 est la permittivité diélectrique du vide ($\epsilon_0 = 8,854 \times 10^{-12}\,\mathrm{F\,m^{-1}}$), tandis que $\epsilon_r(\boldsymbol{r})$ et $\sigma(\boldsymbol{r})$ représentent la permittivité relative et la conductivité au point d'observation $\boldsymbol{r}$. Nous définissons maintenant une fonction contraste normalisée définie dans $\mathcal{D}$ et nulle en dehors de

l'objet : $\boldsymbol{\chi}(\boldsymbol{r}) = (k(\boldsymbol{r})^2 - k_1^2)/k_0^2$, où k_0 est la constante de propagation dans le vide ($k_0^2 = \omega^2 \epsilon_0 \mu_0$).

En appliquant le théorème de Green à l'équation (2.4) et en tenant compte des conditions de continuité des champs et de rayonnement à l'infini, nous sommes conduits à une représentation intégrale du champ électrique constituée de deux équations intégrales couplées, dites équation d'observation et équation de couplage (ou d'état). Elles sont données respectivement par :

$$y(\boldsymbol{r}) = E^{dif}(\boldsymbol{r}) = k_0^2 \int_{\mathcal{D}} G(\boldsymbol{r}, \boldsymbol{r}')\, \chi(\boldsymbol{r}')\, E(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}', \qquad \boldsymbol{r} \in \mathcal{S}, \tag{2.5}$$

$$E(\boldsymbol{r}) = E^{inc}(\boldsymbol{r}) + k_0^2 \int_{\mathcal{D}} G(\boldsymbol{r}, \boldsymbol{r}')\, \chi(\boldsymbol{r}')\, E(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}', \qquad \boldsymbol{r} \in \mathcal{D}, \tag{2.6}$$

où $\boldsymbol{E}^{inc}$ est le champ incident, i.e. le champ observé en l'absence d'objet, $\boldsymbol{E}^{dif}$ est le champ diffracté observé sur le domaine de mesure $\mathcal{S}$, $\boldsymbol{E}$ est le champ total dans l'objet et $G(\boldsymbol{r}, \boldsymbol{r}')$ est la fonction de Green qui est la solution élémentaire de l'équation d'onde scalaire de Helmholtz et qui représente le rayonnement d'une source ponctuelle placée au point $\boldsymbol{r}'$ observé au point $\boldsymbol{r}$ en l'absence d'objet. En milieu homogène, dans le cas bi-dimensionnel qui nous concerne, elle est donnée par :

$$G(\boldsymbol{r}, \boldsymbol{r}') = \frac{i}{4} H_0^1 \left(k_1 |\boldsymbol{r} - \boldsymbol{r}'|\right), \tag{2.7}$$

où H_0^1 est la fonction de Hankel d'ordre 0 et de première espèce.

Formulation contraste source

Nous réécrivons maintenant les équations d'observation et de couplage en faisant apparaître les sources de Huyghens w induites à l'intérieur de l'objet par l'onde incidente ($w(\boldsymbol{r}') = \chi(\boldsymbol{r}')E(\boldsymbol{r}')$) :

$$y(\boldsymbol{r}) = E^{dif}(\boldsymbol{r}) = k_0^2 \int_{\mathcal{D}} G(\boldsymbol{r}, \boldsymbol{r}')\, w(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}', \tag{2.8}$$

$$E(\boldsymbol{r}) = E^{inc}(\boldsymbol{r}) + k_0^2 \int_{\mathcal{D}} G(\boldsymbol{r}, \boldsymbol{r}')\, w(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}'. \tag{2.9}$$

La résolution des équations (2.8) et (2.9) est effectuée à partir de leurs contreparties discrètes obtenues à l'aide de la méthode des moments. Le domaine $\mathcal{D}$ contenant l'objet inconnu est alors discrétisé en $N_{\mathcal{D}}$ pixels élémentaires carrés suffisamment petits pour que l'on puisse considérer le champ électrique et le contraste comme constants sur chacun d'eux (figure (2.2)).

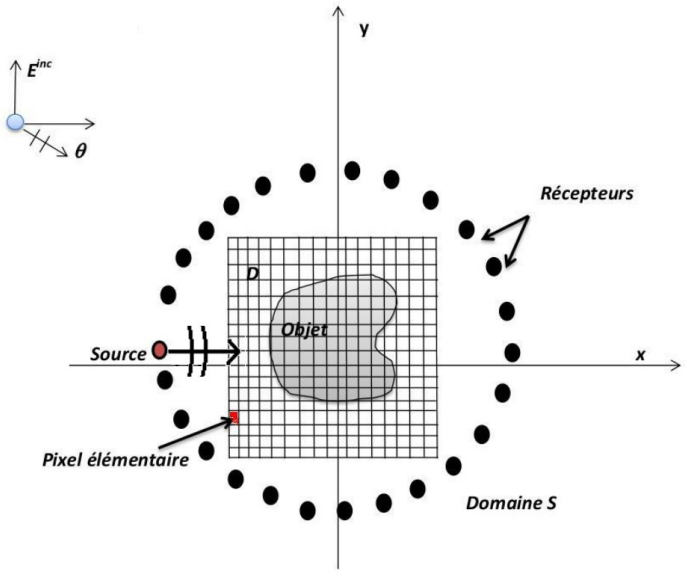


FIGURE 2.2 – Discrétisation du domaine test en pixels élémentaires.

2.3 Modèle discret

2.3.1 Méthode des Moments

La méthode des moments est une méthode couramment utilisée pour transformer des équations fonctionnelles en équations matricielles par projection sur des ensembles de fonctions de base et de fonctions de test [41]. Le problème étant réduit sous la forme $\boldsymbol{AX} = \boldsymbol{B}$, où $\boldsymbol{A}$ est une matrice, $\boldsymbol{X}$ est un vecteur inconnu et $\boldsymbol{B}$ est un vecteur connu, intuitivement, quand la matrice $\boldsymbol{A}$ est inversible, cela nous permet d'obtenir les solutions recherchées.

Principe de la méthode

Soient un opérateur linéaire $\mathcal{L} : \mathcal{G} \rightarrow \mathcal{H}$, et deux fonctions $g \in \mathcal{G}$ et $h \in \mathcal{H}$ telles que $h = \mathcal{L}(g)$. On cherche à déterminer g connaissant $\mathcal{L}$ et h. La fonction g peut être écrite sous la forme $g = \sum_{i=1..I} g_i u_i$, où $\{u_1, u_2 ... u_I\}$ sont des coefficients constants qui constituent un ensemble de fonctions de base u_I. La propriété de linéarité de l'opérateur $\mathcal{L}$ nous donne :

$$h = \mathcal{L}(g) = \mathcal{L}(\sum_{i=1}^{I} g_i u_i) = \sum_{i=1}^{I} g_i \mathcal{L} u_i. \tag{2.10}$$

Le but étant d'approcher l'équation (2.10) par un système linéaire discret, on définit un ensemble de fonctions test $\{t_1, t_2 ... t_N\}$ sur lequel on projette les deux membres de cette équation :

$$\langle t_n, h\rangle_{\mathcal{H}} = \sum_{i=1}^{I} g_i \langle t_n, \mathcal{L}u_i\rangle_{\mathcal{H}}, \quad \forall n = 1, ..., N, \tag{2.11}$$

où $\langle ., .\rangle_{\mathcal{H}}$ est le produit scalaire sur $\mathcal{H}$ tel que $\langle t_n, h\rangle_{\mathcal{H}} = \int_{\mathcal{H}} t_n(x)h^*(x)dx$.

Ainsi, en omettant l'indice $\mathcal{H}$, on obtient l'écriture algébrique $\boldsymbol{h} = \boldsymbol{Lg}$ avec :

$$\boldsymbol{L} = \begin{bmatrix} \langle t_1, \mathcal{L}u_1\rangle & \langle t_1, \mathcal{L}u_2\rangle & \cdots & \langle t_1, \mathcal{L}u_I\rangle \\ \langle t_2, \mathcal{L}u_1\rangle & \langle t_2, \mathcal{L}u_2\rangle & \cdots & \langle t_2, \mathcal{L}u_I\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle t_n, \mathcal{L}u_1\rangle & \langle t_n, \mathcal{L}u_2\rangle & \cdots & \langle t_n, \mathcal{L}u_I\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle t_N, \mathcal{L}u_1\rangle & \langle t_N, \mathcal{L}u_2\rangle & \cdots & \langle t_N, \mathcal{L}u_I\rangle \end{bmatrix}, \quad \boldsymbol{g} = \begin{bmatrix} g_1 \\ \vdots \\ g_i \\ \vdots \\ g_I \end{bmatrix}, \quad \boldsymbol{h} = \begin{bmatrix} \langle t_1, h\rangle \\ \vdots \\ \langle t_n, h\rangle \\ \vdots \\ \langle t_N, h\rangle \end{bmatrix}.$$

Le choix des fonctions de base et des fonctions de test dépend essentiellement de la nature du problème physique à traiter. Notons que si l'on choisit des fonctions test identiques aux fonctions de base, on tombe sur le cas particulier de la "méthode de Galerkin" [40].

2.3.2 Application de la méthode des moments au domaine intégral

Les deux équations couplées (2.5) et (2.6) présentent une propriété de bilinéarité par rapport aux deux variables $\boldsymbol{\chi}$ et $\boldsymbol{E}$. La méthode des moments peut alors être appliquée sur chacune des équations. En effet, si on considère les équations d'observation et d'état comme deux équations linéaires par rapport à la variable $\boldsymbol{E}$, les équations (2.5) et (2.6) peuvent s'écrire comme suit :

$$E^{dif}(\boldsymbol{r}) = k_0^2 \int_{\mathcal{D}} G(\boldsymbol{r}, \boldsymbol{r}')\, \chi(\boldsymbol{r}')\, E(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}' \Rightarrow \boldsymbol{E}^{dif} = \mathcal{L}_1\, \boldsymbol{E}, \tag{2.12}$$

$$E^{inc}(\boldsymbol{r}) = E(\boldsymbol{r}) - k_0^2 \int_{\mathcal{D}} G(\boldsymbol{r}, \boldsymbol{r}')\, \chi(\boldsymbol{r}')\, E(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}' \Rightarrow \boldsymbol{E}^{inc} = \mathcal{L}_2 \boldsymbol{E}, \tag{2.13}$$

où $\mathcal{L}_1$ et $\mathcal{L}_2$ sont deux opérateurs linéaires.

Ensuite pour pouvoir appliquer le principe de la méthode des moments, il faut choisir les fonctions de base $\{u_{1i}\}$ et $\{u_{2i}\}$ et les fonctions tests $\{t_{1i}\}$ et $\{t_{2i}\}$ afin de construire les produits scalaires :

$$\left\langle \int_{\mathcal{D}} k_0^2 G(\boldsymbol{r}, \boldsymbol{r}')\chi(\boldsymbol{r}')u_{1j}(\boldsymbol{r}')d\boldsymbol{r}', t_{1i} \right\rangle_{\mathcal{S}} = \mathcal{L}_{1ij} \tag{2.14}$$

et

$$\left\langle 1 - \int_{\mathcal{D}} k_0^2 G(\boldsymbol{r}, \boldsymbol{r}')\chi(\boldsymbol{r}')u_{2j}(\boldsymbol{r}')d\boldsymbol{r}', t_{2i} \right\rangle_{\mathcal{D}} = \mathcal{L}_{2ij}. \tag{2.15}$$

Ici, nous prendrons comme fonctions de base les fonctions caractéristiques des pixels partitionnant le domaine $\mathcal{D}$ et comme fonctions test des distributions de Dirac δ localisées aux centres de chacun des pixels élémentaires pour l'équation de couplage et aux points de mesure pour l'équation d'observation.

Ainsi, si on note $\boldsymbol{r}_i$ le centre du pixel élémentaire μ_i et si l'on suppose que le contraste χ et le champ électrique E sont constants sur chaque pixel μ_i et égaux à $\chi(\boldsymbol{r}_i)$ et $E(\boldsymbol{r}_i)$ respectivement, les éléments de la matrice de couplage s'écrivent :

$$\mathcal{L}_{2ij} = \delta_{ij} - \chi(\boldsymbol{r}_j) \int_{\mu_j} k_0^2 G(\boldsymbol{r}_i, \boldsymbol{r}') \mathrm{d}\boldsymbol{r}', \tag{2.16}$$

où δ_{ij} est le symbole de Kronecker.

Il s'agit maintenant d'intégrer la fonction de Green sur un pixel élémentaire. Notons qu'une approximation analytique de cette intégrale peut être obtenue en remplaçant le pixel carré par un disque de même surface [40] :

$$\boldsymbol{G}_{ij} = \int_{\mu_j} k_0^2 G(\boldsymbol{r}_i, \boldsymbol{r}') \mathrm{d}\boldsymbol{r}' = \begin{cases} \dfrac{i\pi k_0^2 \Delta}{2k_1} H_0^1\left(k_1 |\boldsymbol{r}_i - \boldsymbol{r}_j|\right) J_1(k_1 \Delta) \text{ si } \boldsymbol{r}_i \neq \boldsymbol{r}_j \\ \dfrac{i\pi k_0^2 \Delta}{2k_1} H_1^1\left(k_1 \Delta\right) - 1 \text{ si } \boldsymbol{r}_i = \boldsymbol{r}_j, \end{cases} \tag{2.17}$$

où J_1 et H_1^1 sont respectivement les fonctions de Bessel et de Hankel de premier ordre et de première espèce et $\Delta = a/\sqrt{\pi}$, où a est le côté de la cellule élémentaire.

Ainsi les équations (2.12) et (2.13) s'écrivent sous la forme algébrique suivante :

$$\boldsymbol{E}^{dif} = \boldsymbol{G}^o \boldsymbol{X} \boldsymbol{E}, \tag{2.18}$$

$$\boldsymbol{E}^{inc} = \boldsymbol{E} - \boldsymbol{G}^c \boldsymbol{X} \boldsymbol{E}, \tag{2.19}$$

où $\boldsymbol{X} = \mathrm{diag}(\boldsymbol{\chi})$ et $\boldsymbol{G}^o$ et $\boldsymbol{G}^c$ sont deux matrices (les exposants o et c indiquant respectivement les matrices d'observation et de couplage) de dimensions égales respectivement à $N_\mathcal{D} \times N_\mathcal{S}$ et $N_\mathcal{D} \times N_\mathcal{D}$, où $N_\mathcal{S}$ est le nombre de points de mesure.

Les dimensions de ces matrices peuvent devenir rapidement importantes dès que le domaine $\mathcal{D}$ dépasse quelques longueurs d'onde. En effet, la convergence des calculs menés à l'aide de la méthode des moments impose une limite supérieure à la taille des pixels à considérer [99, 100, 101]. Des pixels de côté $a \leq \lambda/10$ assurent généralement une bonne convergence des calculs. Réécrivons maintenant les équations (2.18) et (2.19) en tenant compte du fait que les mesures sont effectuées à plusieurs fréquences (indice f) et pour plusieurs positions de la source (indice ϑ) et en introduisant un indice n tenant compte de ces deux dernières ($n = (f, \vartheta)$) :

$$\boldsymbol{E}_n^{dif} = \boldsymbol{G}_f^o \ \boldsymbol{X}_f \boldsymbol{E}_n \,, \tag{2.20}$$

$$\boldsymbol{E}_n^{inc} = (\mathbf{I}_d - \boldsymbol{G}_f^c \boldsymbol{X}_f)\ \boldsymbol{E}_n, \tag{2.21}$$

où $\mathbf{I}_d$ est la matrice identité.

La résolution du problème direct consiste donc, dans un premier temps, à résoudre l'équation (2.21) pour trouver le champ $\boldsymbol{E}_n$, le champ diffracté $\boldsymbol{E}_n^{dif}$ s'en déduisant directement, dans un second temps, à l'aide de l'équation (2.20). La résolution de l'équation (2.21) nécessite l'inversion de la matrice $(\mathbf{I}_d - \boldsymbol{G}_f^c \boldsymbol{X}_f)$ pour chaque fréquence f, ce qui ne pose pas de problème particulier, cette matrice étant généralement bien conditionnée. Cependant, les dimensions de cette dernière pouvant être très importantes, les méthodes d'inversion directes, de type factorisation $\boldsymbol{LU}$ ou élimination de Gauss-Jordan, sont à proscrire car peu efficaces pour la résolution des systèmes de très grande taille et difficiles à mettre en œuvre car elles nécessitent le stockage en mémoire de ces derniers. C'est pourquoi nous avons opté pour un algorithme itératif de gradient bi-conjugué stabilisé (BiCGSTAB), une généralisation naturelle du gradient conjugué. Notons ici que la nature convolutionnelle de l'équation de couplage permet de travailler dans le domaine de Fourier, le calcul des opérateurs étant alors effectué à l'aide de l'algorithme BiCGSTAB combiné avec une transformée de fourrier rapide (BiCGSTAB-FFT). La complexité de cet algorithme est de l'ordre de $O(N \log N)$ et $O(N)$, où N est le rang du système, en termes de temps de calcul et de coût en mémoires respectivement, ce qui est bien plus économique qu'une solution directe [102].

2.4 Génération des données

Comme signalé précédemment, afin d'éviter de commettre un crime inverse, les données du problème inverse sont calculées à partir d'un modèle différent de celui utilisé dans l'inversion. Alors que dans ce dernier, exposé ci-dessus, l'ensemble du sein est partionné en pixels élémentaires et la fonction de Green utilisée est celle d'une configuration 2D en milieu homogène donnée par l'équation (2.7), pour calculer les données seules les inhomogénéités (e.g., $\mathcal{D}_4$, $\mathcal{D}_5$ et $\mathcal{D}_6$) sont discrétisées et la fonction de Green utilisée est celle d'un milieu stratifié cylindrique à deux (fantôme 1 : $\mathcal{D}_1$ et $\mathcal{D}_3$) ou trois couches (fantômes 2, 3 ou 4 : $\mathcal{D}_1$, $\mathcal{D}_2$ et $\mathcal{D}_3$) [39]. Dans ce dernier cas, les matrices $\boldsymbol{G}^c$ et $\boldsymbol{G}^o$ deviennent :

$$G_{ij}^c = k_0^2\,[\,G_{ij}^s + G_{ij}^{ns}\,], \tag{2.22}$$

où

$$G_{ij}^s = \begin{cases} \dfrac{i\pi\Delta}{2\ k_3} J_1(k_3\Delta)\ H_0^1(k_3|\boldsymbol{r}_j - \boldsymbol{r}_i|) & \text{si } \boldsymbol{r}_j \neq \boldsymbol{r}_i, \\ \dfrac{i\pi\Delta}{2\ k_3}\ H_1^1(k_3\Delta) - \dfrac{1}{k_3^2} & \text{si } \boldsymbol{r}_j = \boldsymbol{r}_i, \end{cases} \tag{2.23}$$

$$G_{ij}^{ns} = \sum_{m=0}^{+\infty} \epsilon_m\ \frac{2\pi\Delta}{k_3}\ J_1(k_3\Delta)\ A_m\ J_m(k_3 r_i)\ J_m(k_3 r_j)\ cos(m(\phi_j - \phi_i))$$

et

$$G^o_{ij} = k_0^2 \sum_{m=0}^{+\infty} \epsilon_m \ \frac{2\pi\Delta}{k_3} \ J_1(k_3\Delta) \ B_m \ H^1_m(k_1 r_i) \ J_m(k_3 r_j) \ cos(m(\phi_j - \phi_i)), \tag{2.24}$$

où $(r_i,\, \phi_i)$ et $(r_j,\, \phi_j)$ sont les coordonnées cylindriques du point d'observation $\boldsymbol{r}_i$ et du point source $\boldsymbol{r}_j$, respectivement, et :

$$\epsilon_m = \begin{cases} 1 \text{ si } m = 0 \\ \\ 2 \text{ si } m \neq 0, \end{cases} \tag{2.25}$$

tandis que les coefficients A_m et B_m sont donnés par :

$$\begin{aligned} A_m &= -\frac{i}{4} \ \frac{H^1_m(k_3\alpha)}{J_m(k_3\alpha)} + \ \frac{(\gamma^m_{11} \ H^1_m(k_2\alpha) \ - \ \gamma^m_{12} \ J_m(k_2\alpha))}{2\pi k_3\alpha \ (\gamma^m_{11} \ \gamma^m_{22} - \gamma^m_{12} \ \gamma^m_{21}) \ J_m(k_3\alpha)}, \\ B_m &= \frac{(\gamma^m_{11} \ H^1_m(k_2\beta) \ - \ \gamma^m_{12} \ J_m(k_2\beta))}{2\pi k_3\alpha \ (\gamma^m_{11} \ \gamma^m_{22} - \gamma^m_{12} \ \gamma^m_{21}) \ H^1_m(k_1\beta)}, \end{aligned}$$

où α est le rayon du domaine $\mathcal{D}_3$, β le rayon extérieur du domaine $\mathcal{D}_2$ et les coefficients γ^m_{11}, γ^m_{12}, γ^m_{21} et γ^m_{22} sont égaux respectivement à :

$$\begin{aligned} \gamma^m_{11} &= \left[J_m(k_2\beta) \ \dot{H}^1_m(k_1\beta) - \frac{k_2}{k_1} \dot{J}_m(k_2\beta) \ H^1_m(k_1\beta) \right], \\ \gamma^m_{12} &= \left[H^1_m(k_2\beta) \ \dot{H}^1_m(k_1\beta) - \frac{k_2}{k_1} \dot{H}^1_m(k_2\beta) \ H^1_m(k_1\beta) \right], \\ \gamma^m_{21} &= \left[J_m(k_2\alpha) \ \dot{J}_m(k_3\alpha) - \frac{k_2}{k_3} \dot{J}_m(k_2\alpha) \ J_m(k_3\alpha) \right], \\ \gamma^m_{22} &= \left[H^1_m(k_2\alpha) \ \dot{J}_m(k_3\alpha) - \frac{k_2}{k_3} \dot{H}^1_m(k_2\alpha) \ J_m(k_3\alpha) \right], \end{aligned}$$

où $\dot{J}(.)$ et $\dot{H}(.)$ désignent les dérivées des fonctions de Bessel et Hankel par rapport à leur argument.

Le champ incident $\boldsymbol{E}^{inc}(\boldsymbol{r}_j)$ créé par une source placée en $\boldsymbol{r}_\vartheta = (r_\vartheta,\, \phi_\vartheta)$ dans le domaine $\mathcal{D}_1$ et observé aux centres des pixels $\boldsymbol{r}_j = (r_j,\, \phi_j)$ à l'intérieur du domaine $\mathcal{D}_3$ est, pour sa part, donné par :

$$\boldsymbol{E}^{inc}(\boldsymbol{r}_j) = \sum_{m=0}^{+\infty} \epsilon_m \ C_m J_m(k_3 r_j) H^1_m(k_1 r_\vartheta) cos(m(\phi_j - \phi_\vartheta)), \tag{2.26}$$

où

$$C_m = \ \frac{(\gamma^m_{21} \ H^1_m(k_2\alpha) \ - \ \gamma^m_{22} \ J_m(k_2\alpha))}{2\pi k_1\beta \ (\gamma^m_{11} \ \gamma^m_{22} - \gamma^m_{12} \ \gamma^m_{21}) \ J_m(k_3\alpha)}. \tag{2.27}$$

Notons que les séries infinies apparaissant ci-dessus convergent assez rapidement et une vingtaine de termes suffit, dans les exemples traités ici, à assurer une bonne précision des calculs. Une fois ces substitutions effectuées, le reste des calculs s'effectue comme décrit précédemment. Insistons sur le fait que les champs diffractés calculés de cette façon servirons de données au problème inverse et seront, par la suite, désignés par "les données".

2.5 Configuration de mesure

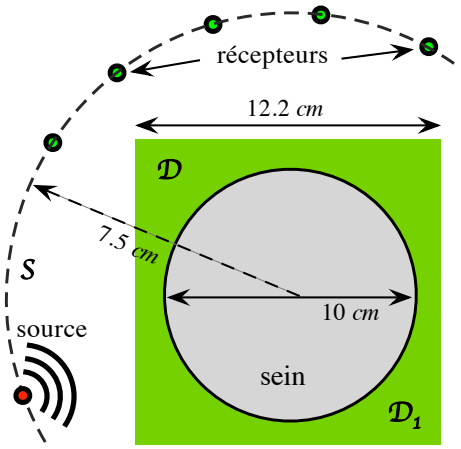


FIGURE 2.3 – Configuration de mesure

Pour effectuer les tests, quatre fantômes de seins placés dans trois milieux d'adaptations $\mathcal{D}_1$ différents ont été considérés (figure (2.4)). Les paramètres des fantômes 1, 2 et 3 ont été définis dans le cadre d'une collaboration informelle avec les universités de Bristol et de Mälardalen et l'Université Polytechnique de Catalogne afin de disposer d'objets de référence pour la comparaison des algorithmes d'inversion. Pour ces fantômes, les différents milieux sont distribués dans des domaines de sections circulaires et sont constitués de graisse (domaine $\mathcal{D}_3$), d'une tumeur (domaine $\mathcal{D}_4$) et, éventuellement de peau (domaine $\mathcal{D}_2$) et de milieux glandulaires (domaines $\mathcal{D}_5$ et $\mathcal{D}_6$). Le fantôme 4 est plus sophistiqué ; il est toujours supposé être de section circulaire, entouré de peau et atteint d'une tumeur, mais le milieu $\mathcal{D}_3$ est un milieu très hétérogène, dont les paramètres varient entre ceux de la graisse et ceux de milieux fibro-glandulaires, construit à partir d'une coupe d'un sein réel obtenue par imagerie par résonance magnétique [97]. Le tableau (2.1) récapitule les différents paramètres électromagnétiques des milieux composants les quatre configurations. Pour faciliter les notations, on associe aux quatre fantômes l'indice i qui indique le milieu d'adaptation ($i = 0, 1, 2, 3, 4$) où $i = 0$ correspond à l'espace libre $(\epsilon_r\,;\sigma) = (1\,;0)$, $i = 1$ correspond au milieu de paramètres $(\epsilon_r\,;\sigma) = (10\,;0,5)$, $i = 2$ correspond au milieu de paramètres $(\epsilon_r;\sigma) = (21,8\,;0,5)$ et $i = 3$ correspond au milieu de paramètres $(\epsilon_r;\sigma) = (35\,;0,5)$.

Finalement, la configuration de mesure est rappelée sur la figure (2.3). On dispose de 64 lignes-sources et de 64 récepteurs régulièrement répartis sur un cercle de rayon $7,5$ cm centré sur l'origine des axes. Le sein est de section circulaire de diamètre égal à 10 cm et le domaine test $\mathcal{D}$ est un carré de 12,2 cm de côté discrétisé en 64×64 pixels carrés de côté égal à $a = 1,9$ mm pour les fantômes 1, 2 et 3 et en 120×120 pixels de côté égal

à $a = 1$ mm pour le fantôme 4. Le dispositif fonctionne à 6 fréquences dans la bande $0,5$ - 3 GHz. Notons ici que, dans la réalité, les différents tissus composant le sein sont généralement fortement dispersifs [93], i.e. leurs permittivité et conductivité dépendent de la fréquence. C'est un aspect que, par souci de simplicité, nous n'aborderons pas dans ce travail. Les paramètres donnés dans la table (2.1) seront donc pris comme constants sur toute la bande de fréquence considérée.

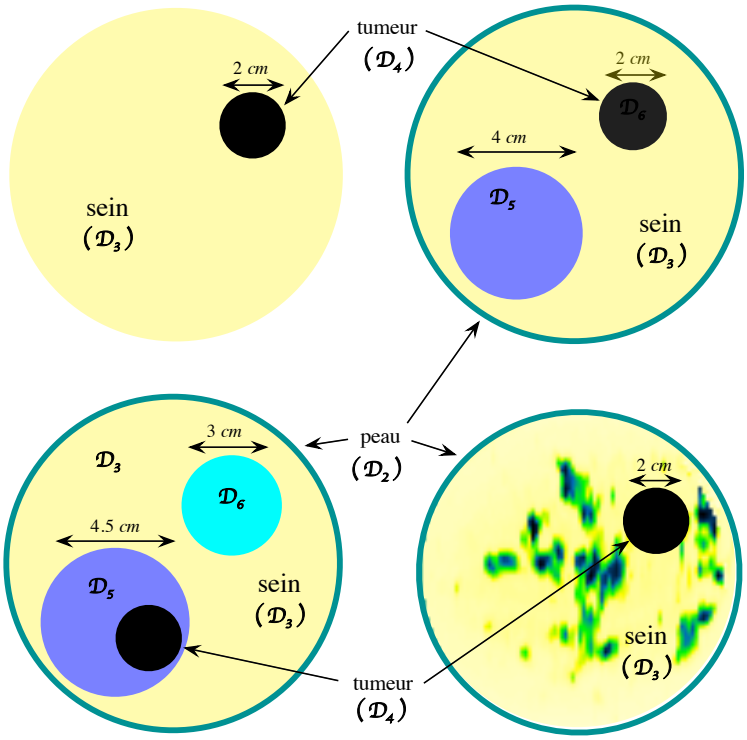


FIGURE 2.4 – Les différents fantômes de sein considérés : fantôme 1 (en haut à gauche), fantôme 2 (en haut à droite), fantôme 3 (en bas à gauche) et fantôme 4 (en bas à droite).

TABLE 2.1 – Les paramètres électromagnétiques des différents milieux et des différents fantômes

Fantômes \ Milieux		$\mathcal{D}_1$			$\mathcal{D}_2$	$\mathcal{D}_3$	$\mathcal{D}_4$	$\mathcal{D}_5$	$\mathcal{D}_6$
Fantôme 1	ϵ_r	10	21	35		6,12	55,3		
	$\sigma(\mathrm{Sm}^{-1})$	0,5	0,5	0,5		0,11	1,57		
	$\varnothing$ (mm)					96	20		
Fantôme 2	ϵ_r	10	21	35	35,7	6,12	55,3	42	
	$\sigma(\mathrm{Sm}^{-1})$	0,5	0,5	0,5	0,32	0,11	1,57	1,11	
	$\varnothing$ (mm)				100	96	20	45	
Fantôme 3	ϵ_r	10	21	35	35,7	6,12	55,3	21,8	42
	$\sigma(\mathrm{Sm}^{-1})$	0,5	0,5	0,5	0,32	0,11	1,57	0,5	1,11
	$\varnothing$ (mm)				100	96	20	45	30
Fantôme 4	ϵ_r	10	21	35	38	[2,46 - 60,6]	55,3		
	$\sigma(\mathrm{Sm}^{-1})$	0,5	0,5	0,5	1,46	[0,01 - 2,28]	1,57		
	$\varnothing$ (mm)				100	96	20		

Algorithme de validation du modèle direct

1 – Entrées :
- $\boldsymbol{N}_S$: vecteur des angles de réception,
- $\boldsymbol{N}_\vartheta$: vecteur des angles d'incidence,
- $\boldsymbol{N}_f$: vecteur des fréquences,
- $\boldsymbol{M}$: matrice des données (tenseur de répartition : $\boldsymbol{M} = \boldsymbol{N}_f \times \boldsymbol{N}_S \times \boldsymbol{N}_\vartheta$),
- paramètres du milieu d'adaptation,
- permittivité diélectrique relative ϵ_r et conductivité σ

de chaque milieu intervenant dans la configuration étudiée.

2 – Calculer :
- le champ incident $\boldsymbol{E}^{inc}$ en absence d'objet aux centres des pixels grâce à l'équation (2.21),
- la fonction du contraste $\boldsymbol{\chi} = k(\boldsymbol{r})^2 - k_1^2)/k_0^2$ en tenant compte du domaine d'intérêt,
- les matrices d'observation $\boldsymbol{G}^o$ et de couplage $\boldsymbol{G}^c$ grâce à l'équation (2.17),
- le champ total dans l'objet $\boldsymbol{E}$ grâce à l'algorithme BiCGSTAB-FFT,
- le champ diffracté $\boldsymbol{y} = \boldsymbol{G}^o \boldsymbol{X} \boldsymbol{E}$.

3 – Comparer :
- les données $\boldsymbol{M}$ à la sortie du modèle $\boldsymbol{y}$ pour toutes les fréquences,

incidences et observations.

4 – Conclure :
- quant à la validation du modèle.

2.6 Validation des modèles

Nous commençons par une validation préliminaire en simulant le problème direct pour un angle de réception et une fréquence donnés. Ensuite on calcule le champ incident en l'absence d'objet et on génère le contraste de l'objet. Puis on procède au calcul des champs diffractés pour toutes les incidences, toutes les fréquences et tous les angles de réception. Ensuite, on compare la sortie de chaque modèle avec les données. D'après les figures (2.5), (2.6), (2.7) et (2.8) nous remarquons que les sorties des modèles coïncident relativement bien avec les données, aussi bien en module qu'en phase, ce que confirment les tableaux (2.2) et (2.3) qui montrent l'erreur relative entre les deux champs calculée pour toutes les positions de source et de récepteur pour chaque fréquence. Cette dernière est généralement faible mais croît avec la fréquence. Ceci s'explique par le fait que la discrétisation, i.e. la taille des pixels, est constante quelle que soit la fréquence. Hors nous avons vu qu'une bonne convergence des calculs menés à l'aide de la méthode des moments nécessite d'indexer cette taille à la longueur d'onde. Par conséquent, plus la fréquence augmente, plus l'erreur liée aux contributions du milieu $\mathcal{D}_2$ et du bord du milieu $\mathcal{D}_3$, décrits exactement pour le calcul des données mais discrétisés dans le modèle direct, devient importante.

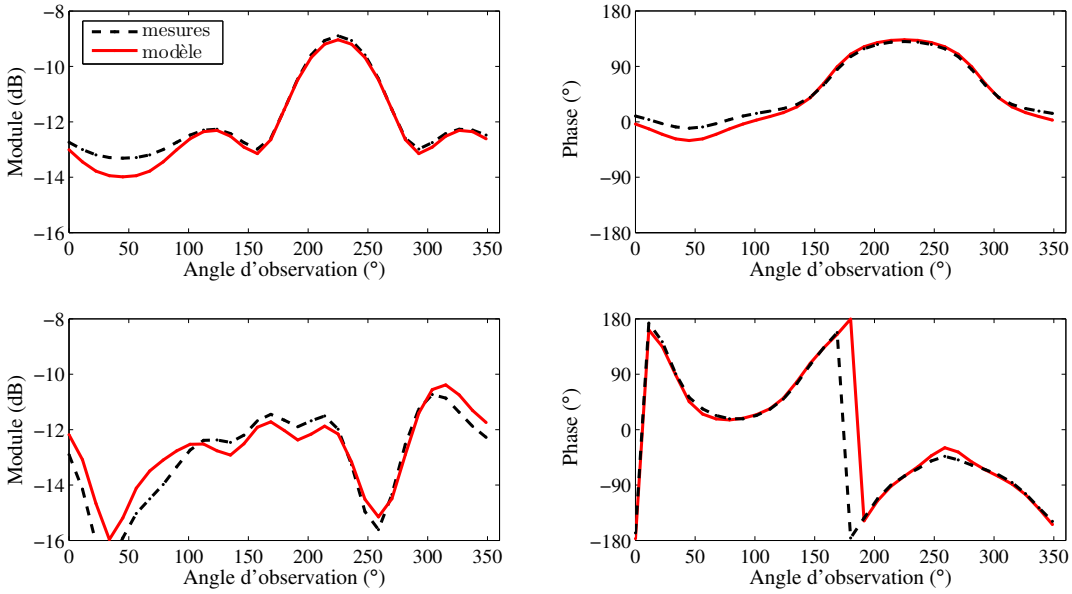


FIGURE 2.5 – Validation du modèle direct pour le fantôme 1-0 pour deux fréquences et deux positions de source différentes : 45° et 1,5 GHz (haut) et 90° et 2,5 GHz (bas). La sortie du modèle (en noir discontinu) est comparée aux données (en rouge continu).

2.7 Conclusions

Dans ce chapitre, nous avons présenté la modélisation du problème direct où, onde incidente et objet étant connus, on s'intéresse au calcul du champ diffracté résultant

TABLE 2.2 – L'erreur relative entre les données et la sortie du modèle pour le fantôme 1-0 en fonction de la fréquence.

Fréquence (GHz)	0,5	1	1,5	2	2,5	3
Erreur %	0,54	1,11	1,69	3,65	5,5	3,02

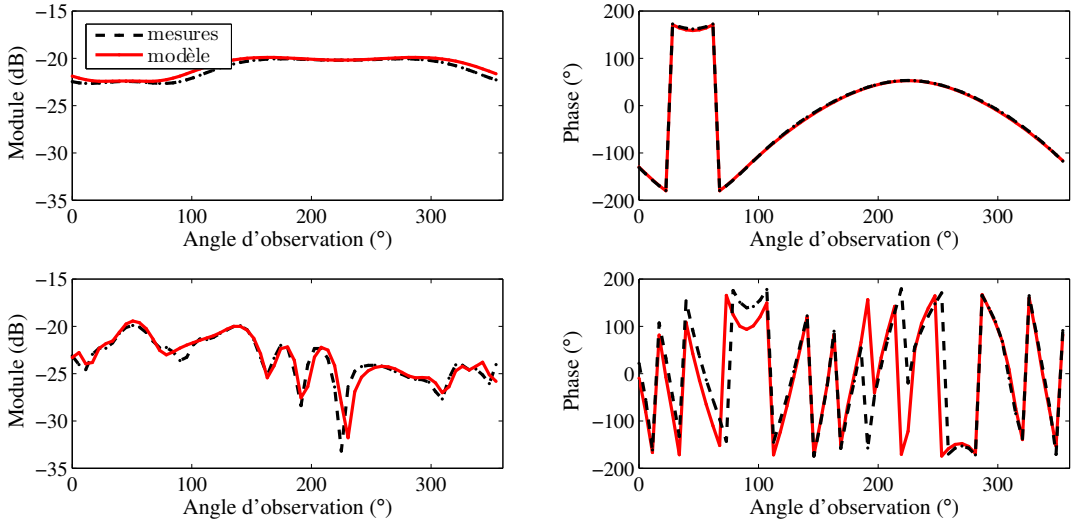


FIGURE 2.6 – Validation du modèle direct pour le fantôme 1-3 pour deux fréquences et deux positions de source différentes : 45° et 1,5 GHz (haut) et 90° et 3 GHz (bas). La sortie du modèle (en noir discontinu) est comparée aux données (en rouge continu).

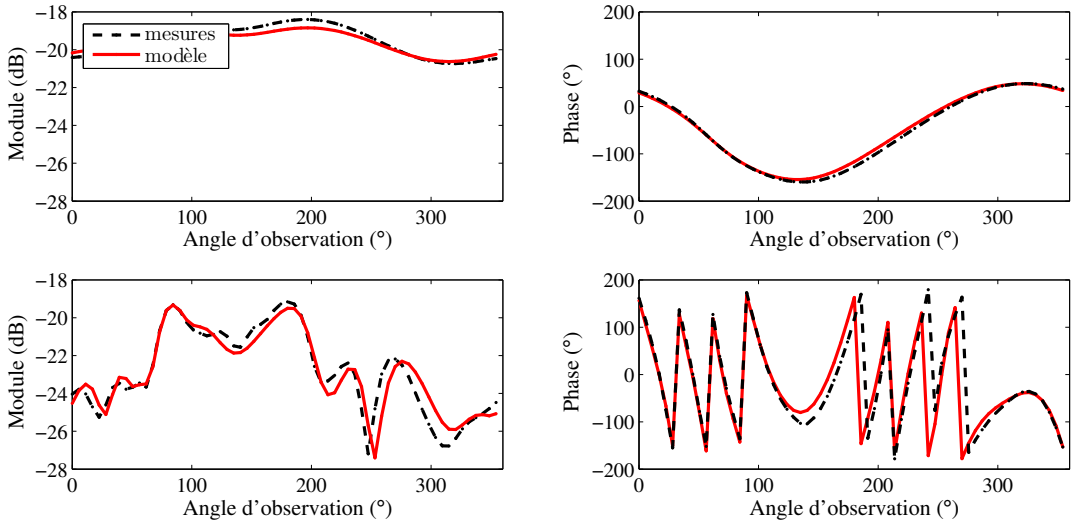


FIGURE 2.7 – Validation du modèle direct pour le fantôme 2-2 pour une position de source de 135° et deux fréquences différentes : 0,5 GHz (haut) et 2,5 GHz (bas). La sortie du modèle (en noir discontinu) est comparée aux données (en rouge continu).

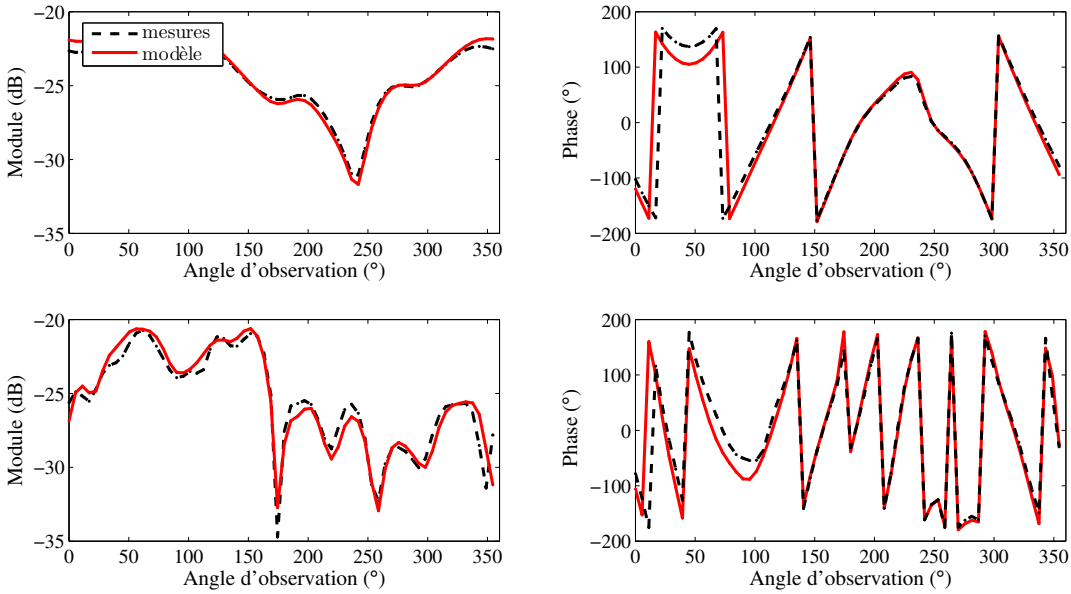


FIGURE 2.8 – Validation du modèle direct pour le fantôme 3-2 pour deux fréquences et deux positions de source différentes : 45° et 1 GHz (haut) et 90° et 3 GHz (bas). La sortie du modèle (en noir discontinu) est comparée aux données (en rouge continu).

TABLE 2.3 – L'erreur relative entre les données et la sortie du modèle pour le fantôme 3-2 en fonction de la fréquence.

Fréquence (GHz)	0,5	1	1,5	2	2,5	3
Erreur %	0,64	0,95	2,12	3,26	6,25	3,96

de leur interaction. La résolution de ce dernier correspond à une première étape indispensable à la résolution du problème inverse où il s'agit de reconstruire l'objet inconnu, l'onde incidente et le champ diffracté étant alors connus. En l'absence de données expérimentales, nous avons validé les calculs de champs électromagnétiques réalisés à l'aide de ce modèle direct par comparaison avec des résultats synthétiques obtenus à l'aide d'un modèle direct différent de celui utilisé pour l'inversion afin d'éviter de commettre un crime inverse. Finalement, nous avons insisté sur la nécessité de développer un code de calcul rapide, car susceptible d'être appelé de façon répétitive lors de l'inversion, et capable de traiter des objets relativement grands devant la longueur d'onde.

3

Inférence bayésienne en imagerie micro-onde

Sommaire

3.1 Introduction 43
3.2 Mise en équation probabiliste du modèle direct 44
3.3 Modèle *a priori* hiérarchique 45
3.3.1 Modèle de Mélange de Gaussiennes conditionnellement Indépendantes (MGI) 48
3.3.2 Modèle de Mélange de Gauss-Markov (MGM) 48
3.4 Cadre quasi non-supervisé 50
3.4.1 Estimation conjointe 52
3.4.2 Estimateur $\mathcal{J}^{MAP}$ (optimisation alternée par point fixe) . . . 52
3.4.3 Discussion 56
3.5 Conclusions 57

3.1 Introduction

Dans le chapitre précédent, nous avons présenté le modèle direct qui nous permet de calculer le champ diffracté en supposant que le contraste et le champ total à l'intérieur de l'objet sont connus. Dans ce chapitre, nous nous intéressons à la formulation du problème inverse, c'est-à-dire la reconstruction du contraste de l'objet à partir des mesures de champ diffracté, dans un cadre probabiliste bayésien.

Le cadre de l'inférence bayésienne, né avec Thomas Bayes et Pierre-Simon de Laplace (figure (3.1)) au cours du $18^{ème}$ siècle, nous fournit un outil important pour la résolution des problèmes inverses. En effet, il permet une introduction aisée d'informations *a priori* adaptées au problème posé, ce qui est très important pour la modélisation probabiliste de tous les paramètres afin d'obtenir un estimateur conjoint de toutes les inconnues qui soit le plus proche possible de la solution désirée.

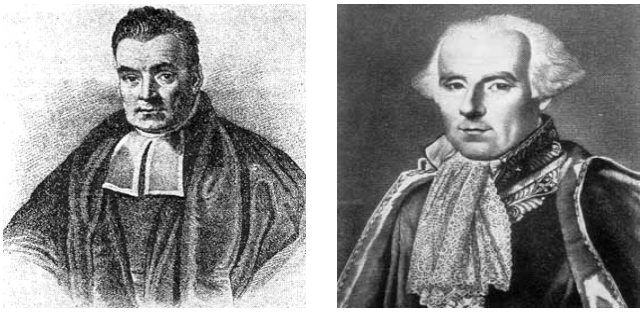


FIGURE 3.1 – Thomas Bayes (1701-1761) (gauche, source : Wikipédia) et Pierre-Simon de Laplace (1749-1827) (droite, source : site de l'Université de Caen http://www.math.unicaen.fr/ reyssat/laplace/).

Les probabilités conditionnelles introduisant le raisonnement de Bayes et sa formulation peuvent être vus à travers l'exemple qui suit, i.e., le dépistage du cancer. On peut dire que dans ce problème nous avons, d'un côté, une hypothèse : "j'ai le cancer" et, de l'autre côté, une observation : "je suis testé positif". Quand on souhaite connaître "la probabilité d'avoir le cancer sachant que l'on est positif", on est en train de chercher la probabilité que l'hypothèse soit vraie étant donnée notre observation. Si on réécrit la règle de Bayes de manière plus abstraite, en considérant une hypothèse H et une observation O, on a :

$$P(H|O) = \frac{P(O|H)}{P(O)} P(H), \tag{3.1}$$

où $P(H|O)$ est la loi *a posteriori*, $P(O|H)$ est la vraisemblance du modèle, $p(O)$ est la loi *a priori* des observations et $p(H)$ la loi *a priori* de l'hypothèse.

La formule de Bayes est aujourd'hui utilisée dans nombreux domaines d'application très concrets. Ici, nous l'utilisons en mode continu où elle constitue le fondement de nos travaux actuels.

3.2 Mise en équation probabiliste du modèle direct

Tout d'abord, une mise en équation probabiliste du modèle direct est nécessaire pour pouvoir appliquer les lois de l'inférence bayésienne. Ainsi, réécrivons les équations d'observation et de couplage en faisant apparaître les sources induites et en introduisant, sous la forme de deux vecteurs $\boldsymbol{\epsilon}$ et $\boldsymbol{\xi}$, deux variables de bruit qui tiennent compte des incertitudes de mesure et des erreurs de modèle :

$$\begin{aligned} \boldsymbol{y}_n &= \boldsymbol{E}_n^{dif} = \boldsymbol{G}_f^o \boldsymbol{w}_n + \boldsymbol{\epsilon}_n, \\ \boldsymbol{w}_n &= \boldsymbol{X}_f \boldsymbol{E}_n^{inc} + \boldsymbol{X}_f \boldsymbol{G}_f^c \boldsymbol{w}_n + \boldsymbol{\xi}_n. \end{aligned} \tag{3.2}$$

Ensuite, définissons les distributions de probabilité de ces variables considérées

comme des bruits blancs gaussiens :

$$p(\boldsymbol{\epsilon}_n) = \mathcal{N}(\boldsymbol{\epsilon}_n|\mathbf{0}, v_\epsilon \mathbf{I}_n) = \frac{1}{(2\pi v_\epsilon)^{\frac{N_\mathcal{S}}{2}}} \exp\left\{-\frac{||\boldsymbol{\epsilon}_n||^2_\mathcal{S}}{2v_\epsilon}\right\}, \tag{3.3}$$

$$p(\boldsymbol{\xi}_n) = \mathcal{N}(\boldsymbol{\xi}_n|\mathbf{0}, v_\xi \mathbf{I}_n) = \frac{1}{(2\pi v_\xi)^{\frac{N_\mathcal{D}}{2}}} \exp\left\{-\frac{||\boldsymbol{\xi}_n||^2_\mathcal{D}}{2v_\xi}\right\}, \tag{3.4}$$

où $N_\mathcal{S}$ et $N_\mathcal{D}$ sont respectivement le nombre de points de mesure et le nombre de pixels partitionnant le domaine test.

Ainsi, connaissant l'équation d'observation et $p(\boldsymbol{\epsilon}_n)$, on peut déduire $p(\boldsymbol{y}|\boldsymbol{\chi}, v_\epsilon)$ et, connaissant l'équation de couplage et $p(\boldsymbol{\xi}_n)$, on peut déduire $p(\boldsymbol{w}|\boldsymbol{\chi}, v_\xi)$. L'étape suivante consiste à déterminer l'information *a priori* adéquate au modèle considéré. D'après la règle de Bayes nous avons :

$$p(\boldsymbol{\chi}, \boldsymbol{w}|\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{w}, v_\epsilon) p(\boldsymbol{w}|\boldsymbol{\chi}, v_\xi) p(\boldsymbol{\chi}), \tag{3.5}$$

où $p(\boldsymbol{\chi})$ est la loi *a priori* du contraste de l'objet recherché, que l'on définit par la suite.

3.3 Modèle *a priori* hiérarchique

Les problèmes inverses de diffraction sont connus pour être mal posés au sens de Hadamard [103], ce qui signifie qu'ils n'admettent pas de solution unique stable. Ainsi, ils doivent être régularisés, ce qui se fait généralement par introduction d'une information *a priori* sur l'objet recherché. Dans une approche déterministe, l'information *a priori* est généralement introduite au travers d'un critère de régularisation [59], ce qui implique plusieurs inconvénients :

— un choix limité car il introduit généralement de la "souplesse" dans la solution,
— la détermination souvent empirique d'un paramètre de régularisation et
— le manque d'outils pour quantifier les incertitudes subsistant sur la solution.

Par contre, les approches bayésiennes ne montrent pas de telles limitations et elles nous permettent d'introduire une information *a priori* adaptée au problème d'une manière très facile en traduisant le contenu de l'objet d'intérêt d'une manière probabiliste.

En effet, dans de nombreux domaines d'application de la tomographie, l'objet est connu pour être composé d'un nombre fini de matériaux homogènes distribués en régions compactes. C'est le cas, par exemple, en contrôle non-destructif où l'objet sous test est généralement un objet artificiel composé de matériaux différents comme de l'air, du métal ou des matériaux composites. C'est aussi parfois le cas en imagerie médicale, certaine parties du corps étant composées par exemple de muscles, d'os, de matière blanche ou de matière grise. La situation est quelque part similaire pour le sein. En effet, bien qu'il soit constitué de tissus très hétérogènes qui montrent une grande variabilité entre les individus, ces tissus peuvent être divisés en quatre groupes différents, un pour les tumeurs [91] et trois pour les tissus normaux qui se distinguent par leur teneur en matières grasses [92], et les propriétés diélectriques moyennes de chacun de ces groupes

peuvent être décrites par des modèles de Debye valables sur une très large bande de fréquence [93].

C'est donc cette information *a priori* que nous voulons introduire ici : le sein est constitué d'un nombre limité de tissus différents (en l'occurrence 4) répartis en régions homogènes et compactes. Cette information *a priori* est introduite à l'aide d'un modèle *a priori* de Gauss-Markov-Potts [104]. Les champs de Markov sont, en effet, un outil utile pour modéliser une dépendance locale contextuelle, comme par exemple ici la dépendance entre les pixels [105]. Ce modèle permet d'introduire une variable cachée représentant une classification (l'appartenance à un type de tissu) et modélisée par un champ de Potts à nombre de classes fixe.

Décrivons maintenant ce modèle. Soient $\mathcal{R}$ l'ensemble des pixels et $\mathcal{V}(\boldsymbol{r})$ un ensemble de pixels voisins du pixel $\boldsymbol{r}$, que l'on choisit comme étant composé des 4 plus proches voisins (figure (3.3)). On associe à chaque pixel $\chi(\boldsymbol{r})$ une variable cachée discrète $z(\boldsymbol{r})$ $\in \boldsymbol{Z} = \{z(\boldsymbol{r}), \boldsymbol{r} \in \mathcal{R}\}$. Chaque variable cachée détermine le tissu (la classe) auquel le pixel appartient. Comme le sein est supposé être composé d'un nombre fini de tissus, ces variables prennent leur valeurs dans l'ensemble fini $E = \{e_1, ..., e_K\}$ où K est le nombre de classes dans l'image segmentée et e_K est le vecteur de dimension K dont la composante numéro k vaut 1 et les autres 0. La figure (3.2) est une représentation 1D de l'association de la variable discrète cachée $\boldsymbol{z}$ au contraste $\boldsymbol{\chi}$. Dans un premier cas (cas a), la corrélation entre les variables est exclusive pour le champ caché : les variables $\boldsymbol{\chi}$ sont conditionnellement indépendantes, ce qui correspond à un champ de Markov caché dans un cas 2D. Dans le second cas (cas b), il existe une corrélation, d'une part, entre les variables $\boldsymbol{\chi}$ et, d'autre part, entre les variables cachées $\boldsymbol{z}$; ce modèle permet par exemple de considérer les textures en segmentation d'image.

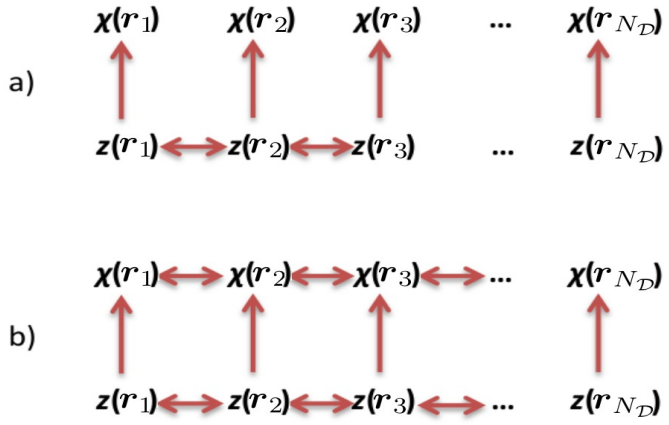


FIGURE 3.2 – Représentation de l'association de la variable discrète cachée $\boldsymbol{z}$ à $\boldsymbol{\chi}$: a) les $\boldsymbol{\chi}$ sont indépendants et les $\boldsymbol{z}$ sont corrélés (processus markovien), b) les $\boldsymbol{\chi}$ et les $\boldsymbol{z}$ suivent un processus markovien.

En plus de la segmentation, on souhaite prendre en compte la dépendance entre les pixels d'une même classe de telle sorte qu'ils se trouvent dans des régions compactes. Cette propriété peut être modélisée par un *a priori* de Potts. Rappelons que le champ de Potts est un cas particulier de champ de Markov. Ce dernier est défini comme suit : soient un ensemble $\mathcal{A} \in \mathbb{R}^2$ et un ensemble de variables aléatoires $\boldsymbol{X} = (X_i,\ i \in \mathcal{A})$, alors $\boldsymbol{X}$ est un champ de Markov pour le voisinage $\mathcal{V}$ si la probabilité conditionnelle de

X_i, sachant toutes les variables $X_{\mathcal{A}-\{i\}}$, ne dépend que des valeurs prises par les variables au voisinage de X_i.

Le champs de Potts est défini par la relation élémentaire suivante :

$$p(z(\boldsymbol{r})|z(\boldsymbol{r}')) \;=\; p(z(\boldsymbol{r})|z_{\mathcal{V}_{\boldsymbol{r}}}) \propto \exp\left\{\lambda \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right\}, \tag{3.6}$$

où λ est le paramètre du modèle, δ est la fonction de Kronecker et $\sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})}$ est la somme de toutes les paires de pixels voisins du pixel $\boldsymbol{r}$.

La version la plus simple de modèle de Potts est le modèle isotrope homogène sans champ externe, dont la distribution de probabilité, pour une réalisation $\boldsymbol{z}$ de $\boldsymbol{Z}$, est donnée par la forme simplifiée suivante :

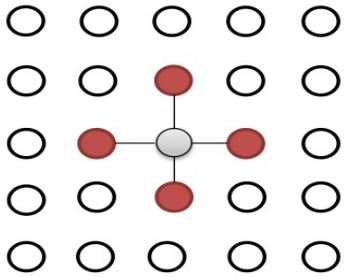


FIGURE 3.3 – Notion de voisinage dans une structure : les quatre pixels rouges forment le voisinage $\mathcal{V}$ du pixel gris.

$$p(\boldsymbol{z}|\lambda) = \frac{1}{T(\lambda)} \exp\left\{\lambda \sum_{\boldsymbol{r}\in\mathcal{R}} \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} \delta\left(z(\boldsymbol{r}) - z(\boldsymbol{r}')\right)\right\}, \tag{3.7}$$

où $T(\lambda)$ est une constante de normalisation qui correspond à la fonction de partition du paramètre λ :

$$T(\lambda) = \sum_k \exp\left\{\lambda \sum_{\boldsymbol{r}\in\mathcal{R}} \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} \delta\left(z(\boldsymbol{r}) - z(\boldsymbol{r}')\right)\right\}. \tag{3.8}$$

Notons que le choix du paramètre de corrélation λ détermine le comportement *a priori* du champ caché et, donc, la taille des régions homogènes de l'image finale : plus λ est grand, plus les zones homogènes sont grandes. Cependant, son estimation est difficile car $T(\lambda)$ n'a pas de forme explicite et la somme est incalculable. Par conséquent, λ est réglé à une valeur légèrement supérieure à une valeur critique $\lambda_c = \log(1+\sqrt{K})$, ce qui donne un bon compromis dans le degré de corrélation [106]. En effet, pour les $\lambda \leq \lambda_c$, le modèle de Potts produit une réalisation avec un grand nombre de petites régions, tandis que, pour les $\lambda \geq \lambda_c$, il produit le contraire, c'est-à-dire un petit nombre de grandes régions.

La seconde clé de l'information *a priori* est l'homogénéité de chaque classe k, ce qui peut être exprimé par une loi de Gauss. En effet, ce choix fait intervenir le contraste moyen du tissu m_k (valeur moyenne de la gaussienne) et la fluctuation autour de cette moyenne v_k (variance de la gaussienne) :

$$p(\chi(\boldsymbol{r})|z(\boldsymbol{r})=k)=\mathcal{N}(\chi(\boldsymbol{r})|m_k,v_k), \qquad k=1,\ldots,K. \tag{3.9}$$

Ainsi le contraste recherché est un mélange de gaussiennes, où chaque loi gaussienne représente un matériau donné, et ses valeurs font partie d'un ensemble continu.

Pour estimer la moyenne m_k et la variance v_k, deux choix sont possibles : i) le mélange de gaussiennes conditionnellement indépendantes (MGI) où les pixels sont *a priori* indépendants et ii) le mélange de Gauss-Markov (MGM) qui est un modèle markovien décrivant la dépendance d'un pixel $\chi(\boldsymbol{r})$ vis-à-vis de ses voisins sur une même région k.

3.3.1 Modèle de Mélange de Gaussiennes conditionnellement Indépendantes (MGI)

Dans ce modèle, l'interaction spatiale entre les pixels d'une même région est simple puisqu'ils sont considérés comme conditionnellement indépendants. Ainsi, chaque pixel $\chi(\boldsymbol{r})$ suit une loi gaussienne dont les paramètres $(m_k(\boldsymbol{r})=m_k, v_k(\boldsymbol{r})=v_k)$ dépendent uniquement de sa région k. Pour un pixel $\chi(\boldsymbol{r})$, on a la loi *a priori* jointe :

$$p(\chi(\boldsymbol{r})|z(\boldsymbol{r}),\boldsymbol{m},\boldsymbol{v})=\prod_{k=1}^{K}(2\pi v_k)^{\frac{N_k}{2}}\exp\left\{-\frac{1}{2v_k}\sum_{\boldsymbol{r}\in\mathcal{R}_k}|\chi(\boldsymbol{r})-m_k(\boldsymbol{r})|^2\right\}, \tag{3.10}$$

où N_k est le nombre de pixels dans une classe k et $\mathcal{R}_k=\{\boldsymbol{r}:z(\boldsymbol{r})=k\}$.

Ce modèle *a priori* ne permet pas de prendre en compte la corrélation entre les pixels d'une même classe. Par conséquent, il ne renforce pas le lien entre ces pixels et, lors de la segmentation, il produit une image divisée en un grand nombre de petites régions. Cela peut être inefficace dans certaines applications où l'on a besoin de segmenter l'image en un petit nombre de grandes régions. Le modèle suivant satisfait cette dernière condition.

3.3.2 Modèle de Mélange de Gauss-Markov (MGM)

L'intérêt des champs de Markov est de pouvoir modéliser des interactions entre les pixels voisins à travers un modèle probabiliste, tout en gardant les pixels des différentes classes indépendants afin de préserver les contours. L'idée ici est de changer la moyenne *a priori* d'un pixel en fonction de ses voisins. On définit la moyenne *a priori* comme étant la moyenne des voisins les plus proches s'ils sont tous dans la même classe, sinon on lui attribue la moyenne de la classe. Mathématiquement, cela s'écrit comme suit :

$$p(\chi(\boldsymbol{r})|z(\boldsymbol{r}))=\mathcal{N}(m_k(\boldsymbol{r}),v_k(\boldsymbol{r})), \qquad k=1,\ldots,K, \tag{3.11}$$

$$\begin{aligned} \text{avec } v_k(\boldsymbol{r}) &= v_k \quad \forall \boldsymbol{r} \in \mathcal{R}_k, \\ m_k(\boldsymbol{r}) &= \begin{cases} m_k \quad \text{si} \quad \mathcal{C}(\boldsymbol{r}) = 1 \\ \dfrac{1}{N_{\mathcal{V}}} \displaystyle\sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \chi(\boldsymbol{r}') \quad \text{si} \quad \mathcal{C}(\boldsymbol{r}) = 0, \end{cases} \\ \mathcal{C}(\boldsymbol{r}) &= 1 - \prod_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}') - z(\boldsymbol{r})), \end{aligned} \tag{3.12}$$

où $N_{\mathcal{V}} = card(\mathcal{V})$.

On peut réécrire la loi jointe du contraste, sachant le champ caché, en tenant compte de ce modèle :

$$\begin{aligned} p(\chi(\boldsymbol{r})|z(\boldsymbol{r}), \boldsymbol{m}, \boldsymbol{v}) &= \prod_{k=1}^{K} \frac{1}{(2\pi v_k)^{N_k/2}} \exp\left(Q\right), \\ Q &= -\sum_{\boldsymbol{r} \in \mathcal{R}_k} \left[\mathcal{C}_k(\boldsymbol{r}) \frac{|\chi(\boldsymbol{r}) - m_k(\boldsymbol{r})|^2}{2v_k} + \frac{(1 - \mathcal{C}_k(\boldsymbol{r}))}{2N_{\mathcal{V}}} \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \frac{|\chi(\boldsymbol{r}) - \chi(\boldsymbol{r}')|^2}{2v_k} \right]. \end{aligned} \tag{3.13}$$

Une fois l'étape du choix du modèle *a priori* adéquat franchie, l'approche bayésienne consiste à construire la distribution *a posteriori* contenant toutes les informations sur les paramètres à estimer.

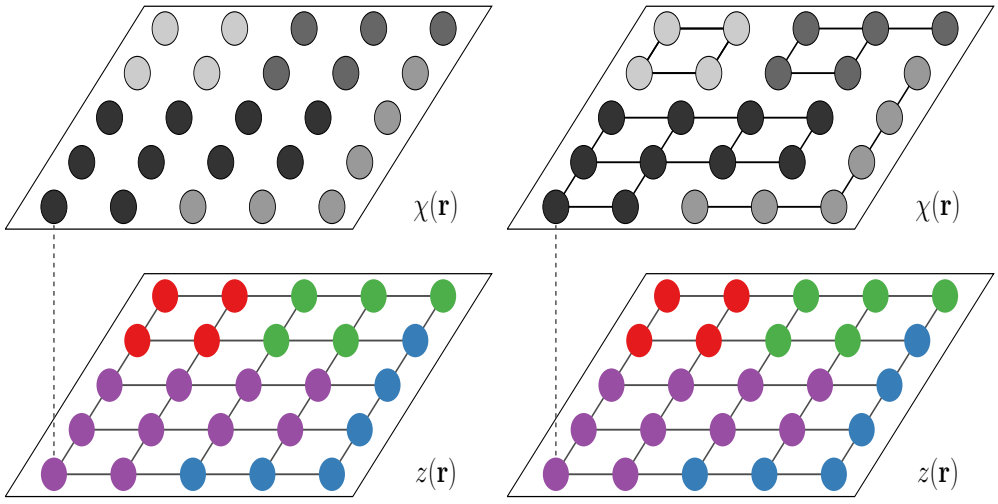


FIGURE 3.4 – Représentations 2D des modèles *a priori* proposés pour les images $\chi(\boldsymbol{r})$ segmentées en K classes. $z(\boldsymbol{r})$ représente ces classes (segmentation). A gauche, le modèle MGI où les pixels de chaque classe sont indépendants et, à droite, le modèle MGM où les pixels de chaque classe sont dépendants.

3.4 Cadre quasi non-supervisé

Dans nos travaux, nous considérons que le nombre de classes composant l'objet à imager est connu à l'avance. Par exemple, pour le modèle 1 (cf. chapitre précédent), le nombre K est fixé à 3, où $k = 1$ correspond au milieu d'adaptation, $k = 2$ correspond au sein et $k = 3$ correspond à la tumeur. Ceci nous conduit à une estimation *a posteriori* dans un cadre semi-supervisé : le contraste $\boldsymbol{\chi}$, les sources induites $\boldsymbol{w}$, la segmentation $\boldsymbol{z}$ et les hyper-paramètres du modèle $\Theta = \{\boldsymbol{m}, \boldsymbol{v}, v_\epsilon, v_\xi\}$ (où $\boldsymbol{m} = \{m_1, m_2, ..., m_K\}$ et $\boldsymbol{v} = \{v_1, v_2, ..., v_K\}$) sont inconnus et sont estimés simultanément. Grâce à la règle de Bayes, la loi jointe *a posteriori* s'écrit :

$$\begin{aligned} p\left(\boldsymbol{\chi}, \boldsymbol{w}, \boldsymbol{z}, \Theta | \boldsymbol{y}\right) &\propto p\left(\boldsymbol{y}|\boldsymbol{w}, v_\epsilon\right) p\left(\boldsymbol{w}|\boldsymbol{\chi}, v_\xi\right) p\left(\boldsymbol{\chi}|\boldsymbol{z}, \boldsymbol{m}, \boldsymbol{v}\right) \\ &\times p\left(\boldsymbol{z}|\lambda\right) p\left(\boldsymbol{m}|\mu_0, \tau_0\right) p\left(\boldsymbol{v}|\eta_0, \phi_0\right) \\ &\times p\left(v_\epsilon|\eta_\epsilon, \phi_\epsilon\right) p\left(v_\xi|\eta_\xi, \phi_\xi\right). \end{aligned} \tag{3.14}$$

Les expressions $p(\boldsymbol{z}|\lambda)$ et $p\left(\boldsymbol{\chi}|\boldsymbol{z}, \boldsymbol{m}, \boldsymbol{v}\right)$ qui apparaissent dans la relation ci-dessus proviennent respectivement des équations (3.7) et (3.9), alors que les expressions $p(\boldsymbol{y}|\boldsymbol{w}, v_\epsilon)$ et $p(\boldsymbol{w}|\boldsymbol{\chi}, v_\xi)$ sont données par :

$$\begin{cases} p(\boldsymbol{y}|\boldsymbol{w}, v_\epsilon) = \dfrac{1}{(2\pi v_\epsilon)^{N_\mathcal{S} N_f N_\vartheta / 2}} \exp\left\{-\dfrac{||\boldsymbol{y} - \boldsymbol{G}^o \boldsymbol{w}||^2_\mathcal{S}}{2v_\epsilon}\right\}, \\ \\ p(\boldsymbol{w}|\boldsymbol{\chi}, v_\xi) = \dfrac{1}{(2\pi v_\xi)^{N_\mathcal{D} N_f N_\vartheta / 2}} \exp\left\{-\dfrac{||\boldsymbol{w} - \boldsymbol{X}\boldsymbol{E}^{inc} - \boldsymbol{X}\boldsymbol{G}^c \boldsymbol{w}||^2_\mathcal{D}}{2v_\xi}\right\}, \end{cases} \tag{3.15}$$

où $||\,.\,||_\mathcal{A}$ est la norme associée au produit scalaire dans $L^2(\mathcal{A})$, $\mathcal{A} = \mathcal{D}$ ou $\mathcal{S}$.

Concernant les hyper-paramètres, les résultats de reconstruction sont moins sensibles à leur distribution *a priori* qu'à celles des paramètres. Ainsi, afin de faciliter les calculs, nous attribuons des lois conjuguées aux distributions *a priori* suivantes afin que les sorties *a posteriori* des modèles restent dans les mêmes familles, i.e. gaussienne ($\mathcal{N}$) pour les moyennes et inverse gamma ($\mathcal{IG}$) pour les variances :

$$\begin{cases} p(m_k) = \mathcal{N}(m_k|\mu_0, \tau_0) \propto \exp\left\{-\dfrac{|m_k - \mu_0|^2}{2\tau_0}\right\}, \\ \\ p(v_k) = \mathcal{IG}(v_k|\eta_0, \phi_0) \propto v_k^{-(\phi_0+1)} \exp\left\{\dfrac{-\eta_0}{v_k}\right\}, \\ \\ p(v_\epsilon) = \mathcal{IG}(v_\epsilon|\eta_\epsilon, \phi_\epsilon) \propto v_\epsilon^{-(\phi_\epsilon+1)} \exp\left\{\dfrac{-\eta_\epsilon}{v_\epsilon}\right\}, \\ \\ p(v_\xi) = \mathcal{IG}(v_\xi|\eta_\xi, \phi_\xi) \propto v_\xi^{-(\phi_\xi+1)} \exp\left\{\dfrac{-\eta_\xi}{v_\xi}\right\}, \end{cases} \tag{3.16}$$

où μ_0, τ_0, η_0, ϕ_0, η_ϵ, ϕ_ϵ, η_ξ et ϕ_ξ sont des méta-hyper-paramètres choisis convenablement afin d'obtenir des distributions *a priori* non informatives, i.e., plates.

Ainsi nous obtenons une structure hiérarchique du modèle *a priori* où l'introduction de la variable cachée et l'accès à une variable du modèle, sachant les paramètres de la loi qu'elle suit, permet de simplifier le calcul d'une loi jointe *a posteriori* généralement difficile à obtenir (figure (3.5)).

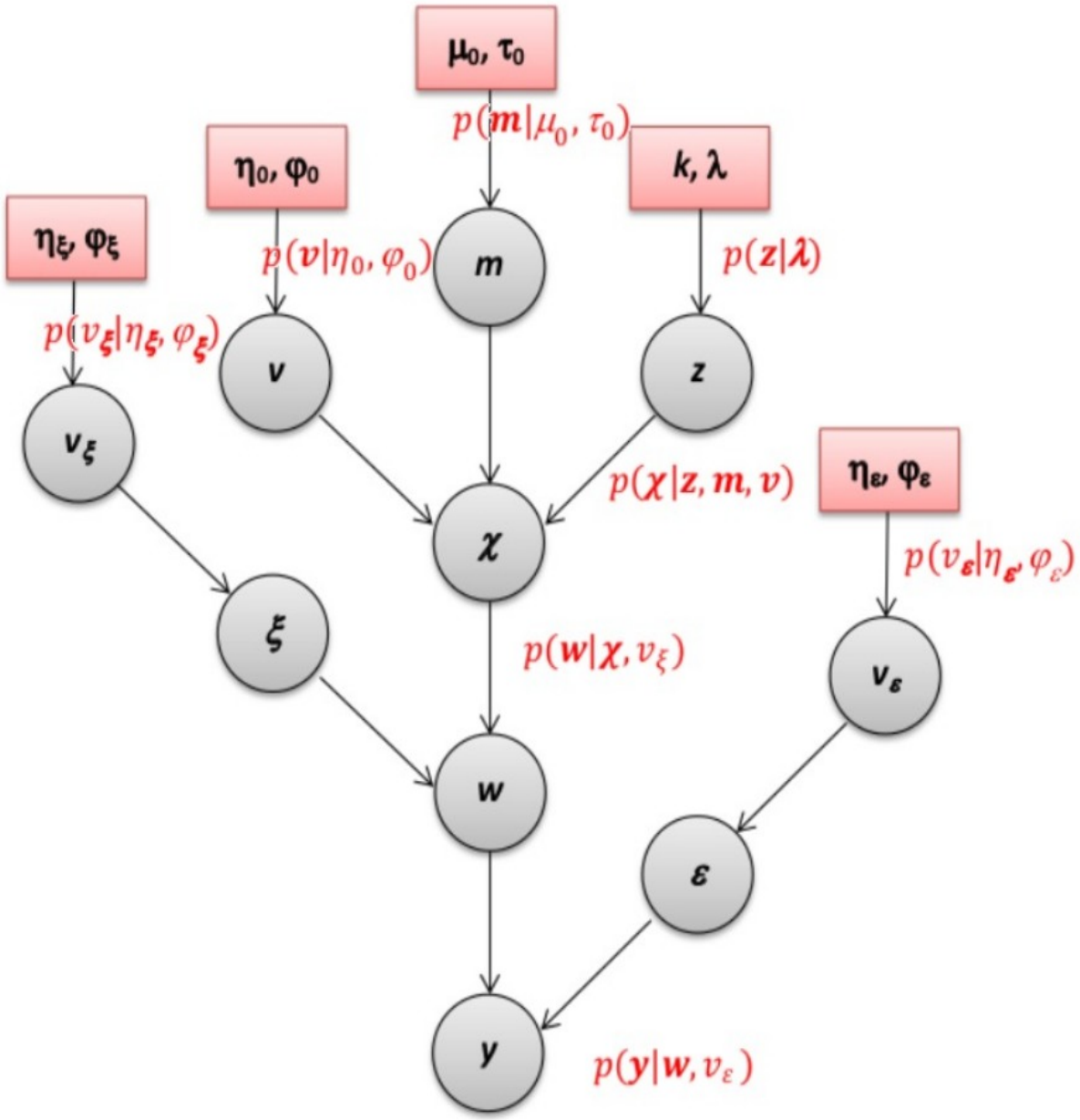


FIGURE 3.5 – Représentation hiérarchique du modèle *a priori* : la distribution du contraste $\boldsymbol{\chi}$, sachant le champ caché $\boldsymbol{z}$, peut suivre un modèle de gaussiennes indépendantes ou un modèle de Gauss-Markov de paramètres $(\boldsymbol{m}, \boldsymbol{v})$. Le champ caché $\boldsymbol{z}$ suit une loi de Potts de paramètres (k, λ). Les hyper-paramètres $(\boldsymbol{m}, \boldsymbol{v}, v_\epsilon, v_\xi)$ sont choisis dans des familles de lois conjuguées et les meta-hyper-paramètres $(\mu_0, \tau_0, \eta_0, \phi_0, \eta_\epsilon, \phi_\epsilon, \eta_\xi, \phi_\xi)$ le sont de façon à obtenir des lois plates "non-informatives".

3.4.1 Estimation conjointe

On développe la loi jointe *a posteriori* des inconnues du modèle et des données en remplaçant chaque terme par son équivalent, cela nous donne :

$$
\begin{aligned}
p\left(\boldsymbol{y},\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\Theta\right) &\propto p\left(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\Theta|\boldsymbol{y}\right)\\
&\propto \frac{1}{(v_\epsilon)^{N_{\mathcal{S}}N_f N_\vartheta/2}}\exp\left\{-\frac{||\boldsymbol{y}-\boldsymbol{G}^o\boldsymbol{w}||^2_{\mathcal{S}}}{2v_\epsilon}\right\}\\
&\times \frac{1}{(v_\xi)^{N_{\mathcal{D}}N_f N_\vartheta/2}}\exp\left\{-\frac{||\boldsymbol{w}-\boldsymbol{X}\boldsymbol{E}^{inc}-\boldsymbol{X}\boldsymbol{G}^c\boldsymbol{w}||^2_{\mathcal{D}}}{2v_\xi}\right\}\\
&\times \prod_{k=1}^{K}\frac{1}{(2\pi v_k)^{N_k/2}}\exp\left\{-\sum_{\boldsymbol{r}\in\mathcal{R}_k}\left(\mathcal{C}_k(\boldsymbol{r})\frac{|\chi(\boldsymbol{r})-m_k(\boldsymbol{r})|^2}{2v_k}\right.\right.\\
&+ \left.\left.\frac{(1-\mathcal{C}_k(\boldsymbol{r}))}{2N_{\mathcal{V}}}\sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})}\frac{|\chi(\boldsymbol{r})-\chi(\boldsymbol{r}')|^2}{2v_k}\right)\right\}\\
&\times \exp\left\{\lambda\sum_{\boldsymbol{r}}\sum_{\boldsymbol{r}'}\delta\left(z(\boldsymbol{r})-z(\boldsymbol{r}')\right)\right\}\\
&\times \prod_{k=1}^{K}\exp\left\{\frac{|m_k-\mu_0|^2}{2\tau_0}\right\}v_k^{-\phi_0-1}\exp\left\{-\frac{\eta_0}{v_k}\right\}\\
&\times v_\epsilon^{-\phi_\epsilon-1}\exp\left\{-\frac{\eta_\epsilon}{v_\epsilon}\right\}v_\xi^{-\phi_\xi-1}\exp\left\{-\frac{\eta_\xi}{v_\xi}\right\},
\end{aligned}
\tag{3.17}
$$

Une fois l'expression de la loi *a posteriori* (3.17) obtenue, une manière naturelle de définir une solution du problème consiste à déterminer un estimateur du maximum *a posteriori* de la loi jointe ($\mathcal{J}^{MAP}$) ou de sa moyenne *a posteriori* ($\mathcal{J}^{MP}$). L'estimateur $\mathcal{J}^{MAP}$, qui transforme le problème d'estimation en problème d'optimisation, est souvent employé pour la résolution des problèmes inverses linéaires. Bien que dans notre cas cet estimateur soit difficilement calculable, nous donnons tout de même les expressions des estimateurs obtenus par le biais de cette méthode et nous discutons ensuite leur validité dans le paragraphe suivant.

3.4.2 Estimateur $\mathcal{J}^{MAP}$ (optimisation alternée par point fixe)

La solution $\mathcal{J}^{MAP}$, i.e. $(\widehat{\boldsymbol{\chi}},\widehat{\boldsymbol{w}},\widehat{\boldsymbol{z}},\widehat{\Theta}) = \arg\max\mathcal{J}^{MAP}(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\Theta)$, peut être obtenue par une optimisation alternée du critère $\mathcal{J}^{MAP}$ par rapport à chaque composante de ce

dernier. Tout d'abord écrivons son expression :

$$\begin{aligned}
\mathcal{J}^{MAP}(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\Theta) &\propto -\log p(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\Theta|\boldsymbol{y}) \\
&\propto \frac{N_{\mathcal{S}}N_f N_\vartheta}{2}\log(v_\epsilon)+\frac{1}{2v_\epsilon}||\boldsymbol{y}-\boldsymbol{G}^o\boldsymbol{w}||^2_{\mathcal{S}} \\
&+ \frac{N_{\mathcal{D}}N_f N_\vartheta}{2}\log(v_\xi)+\frac{1}{2v_\xi}||\boldsymbol{w}-\boldsymbol{X}\boldsymbol{E}^{inc}-\boldsymbol{X}\boldsymbol{G}^c\boldsymbol{w}||^2_{\mathcal{D}} \\
&+ \sum_k\left[\frac{N_k\log(v_k)}{2}+\sum_{\boldsymbol{r}}\left(\frac{\mathcal{C}_k(\boldsymbol{r})|\chi(\boldsymbol{r})-m_k(\boldsymbol{r})|^2}{2v_k}+\frac{(1-\mathcal{C}_k(\boldsymbol{r}))}{2N_{\mathcal{V}}}\right.\right. \\
&\times \left.\left.\sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})}\frac{|\chi(\boldsymbol{r})-\chi(\boldsymbol{r}')|^2}{2v_k}\right)\right]-\boldsymbol{\lambda}\sum_{\boldsymbol{r}}\sum_{\boldsymbol{r}'}\delta(z(\boldsymbol{r})-z(\boldsymbol{r}')) \\
&+ \frac{\sum_k|m_k-\mu_0|^2}{2\tau_0}+\sum_k\left(\frac{\eta_0}{v_k}+(\phi_0+1)\log(v_k)\right) \\
&+ \frac{\eta_\epsilon}{v_\epsilon}+(\phi_\epsilon+1)\log(v_\epsilon)+\frac{\eta_\xi}{v_\xi}+(\phi_\xi+1)\log(v_\xi). \qquad (3.18)
\end{aligned}$$

Ensuite, déterminons les solutions $\widehat{\boldsymbol{x}}=\{\widehat{\boldsymbol{\chi}},\widehat{\boldsymbol{w}},\widehat{\boldsymbol{z}},\widehat{\Theta}\}$:

— les sources induites :

$$\bigtriangledown_{\boldsymbol{w}}\mathcal{J}^{MAP} = \bigtriangledown_{\boldsymbol{w}}\mathcal{J}_0(\boldsymbol{w}), \qquad (3.19)$$

où $\bigtriangledown$ désigne le gradient et

$$\begin{aligned}
\bigtriangledown_{\boldsymbol{w}}\mathcal{J}_0(\boldsymbol{w}) &= \left(\frac{1}{2v_\epsilon}||\boldsymbol{y}-\boldsymbol{G}^o\boldsymbol{w}||^2_{\mathcal{S}}+\frac{1}{2v_\xi}||\boldsymbol{w}-\boldsymbol{X}\boldsymbol{E}^{inc}-\boldsymbol{X}\boldsymbol{G}^c\boldsymbol{w}||^2_{\mathcal{D}}\right) \\
&= v_\epsilon^{-1}\boldsymbol{G}^{oH}\Big(\boldsymbol{y}-\boldsymbol{G}^o\boldsymbol{w}\Big)+v_\xi^{-1}\boldsymbol{Z}^H\Big(\boldsymbol{w}\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{E}^{inc}\Big) \qquad (3.20) \\
&= v_\epsilon^{-1}\boldsymbol{G}^{oH}\boldsymbol{y}-v_\epsilon^{-1}\boldsymbol{G}^{oH}\boldsymbol{G}^o\boldsymbol{w}+v_\xi^{-1}\boldsymbol{Z}^H\boldsymbol{w}\boldsymbol{Z}-v_\xi^{-1}\boldsymbol{Z}^H\boldsymbol{X}\boldsymbol{G}^c,
\end{aligned}$$

où l'exposant H désigne l'opérateur adjoint et $\boldsymbol{Z}=(\mathbf{I}-\boldsymbol{X}\boldsymbol{G}^c)$. L'estimateur MAP des sources induites $\widehat{\boldsymbol{w}}$ est alors la solution de $\bigtriangledown_{\boldsymbol{w}}\mathcal{J}_0(\boldsymbol{w})=0$. Il est donné par :

$$\boxed{\widehat{\boldsymbol{w}}=\left(v_\xi^{-1}\boldsymbol{Z}^H\boldsymbol{Z}-v_\epsilon^{-1}\boldsymbol{G}^{oH}\boldsymbol{G}^o\right)^{-1}\left(v_\xi^{-1}\boldsymbol{Z}^H\boldsymbol{X}\boldsymbol{E}^{inc}-v_\epsilon^{-1}\boldsymbol{G}^{oH}\boldsymbol{y}\right),} \qquad (3.21)$$

où $\left(v_\xi^{-1}\boldsymbol{Z}^H\boldsymbol{Z}-v_\epsilon^{-1}\boldsymbol{G}^{oH}\boldsymbol{G}^o\right)$ est une matrice de grande taille non hermitienne dont l'inversion n'est pas facile.

— le contraste :

$$\bigtriangledown_{\boldsymbol{\chi}}\mathcal{J}^{MAP} = \bigtriangledown_{\boldsymbol{\chi}}\mathcal{J}_0(\boldsymbol{\chi}), \qquad (3.22)$$

avec

$$\begin{aligned}
\bigtriangledown_{\boldsymbol{\chi}}\mathcal{J}_0(\boldsymbol{\chi}) &= \bigtriangledown_{\boldsymbol{\chi}}\left(\frac{1}{2v_\xi}||\boldsymbol{w}-\boldsymbol{X}\boldsymbol{E}||^2_{\mathcal{D}}+\sum_k\sum_{\boldsymbol{r}}\frac{|\chi(\boldsymbol{r})-m_k(\boldsymbol{r})|^2}{2v_k}\right) \\
&= v_\xi^{-1}\boldsymbol{X}\boldsymbol{E}^H\boldsymbol{E}+\sum_k v_k^{-1}\boldsymbol{X}+v_\xi^{-1}\boldsymbol{E}^H\boldsymbol{w}-\sum_k v_k^{-1}m_k, \qquad (3.23)
\end{aligned}$$

où $\boldsymbol{E} = \boldsymbol{E}^{inc} + \boldsymbol{G}^c\boldsymbol{w}$. L'estimateur $\widehat{\chi}(\boldsymbol{r})$ est la solution de $\bigtriangledown_{\boldsymbol{\chi}}\mathcal{J}_0(\boldsymbol{\chi}) = 0$, c'est-à-dire :

$$\boxed{\widehat{\chi}(\boldsymbol{r}) = \left(v_\xi^{-1}\boldsymbol{E}^H\boldsymbol{E} + v_k^{-1}\right)^{-1}\left(v_k^{-1}m_k - v_\xi^{-1}\boldsymbol{w}\boldsymbol{E}^H\right),} \tag{3.24}$$

où, de même que pour l'équation de mise à jour précédente, $(v_\xi^{-1}\boldsymbol{E}^H\boldsymbol{E} + v_k^{-1})$ est une matrice de grande taille très difficile à inverser.

— le champ caché :

$$\begin{aligned}\widehat{z}(\boldsymbol{r}) &= \arg\max_{\{z(\boldsymbol{r})=1,2,\ldots,K\}}\mathcal{J}^{MAP}(\boldsymbol{\chi},\boldsymbol{w},z(\boldsymbol{r}),\Theta)\\ &= \arg\max_{\{z(\boldsymbol{r})=1,2,\ldots,K\}}\Bigg[\frac{\log(v_k)}{2} + \frac{v_k^{-1}|\chi(\boldsymbol{r}) - m_k(\boldsymbol{r})|^2}{2}\\ &- \boldsymbol{\lambda}\sum_{\boldsymbol{r}}\sum_{\boldsymbol{r}'}\delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\Bigg],\end{aligned} \tag{3.25}$$

où la connaissance du $\boldsymbol{z}$ qui maximise le critère $\mathcal{J}^{MAP}(\boldsymbol{\chi},\boldsymbol{w},z(\boldsymbol{r}),\Theta)$ nécessite de parcourir tout les $\boldsymbol{z}$ couvrant toutes les classes, pixel par pixel, ce qui entraîne un coût de calcul important.

— la variance du bruit d'observation :

$$\begin{aligned}\frac{\partial\mathcal{J}^{MAP}(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\Theta)}{\partial v_\epsilon} &= \frac{\partial}{\partial v_\epsilon}\Bigg[\frac{N_{\mathcal{S}}N_fN_\vartheta}{2}\log(v_\epsilon) + \frac{1}{2v_\epsilon}||\boldsymbol{y} - \boldsymbol{G}^o\boldsymbol{w}||^2_{\mathcal{S}}\\ &+ \frac{\eta_\epsilon}{v_\epsilon} + (\phi_\epsilon + 1)\log(v_\epsilon)\Bigg],\\ &= \frac{1}{v_\epsilon}\left(\frac{N_{\mathcal{S}}N_fN_\vartheta}{2} + (\phi_\epsilon + 1) - \frac{||\boldsymbol{y} - \boldsymbol{G}^o\boldsymbol{w}||^2_{\mathcal{S}}}{2v_\epsilon} - \frac{\eta_\epsilon}{v\epsilon}\right)\end{aligned} \tag{3.26}$$

L'estimateur $\widehat{v_\epsilon}$ est la solution de $\dfrac{\partial\mathcal{J}^{MAP}(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\Theta)}{\partial v_\epsilon} = 0$, c'est-à-dire :

$$\boxed{\widehat{v_\epsilon} = \left(\frac{N_{\mathcal{S}}N_fN_\vartheta}{2} + \phi_\epsilon + 1\right)^{-1}\left(\frac{||\boldsymbol{y} - \boldsymbol{G}^o\boldsymbol{w}||^2_{\mathcal{S}}}{2} + \eta_\epsilon\right).} \tag{3.27}$$

— la variance du bruit de couplage :

$$\begin{aligned}\frac{\partial\mathcal{J}^{MAP}(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\Theta)}{\partial v_\xi} &= \frac{\partial}{\partial v_\xi}\Bigg[\frac{N_{\mathcal{D}}N_fN_\vartheta}{2}\log(v_\xi)\\ &+ \frac{1}{2v_\xi}||\boldsymbol{w} - \boldsymbol{X}\boldsymbol{E}||^2_{\mathcal{D}} + \frac{\eta_\xi}{v_\xi} + (\phi_\xi + 1)\log(v_\xi)\Bigg],\\ &= \frac{1}{v_\xi}\left(\frac{N_{\mathcal{D}}N_fN_v}{2} + (\phi_\xi + 1) - \frac{||\boldsymbol{w} - \boldsymbol{X}\boldsymbol{E}||^2_{\mathcal{D}}}{2v_\xi} - \frac{\eta_\xi}{v_\xi}\right).\end{aligned} \tag{3.28}$$

L'estimateur $\widehat{v_\xi}$ est la solution de $\dfrac{\partial\mathcal{J}^{MAP}(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\Theta)}{\partial v_\xi} = 0$, c'est-à-dire :

$$\boxed{\widehat{v_\xi} = \left(\frac{N_{\mathcal{D}} N_f N_v}{2} + \phi_\xi + 1\right)^{-1} \left(\frac{||\boldsymbol{w} - \boldsymbol{X}\boldsymbol{E}||^2_{\mathcal{D}}}{2} + \eta_\xi\right).} \tag{3.29}$$

— la variance des classes :

$$\begin{aligned}
\frac{\partial \mathcal{J}^{MAP}(\boldsymbol{\chi}, \boldsymbol{w}, \boldsymbol{z}, \Theta)}{\partial v_k} &= \frac{\partial}{\partial v_k}\left[\frac{\sum_k N_k \log(v_k)}{2} + \sum_k \sum_{\boldsymbol{r}} \frac{|\chi(\boldsymbol{r}) - m_k(\boldsymbol{r})|^2}{2v_k}\right. \\
&+ \left.\sum_k \left(\frac{\eta_0}{v_k} + (\phi_0 + 1)\log(v_k)\right)\right], \\
&= \frac{1}{v_k}\left[\frac{\sum_k N_k}{2} - \frac{\sum_k \sum_r |\chi(\boldsymbol{r}) - m_k(\boldsymbol{r})|^2}{2v_k}\right. \\
&- \left.\frac{\sum_k \eta_0}{v_k} + (\phi_0 + 1)\right].
\end{aligned} \tag{3.30}$$

L'estimateur $\widehat{v_k}$ est la solution de $\dfrac{\partial \mathcal{J}^{MAP}(\boldsymbol{\chi}, \boldsymbol{w}, \boldsymbol{z}, \Theta)}{\partial v_k} = 0$, c'est-à-dire :

$$\boxed{\widehat{v_k} = \left(\frac{N_k}{2} + \phi_0 + 1\right)^{-1} \left(\frac{\sum_{\boldsymbol{r}} |\chi(\boldsymbol{r}) - m_k(\boldsymbol{r})|^2}{2} + k\eta_0\right).} \tag{3.31}$$

— la moyenne des classes :

$$\begin{aligned}
\frac{\partial \mathcal{J}^{MAP}(\boldsymbol{\chi}, \boldsymbol{w}, \boldsymbol{z}, \Theta)}{\partial m_k} &= \frac{\partial}{\partial m_k}\left(\sum_k \Big(\sum_{\boldsymbol{r}} \frac{|\chi(\boldsymbol{r}) - m_k(\boldsymbol{r})|^2}{2v_k} + \frac{|m_k - \mu_0|^2}{2\tau_0}\Big)\right), \\
&= v_k^{-1} \sum_{\boldsymbol{r}} \Big(m_k(\boldsymbol{r}) - \chi(\boldsymbol{r})\Big) + \tau_0^{-1}\Big(\sum_k m_k - \mu_0\Big).
\end{aligned} \tag{3.32}$$

L'estimateur $\widehat{m_k}$ est la solution de $\dfrac{\partial \mathcal{J}^{MAP}(\boldsymbol{\chi}, \boldsymbol{w}, \boldsymbol{z}, \Theta)}{\partial m_k} = 0$, c'est-à-dire :

$$\boxed{\widehat{m_k} = \left(\tau_0^{-1} + v_k^{-1}\right)^{-1} \left(v_k^{-1} \sum_{\boldsymbol{r}} \chi(\boldsymbol{r}) + \mu_0 \tau_0^{-1}\right).} \tag{3.33}$$

Pour récapituler, l'algorithme d'optimisation alternée est donné par :

Optimisation alternée ($\mathcal{J}^{MAP}$)

1 – Initialiser $\boldsymbol{\chi}^{(0)}, \boldsymbol{w}^{(0)}, \boldsymbol{z}^{(0)}, \Theta^{(0)}$.

2 – Mise à jour des paramètres :

Itérer

- estimer $\widehat{\boldsymbol{z}}$ en utilisant l'équation (3.25),
- estimer $\widehat{v_\epsilon}$ en utilisant l'équation (3.27),
- estimer $\widehat{v_\xi}$ en utilisant l'équation (3.29),
- estimer $\widehat{v_k}$ en utilisant l'équation (3.31),
- estimer $\widehat{m_k}$ en utilisant l'équation (3.33),
- estimer $\widehat{\boldsymbol{w}}$ en utilisant l'équation (3.21),
- estimer $\widehat{\boldsymbol{\chi}}$ en utilisant l'équation (3.24),

jusqu'à convergence. (3.34)

3.4.3 Discussion

L'une des principales difficultés pour trouver l'estimateur joint *a posteriori* est la dépendance mutuelle entre les différentes variables qui rend les calculs souvent complexes. De plus, si on regarde de près les équations de mise à jour de l'algorithme d'optimisation alternée (3.34), bien que la mise à jour des hyper-paramètres par le biais de ce dernier se fasse d'une manière analytique, celle des paramètres rencontre des difficultés. En effet, on peut voir, par exemple dans les termes de gauche des équations (3.21) et (3.24), l'apparition de matrices de très grandes dimensions dont l'inversion n'est pas triviale. Cependant, notons que les critères $\mathcal{J}_0(\boldsymbol{\chi})$ et $\mathcal{J}_0(\boldsymbol{w})$ ont une forme quadratique et par conséquent leur résolution en pratique peut se faire en les optimisant par un algorithme de type gradient.

En ce qui concerne l'équation de mise à jour du champ caché (3.25), on voit qu'une optimisation n'est pas possible car l'expression obtenue n'est pas convexe et cela représente un problème très difficile de plus à résoudre du fait que l'opération de recherche du maximum est coûteuse.

Enfin, nous n'oublions pas de mettre l'accent sur le fait que, lors des mises à jour des variables, les incertitudes ne sont pas prises en compte et aucune information sur les erreurs de modèle et de mesure n'est transmise, ce qui constitue un des inconvénients de cet algorithme d'estimation ponctuelle.

Ainsi, la méthode $\mathcal{J}^{MAP}$ n'est pas efficace pour la résolution de notre problème inverse non linéaire. Par conséquent, une approximation souple de la distribution jointe *a posteriori* est souhaitable et susceptible de donner des estimations plus précises.

Une méthode d'approximation efficace réside dans l'échantillonnage stochastique de Monte Carlo par chaîne de Markov. Cette technique a été testée en imagerie micro-onde et en imagerie optique dans les travaux précédents de Féron [89] et de Ayasso [90]. Les

applications ciblées ont porté sur des objets plongés en milieu homogène ou stratifié [107] dont le contraste était à valeurs réelles, les milieux étant considérés comme sans pertes. Dans nos travaux, l'application visée est plus complexe étant donné que le contraste recherché est à valeurs complexes car les tissus du sein présentent de fortes pertes.

3.5 Conclusions

Dans ce chapitre, nous avons posé le problème inverse d'imagerie micro-onde dans le cadre de l'estimation bayésienne. L'intérêt de cette approche réside dans les aspects suivants :

- prise en compte des erreurs de modèle et des incertitudes de mesure,
- exploitation de la structure à données cachées pour développer un modèle d'estimation hiérarchique en introduisant des informations *a priori* pour le modèle (Gauss-Markov-Potts) et pour les hyper-paramètres (distributions plates non informatives),
- écriture de la loi conjointe *a posteriori* contenant toutes les informations sur les paramètres à estimer en fonction de sa vraisemblance et des *a priori*,
- définition d'un estimateur de la loi jointe *a posteriori*.

Ensuite, diverses techniques de calcul peuvent être employées pour trouver les estimées. Lorsque l'expression de la loi jointe *a posteriori* n'est pas complexe, les estimateurs du type maximum *a posteriori* (MAP) et moyenne *a posteriori* (MP) peuvent être appliqués, mais cela n'est pas le cas dans le cadre de cette thèse. Ainsi, une approximation de la loi jointe *a posteriori* doit être recherchée. Le chapitre suivant porte sur l'approximation bayésienne variationnelle et montre l'avantage d'utiliser cette approche dans la recherche de la solution.

4

Méthodes bayésiennes variationnelles

Sommaire

4.1 Introduction 59
4.2 Principe de la méthode bayésienne variationnelle 60
4.3 Approche bayésienne variationnelle en imagerie micro-onde 62
4.3.1 Équations de mise à jour 63
4.3.2 Déroulement de l'algorithme 66
4.4 Méthode du gradient pour l'approche bayésienne variationnelle 67
4.4.1 Principe de la méthode 68
4.4.2 Équations de mise à jour 70
4.4.3 Pas de descente 71
4.4.4 Déroulement de l'algorithme 72
4.5 Comparaison théorique entre BV, G-BV et $\mathcal{J}^{MAP}$ 73
4.6 Conclusions 74

4.1 Introduction

Deux classes de méthodes qui approchent les lois a *posteriori* sont connues : les méthodes analytiques comme l'approximation de Laplace [108] et les méthodes numériques comme l'échantillonnage MCMC [109]. Mais, l'apparition récente des méthodes variationnelles et leur succès dû à leur rapidité d'exécution dans des cas difficiles à traiter avec les outils classiques ont fait de ces méthodes un réel centre d'intérêt et leur application commence à se propager dans plusieurs domaines. Leur apport considérable par rapport aux méthodes stochastiques a déjà été prouvé en imagerie optique et en imagerie micro-onde [110, 111].

La méthodologie bayésienne variationnelle consiste à chercher la solution d'un problème d'optimisation concave dans un espace de densités de probabilité, en approchant la vraie loi *a posteriori* jointe des inconnues du modèle par une loi plus simple à mettre en œuvre, à savoir une loi séparable. La séparation peut être faible (partielle) ou forte (totale) selon la complexité des calculs liés au problème traité. La séparation forte est souvent plus utilisée car elle permet de faciliter le calcul des lois approchantes en éliminant la dépendance *a posteriori* entre les variables. Ce dernier point est un vrai obstacle pour déterminer des estimateurs efficaces des inconnues du modèle, comme montré dans le chapitre précédent.

La présentation de cette méthodologie fait l'objet de ce chapitre, son extension en une approche variationnelle de type gradient et son application au problème de l'imagerie micro-onde du sein constituent les majeures contributions de ces travaux.

4.2 Principe de la méthode bayésienne variationnelle

L'approximation bayésienne variationnelle a été introduite dans le domaine de l'inférence bayésienne [112, 113]. Son apparition dans le domaine des problèmes inverses est relativement récente avec une première application en restauration d'image [114], puis dans les problèmes de séparation de sources [115] et dans les problèmes d'imagerie hyperspectrale [116]. Elle représente une approximation déterministe de la loi *a posteriori*, où l'idée est d'approcher la loi jointe par une loi plus simple à mettre en œuvre, par exemple une loi séparable. La forme de cette loi approchante est obtenue en minimisant la distance de Kullback-Leibler (KL). Notons que le choix d'approximer une loi conjointe de plusieurs variables par une loi séparable n'est pas nouveau et est apparu dès la fin des années 1990 [117, 118, 119]. Dans ce qui suit, nous présentons le principe de cette approche d'une manière générale.

Soient un ensemble de variables observées Y et un jeu de paramètres θ. Considérons un modèle statistique paramétrique $p(Y|\theta)$. L'apprentissage bayésien s'attache à l'optimisation de l'évidence $p(Y)$, où les paramètres θ ont été marginalisés. La loi *a priori* qui modélise son incertitude est donnée par le paramètre $\pi(\theta)$.

La règle de Bayes nous donne :

$$p(\theta|Y) = \frac{p(Y|\theta)\pi(\theta)}{p(Y)}, \tag{4.1}$$

soit encore, en notant que $p(Y|\theta)\pi(\theta) = p(Y,\theta)$:

$$p(Y) = \frac{p(Y,\theta)}{p(\theta|Y)}, \tag{4.2}$$

où $p(\theta|Y)$ est la *loi a posteriori* du paramètre (on peut consulter [120] pour une présentation plus détaillée du cadre bayésien).

En théorie des probabilités et en théorie de l'information, la divergence de Kullback-Leibler est une mesure de similarité entre deux distributions de probabilités. Elle a les propriétés suivantes :

— elle est toujours positive ($\mathrm{KL}(p\|q) > 0,\ \forall\ p, q$) et elle s'annule quand les deux distributions sont identiques,
— elle est non symétrique ($\mathrm{KL}(p\|q) \neq \mathrm{KL}(q\|p)$) et
— elle est convexe ($\mathrm{KL}(\alpha q_1 + (1-\alpha)q_2)\|p) \leq \alpha \mathrm{KL}(q_1\|p) + (1-\alpha)\mathrm{KL}(q_2\|p)$).

De plus, on a :

$$\begin{aligned}
\mathrm{KL}(q(\theta)|p(\theta|Y)) &= \int q(\theta) \ln \frac{q(\theta)}{p(\theta|Y)} d\theta, \\
&= \int q(\theta) \ln \frac{q(\theta)p(Y)}{p(Y,\theta)} d\theta, \\
&= \ln p(Y) - \int q(\theta) \ln \frac{p(Y,\theta)}{q(\theta)} d\theta, \\
&= \ln p(Y) - \mathcal{F}(q), \qquad (4.3)
\end{aligned}$$

où $\mathcal{F}$ est l'énergie libre négative que l'on peut développer ainsi :

$$\begin{aligned}
\mathcal{F}(q(\theta)) &= \int q(\theta) \ln p(Y|\theta) d\theta - \int q(\theta) \ln q(\theta) d\theta, \\
&= \Big\langle \log p(Y,\theta) \Big\rangle_{q(\theta)} + \mathcal{H}(q), \qquad (4.4)
\end{aligned}$$

où $\mathcal{H}(q) = -\int q(\theta) \ln q(\theta) d\theta$ est l'entropie de la loi approchante et $\langle . \rangle_{q(\theta)}$ signifie la moyenne par rapport à $q(\theta)$.

A ce stade, on peut faire quelques remarques :

1. la divergence de Kullback-Leibler n'a pas de forme explicite car elle dépend de la loi *a posteriori* qui est inconnue et que l'on cherche justement à estimer,
2. $\mathcal{F}(q(\theta))$ représente une borne inférieure de la log-vraisemblance $\ln p(Y)$, c'est-à-dire que $\ln p(Y) \geq \mathcal{F}(q(\theta))$ car la divergence de Kullback-Leiber est positive,
3. minimiser la divergence de Kullback-Leibler revient à maximiser l'énergie libre négative qui peut alors être choisie comme un critère de sélection des modèles. En effet, comme la divergence de Kullback-Leiber est convexe et que la somme des deux est constante, l'énergie libre est concave relativement à la densité de probabilité q. L'idée de maximiser l'énergie libre négative vient de la physique statistique [121] où il s'agissait d'approcher un potentiel de Gibbs dans les travaux de Feynman [122].

Ainsi, trouver la forme optimale de la densité de probabilité q revient à résoudre le problème d'optimisation suivant :

$$\tilde{q}^{opt} = \arg\min_q \mathrm{KL}(q||p) = \arg\max_q \mathcal{F}(q(\theta)). \qquad (4.5)$$

Cependant, la clé importante de l'approche bayésienne variationnelle réside dans le choix de la forme approchante. Cette dernière est choisie la plus simple possible comme

une loi séparable $q(\theta) = \prod_i q_i(\theta_i)$. Par exemple, dans le cas de deux paramètres, en choisissant $q(\theta) = q_1(\theta_1)q_2(\theta_2)$ on obtient :

$$(\widetilde{q_1}, \widetilde{q_2}) = \arg\min_{q_1,q_2} \mathrm{KL}(q_1, q_2||p) = \arg\max_{q_1,q_2} \mathcal{F}(q_1(\theta_1), q_2(\theta_2)). \tag{4.6}$$

Puisque $\mathrm{KL}(q_1, q_2||p)$ est convexe en q_1 quand q_2 est fixé et inversement, la solution de l'équation (4.6) peut être obtenue itérativement comme suit :

$$\widetilde{q_1} = \arg\min_{q_1} \mathrm{KL}(q_1, \widetilde{q_2}||p) = \arg\max_{q_1} \mathcal{F}(q_1(\theta_1), \widetilde{q_2}(\theta_2)), \tag{4.7}$$
$$\widetilde{q_2} = \arg\min_{q_2} \mathrm{KL}(\widetilde{q_1}, q_2||p) = \arg\max_{q_2} \mathcal{F}(\widetilde{q_1}(\theta_1), q_2(\theta_2)). \tag{4.8}$$

Les solutions analytiques d'optimisation de ces deux équations sont données par (cf. annexe A) :

$$\widetilde{q_1}(\theta_1) \propto \exp\left\{\langle \log p(Y, \theta)\rangle_{\widetilde{q_2}}\right\}, \tag{4.9}$$
$$\widetilde{q_2}(\theta_2) \propto \exp\left\{\langle \log p(Y, \theta)\rangle_{\widetilde{q_1}}\right\}, \tag{4.10}$$

ce qui s'écrit d'une manière générale :

$$\boxed{\widetilde{q_i}(\theta_i) = \frac{1}{\kappa_i} \exp\left\{\langle \log p(Y, \theta)\rangle_{\prod_{i\neq j} \widetilde{q_j}(\theta_j)}\right\},} \tag{4.11}$$

où κ_i est une constante de normalisation.

La solution obtenue par l'approche bayésienne variationnelle possède une forme itérative où les paramètres de forme sont mutuellement dépendants. Le calcul itératif des q_i impose la connaissance de tous les q_j pour $j \neq i$. Cependant, même si cette solution est obtenue analytiquement, elle n'a pas de forme explicite. C'est pourquoi, le choix de l'appartenance des paramètres du modèle étudié à des familles de lois exponentielles conjuguées est recommandé. En effet, cela nous permet de ne remettre à jour que les paramètres de ces lois en garantissant que ces formes paramétriques ne changent pas au cours des itérations [118, 123].

D'autre part, pour évaluer la convergence de l'algorithme, on peut vérifier le comportement de l'énergie libre négative dont la courbe doit se stabiliser lors de la convergence.

4.3 Approche bayésienne variationnelle en imagerie micro-onde

En projetant le principe de la méthode bayésienne variationnelle sur notre problème d'imagerie micro-onde, on cherche à approcher la vraie loi *a posteriori* $p(\boldsymbol{u}|\boldsymbol{y})$, où $\boldsymbol{u} = \{\boldsymbol{\chi}, \boldsymbol{w}, \boldsymbol{z}, \Theta\}$ avec $\Theta = \{\boldsymbol{m}, \boldsymbol{v}, v_\epsilon, v_\xi\}$, par un produit $q(\boldsymbol{u}) = \prod_i q(\boldsymbol{u}_i)$ de lois *a posteriori* approchantes. L'approximation optimale est obtenue en maximisant l'énergie libre négative :

$$\mathcal{F}(\boldsymbol{q}) = \int_{\mathcal{R}^N} q(\boldsymbol{u}) \ln \frac{p(\boldsymbol{y}, \boldsymbol{u})}{q(\boldsymbol{u})} \, \mathrm{d}\boldsymbol{u}. \tag{4.12}$$

Ensuite, en tenant compte de la séparabilité $q(\boldsymbol{u}) = \prod_i q(\boldsymbol{u}_i)$, le but de l'approche bayésienne variationnelle est de trouver $q^{opt}(\boldsymbol{u})$ tel que :

$$\begin{aligned} q^{opt}(\boldsymbol{u}) &= \arg\max_{\boldsymbol{q}} \mathcal{F}(q(\boldsymbol{u})), \\ &\propto \exp\left\{\langle \log(p(\boldsymbol{y},\boldsymbol{u}))\rangle_{\prod_{j\neq i} q(\boldsymbol{u}_j)}\right\}. \end{aligned} \tag{4.13}$$

Nous appliquons alors ce principe pour approcher la distribution *a posteriori* (3.17) par :

$$q(\boldsymbol{u}) = q(v_\epsilon)q(v_\xi)\prod_k q(v_k)q(m_k)\prod_i q(\chi_i)q(z_i)\prod_j q(w_j), \tag{4.14}$$

où le choix de cette forme de séparation permet de garder des liens forts dans l'approximation et de simplifier, par conséquent, les expressions des paramètres de forme des lois approchantes sans trop détériorer la qualité de l'approximation.

4.3.1 Équations de mise à jour

Ensuite, la forme optimale (équation (4.13)) est recherchée. Par exemple, pour les courants induits nous avons :

$$\widetilde{q}(w(\boldsymbol{r}))) \propto \exp\left\{\langle \log(p(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\boldsymbol{\psi},\boldsymbol{y}))\rangle_{\widetilde{q}(/w(r))}\right\}, \tag{4.15}$$

où $\widetilde{q}(/w(\boldsymbol{r})) = \prod_{\boldsymbol{r}'\neq\boldsymbol{r}} q(w(\boldsymbol{r}'))q(\boldsymbol{\chi})q(\boldsymbol{z})q(\boldsymbol{\psi})$.

Il en va de même pour les autres paramètres, ce qui nous permet enfin d'obtenir, après calculs, les distributions paramétriques suivantes [90] :

$$\begin{aligned} \widetilde{q}(\boldsymbol{w}) &= \prod_{\boldsymbol{r}} \widetilde{q}(w(\boldsymbol{r})) = \mathcal{N}(\widetilde{\boldsymbol{m}}_w, \widetilde{\boldsymbol{V}}_w), & \widetilde{q}(\boldsymbol{\chi}) &= \prod_{\boldsymbol{r}} \widetilde{q}(\chi(\boldsymbol{r})) = \mathcal{N}(\widetilde{\boldsymbol{m}}_\chi, \widetilde{\boldsymbol{V}}_\chi), \\ \widetilde{q}(m_k) &= \mathcal{N}(\widetilde{\mu}_k, \widetilde{\tau}_k), & \widetilde{q}(v_k) &= \mathcal{IG}(\widetilde{\eta}_k, \widetilde{\phi}_k), \\ \widetilde{q}(v_\epsilon) &= \mathcal{IG}(\widetilde{\eta}_\epsilon, \widetilde{\phi}_\epsilon), & \widetilde{q}(v_\xi) &= \mathcal{IG}(\widetilde{\eta}_\xi, \widetilde{\phi}_\xi), \end{aligned}$$

$$\widetilde{q}(z(\boldsymbol{r})) = \prod_{\boldsymbol{r}} \widetilde{\zeta_k}(\boldsymbol{r}), \tag{4.16}$$

où les paramètres tildés sont les paramètres de mise à jour. Ils sont donnés ci dessous à l'itération i :

— pour les courant induits

$$\begin{aligned} \widetilde{\boldsymbol{V}}_w^{(i)} &= \left[\mathbf{Diag}\left(\overline{v_\epsilon^{-1}}\Gamma^o + \overline{v_\xi^{-1}}\overline{\Gamma^c}\right)\right]^{-1}, \\ \widetilde{\boldsymbol{m}}_w^{(i)} &= \widetilde{\boldsymbol{m}}_w + \widetilde{\boldsymbol{V}}_w^{(i)}\Bigg[\overline{v_\epsilon^{-1}}\boldsymbol{G}^{o^H}\left(\boldsymbol{y} - \boldsymbol{G}^o\widetilde{\boldsymbol{m}}_w\right) \\ &+ \overline{v_\xi^{-1}}\left(\mathbf{I} - \boldsymbol{G}^{c^H}\widetilde{\boldsymbol{m}}_\chi^H\right)\left(\widetilde{\boldsymbol{m}}_\chi(\boldsymbol{E}^{inc} + \boldsymbol{G}^c\widetilde{\boldsymbol{m}}_w) - \widetilde{\boldsymbol{m}}_w\right) \\ &+ \overline{v_\xi^{-1}}\widetilde{\boldsymbol{V}}_\chi\left(\boldsymbol{E}^{inc} + \boldsymbol{G}^c\widetilde{\boldsymbol{m}}_w\right)\Bigg], \end{aligned} \tag{4.17}$$

où les paramètres tildés sans exposant sont obtenus à l'itération $(i-1)$, $\overline{u} = \mathrm{E}(u)_q$ et l'exposant H indique le transposé conjugué tandis que Γ^o et Γ^c sont tels que :

$$\begin{aligned} \Gamma^o(\boldsymbol{r}) &= \sum_{\boldsymbol{r}'} |G^o(\boldsymbol{r}', \boldsymbol{r})|^2, \\ \overline{\Gamma^c}(\boldsymbol{r}) &= 1 - 2\Re e(G^c(\boldsymbol{r}, \boldsymbol{r})\widetilde{m}_\chi(\boldsymbol{r})) + (\widetilde{m}_\chi^2(\boldsymbol{r}) + \widetilde{V}_\chi(\boldsymbol{r})) \sum_{\boldsymbol{r}'} |G^c(\boldsymbol{r}', \boldsymbol{r})|^2, \end{aligned} \tag{4.18}$$

— pour le contraste

$$\begin{aligned} \widetilde{\boldsymbol{V}}_\chi^{(i)} &= \left[\mathbf{Diag}\left(\overline{v_\xi^{-1}}\,\overline{\boldsymbol{E}^2} + \overline{\boldsymbol{V}_\chi^{-1}}\right)\right]^{-1}, \\ \widetilde{\boldsymbol{m}}_\chi^{(i)} &= \widetilde{\boldsymbol{V}}_\chi^{(i)} \left(\sum_k \overline{v_k^{-1}}\widetilde{\zeta}_k \circ \widetilde{\boldsymbol{m}}_\chi^\dagger + \overline{v_\xi^{-1}}\,\overline{\boldsymbol{w}\boldsymbol{E}^*}\right), \end{aligned} \tag{4.19}$$

où * désigne le complexe conjugué, $\circ$ représente le produit de Hadamard et $\overline{\boldsymbol{w}\boldsymbol{E}^*}$ est la moyenne du vecteur $\boldsymbol{w}\boldsymbol{E}^*$:

$$\begin{aligned} \overline{wE^*}(\boldsymbol{r}) &= \sum_{N_f N_\nu} E^{inc*}(\boldsymbol{r})\widetilde{m}_w(\boldsymbol{r}) + \widetilde{m}_w(\boldsymbol{r}) \sum_{N_\mathcal{D}} G^{c*}(\boldsymbol{r}, \boldsymbol{r}')\widetilde{m}_w^*(\boldsymbol{r}') \\ &+ G^{c*}(\boldsymbol{r}, \boldsymbol{r})\widetilde{V}_w(\boldsymbol{r}) \end{aligned} \tag{4.20}$$

et $\overline{\boldsymbol{V}_\chi^{-1}}$, $\overline{\boldsymbol{E}^2}$ sont des matrices diagonales dont les éléments sont donnés par :

$$\overline{V_\chi^{-1}}(\boldsymbol{r}, \boldsymbol{r}) = \sum_k \widetilde{\boldsymbol{\zeta}}_k(\boldsymbol{r})\overline{v_k^{-1}}, \tag{4.21}$$

$$\begin{aligned} \overline{E^2}(\boldsymbol{r}) &= \sum_{N_f N_\vartheta} |E^{inc}(\boldsymbol{r})|^2 + 2\Re e\left(E^{inc*}(\boldsymbol{r})G^c\widetilde{m}_w(\boldsymbol{r})\right) \\ &+ |\sum_{\boldsymbol{r}'} G^c(\boldsymbol{r}, \boldsymbol{r}')\widetilde{m}_w(\boldsymbol{r})|^2 + \sum_{\boldsymbol{r}'} |G^c(\boldsymbol{r}, \boldsymbol{r}')|^2 \widetilde{V}_w(\boldsymbol{r}), \end{aligned} \tag{4.22}$$

tandis que le terme $\widetilde{\boldsymbol{m}}_\chi^\dagger$ est donné par :

$$\widetilde{\boldsymbol{m}}_\chi^\dagger = \widetilde{\mathcal{C}}_k(\boldsymbol{r})\widetilde{m}_k + \frac{1 - \widetilde{\mathcal{C}}_k(\boldsymbol{r})}{N_\mathcal{V}} \sum_{r' \in \mathcal{V}(r)} \widetilde{m}_\chi(\boldsymbol{r}'), \tag{4.23}$$

où $\mathcal{C}$ est défini dans l'équation (3.12),

— pour le champ caché

$$\begin{aligned} \widetilde{\boldsymbol{\zeta}}_k^{(i)} &\propto \exp\left\{ -\frac{1}{2}\left(\Psi(\widetilde{\eta}_k) + \log(\widetilde{\phi}_k) + \overline{v_k^{-1}}\Big((|\widetilde{m}_\chi(\boldsymbol{r})|^2 + \widetilde{m}_\chi^{\dagger 2}(\boldsymbol{r})\right.\right. \\ &- \left.\left. 2\Re e(\widetilde{m}_\chi^\dagger(\boldsymbol{r})\widetilde{m}_\chi^*(\boldsymbol{r})) + \widetilde{V}_\chi(\boldsymbol{r})\Big) - \lambda \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \widetilde{\boldsymbol{\zeta}}_k(\boldsymbol{r}')\right)\right\}, \end{aligned} \tag{4.24}$$

où Ψ est la fonction digamma, λ le paramètre de Potts (cf. équation (3.6)) et :

$$\widetilde{m}_\chi^{\dagger 2} = \frac{1 - \widetilde{\mathcal{C}}_k(\boldsymbol{r})}{N_\mathcal{V}^2}\left(\Big|\sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \widetilde{m}_\chi(\boldsymbol{r}')\Big|^2 + \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \widetilde{V}_\chi(\boldsymbol{r}')\right) + \widetilde{\mathcal{C}}_k(\boldsymbol{r})\left(\widetilde{m}_k^2 + \widetilde{\tau}_k\right), \tag{4.25}$$

— pour la variance du bruit d'observation

$$\begin{aligned}\widetilde{\phi}_{\epsilon}^{(i)} &= \phi_{\epsilon} + \frac{N_f N_\vartheta N_{\mathcal{S}}}{2}, \\ \widetilde{\eta}_{\epsilon}^{(i)} &= \eta_{\epsilon} + \frac{1}{2}\Big(||\boldsymbol{y}||_{\mathcal{S}}^2 + ||\boldsymbol{G}^o \widetilde{\boldsymbol{m}}_w||_{\mathcal{S}}^2 - 2\Re e(\boldsymbol{y}^H \boldsymbol{G}^o \widetilde{\boldsymbol{m}}_w) + ||\boldsymbol{G}^{o2} \widetilde{\boldsymbol{v}}_w||_1 \Big),\end{aligned} \tag{4.26}$$

— pour la variance du bruit de couplage

$$\begin{aligned}\widetilde{\phi}_{\xi}^{(i)} &= \phi_{\xi} + \frac{N_f N_\vartheta N_{\mathcal{D}}}{2}, \\ \widetilde{\eta}_{\xi}^{(i)} &= \eta_{\xi} + \frac{1}{2}\Big(||\widetilde{\boldsymbol{m}}_w||_{\mathcal{D}}^2 + ||\widetilde{\boldsymbol{V}}_w||_{\mathcal{D}}^2 + || \left(\widetilde{\boldsymbol{m}}_\chi^2 + \widetilde{\boldsymbol{V}}_\chi \right) \overline{\boldsymbol{E}^2} ||_1 \\ &- 2\Re e \left(\widetilde{\boldsymbol{m}}_\chi^H \overline{\boldsymbol{w}\boldsymbol{E}^*} \right) \Big),\end{aligned} \tag{4.27}$$

où $\widetilde{\boldsymbol{m}}_\chi^2 = \widetilde{\boldsymbol{m}}_\chi^H \widetilde{\boldsymbol{m}}_\chi$,

— pour la variance des classes

$$\begin{aligned}\widetilde{\phi}_k^{(i)} &= \phi_0 + \frac{1}{2} \sum_{\boldsymbol{r}} \widetilde{\zeta}_k(\boldsymbol{r}), \\ \widetilde{\eta}_k^{(i)} &= \eta_0 + \frac{1}{2} \sum_{\boldsymbol{r}} \widetilde{\zeta}_k(\boldsymbol{r}) \Bigg(|\widetilde{m}_\chi(\boldsymbol{r})|^2 + \widetilde{V}_\chi(\boldsymbol{r}) + \widetilde{\mathcal{C}}_k(\boldsymbol{r}) \Big(\widetilde{m}_k^2 + \widetilde{\tau}_k \Big) \\ &+ \frac{1 - \widetilde{\mathcal{C}}_k(\boldsymbol{r})}{2N_{\mathcal{V}}} \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} |\widetilde{m}_\chi(\boldsymbol{r}')|^2 + \widetilde{V}_\chi(\boldsymbol{r}') - 2\Re e(\widetilde{m}_\chi^\dagger(\boldsymbol{r}) \widetilde{m}_\chi^*(\boldsymbol{r})) \Bigg)\end{aligned} \tag{4.28}$$

— et pour la moyenne des classes

$$\begin{aligned}\widetilde{\tau}_k^{(i)} &= \left(\tau_0^{-1} + \overline{v_k^{-1}} \sum_{\boldsymbol{r}} \widetilde{\zeta}_k(\boldsymbol{r}) \right)^{-1}, \\ \widetilde{\mu}_k^{(i)} &= \widetilde{\tau}_k \Big(\frac{\mu_0}{\tau_0} + \overline{v_k^{-1}} \sum_{\boldsymbol{r}} \widetilde{\mathcal{C}}_k(\boldsymbol{r}) \widetilde{\zeta}_k(\boldsymbol{r}) \widetilde{m}_\chi(\boldsymbol{r}) \Big).\end{aligned} \tag{4.29}$$

Ainsi, on voit que ces paramètres sont mutuellement dépendants et doivent être mis à

jour d'une manière itérative.

Algorithme bayésien variationnel (BV)

1 – Initialiser $\boldsymbol{\chi}^{(0)}, \boldsymbol{w}^{(0)}, \boldsymbol{z}^{(0)}, \Theta^{(0)}$.

2 – Mise à jour des paramètres :

itérer :

- calculer $\widetilde{\boldsymbol{\zeta}_k}$ en utilisant l'équation (4.24),
- calculer $(\widetilde{\eta}_\epsilon, \widetilde{\phi}_\epsilon)$ en utilisant l'équation (4.26),
- calculer $(\widetilde{\eta}_\xi, \widetilde{\phi}_\xi)$ en utilisant l'équation (4.27),
- calculer $(\widetilde{\eta}_k, \widetilde{\phi}_k)$ en utilisant l'équation (4.28),
- calculer $(\widetilde{\tau}_k, \widetilde{\mu}_k)$ en utilisant l'équation (4.29),
- calculer $(\widetilde{\boldsymbol{m}}_w, \widetilde{\boldsymbol{v}}_w)$ en utilisant l'équation (4.17),
- calculer $(\widetilde{\boldsymbol{m}}_\chi, \widetilde{\boldsymbol{v}}_\chi)$ en utilisant l'équation (4.19),

jusqu'à convergence. (4.30)

4.3.2 Déroulement de l'algorithme

Initialisation

Dans l'algorithme bayésien variationnel, la mise à jour des paramètres des lois approchantes se fait d'une manière itérative. L'étape d'initialisation des paramètres $\boldsymbol{w}$, $\boldsymbol{\chi}$, $\boldsymbol{z}$ et Θ est très importante afin de garantir une bonne convergence de l'algorithme, car une initialisation arbitraire pourrait conduire à une solution locale. Ainsi, l'initialisation des courants induits $\boldsymbol{w}^{(0)}$ et du contraste $\boldsymbol{\chi}^{(0)}$ est obtenue en rétropropageant le champ diffracté du domaine de mesure $\mathcal{S}$ vers le domaine test $\mathcal{D}$ [124]. Ensuite, sachant $\boldsymbol{w}^{(0)}$ et $\boldsymbol{\chi}^{(0)}$, la valeur initiale du champ caché $\boldsymbol{z}^{(0)}$ est donnée par l'algorithme de segmentation K-means (*K-means* clustering) [125] avec des estimateurs empiriques pour les hyper-paramètres $\Theta^{(0)}$. Cependant, comme le contraste recherché est à valeurs complexes, mais que sa partie réelle et sa partie imaginaire sont censées avoir la même segmentation puisqu'elles correspondent au même tissu dans chaque région, nous effectuons, tout d'abord, la segmentation de la partie réelle et cette dernière est utilisée pour initialiser la partie imaginaire. L'initialisation du nombre de classes est effectuée d'une manière arbitraire. En effet, en l'initialisant avec un nombre supérieur au nombre réel, il a été montré par l'expérience [90] que ce nombre finit par converger vers le nombre optimal de classes selon deux chemins possibles :

— si le nombre de pixels dans une classe donnée diminue jusqu'à une valeur de seuil bas, cette classe finit par disparaître,
— si les moyennes de deux classes différentes sont très proches l'une de l'autre, alors ces deux classes finissent par fusionner.

Pour résumer cela, l'initialisation des différentes variables se déroule comme suit. On commence par l'initialisation des courants induits :

$$\boldsymbol{w}_n^{(0)} = \Gamma \boldsymbol{G}_f^{o\dagger} \boldsymbol{y}_n, \tag{4.31}$$

où Γ est une constante de normalisation déterminée en minimisant l'erreur dans l'équation d'observation :

$$\sum_n \| \ \boldsymbol{y}_n - \Gamma \ \boldsymbol{G}_f^o \boldsymbol{G}_f^{o\dagger} \boldsymbol{y}_n \|_{\mathcal{S}}^2$$

et $\boldsymbol{G}^{o\dagger}$ est l'opérateur adjoint à $\boldsymbol{G}^o$ agissant de $L^2(\mathcal{S})$ vers $L^2(\mathcal{D})$, tel que :

$$< \boldsymbol{w}_n \ , \ \boldsymbol{G}_f^{o\dagger} \boldsymbol{y}_n >_{\mathcal{D}} = < \boldsymbol{G}_f^o \boldsymbol{w}_n \ , \ \boldsymbol{y}_n >_{\mathcal{S}} .$$

L'initialisation du champ total est alors effectuée à l'aide de l'équation de couplage :

$$\boldsymbol{E}_n^{(0)} = \boldsymbol{E}_n^{inc} + \boldsymbol{G}_f^c \boldsymbol{w}_n^{(0)}, \tag{4.32}$$

et le contraste initial $\boldsymbol{\chi}_f^{(0)}$ est obtenu en minimisant l'erreur dans l'équation constitutive des sources :

$$\boldsymbol{w}_n = \boldsymbol{X}_f \boldsymbol{E}_n.$$

En tenant compte du fait que le contraste recherché dans le cadre de notre application est à valeurs complexes, nous l'écrivons :

$$\boldsymbol{\chi}_f^{(0)} = \boldsymbol{\chi}_r^{(0)} + i \ \boldsymbol{\chi}_i^{(0)} \eta_f, \tag{4.33}$$

où $\eta_f = \dfrac{1}{\omega_f \, \epsilon_0}$, ce qui conduit à l'initialiser de la manière suivante :

$$\boldsymbol{\chi}_r^{(0)} = \frac{\sum_n \Re e(\boldsymbol{w}_n^{(0)} \boldsymbol{E}_n^{*(0)})}{\sum_n |\boldsymbol{E}_n^{(0)}|^2}, \tag{4.34}$$

$$\boldsymbol{\chi}_i^{(0)} = \frac{\sum_n \eta_f \, \Im m(\boldsymbol{w}_n^{(0)} \boldsymbol{E}_n^{*(0)})}{\sum_n \eta_f^2 \, |\boldsymbol{E}_n^{(0)}|^2}. \tag{4.35}$$

Les estimations initiales des variances des bruits d'observation et de couplage sont données par :

$$v_{\epsilon_n}^{(0)} = \frac{||\boldsymbol{y}_n - \boldsymbol{G}_f^o \boldsymbol{w}_n^{(0)}||_{\mathcal{S}}^2}{N_{\mathcal{S}}}, \qquad v_{\xi_n}^{(0)} = \frac{||\boldsymbol{w}_n^{(0)} - \boldsymbol{X}_f^{(0)} \boldsymbol{E}_n^{(0)}||_{\mathcal{D}}^2}{N_{\mathcal{D}}}. \tag{4.36}$$

Convergence

Une fois initialisé, l'algorithme peut se dérouler comme décrit dans l'encadré (4.30). Partant des sources $\boldsymbol{w}^{(i-1)}$, du contraste $\boldsymbol{\chi}^{(i-1)}$, du champ caché $\boldsymbol{z}^{(i-1)}$ et des hyperparamètres $\Theta^{(i-1)}$ à l'itération $(i-1)$, la mise à jour des variables à l'itération i se fait à travers la mise à jour des paramètres des lois de probabilité assignées à ces variables

à l'aide des équations indiquées dans (4.30). Le processus est itéré jusqu'à atteindre la convergence. Cette dernière est observée empiriquement en suivant l'évolution des variables et des hyper-paramètres au cours des itérations et en évaluant l'énergie libre négative qui peut être calculée avec les paramètres de mise à jour des lois.

Pour récapituler, l'approche bayésienne variationnelle conduit à une forme paramétrique des lois approchantes qui sont mutuellement dépendantes. Elle permet d'obtenir des valeurs optimales des estimateurs en les mettant à jour un par un d'une manière itérative jusqu'à convergence. Cependant, le fait de mettre à jour les variables séparément augmente considérablement le temps de calcul. C'est pourquoi, une idée, pour accélérer la convergence de l'algorithme, est de transposer une méthode d'optimisation déterministe à l'algorithme bayésien variationnel [126]. Dans la section suivante, nous présentons une méthode de gradient développée pour la méthode bayésienne variationnelle afin d'en améliorer la convergence. Ceci est une des contributions majeures de nos travaux.

4.4 Méthode du gradient pour l'approche bayésienne variationnelle

Les algorithmes à direction de descente sont des algorithmes d'optimisation déterministes itératifs qui peuvent s'appliquer aux problèmes avec ou sans contraintes et forment une classe d'algorithmes faciles d'utilisation fournissant généralement des résultats de convergence satisfaisants pour peu qu'ils soient initialisés dans un état pas trop éloigné de la solution. L'algorithme de descente le plus simple est celui du gradient. Ce dernier a été adapté au cadre bayésien variationnel et des résultats intéressants en ce qui concerne sa convergence ont été obtenus en tomographie X, en décomposition de dictionnaire [126] et, plus récemment, en imagerie astrophysique [127]. L'originalité ici est d'adapter cette technique à la tomographie micro-onde et de le tester pour la détection du cancer du sein où le contraste est à valeurs complexes, ce qui signifie que deux images doivent être reconstruites : une carte de permittivité et une de conductivité.

4.4.1 Principe de la méthode

On rappelle que le problème d'optimisation bayésien variationnel consiste à déterminer la densité q^{opt} appartenant à l'espace de densité de probabilité maximisant l'énergie libre négative $\mathcal{F}(q)$:

$$q^{opt} = \arg\max_{q\in\Omega} \mathcal{F}(q), \tag{4.37}$$

où $\mathcal{F}$ est tel que $\mathcal{F}(q) = \int_{\mathcal{R}^n} f(q, \boldsymbol{u})d\boldsymbol{u}$ et Ω vérifie $\Omega = \{q \geq 0 \,;\, \int_{\mathcal{R}^n} q(\boldsymbol{u})d\boldsymbol{u} = 1\}$.

La clé de l'algorithme du gradient de type bayésien variationnel, présenté dans [126], réside essentiellement dans un résultat de la théorie de la mesure qui est le théorème de "Radon-Nikodym" [128]. Ce dernier permet de conserver la structure de l'espace des densités de probabilités et d'avoir la relation suivante :

$$q^{(i+1)}(d\boldsymbol{u}) = h^{(i)}(\boldsymbol{u})q^{(i)}(d\boldsymbol{u}),\ \forall \boldsymbol{u}\ \in \mathcal{R}^n, \tag{4.38}$$

où la fonction h est à construire.

Pour ce faire, une approche classique est de considérer chaque densité q comme étant une fonction de $\mathcal{L}^1(\mathcal{R}^n)$ pour pouvoir appliquer l'algorithme de minimisation par descente de gradient classique. Pour construire une méthode itérative vérifiant à la fois le schéma d'une descente de gradient, c'est à dire $q^{(i+1)} = q^{(i)} + \alpha_i \Delta\mathcal{F}(q^{(i)})$ où α_i est le pas de descente à l'itération i, et la relation (4.38), une idée est d'utiliser la méthode du "gradient exponentialisé" [129]. Cette dernière, déjà étudiée dans le domaine de l'apprentissage automatique, nous permet d'avoir h sous la forme $h^{(i)}(\boldsymbol{u}) = C \exp(\alpha_i df(q^{(i)}, \boldsymbol{u}))$, où C est une constante de normalisation telle que $\int_{\mathcal{R}^n} q^{(i+1)} = 1$. Ce choix précis de forme exponentielle pour h est bien adapté à la structure de descente du gradient ainsi qu'au contexte considéré, c'est-à-dire qu'il vérifie les contraintes de positivité et d'intégrabilité que requiert l'espace des densités de probabilité. Notons aussi que, lorsque $\exp(\alpha_i df(q^{(i)}, \boldsymbol{u}))$ est non intégrable, on impose que $h^{(i)}(\boldsymbol{u}) = \infty$.

Ainsi, en ayant construit $q^{(i)}$, $i \geq 0$, suivant ce qui précède, où $\forall\ \alpha\ \geq 0$, q^α est donné par :

$$\forall\ \boldsymbol{u}\ \in \mathcal{R}^n,\ q^\alpha(\boldsymbol{u}) = C\ q^{(i)}(\boldsymbol{u}) \exp[\alpha_i df(q^{(i)}, \boldsymbol{u})], \tag{4.39}$$

avec

$$\forall\ l\ \in \{1, ..., n\},\ \forall \boldsymbol{u} \in \mathcal{R}^n,\ d_l f(q_l, \boldsymbol{u}) = \langle \log p(\boldsymbol{y}, \boldsymbol{u}) \rangle_{\prod_{j \neq l} q_j^{(i)}(\boldsymbol{u}_j)} - \log q_l^{(i)} - 1, \tag{4.40}$$

et, en substituant (4.40) dans (4.39), on obtient :

$$\begin{aligned}
\forall\ \boldsymbol{u}\ \in \mathcal{R}^n,\ q^\alpha(\boldsymbol{u}) &= C\ q^{(i)}(\boldsymbol{u}) \Big(\prod_i \exp d_i f(q_l^k, \boldsymbol{u}_l) \Big)^\alpha \\
&= C\ q^{(i)}(\boldsymbol{u}) \Big(\prod_l \exp \left(\langle \log p(\boldsymbol{y}, \boldsymbol{u}) \rangle_{\prod_{j \neq l} q_j(\boldsymbol{u}_j)} - \log q_l^{(i)} - 1 \right) \Big)^\alpha \\
&= C\ q^{(i)}(\boldsymbol{u}) \Big(\prod_l \frac{\exp \langle \log p(\boldsymbol{y}, \boldsymbol{u}) \rangle_{\prod_{j \neq l} q_j^{(i)}(\boldsymbol{u}_j)}}{q_l^{(i)}(\boldsymbol{u}_l)} \Big)^\alpha \\
&= \tilde{C}\ q^{(i)}(\boldsymbol{u}) \Big(\prod_l \frac{q_l^r(\boldsymbol{u}_l)}{q_l^{(i)}(\boldsymbol{u}_l)} \Big)^\alpha,
\end{aligned} \tag{4.41}$$

où $\tilde{C}$ est la constante de normalisation et q^r une mesure intermédiaire donnée par :

$$q_l^r(\boldsymbol{u}_l) = \exp \langle \log p(\boldsymbol{y}, \boldsymbol{u}) \rangle_{\prod_{j \neq l} q_j^{(i)}(\boldsymbol{u}_j)},\ \forall\ l\ \in\ \{1, ..., n\}. \tag{4.42}$$

Finalement, à l'étape $(i+1)$ et en tenant compte de la séparabilité $q(\boldsymbol{u}) = \prod_l q_l(\boldsymbol{u})$, l'équation de mise à jour de l'algorithme bayésien variationnel de type gradient s'écrit [126] :

$$q^{(i+1)}(\boldsymbol{u}) = C\ q^{(i)}(\boldsymbol{u}) \Big(\frac{q^r(\boldsymbol{u})}{q^{(i)}(\boldsymbol{u})} \Big)^\alpha, \tag{4.43}$$

où α est le pas optimal de l'algorithme dont la détermination est souvent coûteuse, ce qui a poussé les auteurs à adopter plutôt un pas sous optimal.

Ainsi, en combinant les équations (4.42) et (4.43), nous obtenons l'équation clé de cette nouvelle technique permettant de mettre à jour les paramètres des lois approchantes simultanément avec un "pas de descente" optimal qui reste à déterminer :

$$\boxed{\widetilde{q}_l^{(i+1)}(\boldsymbol{u}_l) \propto \left(\widetilde{q}_l^{(i)}(\boldsymbol{u}_l)\right)^{(1-\alpha)} \exp\left\{\alpha\Big\langle \ln(p(\boldsymbol{y},\boldsymbol{u}))\Big\rangle_{\prod_{l\neq l'}\widetilde{q}_{l'}^{(i)}(u_j)}\right\}.} \tag{4.44}$$

En choisissant la même forme de séparation qu'avec l'approche précédente (équation (4.14)), on obtient les lois marginales approchées suivantes :

$$\begin{aligned}
&\widetilde{q}(\boldsymbol{w}) = \mathcal{N}(\widetilde{\boldsymbol{m}}_w, \widetilde{\boldsymbol{V}}_w), && \widetilde{q}(\boldsymbol{\chi}) = \mathcal{N}(\widetilde{\boldsymbol{m}}_\chi, \widetilde{\boldsymbol{V}}_\chi),\\
&\widetilde{q}(m_k) = \mathcal{N}(\widetilde{\mu}_k, \widetilde{\tau}_k), && \widetilde{q}(v_k) = \mathcal{IG}(\widetilde{\eta}_k, \widetilde{\phi}_k),\\
&\widetilde{q}(v_\epsilon) = \mathcal{IG}(\widetilde{\eta}_\epsilon, \widetilde{\phi}_\epsilon), && \widetilde{q}(v_\xi) = \mathcal{IG}(\widetilde{\eta}_\xi, \widetilde{\phi}_\xi),\\
&\widetilde{q}(\boldsymbol{z}) = \widetilde{\boldsymbol{\zeta}}_k, &&
\end{aligned} \tag{4.45}$$

où les paramètres de forme sont donnés par la suite (cf. annexe B). Notons que chaque variable nécessite un pas de descente différent de celui des autres variables qui sera noté α_ρ, où $\rho = \boldsymbol{w}, \boldsymbol{\chi}, \boldsymbol{z}, v_\epsilon, v_\xi, v_k$ ou m_k.

4.4.2 Équations de mise à jour

— des courants induits :

$$\widetilde{\boldsymbol{V}}_w^{(i)} = \left[(1-\alpha_w)\,\widetilde{\boldsymbol{V}}_w^{-1} + \alpha_w\,\mathbf{Diag}\left(\overline{v_\epsilon^{-1}}\Gamma^o + \overline{v_\xi^{-1}}\overline{\Gamma^c}\right)\right]^{-1}, \tag{4.46}$$

$$\begin{aligned}
\widetilde{\boldsymbol{m}}_w^{(i)} = {}& \widetilde{\boldsymbol{m}}_w + \alpha_w\,\widetilde{\boldsymbol{V}}_w^{(i-1)}\Bigg[\overline{v_\epsilon^{-1}}\boldsymbol{G}^{o^H}\left(\boldsymbol{y} - \boldsymbol{G}^o\widetilde{\boldsymbol{m}}_w\right)\\
&+ \overline{v_\xi^{-1}}\Big(\widetilde{\boldsymbol{m}}_\chi\boldsymbol{E}^{inc} - \boldsymbol{G}^{c^H}\left(\widetilde{\boldsymbol{m}}_\chi^2 + \widetilde{\boldsymbol{V}}_\chi\right)\boldsymbol{E}^{inc}\\
&- \widetilde{\boldsymbol{m}}_w + \widetilde{\boldsymbol{m}}_\chi\boldsymbol{G}^c\widetilde{\boldsymbol{m}}_w + \boldsymbol{G}^{c^H}\widetilde{\boldsymbol{m}}_\chi^*\widetilde{\boldsymbol{m}}_w - \boldsymbol{G}^{c^H}\left(\widetilde{\boldsymbol{m}}_\chi^2 + \widetilde{\boldsymbol{V}}_\chi\right)\boldsymbol{G}^c\widetilde{\boldsymbol{m}}_w\Big)\Bigg],
\end{aligned}$$

— du contraste :

$$\widetilde{\boldsymbol{V}}_\chi^{(i)} = \left[(1-\alpha_\chi)\,\widetilde{\boldsymbol{V}}_\chi^{-1} + \alpha_\chi\left(\mathbf{Diag}\left(\overline{v_\xi^{-1}\boldsymbol{E}^2} + \overline{\boldsymbol{V}_\chi^{-1}}\right)\right)\right]^{-1}, \tag{4.47}$$

$$\widetilde{\boldsymbol{m}}_\chi^{(i)} = \alpha_\chi\,\widetilde{\boldsymbol{V}}_\chi^{(i-1)}\left(\sum_k \overline{v_k^{-1}}\widetilde{\boldsymbol{\zeta}}_k \circ \widetilde{\boldsymbol{m}}_\chi^\dagger + \overline{v_\xi^{-1}\boldsymbol{w}} \circ \boldsymbol{E}^*\right),$$

— du champ caché :

$$\begin{aligned}
\widetilde{\zeta}_k^{(i)} \propto {}& \widetilde{\zeta}_k^{(1-\alpha_z)} \exp\Bigg\{-\frac{\alpha_z}{2}\Bigg(\Psi(\widetilde{\eta}_k) + \log(\widetilde{\phi}_k) + \overline{v_k^{-1}}\Big((|\widetilde{m}_\chi(\boldsymbol{r})|^2 + \widetilde{m}_\chi^{\dagger 2}(\boldsymbol{r})\\
&- 2\Re e(\widetilde{m}_\chi^\dagger(\boldsymbol{r})\widetilde{m}_\chi^*(\boldsymbol{r})) + \widetilde{V}_\chi(\boldsymbol{r})\Big) - \lambda\sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})}\widetilde{\boldsymbol{\zeta}}_k(\boldsymbol{r}')\Bigg)\Bigg\},
\end{aligned} \tag{4.48}$$

— de la variance de l'erreur d'observation :

$$\begin{aligned}\widetilde{\phi}_{\epsilon}^{(i)} &= (1-\alpha_{v_\epsilon})\,\widetilde{\phi}_{\epsilon} + \alpha_{v_\epsilon}\left(\phi_\epsilon + \frac{N_f N_\vartheta N_{\mathcal{S}}}{2}\right), \qquad (4.49)\\ \widetilde{\eta}_{\epsilon}^{(i)} &= \frac{1-\alpha_{v_\epsilon}}{\widetilde{\eta}_\epsilon} + \alpha_{v_\epsilon}\left[\eta_\epsilon + \frac{1}{2}\Big(||\boldsymbol{y}||_{\mathcal{S}}^2 + ||\boldsymbol{G}^o\widetilde{\boldsymbol{m}}_w||_{\mathcal{S}}^2 - 2\Re e(\boldsymbol{y}^H\boldsymbol{G}^o\widetilde{\boldsymbol{m}}_w)\right.\right.\\ &+ \left.\left.||\boldsymbol{G}^{o2}\widetilde{\boldsymbol{v}}_w||_1\Big)\right],\end{aligned}$$

— de la variance de l'erreur de couplage :

$$\begin{aligned}\widetilde{\phi}_{\xi}^{(i)} &= \left(1-\alpha_{v_\xi}\right)\widetilde{\phi}_{\xi} + \alpha_{v_\xi}\left(\phi_\xi + \frac{N_f N_\vartheta N_{\mathcal{D}}}{2}\right)\left[\alpha_{v_\xi}\,\boldsymbol{d}_\chi + \left(1-\alpha_{v_\xi}\right)\widetilde{\boldsymbol{m}}_\chi\right], \qquad (4.50)\\ \widetilde{\eta}_{\xi}^{(i)} &= \frac{1-\alpha_{v_\xi}}{\widetilde{\eta}_\xi} + \alpha_{v_\xi}\left(\eta_\xi + \frac{1}{2}\Big[||\widetilde{\boldsymbol{m}}_w||_{\mathcal{D}}^2 + ||\widetilde{\boldsymbol{V}}_w||_{\mathcal{D}}^2 + ||\left(\widetilde{\boldsymbol{m}}_\chi^2 + \widetilde{\boldsymbol{V}}_\chi\right)\overline{\boldsymbol{E}^2}||_1\right.\right.\\ &- \left.\left.2\Re e\left(\widetilde{\boldsymbol{m}}_\chi^H\overline{\boldsymbol{w}\circ\boldsymbol{E}^*}\right)\Big]\right),\end{aligned}$$

— de la variance des classes :

$$\begin{aligned}\widetilde{\phi}_{k}^{(i)} &= (1-\alpha_{v_k})\,\widetilde{\phi}_k + \alpha_{v_k}\left(\phi_0 + \frac{\sum \boldsymbol{r}\widetilde{\zeta}_k(\boldsymbol{r})}{2}\right), \qquad (4.51)\\ \widetilde{\eta}_{k}^{(i)} &= \frac{1-\alpha_{v_k}}{\widetilde{\eta}_k} + \alpha_{v_k}\left(\eta_0 + \frac{1}{2}\sum_{\boldsymbol{r}}\widetilde{\zeta}_k(\boldsymbol{r})\left[|\widetilde{m}_\chi(\boldsymbol{r})|^2 + \widetilde{V}_\chi(\boldsymbol{r}) + \widetilde{\mathcal{C}}_k(\boldsymbol{r})\left(\widetilde{m}_k^2 + \widetilde{\tau}_k\right)\right.\right.\\ &+ \left.\left.\frac{1-\widetilde{\mathcal{C}}_k(\boldsymbol{r})}{2N_{\mathcal{V}}}\sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})}|\widetilde{m}_\chi(\boldsymbol{r}')|^2 + \widetilde{V}_\chi(\boldsymbol{r}') - 2\Re e\left(\widetilde{m}_\chi^\dagger(\boldsymbol{r})\widetilde{m}_\chi^*(\boldsymbol{r})\right)\right]\right)\end{aligned}$$

— et de la moyenne des classes :

$$\begin{aligned}\widetilde{\tau}_{k}^{(i)} &= \left[(1-\alpha_{m_k})\,\widetilde{\tau}_k + \alpha_{m_k}\left(\tau_0^{-1} + \overline{v_k^{-1}}\sum_{\boldsymbol{r}}\widetilde{\zeta}_k(\boldsymbol{r})\right)\right]^{-1}, \qquad (4.52)\\ \widetilde{\mu}_{k}^{(i)} &= \widetilde{\tau}_k^n\left((1-\alpha_{m_k})\,\widetilde{\mu}_k + \alpha_{m_k}\left[\widetilde{\tau}_k\left(\frac{\mu_0}{\tau_0} + \overline{v_k^{-1}}\sum_{\boldsymbol{r}}\widetilde{\mathcal{C}}_k\widetilde{\zeta}_k(\boldsymbol{r})\widetilde{m}_\chi(\boldsymbol{r})\right)\right]\right).\end{aligned}$$

4.4.3 Pas de descente

La définition du pas de descente est une étape très importante pour garantir la rapidité de convergence de l'algorithme vers la solution optimale. Ici, nous avons sept pas de descente différents. On peut choisir de les déterminer tous ou bien procéder autrement. Ici, seul le pas de descente correspondant aux sources induites sera calculé (cf. annexe C), tandis que les pas de descente de $\boldsymbol{\chi}$, $\boldsymbol{z}$, v_ϵ, v_ξ, v_k et m_k seront fixés à 1 afin d'accélérer le processus de convergence. En effet, on peut remarquer dans l'équation (4.46) que la mise à jour de la moyenne des sources induites est une somme de deux termes dont l'un est la moyenne à l'itération précédente, ce qui crée une dépendance

entre les itérées de cette variable, dépendance qui n'existe pas pour les autres variables $\boldsymbol{\chi}$, $\boldsymbol{z}$, v_ϵ, v_ξ, v_k et m_k.

D'autre part, la recherche du pas de descente optimal, généralement effectuée de façon linéaire, pourrait être très lente. Réciproquement, utiliser un pas fixe pourrait mener à de mauvais résultats. Néanmoins, des méthodes comme la méthode de Newton offrent souvent de meilleurs résultats. C'est pourquoi, le pas de descente α_w peut être calculé d'une manière optimale afin de garantir une convergence rapide des paramètres de forme vers les estimateurs par le biais de la méthode de Newton appliquée à l'énergie libre négative $\mathcal{F}$ qui s'écrit (cf. annexe C) :

$$\mathcal{F}(\boldsymbol{q}) = \langle p(\boldsymbol{y}, \boldsymbol{\chi}, \boldsymbol{w}, \boldsymbol{z}, \Theta|\mathcal{M})\rangle_{\boldsymbol{q}} + \mathcal{H}(q(\boldsymbol{\chi}, \boldsymbol{w}, \boldsymbol{z}, \Theta)). \tag{4.53}$$

Le pas de descente optimal s'écrit alors :

$$\alpha_w^{opt} = \nabla\mathcal{F}(\alpha_w)/\nabla^2\mathcal{F}(\alpha_w)\Big|_{\alpha_w=0}, \tag{4.54}$$

avec $\nabla\mathcal{F} = \partial\mathcal{F}/\partial\alpha_w$ et $\nabla^2\mathcal{F} = \partial^2\mathcal{F}/\partial^2\alpha_w$, ce qui conduit à l'expression analytique :

$$\alpha_w^{opt} = \frac{\dot{\widetilde{\boldsymbol{m}}}_{w0}\boldsymbol{d}_w + \frac{1}{2}\sum_i \boldsymbol{S}_{w_i}^2}{\ddot{\widetilde{\boldsymbol{m}}}_{w0}\boldsymbol{d}_w + \dot{\widetilde{\boldsymbol{m}}}_{w0}\dot{\boldsymbol{d}}_w + \frac{1}{2}\sum_i \boldsymbol{S}_{w_i}^3}, \tag{4.55}$$

où $\dot{}$ et $\ddot{}$ désignent respectivement les dérivées première et seconde et :

$$\begin{aligned}
\boldsymbol{S}_w &= \mathbf{Diag}\left(\overline{v_\epsilon^{-1}}\Gamma^o + \overline{v_\xi^{-1}}\overline{\Gamma^c}\right)\widetilde{\boldsymbol{V}}_w - 1, \\
\boldsymbol{d}_w &= \Bigg[\overline{v_\epsilon^{-1}}\boldsymbol{G}^{o^H}(\boldsymbol{y} - \boldsymbol{G}^o\widetilde{\boldsymbol{m}}_w) + \overline{v_\xi^{-1}}\Big(\widetilde{\boldsymbol{m}}_\chi\boldsymbol{E}^{inc} - \boldsymbol{G}^{c^H}\left(\widetilde{\boldsymbol{m}}_\chi^2 + \widetilde{\boldsymbol{V}}_\chi\right)\boldsymbol{E}^{inc} \\
&- \widetilde{\boldsymbol{m}}_w + \widetilde{\boldsymbol{m}}_\chi\boldsymbol{G}^c\widetilde{\boldsymbol{m}}_w + \boldsymbol{G}^{c^H}\widetilde{\boldsymbol{m}}_\chi^*\widetilde{\boldsymbol{m}}_w - \boldsymbol{G}^{c^H}\left(\widetilde{\boldsymbol{m}}_\chi^2 + \widetilde{\boldsymbol{V}}_\chi\right)\boldsymbol{G}^c\widetilde{\boldsymbol{m}}_w\Big)\Bigg].
\end{aligned}$$

Finalement l'algorithme du gradient, comme approche bayésienne variationnelle, se

résume ainsi :

Gradient comme bayésien variationnel (G-BV)

1 – Initialiser $\boldsymbol{\chi}^{(0)}, \boldsymbol{w}^{(0)}, \boldsymbol{z}^{(0)}, \Theta^{(0)}$.

2 – Mise à jour des paramètres :

itérer :

- calculer α_w^{opt} en utilisant l'équation (4.55),
- calculer $\widetilde{\boldsymbol{\zeta}_k}$ en utilisant l'équation (4.48),
- calculer $(\widetilde{\eta}_\epsilon, \widetilde{\phi}_\epsilon)$ en utilisant l'équation (4.49),
- calculer $(\widetilde{\eta}_\xi, \widetilde{\phi}_\xi)$ en utilisant l'équation (4.50),
- calculer $(\widetilde{\eta}_k, \widetilde{\phi}_k)$ en utilisant l'équation (4.51),
- calculer $(\widetilde{\tau}_k, \widetilde{\mu}_k)$ en utilisant l'équation (4.52),
- calculer $(\widetilde{\boldsymbol{m}}_w, \widetilde{\boldsymbol{v}}_w)$ en utilisant l'équation (4.46),
- calculer $(\widetilde{\boldsymbol{m}}_\chi, \widetilde{\boldsymbol{v}}_\chi)$ en utilisant l'équation (4.47),

jusqu'à convergence. (4.56)

4.4.4 Déroulement de l'algorithme

Après avoir donné les bases théoriques de notre approche, voyons sa mise en œuvre pratique. D'abord, la phase d'initialisation, toujours nécessaire en raison du caractère itératif du problème, se déroule exactement comme pour l'algorithme bayésien variationnel décrit précédemment. On utilise les mêmes techniques pour initialiser les sources induites et le contraste ainsi que le champ caché et les hyper-paramètres.

Il faut cependant noter que, comme nous sommes dans un cadre quasi non supervisé, le paramètre d'adéquation aux données (variance du bruit d'observation, équation (4.49)) risque d'être surestimé par rapport à la variance du bruit de couplage (équation (4.50)). Pour contourner cette difficulté, une idée est d'estimer la variance du bruit d'observation et d'écrire ensuite la variance du bruit de couplage en fonction de celle-ci : $v_\xi = \gamma\, v_\epsilon$, où γ est un facteur de normalisation calculé à partir de l'initialisation des variances des bruits d'observation et de couplage (4.36) de façon à faire intervenir les données, la variable d'intérêt et le champ incident au cours des itérations :

$$\gamma = \frac{||\boldsymbol{y}||_{\mathcal{S}}^2}{||\boldsymbol{X}\boldsymbol{E}^{inc}||_{\mathcal{D}}^2}. \tag{4.57}$$

4.5 Comparaison théorique entre BV, G-BV et $\mathcal{J}^{MAP}$

Avant de conclure, une comparaison théorique s'impose entre les trois approches bayésiennes présentées dans cette thèse, l'algorithme d'optimisation alternée $\mathcal{J}^{MAP}$ discuté dans le chapitre précédent et les deux approches bayésiennes, l'une classique et l'autre de type gradient, présentées ci dessus. Si nous regardons les équations de mise à

jour une à une dans les algorithmes (3.34), (4.30) et (4.56), nous constatons que les mises à jour des paramètres des lois approchantes des deux approches variationnelles sont quasiment les mêmes concernant les moyennes et les variances, sauf en ce qui concerne les courants induits dans la mise à jour desquels apparaît le pas de descente optimal α_w qui doit être déterminé à son tour. En ce qui concerne les mises à jour des paramètres par la méthode $\mathcal{J}^{MAP}$, pour chaque variable nous avons une seule expression, équivalente à la moyenne pour les approches variationnelles, ce qui constitue un point faible de cette approche qui ne procure pas un intervalle de confiance contrairement aux deux autres où une meilleure précision est apportée grâce aux variances. Néanmoins, on peut remarquer qu'il existe une certaine ressemblance entre les expressions des estimées. Par exemple, une comparaison entre les équations (3.24) et (4.19), rappelées ci-dessous, montre que le produit apparaissant dans la première apparaît sous une autre forme dans la seconde, entre la variance et la moyenne avec quelques termes en plus ou en moins :

$$\begin{aligned}
\widehat{\chi}(\boldsymbol{r}) &= \left(v_\xi^{-1}\boldsymbol{E}^H\boldsymbol{E} + v_k^{-1}\right)^{-1} \times \left(v_k^{-1}m_k - v_\xi^{-1}\boldsymbol{w}\boldsymbol{E}^H\right), \\
&= A \times B, \qquad (3.24) \\
\\
\widetilde{\boldsymbol{m}}_\chi^{(i)} &= \widetilde{\boldsymbol{V}}_\chi^{(i)} \times \left(\sum_k \overline{v_k^{-1}}\widetilde{\zeta}_k \circ \widetilde{\boldsymbol{m}}_\chi^\dagger + \overline{v_\xi^{-1}}\,\overline{\boldsymbol{w}\boldsymbol{E}^*}\right), \\
&= \left[\mathbf{Diag}\left(\overline{v_\xi^{-1}}\,\overline{\boldsymbol{E}^2} + \sum_k \widetilde{\boldsymbol{\zeta}}_k(\boldsymbol{r})\overline{v_k^{-1}}\right)\right]^{-1} \times \left(\sum_k \overline{v_k^{-1}}\widetilde{\zeta}_k \circ \widetilde{\boldsymbol{m}}_\chi^\dagger + \overline{v_\xi^{-1}}\,\overline{\boldsymbol{w}\boldsymbol{E}^*}\right), \\
&= C \times D. \qquad (4.19)
\end{aligned}$$

En comparant deux à deux A et C, d'une part, et B et D, d'autre part, on peut voir que A et B se démarquent de C et D par l'absence des opérateurs **Diag** et $\sum_k \widetilde{\boldsymbol{\zeta}}_k(\boldsymbol{r})$ pour A et du terme $\sum_k \overline{v_k^{-1}}\widetilde{\zeta}_k \circ \widetilde{\boldsymbol{m}}_\chi^\dagger$ pour B, ainsi que par la non prise en compte des moyennes par rapport à la densité q.

Finalement, l'algorithme $\mathcal{J}^{MAP}$ est certainement moins coûteux en calculs que les deux algorithmes bayésiens variationnels, mais son implémentation, telle qu'elle est présentée, n'est pas possible vu l'apparition de matrices de très grandes dimensions qui nécessitent des algorithmes d'optimisation globale et l'impossibilité d'une optimisation du champ caché qui requiert une méthode de type échantillonnage. Pour ce qui est des algorithmes BV et G-BV, l'initialisation des paramètres et des hyper-paramètres se passe comme décrit ci dessus. Ensuite le processus itératif commence et, pour pouvoir évaluer la convergence des deux méthodes, une attention particulière est portée à l'évolution des parties réelles et imaginaires des courants induits, du contraste, des moyennes et variances des classes et des moyennes et variances des bruits d'observation et de couplage.

4.6 Conclusions

Dans ce chapitre, nous avons présenté la méthode bayésienne variationnelle aux niveaux fondamental et applicatif. Avec cette approche, d'un point de vue théorique, la

mise à jour des variables doit se faire d'une manière itérative jusqu'à la convergence. Du point de vue pratique, elle passe par la mise à jour des paramètres des lois de probabilité assignées à ces variables. Cette méthode se distingue des méthodes stochastiques par sa rapidité de convergence et, d'après diverses expériences [110], elle présente une vraie alternative à ces dernières. L'incorporation de variables cachées, lors du choix du modèle *a priori*, et la modélisation hiérarchique qui en résulte jouent un rôle important pour simplifier l'implémentation de l'algorithme. Celle-ci a cependant nécessité, ici, une attention particulière car le contraste recherché est à valeurs complexes.

En dernier lieu, pour améliorer la qualité de reconstruction et la rapidité de convergence, nous avons présenté un nouvel algorithme basé sur un principe bayésien variationnel couplé à une méthode de descente de gradient.

Afin de tester l'efficacité de ces deux approches, nous les avons appliquées à notre problème d'imagerie et les résultats de reconstruction sont présentés dans le prochain chapitre.

5

Résultats d'inversion

Sommaire

5.1 Introduction 75
5.2 Résultats de reconstruction 76
5.2.1 Qualité de reconstruction 77
5.2.2 Complexité algorithmique 87
5.3 Données bruitées 88
5.4 Paramètre de corrélation 91
5.5 MGI vs MGM 92
5.6 Conclusions 92

5.1 Introduction

Dans ce chapitre, nous présentons les différents résultats de simulations obtenus par le biais des méthodes bayésiennes variationnelles décrites dans le chapitre 4, appliquées aux différents fantômes de sein (1,2,3 et 4) décrits dans le chapitre 2. Une des principales difficultés rencontrées pour l'imagerie du sein est liée au milieu d'adaptation. En effet, les valeurs de permittivité diélectrique et de conductivité de ce dernier influent sur la bonne pénétration des ondes dans le sein et, par conséquent, sur la reconstruction des différentes composantes de celui-ci. Notons que la fréquence de l'onde interrogatrice joue également un rôle dans la qualité de la reconstruction. Ici, les images sont reconstruites à l'aide de six fréquences utilisées simultanément dans le processus d'inversion et réparties régulièrement dans la bande $[0.5 - 3]$ GHz.

Une autre difficulté réside dans le choix du nombre de classes segmentant l'objet. Nous avons déjà signalé que ce nombre pouvait être choisi comme grand au départ du processus et qu'il convergeait alors vers le nombre optimal durant les itérations.

Ceci a été montré par l'expérience mais n'a pas encore été prouvé théoriquement. C'est pourquoi, pour les fantômes complexes étudiés, nous avons testé nos algorithmes avec divers nombres de classes, jusqu'au choix final de 4. Ce choix a été motivé par les travaux publiés dans [91, 92, 93] où il est dit que les tissus mammaires peuvent être classifiés en 4 groupes selon leur propriétés diélectriques, dont 3 correspondent aux tissus sains, alors classifiés selon leur contenu en graisse, et le quatrième correspond aux tissus cancéreux.

Le choix du paramètre de corrélation du modèle de Potts joue, lui aussi, un rôle important dans la qualité de reconstruction surtout pour les fantômes de sein les plus sophistiqués. Notons que le fait de fixer *a priori* le nombre de classes et la valeur du paramètre de corrélation de Potts conduit à des méthodes bayésiennes quasi non supervisées. Notons finalement aussi que, pour les 3 premiers fantômes, seul le modèle *a priori* de mélange de gaussiennes indépendantes (MGI) a été employé vu l'homogénéité des parties composant le sein. Pour le fantôme issu d'une IRM d'un vrai sein, le modèle de mélange de Gauss-Markov (MGM) a également été testé afin de mieux décrire l'hétérogénéité des tissus mammaires considérés.

Dans le reste du chapitre nous présentons les résultats de reconstruction d'une manière variée afin de montrer le maximum de points forts des algorithmes d'estimation bayésienne considérés, en les comparant à une approche déterministe d'inversion contraste source (CSI). Par ailleurs, nous discuterons également leurs limitations à travers les remarques faites au fur et à mesure de l'exposé des résultats obtenus.

5.2 Résultats de reconstruction

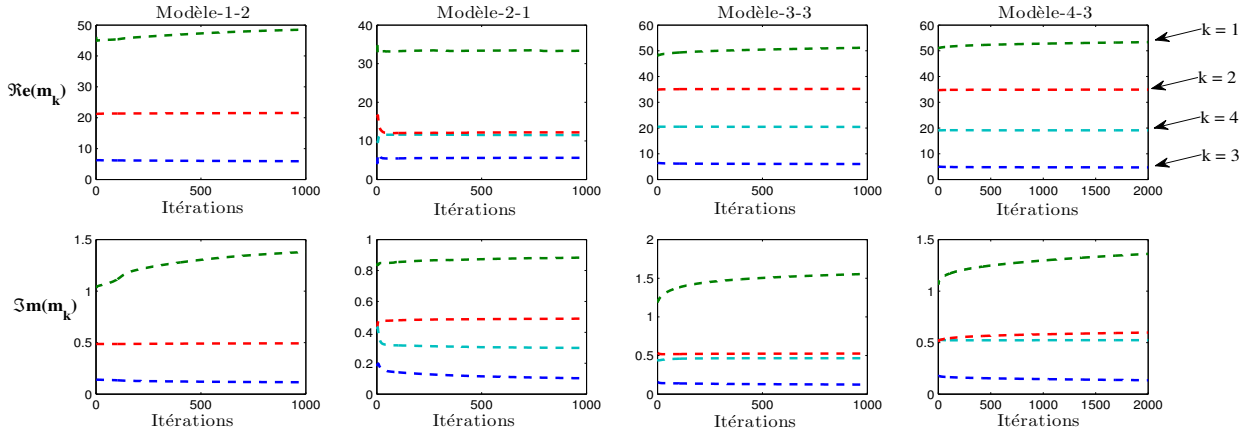


FIGURE 5.1 – Exemples de comportement des parties réelles et imaginaires des moyennes des classes pour 4 fantômes différents.

Les méthodes présentées dans le chapitre précédent ont été appliquées aux configurations décrites dans le chapitre 2. Le domaine test est divisé en 64×64 pixels élémentaires de coté $\delta = 1,9$ mm pour les 3 premiers fantômes et en 120×120 pixels élémentaires de coté $\delta = 1$ mm pour le dernier fantôme. L'algorithme CSI est arrêté après 500 itérations pour les 3 premiers fantômes et après 1000 itérations pour le dernier. Notons que cet algorithme commence à se stabiliser relativement rapidement (environ 300 et

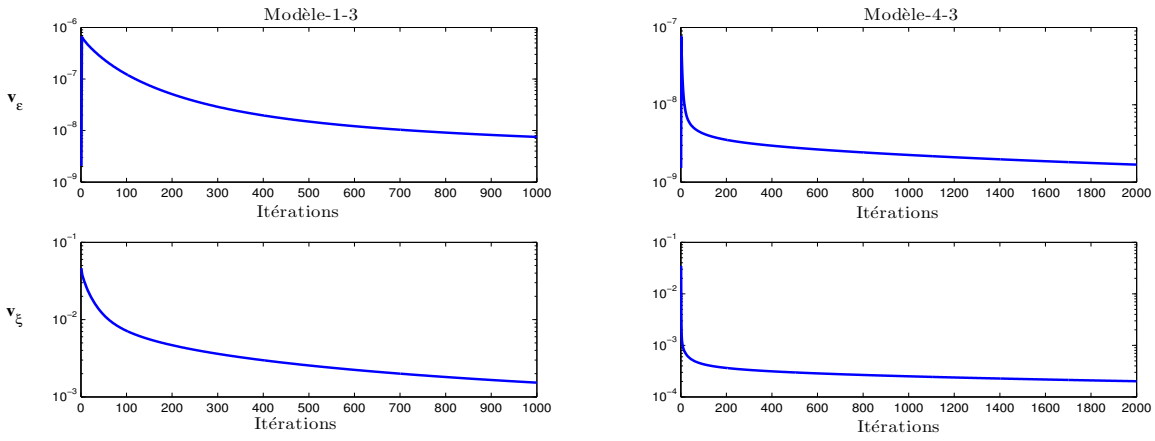


FIGURE 5.2 – Exemples de comportement des variances des bruits d'observation et de couplage pour 2 fantômes différents.

600 itérations dans les cas précédents) par rapport aux autres approches. En effet, les algorithmes bayésiens variationnels simple (BV) et du type gradient (G-BV) sont arrêtés après 1000 itérations pour les 3 premiers fantômes et après 2000 pour le dernier. Leur convergence est estimée par le comportement des hyper-paramètres au cours des itérations (figures (5.1) et (5.2)). Par ailleurs, les tableaux (5.1) et (5.2) résument respectivement les paramètres des configurations et les valeurs assignées aux paramètres des algorithmes bayésiens.

TABLE 5.1 – Récapitulatif des paramètres des configurations pour les 4 fantômes.

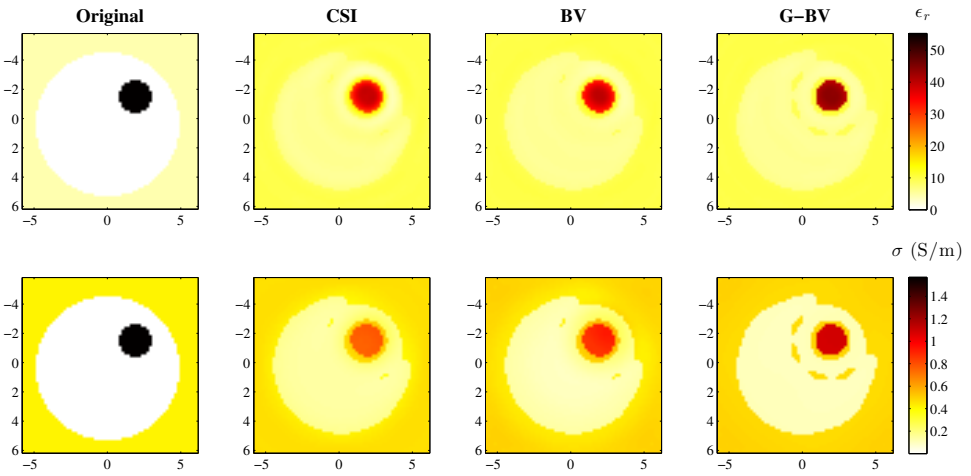
$N_{\mathcal{D}}$	δ (mm)	$N_f \times N_{\mathcal{S}} \times N_\vartheta$	F (GHz)
64×64 et 120×120	$1,9$ et 1	$6 \times 64 \times 64$	$0,5 - 3$

TABLE 5.2 – Récapitulatif des paramètres des algorithmes BV et G-BV.

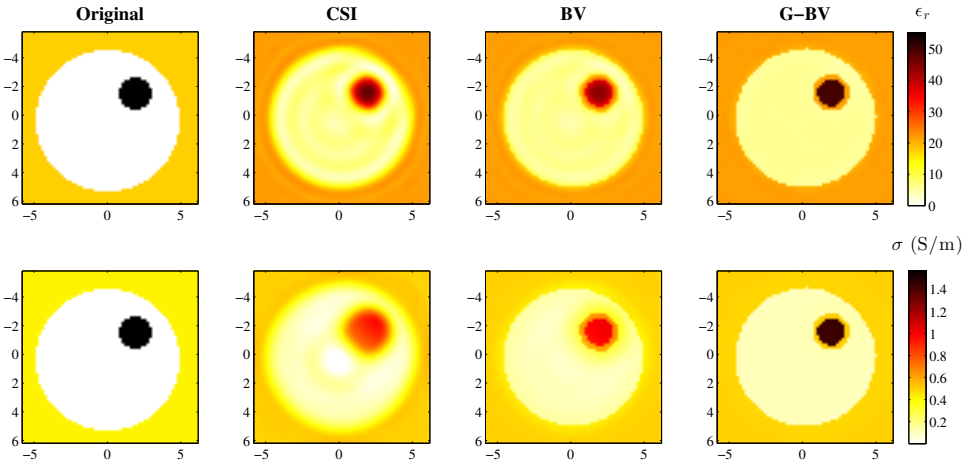
λ	μ_k	τ_k	η_k	ϕ_k
1 et 10	$0,5$	10^{-4}	500	10^{-1}
K	η_ϵ	ϕ_ϵ	η_ξ	ϕ_ξ
3 et 4	10^{-7}	1	5×10^{-4}	5

5.2.1 Qualité de reconstruction

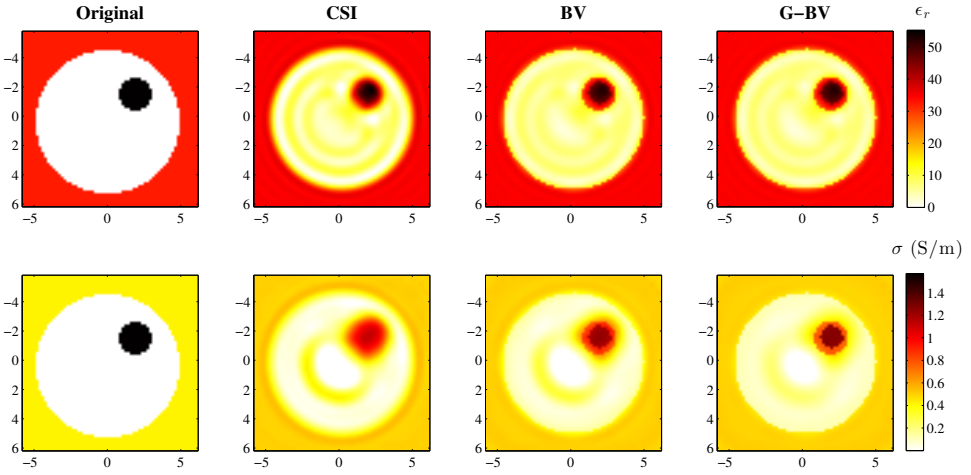
Les figures (5.3), (5.4), (5.5) et (5.6) présentent les résultats de reconstruction des permittivités diélectriques et conductivités des objets placés dans différents milieux d'adaptation et reconstruits par le biais des l'algorithmes BV et G-BV, comparés aux résultats obtenus par la méthode CSI. Pour les 3 premiers fantômes, les 3 algorithmes parviennent à retrouver quasiment toutes les régions correspondants aux différents tissus



(a) Fantôme 1-1

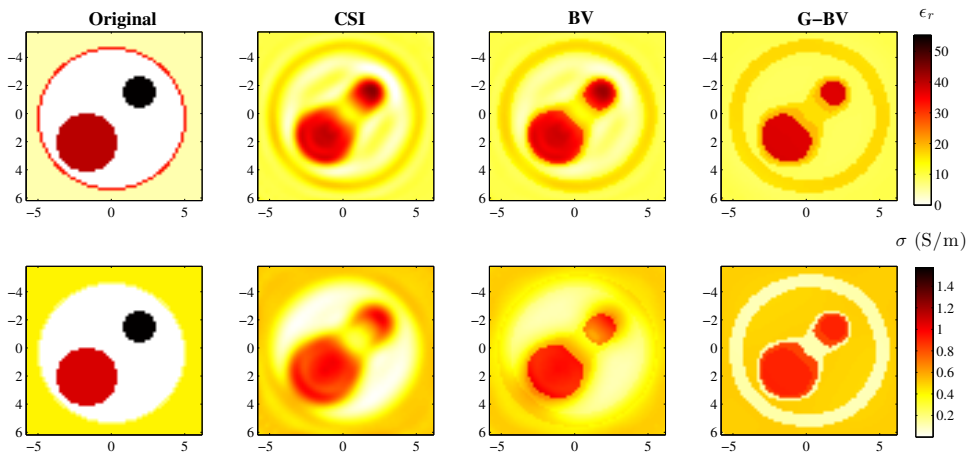


(b) Fantôme 1-2

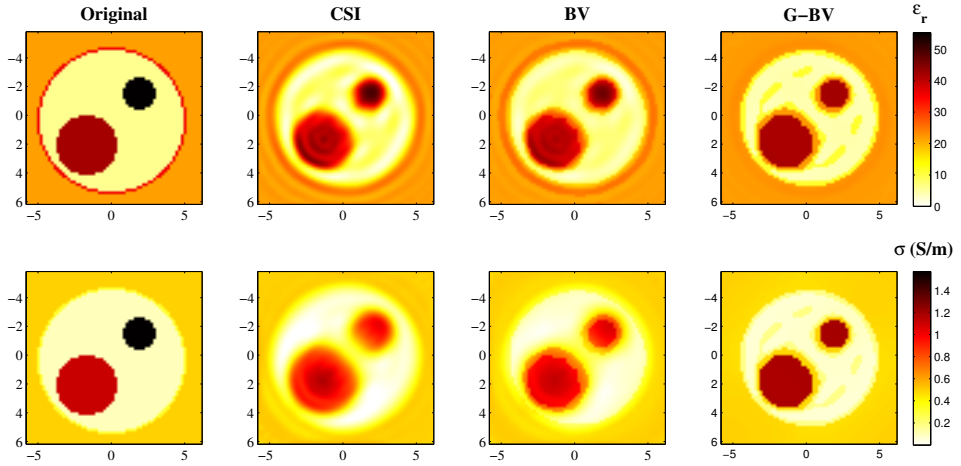


(c) Fantôme 1-3

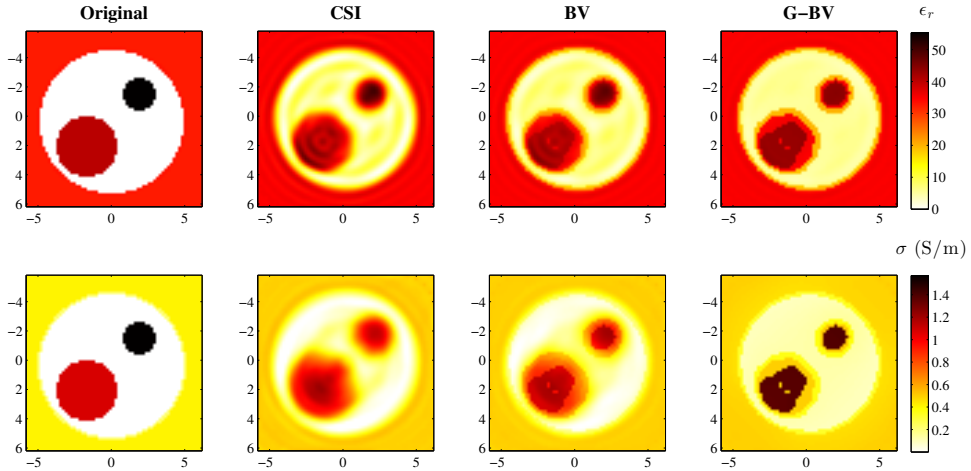
FIGURE 5.3 – La permittivité ϵ_r et la conductivité σ reconstruites à l'aide de CSI (2^e colonne), BV (3^e colonne) et G-BV (4^e colonne) comparées aux valeurs réelles (1^{re} colonne).



(a) Fantôme 2-1

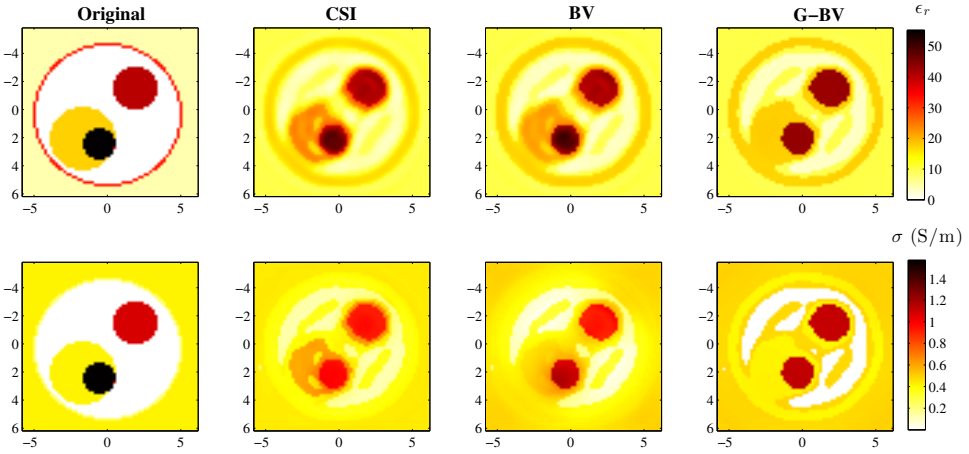


(b) Fantôme 2-2

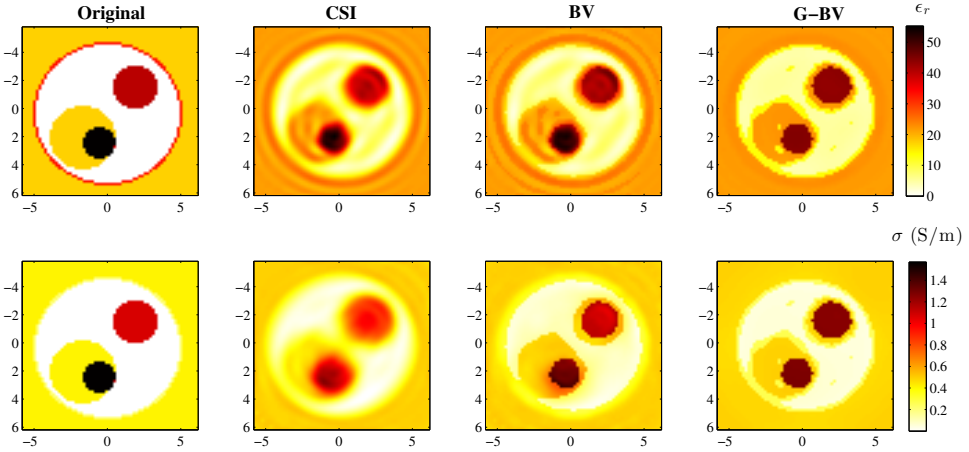


(c) Fantôme 2-3

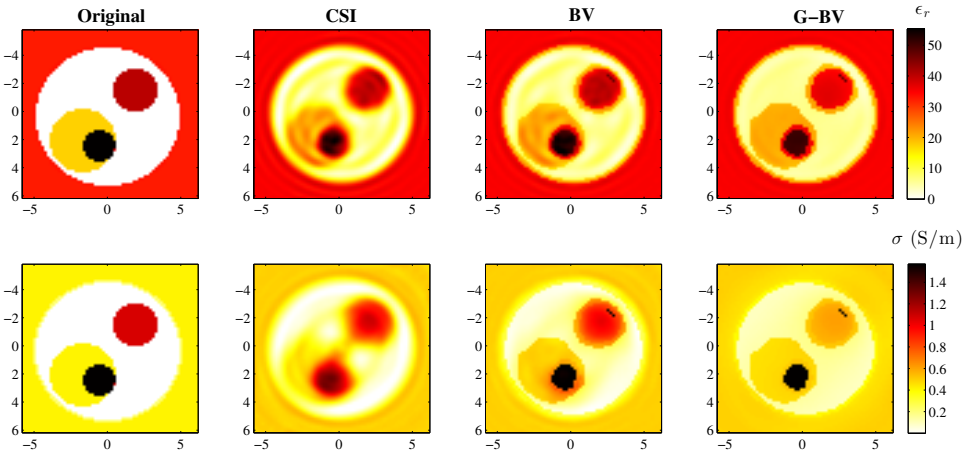
FIGURE 5.4 – La permittivité ϵ_r et la conductivité σ reconstruites à l'aide de CSI (2^e colonne), BV (3^e colonne) et G-BV (4^e colonne) comparées aux valeurs réelles (1^{re} colonne) pour les fantômes 2-1, 2-2 et 2-3.



(a) Fantôme 3-1

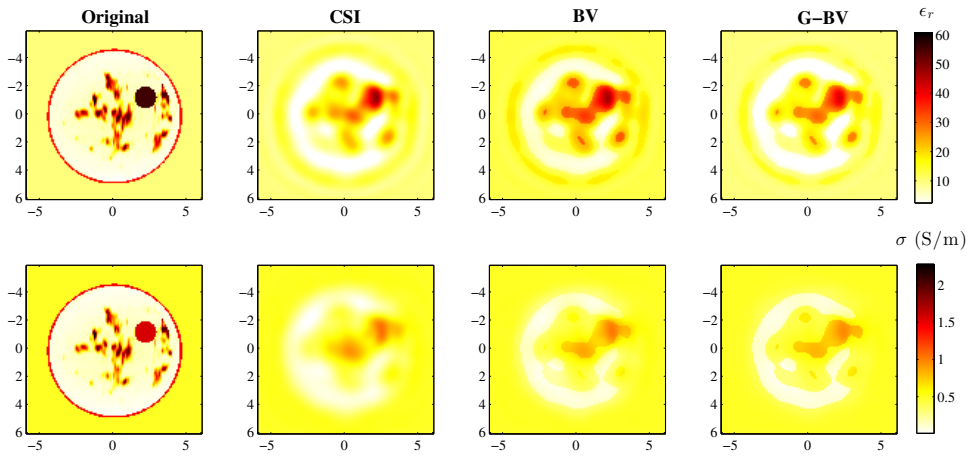


(b) Fantôme 3-2

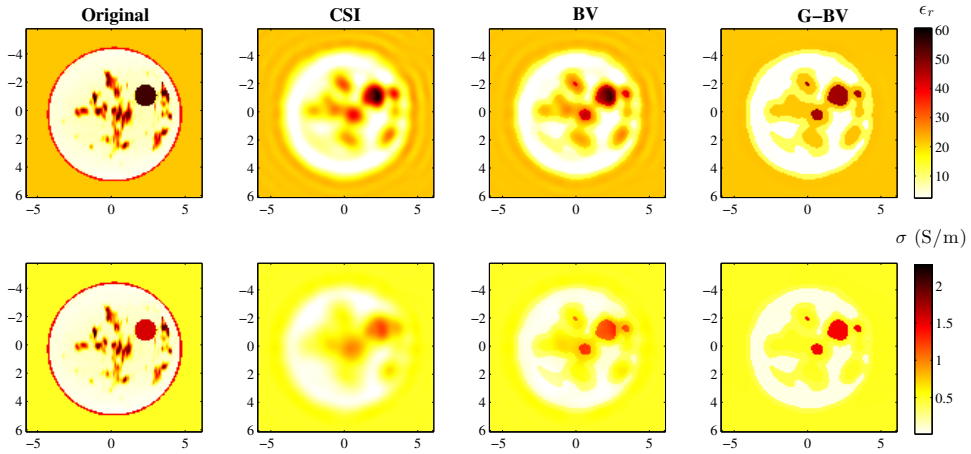


(c) Fantôme 3-3

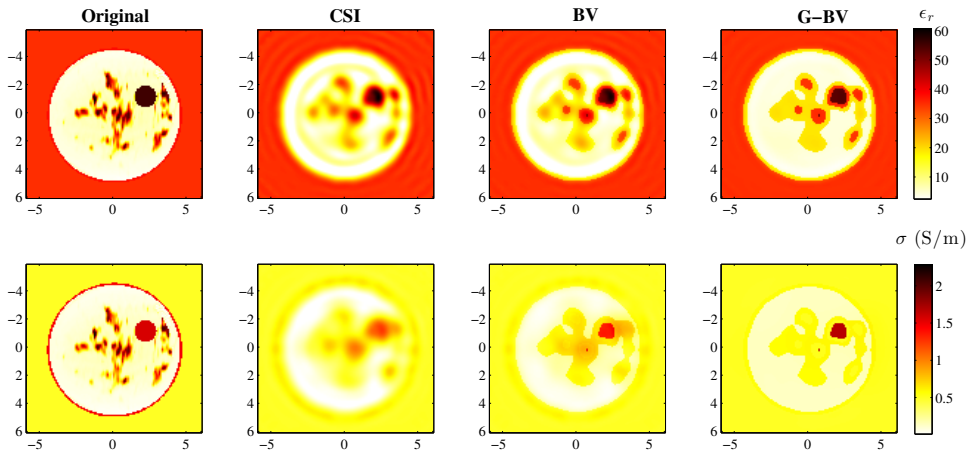
FIGURE 5.5 – La permittivité ϵ_r et la conductivité σ reconstruites à l'aide de CSI (2^e colonne), BV (3^e colonne) et G-BV (4^e colonne) comparées aux valeurs réelles (1^{re} colonne) pour les fantômes 3-1, 3-2 et 3-3.



(a) Fantôme 4-1



(b) Fantôme 4-2



(c) Fantôme 4-3

FIGURE 5.6 – La permittivité ϵ_r et la conductivité σ reconstruites à l'aide de CSI (2^e colonne), BV (3^e colonne) et G-BV (4^e colonne) comparées aux valeurs réelles (1^{re} colonne) pour les fantômes 4-1, 4-2 et 4-3.

présents, i.e. la graisse, les milieux glandulaires et la tumeur. Seule la peau n'a pas été retrouvée d'une manière efficace dans les cas où ses paramètres électromagnétiques se rapprochent de ceux du milieu d'adaptation. Mais, généralement, la forme ainsi que la localisation des différentes régions correspondant aux différents tissus sont relativement bien estimées, avec de meilleures précision et résolution pour les algorithmes bayésiens. Notons que, dans la plupart des cas, la méthode G-BV réussit à retrouver avec une meilleure exactitude les cartes de permittivité et de conductivité que la méthode BV, mais n'est pas plus rapide quand il s'agit des fantômes les plus complexes (table (5.3)). Notons que l'efficacité des divers algorithmes est estimée numériquement au travers le calcul du rapport signal sur bruit (PSNR) donné par :

$$\mathrm{PSNR\,(dB)} = 10\log_{10}\left(\frac{N_{\mathcal{D}}\,|\chi_{max}|^2}{\sum_{i=1}^{N_{\mathcal{D}}}|\chi(\boldsymbol{r}_i)-\chi_a(\boldsymbol{r}_i)|^2}\right), \tag{5.1}$$

où l'indice a désigne la valeur actuelle et χ_{max} désigne la valeur maximale que prend le contraste.

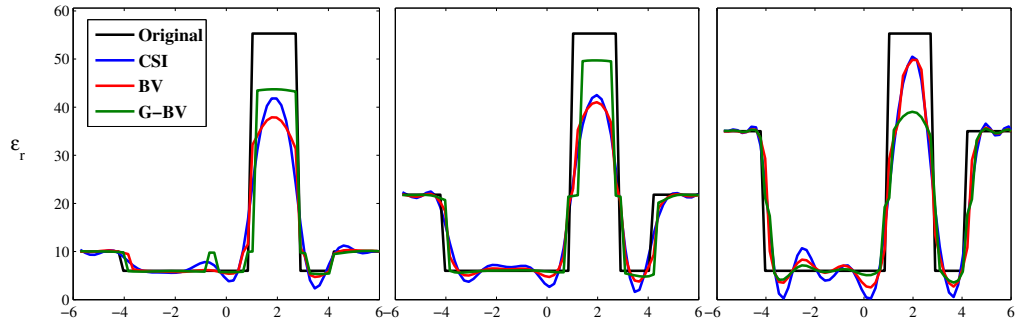
TABLE 5.3 – Comparaison des différentes approches en termes de temps de calcul et de PSNR pour 4 fantômes différents.

Fantômes	Critères	CSI	BV	G-BV
Fantôme 1-2	CPU (min)	26,75	29	23,17
	PSNR (dB)	71,59	69,44	84,17
Fantôme 2-1	CPU (min)	20,50	22,50	25,80
	PSNR (dB)	49,25	50,61	52,84
Fantôme 3-3	CPU (min)	21,69	26	37,6
	PSNR (dB)	39,14	46,03	47,10
Fantôme 4-3	CPU (min)	73,06	76,58	82,13
	PSNR (dB)	44,50	51,46	49,70

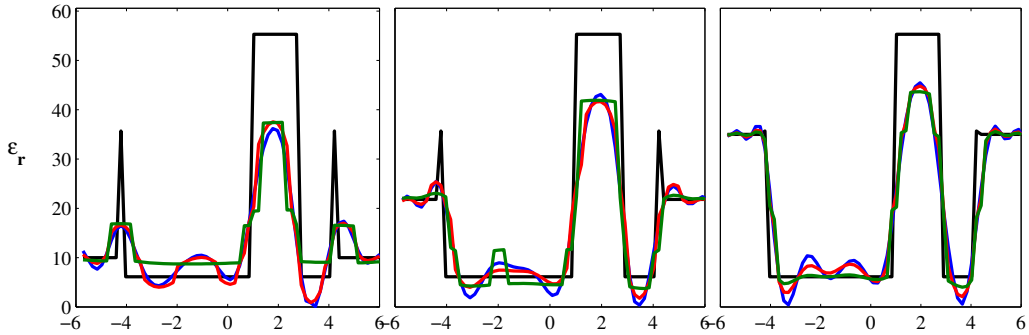
TABLE 5.4 – Le nombre de pixels mal classés par rapport à la segmentation de référence pour 4 fantômes différents.

Fantôme	CSI	BV	G-BV
Fantôme 1-2	341	311	301
Fantôme 2-3	622	528	519
Fantôme 3-2	528	519	503
Fantôme 4-3	4510	3267	3272

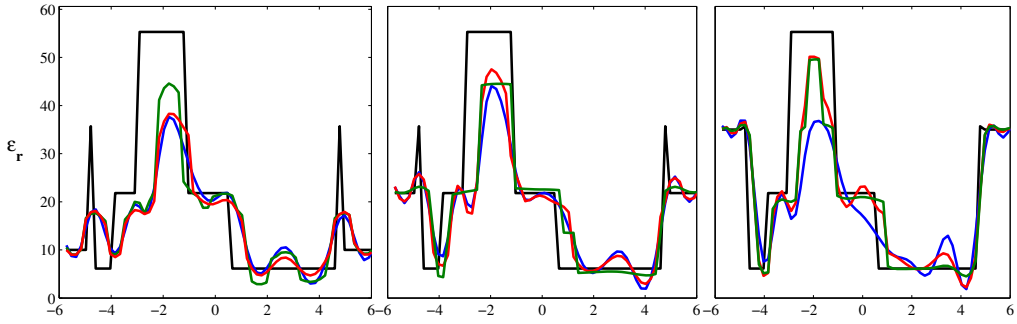
Quant au fantôme 4, les parties glandulaires ne sont pas reconstruites de manière satisfaisante. En effet, il est difficile de retrouver le contraste correspondant aux zones glandulaires de permittivité et conductivité faibles à moyennes. Seules quelques zones présentant des contrastes élevés sont retrouvées et, en particulier, les 3 méthodes réussissent à imager la tumeur. Les figures (5.7) et (5.8) présentent les profils reconstruits le long d'une ligne traversant le centre de la tumeur et confirme la supériorité des mé-



(a) Fantôme 1

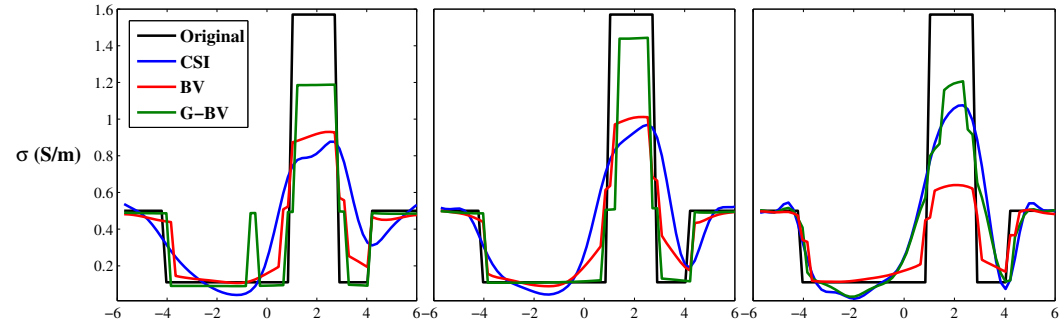


(b) Fantôme 2

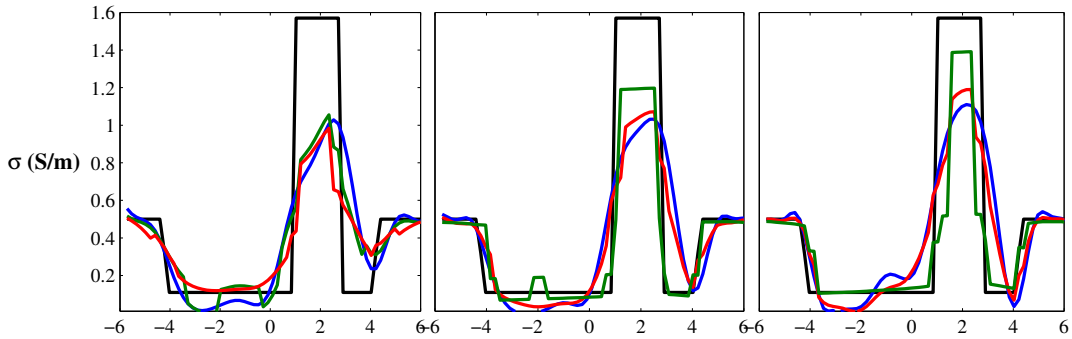


(c) Fantôme 3

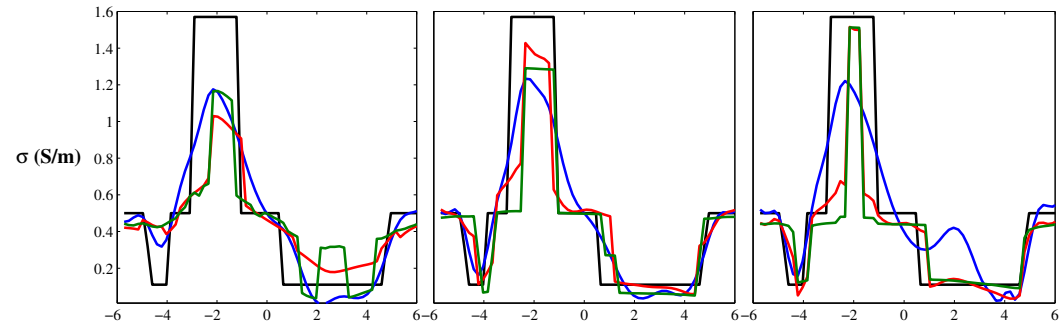
FIGURE 5.7 – Profils de permittivité reconstruits le long d'une ligne qui passe par le centre de la tumeur pour les différents fantômes (gauche : milieu d'adaptation 1, centre : milieu d'adaptation 2, droite : milieu d'adaptation 3).



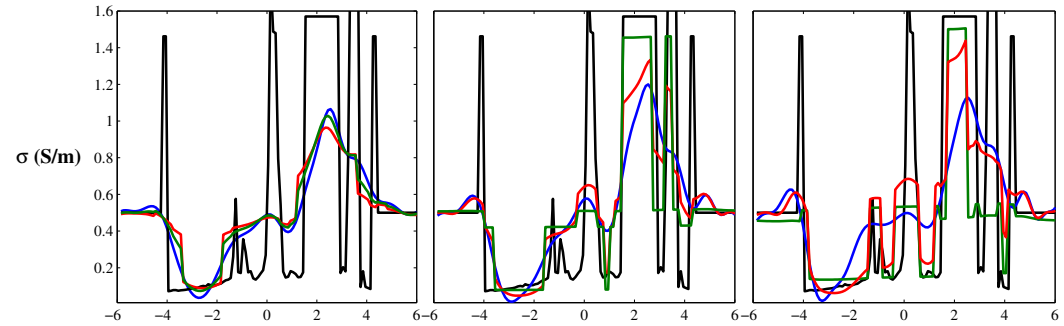
(a) Fantôme 1



(b) Fantôme 2

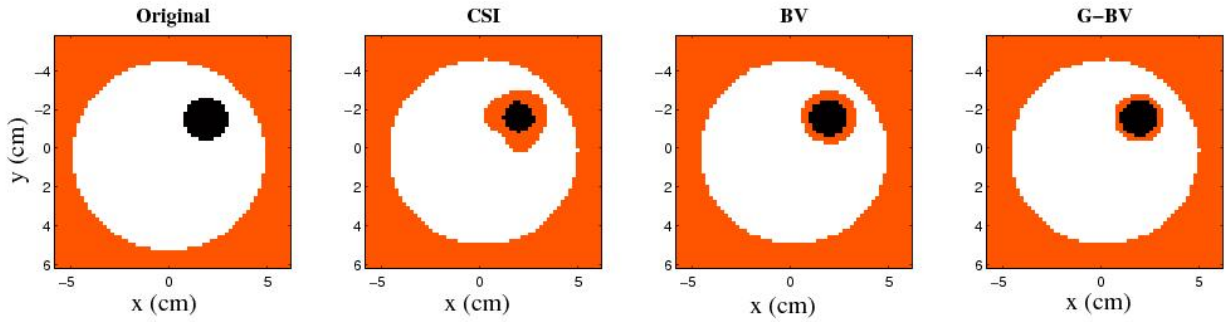


(c) Fantôme 3

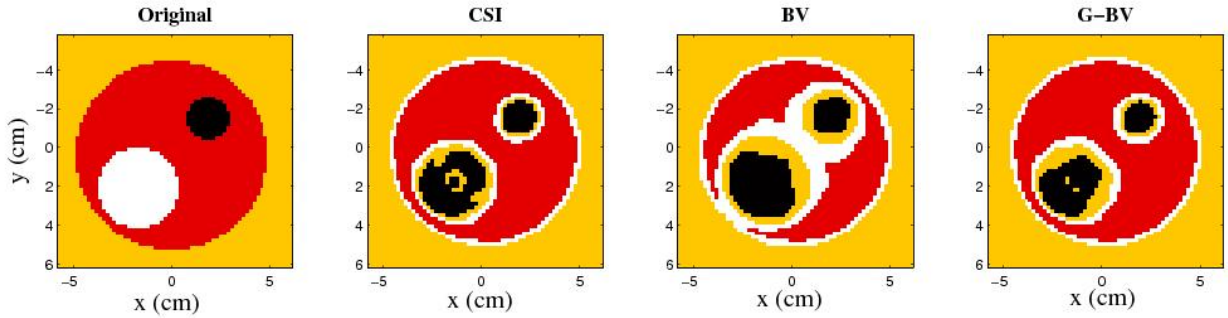


(d) Fantôme 4

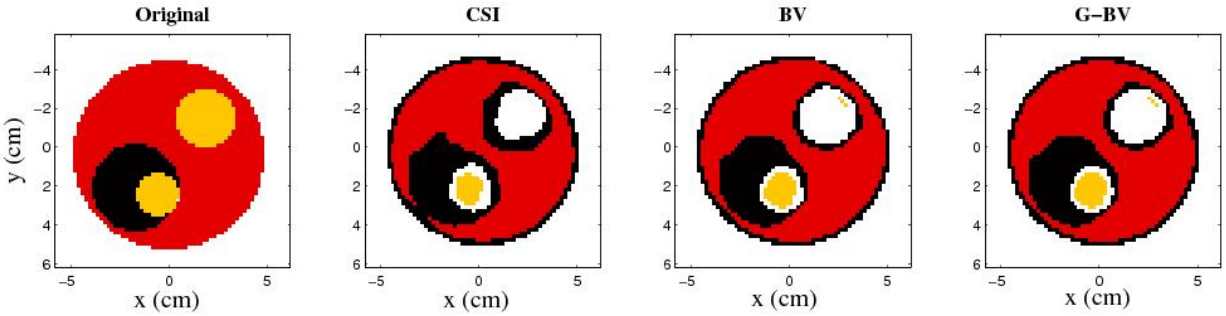
FIGURE 5.8 – Profils de conductivité reconstruits le long d'une ligne qui passe par le centre de la tumeur pour les différents fantômes (gauche : milieu d'adaptation 1, centre : milieu d'adaptation 2, droite : milieu d'adaptation 3).



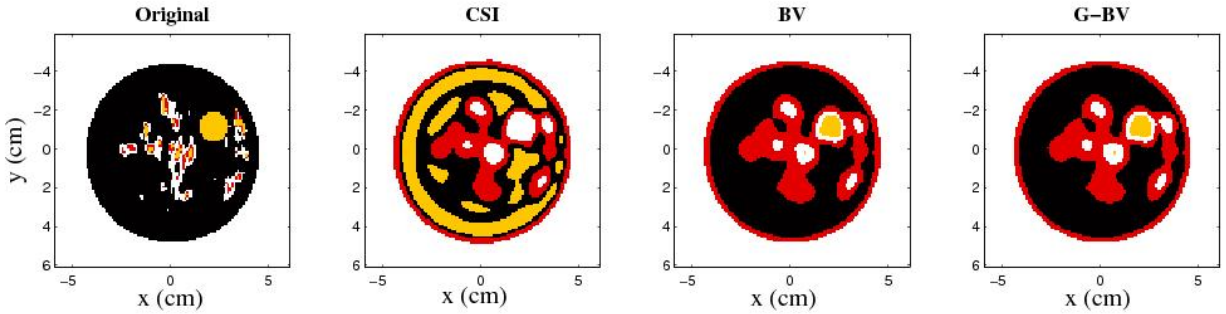
(a) Fantôme 1-2



(b) Fantôme 2-3



(c) Fantôme 3-2



(d) Fantôme 4-3

FIGURE 5.9 – Champs cachés obtenus par le biais des approches BV et G-BV comparés aux segmentations des images réelles et de celles obtenues par CSI par le biais de la méthode K-means (K=4) pour 4 fantômes différents.

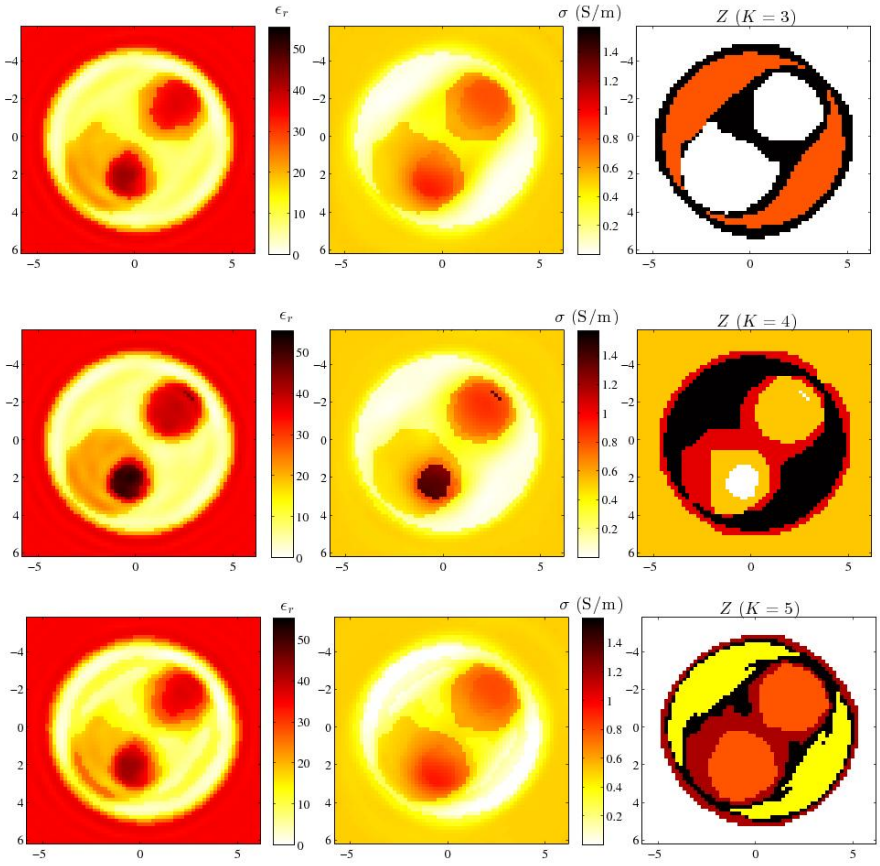


FIGURE 5.10 – Résultats de reconstruction du fantôme 3-3 après 1000 itérations par le biais de la méthode BV utilisant trois nombres de classes différents.

thodes bayésiennes par rapport à la méthode déterministe, allant ainsi dans le sens des constations faites sur les images obtenues.

D'autre part, la figure (5.9) montre les champs cachés correspondants à 4 fantômes différents comparés à la segmentation initiale obtenue au moyen de l'algorithme *K-means*. Il apparaît que la méthode G-BV offre une meilleure segmentation que la méthode BV. Le tableau (5.4), qui dénombre les pixels différents de ceux de la segmentation initiale, confirme cette constatation. Notons que divers tests ont été effectués avec des nombres de classes différents ; les résultats sont présentés sur la figure (5.10). Cette dernière montre les résultats obtenus pour le fantôme 3-3 par le biais de l'approche BV au bout de 1000 itérations. On constate que la segmentation offerte par $K = 4$ est plus significative que celles obtenues pour $K = 3$ et $K = 5$. Le nombre de classes à donc été fixé à $K = 4$ pour la suite de l'étude.

5.2.2 Complexité algorithmique

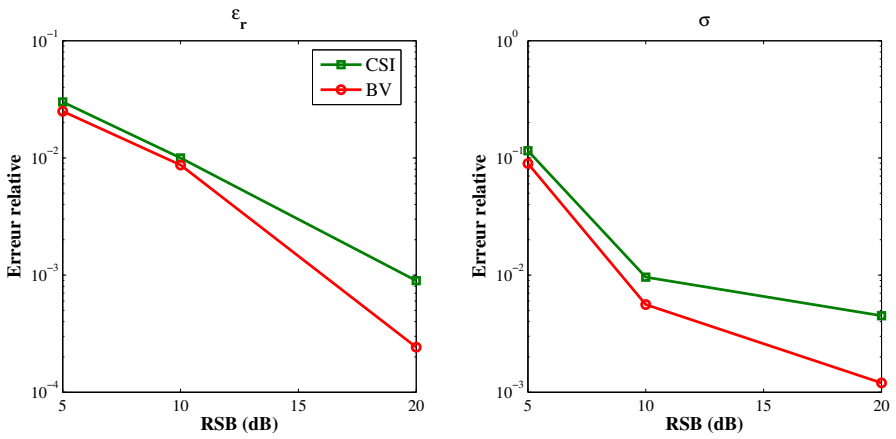
TABLE 5.5 – Tableau résumant le nombre d'itérations et le temps de calcul nécessaires pour assurer la convergence des 3 algorithmes pour le cas du fantôme 1-1.

Critères \ Algorithmes	CSI	BV	G-BV
Nombre itérations pour stabilisation	≈ 300	≈ 600	≈ 600
Nombre itérations pour convergence	≈ 500	≈ 1000	≈ 1000
Temps (s)/ itération	$\approx 4,50$	$\approx 1,50$	$\approx 1,23$
Temps (min) pour convergence	≈ 37	≈ 25	≈ 20

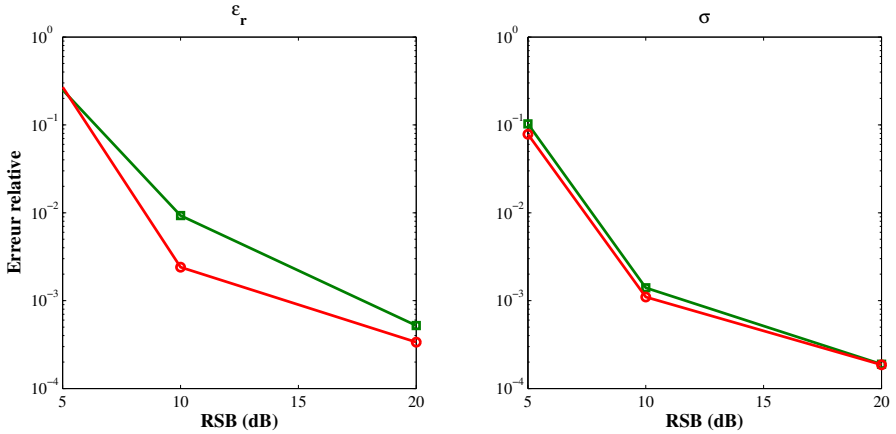
TABLE 5.6 – Tableau résumant le nombre d'itérations et le temps de calcul nécessaires pour assurer la convergence des 3 algorithmes pour le cas du fantôme 4-2.

Critères \ Algorithmes	CSI	BV	G-BV
Nombre itérations pour stabilisation	≈ 600	≈ 1400	≈ 1400
Nombre itérations pour convergence	≈ 1000	≈ 2000	≈ 2000
Temps (s)/ itération	$\approx 13,85$	$\approx 5,98$	$\approx 7,80$
Temps (min) pour convergence	≈ 75	≈ 81	≈ 90

Du point de vue de la complexité algorithmique, les trois algorithmes (BV, G-BV et CSI), codés avec Matlab®, nécessitent pratiquement le même stockage en mémoire avec un peux plus d'espace pour les méthodes bayésiennes, dû à la présence du champ caché et des hyper-paramètres, et, même si la méthode CSI nécessite moins d'itérations pour atteindre la convergence que les méthodes bayésiennes, elle consomme beaucoup plus de temps par itération. Par exemple, pour le fantôme 1-3, CSI nécessite $4,05$ s/itération contre $1,75$ s/itération pour la méthode BV et $1,39$ s/itération pour la méthode G-BV. Les tableaux (5.5) et (5.6) montrent deux autres exemples de convergence des 3 algorithmes qui concernent la reconstruction des fantômes 1-1 et 4-2.



(a) Fantôme 3-1

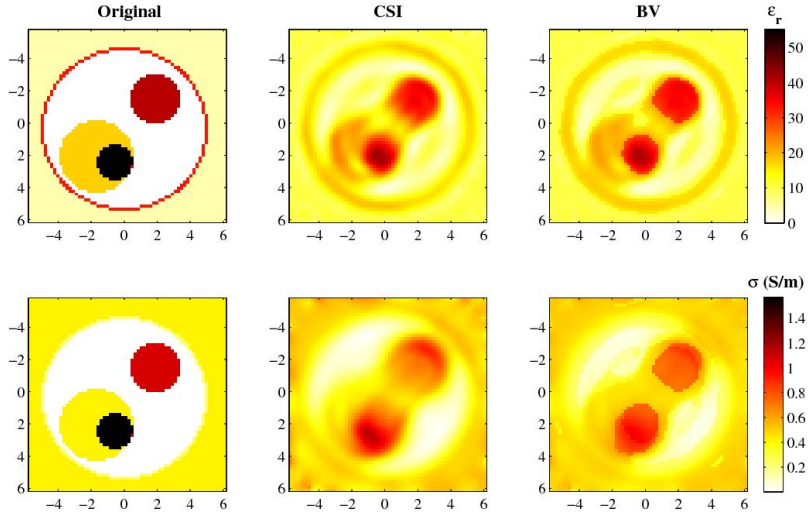


(b) Fantôme 4-3

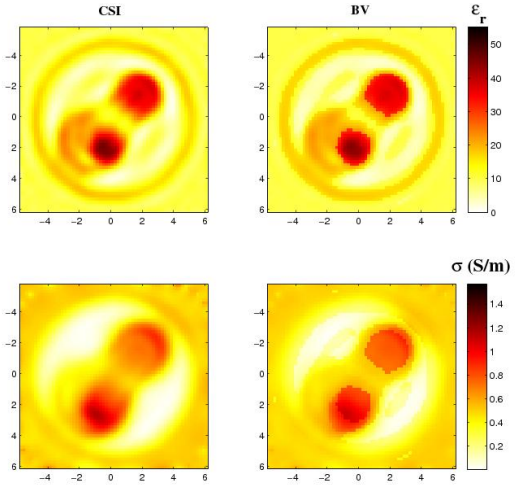
FIGURE 5.11 – Erreur relative de reconstruction des fantômes 3-1 et 4-3 en fonction du rapport signal sur bruit.

5.3 Données bruitées

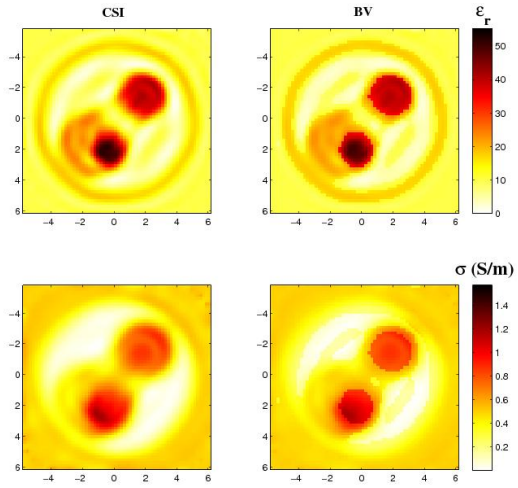
Afin de tester la sensibilité des méthodes bayésiennes vis à vis du bruit de mesure, nous avons bruité deux jeux de données (fantômes 3-1 et 4-3) avec des bruits additifs blancs gaussiens de rapports signal à bruit égaux à 5 dB, 10 dB et 20 dB, définis par rapport au champ diffracté. Bien sûr, ceci n'est pas une modélisation réaliste du bruit pouvant contaminer des données expérimentales acquises en situation contrôlée, où le bruit n'est pas forcément gaussien et où le processus de bruitage intervient en deux temps, d'une part lors de l'acquisition du champ incident en l'absence d'objet et, d'autre part, lors de l'acquisition du champ total en présence de celui-ci, le champ diffracté étant alors obtenu en faisant la différence entre ces deux champs. Cependant ce bruit permet tout de même de montrer la robustesse des méthodes développées face à un bruit de mesure que l'on peut quantifier (cf. figures (5.11), (5.12) et (5.13)). Rappelons que, comme souligné dans le chapitre 2, un bruit de modèle est déjà systématiquement pris en compte lors de la génération des données.



(a) SNR = 5 dB

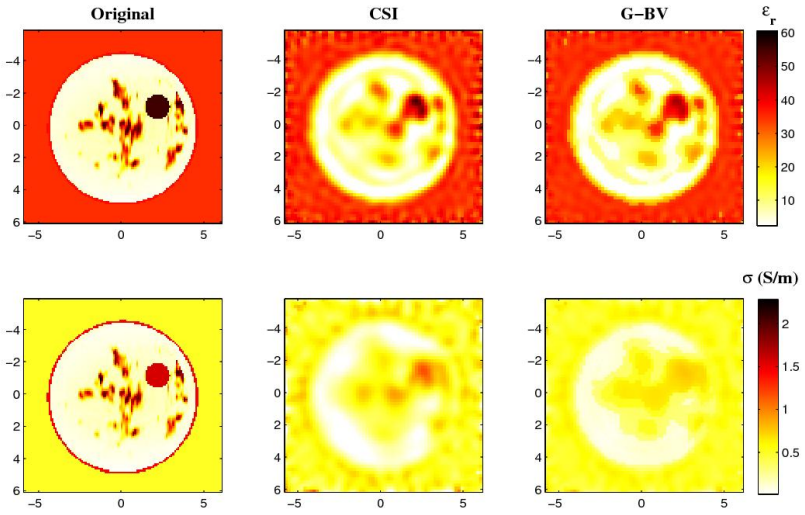


(b) SNR = 10 dB

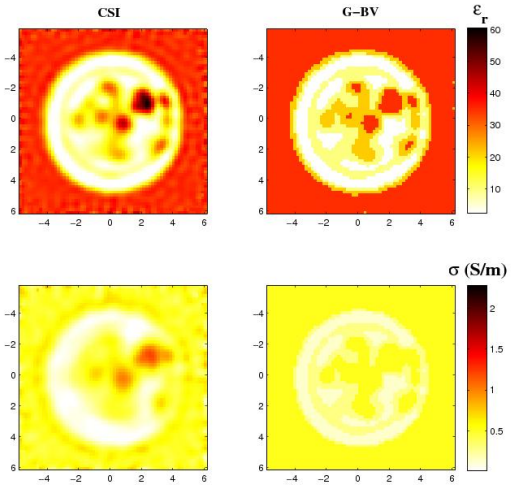


(c) SNR = 20 dB

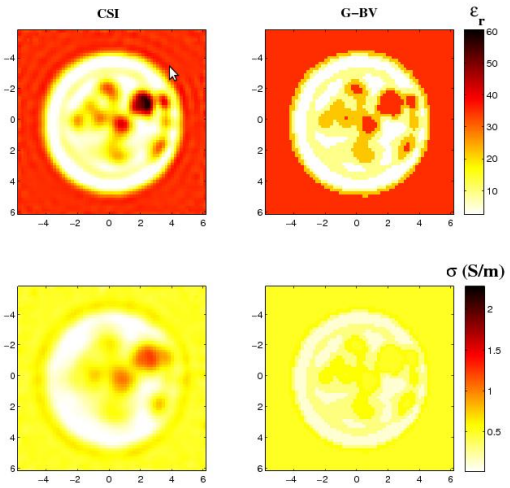
FIGURE 5.12 – Résultats de reconstruction du fantôme 3-1 à l'aide de la méthode BV pour des données de champ diffracté bruitées avec des rapports signal à bruit de 5 dB, 10 dB et 20 dB.



(a) SNR = 5 dB



(b) SNR = 10 dB



(c) SNR = 20 dB

FIGURE 5.13 – Résultats de reconstruction du fantôme 4-3 à l'aide de la méthode BV pour des données de champ diffracté bruitées avec des rapports signal à bruit de 5 dB, 10 dB et 20 dB.

5.4 Paramètre de corrélation

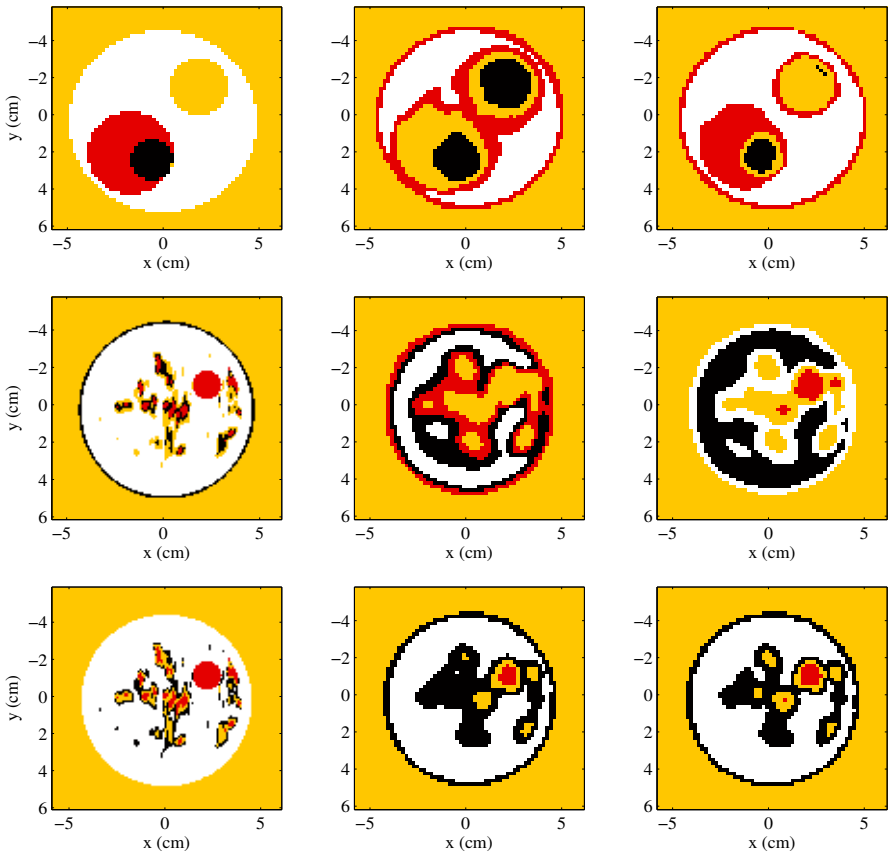


FIGURE 5.14 – Champs cachés correspondant aux fantômes 3-3 (haut), 4-2 (milieu) et 4-3 (bas) obtenus avec deux valeurs différentes du paramètre de Potts : $\lambda = 1$ (milieu), $\lambda = 10$ (droite) comparés aux segmentations des objets réels (gauche) par la méthode K-means ($K = 4$).

La figure (5.14) présente une étude de l'influence du paramètre de Potts, qui contrôle la corrélation entre les pixels voisins des différents classes, sur les résultats de segmentation comme indiqué dans le paragraphe (3.3). Lorsque la valeur de ce paramètre est inférieure à la valeur critique, les régions segmentées ne sont pas complètement homogènes et peuvent présenter quelques granularités, tandis que lorsque sa valeur est supérieure à la valeur critique, la segmentation des régions est plus uniforme et homogène. Ceci illustre effectivement les propriétés d'un champ de Potts : plus la valeur du paramètre de corrélation est grande, plus les régions segmentées sont grandes et homogènes. Sur la figure (5.14) les résultats obtenus pour le fantôme 3-3 sont bien meilleurs pour $\lambda = 10$ que pour les valeurs inférieures, cependant, pour le fantôme 4-2 avec cette valeur de λ, certaines petites régions clés sont masquées dans la segmentation alors qu'elles apparaissent pour $\lambda = 1$.

5.5 MGI vs MGM

Afin de tester l'apport des modèles *a priori* présentés dans le chapitre 3, c'est-à-dire le mélange de gaussiennes conditionnellement indépendantes (MGI) et le mélange de Gauss-Markov (MGM), nous utilisons le fantôme 4 mais cette fois-ci en absence de tumeur, vu l'hétérogénéité des tissus qui le composent. En effet, le modèle MGI ne permet pas de tenir compte de la corrélation entre les pixels d'une même classe, tandis que le modèle MGM renforce cette corrélation tout en préservant les contours séparant les différentes régions. La figure (5.15) présente les résultats obtenus pour le fantôme 4-2. La différence entre les reconstructions obtenues à l'aide des deux modèles n'est pas très significative ; elle l'est un peu plus entre ces dernières et les reconstructions obtenues avec l'algorithme CSI. Néanmoins, l'*a priori* MGM permet d'obtenir une meilleure précision dans la reconstruction des zones glandulaires (cf. table 5.7).

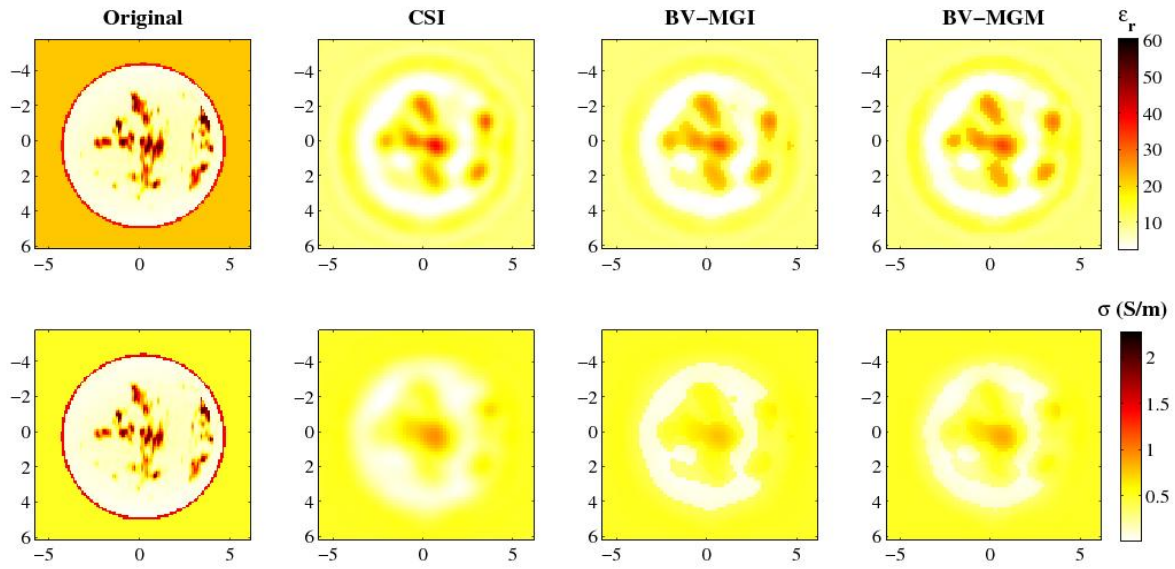


FIGURE 5.15 – Résultats de reconstruction du fantôme 4-2 en absence de tumeur par le biais des méthodes CSI, BV-MGI et BV-MGM comparés au fantôme d'origine.

TABLE 5.7 – Comparaison des approches CSI, BV-MGI et BV-MGM en termes de PSNR pour la reconstruction du fantôme 4-2 en absence de tumeur.

Critère	CSI	BV-MGI	BV-MGM
PSNR (dB)	52, 21	53, 70	54, 39

5.6 Conclusions

Dans ce chapitre, nous avons testé les méthodes bayésiennes discutées dans cette thèse dans le cadre de l'application de la tomographie micro-onde à l'imagerie du cancer

du sein. La diversité des fantômes considérés permet d'effectuer certaines constatations. Les méthodes bayésiennes réussissent beaucoup mieux à imager les objets et à retrouver les tumeurs et les zones glandulaires que la méthode déterministe CSI. Ceci est dû au fait qu'elles permettent d'apporter beaucoup plus d'informations *a priori* grâce au modèle de Gauss-Markov-Potts et à la modélisation probabiliste de tous les autres paramètres et hyper-paramètres. D'autre part, l'approche G-BV n'a pas réussi à gagner en ordre de grandeur de temps de calcul d'une manière absolue par rapport à la méthode BV classique. Cela nécessite de revoir le pré-conditionnement du problème ou d'essayer d'adapter à l'algorithme BV une méthode d'optimisation plus rapide, puisque la descente de gradient ne s'est pas avérée optimale [130].

6

Bilan et perspectives

Sommaire

6.1 Bilan de l'étude 95
6.2 Perspectives du travail 97
6.2.1 Autre modèle *a priori* 97
6.2.2 Autre modélisation du bruit de mesure 98
6.2.3 Mises à jour non simultanées des paramètres de forme 98
6.2.4 Approche totalement non supervisée 99
6.2.5 Prise en compte du caractère dispersif des tissus 99
6.2.6 Taux de fausse alarme 100

6.1 Bilan de l'étude

DANS le cadre de cette thèse, nous nous sommes intéressés au problème d'imagerie du cancer du sein en utilisant la technique de tomographie micro-onde. Cette dernière correspond mathématiquement à un problème inverse non linéaire et mal posé dont la résolution repose sur deux grands axes : la modélisation du problème direct et l'inversion. Ici, nous traitons cette dernière dans un cadre bayésien où le caractère mal posé du problème inverse est surmonté en introduisant de l'information dans les lois *a priori* vérifiées par les inconnues.

Tout d'abord, le problème direct est résolu dans un cadre discret fondé sur l'application de la méthode des moments aux équations intégrales vérifiées par le champ électrique. Pour valider le modèle direct, nous avons développé deux modèles différents, l'un basé sur la symétrie cylindrique apparaissant dans les fantômes étudiés et destiné à la génération des données du problème inverse et l'autre, plus général, utilisé lors de l'inversion. Ceci nous a permis d'éviter de commettre un crime inverse en introduisant

un bruit de modèle systématique. Les résultats obtenus à l'aide de ces deux modèles ont été comparés et l'erreur relative a été estimée en fonction de la fréquence. Généralement, ces résultats coïncident relativement bien, ce qui nous a permis de valider le modèle direct et de nous tourner vers l'inversion.

Pour ce faire, nous avons choisi un cadre d'estimation probabiliste bayésien. Après avoir écrit la relation probabiliste qui lie toutes les variables du modèle, connues et inconnues, grâce à la règle de Bayes, nous avons considéré un modèle *a priori* de "Gauss-Markov-Potts" qui prend en compte le fait que le sein est supposé être composé d'un nombre fini de tissus différents et que ceux-ci sont répartis dans des zones homogènes et compactes. Ensuite, nous nous sommes proposé d'estimer la loi conjointe *a posteriori* des inconnues connaissant les données et le modèle.

Le point de départ de notre réflexion concernant cette estimation conjointe est un algorithme d'optimisation alternée que nous avons adapté à notre cadre afin de déterminer les différents estimateurs du modèle. Cependant, du fait de la difficulté du calcul des estimateurs et de la non-prise en compte des incertitudes dans ces calculs, il s'est avéré que cette méthode ne peut pas aboutir pratiquement à une solution efficace. Une méthode bayésienne variationnelle (BV) est alors utilisée pour approcher la loi *a posteriori* conjointe des inconnues et des paramètres du modèle par un produit de lois marginales dont les expressions optimales sont obtenues en minimisant la divergence de Kullback-Leibler.

L'un des points délicats par rapport aux travaux précédents qui ont fait appel à la méthode BV, qui constitue une première grande contribution de cette thèse, est la nature complexe ainsi que l'amplitude importante du contraste considéré ici. En effet, les tissus du sein présentent de forts contrastes de permittivité et conductivité qui doivent être retrouvés simultanément lors de la phase de reconstruction. La prise en compte de cela a nécessité une attention particulière lors de la mise en œuvre théorique et pratique de l'algorithme BV, en particulier lors de l'initialisation du champ caché, des hyperparamètres, du contraste et des courants induits, ainsi que lors de l'implémentation des paramètres de mises à jour. Dans ce contexte, la programmation de cette technique était loin d'être triviale. Les résultats obtenus montrent l'efficacité de l'approche développée et une amélioration de la précision des valeurs obtenues comparée à celle donnée par la méthode d'inversion déterministe contraste source (CSI) qui a déjà été appliquée dans le cadre de l'imagerie du cancer du sein par micro-ondes. Il reste cependant à améliorer certains points, comme la qualité de reconstruction ainsi que la vitesse de convergence de la technique BV, ce qui peut être fait en lui adaptant un algorithme d'optimisation.

Partant de ce principe, la deuxième grande contribution de la thèse réside dans le développement de la méthode de gradient pour l'approche bayésienne variationnelle (G-BV) où, inspirés de travaux récents pour la résolution de problèmes inverses linéaires de grandes dimensions, nous avons adapté la technique de descente de gradient à notre approche BV. Cette nouvelle technique introduit la notion de pas de descente qui intervient dans la mise à jour des différentes variables. Les résultats de reconstruction obtenus par le biais de cette technique montrent bien que cette dernière améliore la qualité de reconstruction : les cartes de permittivité et de conductivité ont une meilleure résolution que celles données par l'approche déterministe CSI d'un coté et par l'approche BV classique de l'autre. Cependant, ses performances varient d'un modèle à un autre. Notamment concernant la vitesse de convergence de la méthode, qui s'est révélée souvent plus rapide

que l'approche BV, mais pas systématiquement. En effet, une convergence plus lente que celle de l'approche classique a été notée pour les fantômes les plus sophistiqués. Notons aussi que cet algorithme a nécessité également une attention particulière lors de son implémentation.

L'application ciblée dans cette thèse constitue également une contribution importante de ces travaux. En effet, l'imagerie du cancer du sein est actuellement principalement effectuée en rayons X, des rayonnements ionisants et donc nocifs pour le corps humain, éventuellement complétés par une imagerie par résonance magnétique. Ces deux modes d'imagerie, malgré la grande résolution spatiale qu'ils procurent, demeurent coûteux et ne bénéficient pas du contraste important entre tissus sains et tissus cancéreux qui existe dans le domaine des micro-ondes. Tous ces facteurs réunis font que ces dernières représentent une véritable alternative aux technique d'imagerie classiques destinées aux applications biomédicales.

Finalement, sur le plan académique, les résultats obtenus lors de cette thèse ont fait l'objet de plusieurs publications dans des revues scientifiques et des actes de conférences internationales.

6.2 Perspectives du travail

Plusieurs problématiques concernant la résolution du problème d'imagerie micro-onde dans un cadre bayésien se présentent. Durant les travaux de thèse nous avons pu répondre à certaines questions. Cependant, d'autres idées peuvent être exploitées et plusieurs améliorations peuvent être apportées aux méthodes proposées.

6.2.1 Autre modèle *a priori*

La prise en compte d'un nouveau modèle *a priori* pour reconstruire notre objet par imagerie micro-onde est envisageable. En effet, comme le sein que nous voulons reconstruire est censé être composé d'un nombre fini de tissus repartis dans en zones homogènes et compactes, une idée, pour exploiter cette information, est de considérer le gradient de l'image recherchée comme étant parcimonieux. Notons que le gradient de l'image, défini par $\bigtriangledown\chi(\boldsymbol{r}) = \left|\chi(\boldsymbol{r}) - |\mathcal{V}(\boldsymbol{r})|^{-1} \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \chi(\boldsymbol{r}')\right|$, est créé généralement à partir de l'image d'origine par convolution avec un filtre. Chaque pixel du gradient de l'image correspond à la variation du contraste de ce pixel dans l'image d'origine par rapport à ses voisins. Dans une configuration 2D, pour obtenir le gradient dans toutes les directions il faut faire le calcul selon l'axe des abscisses et l'axe des ordonnées. L'histogramme du module est une distribution à queue lourde comme une distribution de Laplace, une distribution Bernouilli gaussienne, un mélange de gaussiennes, une distribution de Cauchy ou une distribution α-stable [131].

Par exemple, l'information parcimonieuse peut être modélisée par un mélange de gaussiennes dont l'avantage principal est qu'il peut aisément être écrit comme une distribution gaussienne conditionnée par une variable cachée $\boldsymbol{z}$ ($\boldsymbol{z} = \prod_{\boldsymbol{r}} z(\boldsymbol{r}), \boldsymbol{r} = 1, ..., N_{\mathcal{D}}$) telle que $\chi(\boldsymbol{r}) \sim \mathcal{N}(0, v_s z(\boldsymbol{r})^{-1})$ et $z(\boldsymbol{r}) \sim \mathcal{G}(\alpha_0, \beta_0)$. Les probabilités du contraste et du champ caché s'écrivent alors comme suit :

$$p(\boldsymbol{\chi}|\boldsymbol{z}, v_s) \propto (2\pi v_s)^{-\frac{N}{2}} \prod_{\boldsymbol{r}} z(\boldsymbol{r})^{\frac{1}{2}} \exp\left\{-\frac{\sum_{\boldsymbol{r}} \sum_{\boldsymbol{r}' \in \mathcal{V}_{\boldsymbol{r}}} z(\boldsymbol{r})(\chi(\boldsymbol{r}) - \chi(\boldsymbol{r}'))^2}{2v_s}\right\}, \quad (6.1)$$

$$p(\boldsymbol{z}|\alpha_0, \beta_0) \propto \mathcal{G}(\alpha_0, \beta_0) \propto \prod_{\boldsymbol{r}} z(\boldsymbol{r})^{\beta_0 - 1} \exp\left\{-\frac{\sum_{\boldsymbol{r} \in \mathcal{R}_k} z(\boldsymbol{r})}{\alpha_0}\right\}. \quad (6.2)$$

Ensuite, en prenant en compte la nouvelle distribution *a priori*, on réécrit la loi jointe *a posteriori* à laquelle on applique l'approximation bayésienne variationnelle afin de déterminer les estimateurs. Cependant, les questions d'efficacité et de robustesse de cette méthode restent ouvertes.

6.2.2 Autre modélisation du bruit de mesure

Dans ce travail, les bruits de modèle et de mesure ont été modélisés par des bruits blancs additifs gaussiens car, d'une part, cela permet de faciliter les calculs des lois approchantes et des mises à jour des paramètres et, d'autre part, en l'absence d'information *a priori* sur le bruit, sauf peut être le rapport signal à bruit qui est relié à sa variance, c'est l'*a priori* le moins informatif qui puisse être choisi. Notons cependant que, pour beaucoup de dispositifs expérimentaux, ce modèle ne correspond certainement pas à la réalité et, d'autre part, une bonne modélisation du bruit à une influence décisive sur la qualité des résultats de l'inversion [132]. Pour répondre à ce problème, il est tout-à-fait possible de considérer d'autres modèles de bruit si nous disposons d'informations *a priori* sur ce dernier, au prix cependant d'un coup de calcul bien plus élevé. Par exemple, au lieu d'un bruit additif gaussien, nous pourrions considérer un couplage signal-bruit multiplicatif comme c'est le cas pour le bruit de speckle [133] rencontré en imagerie ultra-sonore et en imagerie radar SAR.

Dans ce cas, le modèle direct s'écrit :

$$\begin{aligned} \boldsymbol{y}_n &= \left(\boldsymbol{G}_f^o \, \boldsymbol{w}_n\right) \times \boldsymbol{\epsilon}_n \,, \\ \boldsymbol{w}_n &= \left(\boldsymbol{X}_f \, \boldsymbol{E}_n^{inc} + \boldsymbol{X}_f \, \boldsymbol{G}_f^c \, \boldsymbol{w}_n\right) \times \boldsymbol{\xi}_n. \end{aligned} \quad (6.3)$$

6.2.3 Mises à jour non simultanées des paramètres de forme

Une autre question est quelle forme de séparation doit prendre la loi approchant la loi conjointe *a posteriori* recherchée lors de l'application de la méthode bayésienne variationnelle. La séparation forte choisie dans ce travail conduit à des mises à jour simultanées de tous les paramètres des lois approchantes, ce qui est très coûteux en temps de calcul. Si on choisit une séparation faible de la forme :

$$q(\boldsymbol{u}) = q(\boldsymbol{w})q(\boldsymbol{\chi})q(\boldsymbol{z})q(v_\epsilon)q(v_\xi)\prod_k q(v_k)q(m_k),$$

les sources induites peuvent être mises à jour par une méthode de type gradient, par exemple, et les autres paramètres par la technique BV, ce qui pourrait accélérer la convergence de l'algorithme.

6.2.4 Approche totalement non supervisée

L'approche d'inversion bayésienne proposée ici est quasi non supervisée, le quasi provenant du fait que tous les paramètres du modèle sont recherchés en même temps que les inconnues, excepté deux que l'on suppose connus à l'avance, i.e. le nombre K de classes (tissus) composant l'image à reconstruire et le paramètre de corrélation de Potts. En ce qui concerne le premier, comme souligné précédemment, connaître le nombre de classes n'est pas impératif. Si celui-ci n'est pas connu, K peut être initialisé à un nombre élevé et, au cours des itérations, il finit par converger vers le nombre optimal de classes selon deux chemins possibles :

— si le nombre de pixels dans une classe donnée diminue jusqu'à une valeur de seuil bas, alors cette classe finit par disparaître,
— si les moyennes de deux classes différentes sont très proches l'une de l'autre, alors ces deux classes finissent par fusionner.

Ceci a été montré par l'expérience [90] mais n'a pas encore été démontré théoriquement. Il serait cependant mieux de disposer d'une approche totalement non supervisée où le nombre de classes K serait estimé en même temps que les autres paramètres, ainsi que le paramètre de corrélation de Potts. L'estimation de ce dernier fait déjà l'objet d'un important sujet de discussion dans la littérature et est un problème bien connu en segmentation d'image [134, 135].

6.2.5 Prise en compte du caractère dispersif des tissus

Dans nos travaux, nous avons considéré un dispositif micro-onde multifréquentiel et, dans le processus d'inversion, les données correspondant aux différentes fréquences ont été traitées simultanément sans tenir compte de la nature dispersive des tissus composant le sein. Hors il est connu que ces derniers sont très dispersifs, c'est-à-dire que leurs permittivité et conductivité dépendent fortement de la fréquence. Pour pouvoir appliquer notre méthode d'inversion dans des conditions réalistes, il est donc impératif de tenir compte de cette propriété. Comme signalé précédemment, pour le sein, quatre groupes de tissus (sains ou cancéreux) ont été identifiés et les dépendances en fréquence des permittivités et conductivités moyennes de chacun d'entre eux ont été décrites par des modèles de Debye à un pôle [93] (figure (6.1)). Un tel modèle, basé sur la théorie de la relaxation de Debye, permet de décrire l'évolution de la permittivité complexe ($\epsilon = \epsilon_r + i\sigma/\epsilon_0\omega$) d'un matériau en fonction de la fréquence à partir des valeurs limites de la constante diélectrique en régime statique ϵ_s et à haute fréquence ϵ_∞, d'une constante de relaxation τ (équivalente à un temps) et de la conductivité ionique σ_i sous la forme suivante :

$$\epsilon = \epsilon' + i\epsilon'', \tag{6.4}$$

où

$$\epsilon' = \epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{1 + \omega^2\tau^2}, \qquad \epsilon'' = \frac{\sigma_i}{\omega\epsilon_0} + \frac{(\epsilon_s - \epsilon_\infty)\omega\tau}{1 + \omega^2\tau^2}. \tag{6.5}$$

Tenir compte de la dipersivité des tissus dans le processus d'inversion revient à chercher comme inconnues non plus la permittivité relative ϵ_r et la conductivité σ qui dépendent de la fréquence, mais les invariants ϵ_s, ϵ_∞, τ et σ_i, ce qui complique les choses car cela représente quatre inconnues au lieu de deux [62, 136].

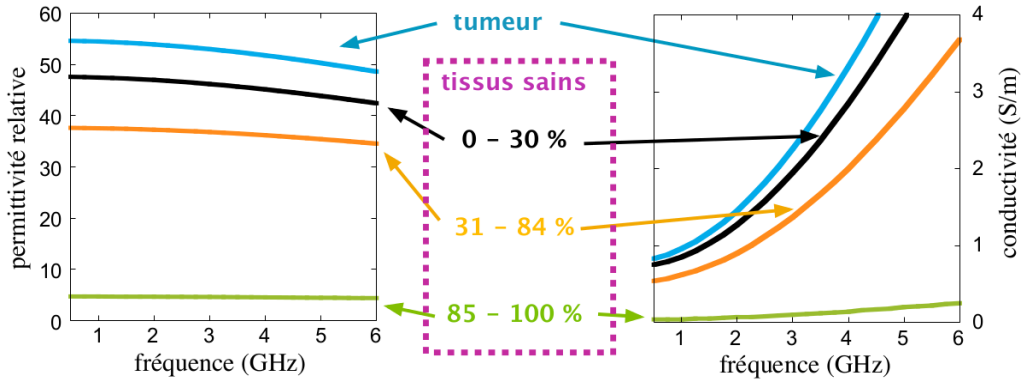


FIGURE 6.1 – Variation en fréquence de la permittivité relative et de la conductivité décrite par un modèle de Debye.

6.2.6 Taux de fausse alarme

Enfin, deux tâches restent à accomplir avant que notre méthode d'inversion puisse être utilisée par un imageur micro-onde en situation de "terrain clinique". Il s'agit d'étudier ses performances, d'une part, sur des données expérimentales acquises en situation contrôlée sur des fantômes physiques de sein et, d'autre part, sur des données issues de base de données comportant des cas de tumeurs difficiles à détecter (qui restent à construire) afin de pouvoir comparer ces performances à celles d'autres méthodes d'inversion en termes de taux de fausses alarmes, positives ou négatives, ce qui permettrait de conclure quant à la robustesse des méthodes bayésiennes variationnelles en présence de véritables facteurs cliniques.

A

Solution analytique de l'approche bayésienne variationnelle

Trouver la forme optimale de la densité de probabilité q qui correspond à la solution analytique de la formulation bayésienne variationnelle revient à résoudre le problème d'optimisation suivant :

$$\widehat{q}^{opt} = \arg\min_q \mathrm{KL}(q||p) = \arg\max_q \mathcal{F}(q(\theta)), \tag{A.1}$$

où KL est la divergence de Kullback-Leibler entre p et q et $\mathcal{F}$ est l'énergie libre négative.

Néanmoins, le choix de la forme approchante est crucial. Cette dernière est choisie la plus simple possible comme une loi séparable $q(\theta) = \prod_i q_i(\theta_i)$. Par exemple soit $q(\theta) = q_1(\theta_1)q_2(\theta_2)$, alors on obtient :

$$(\widehat{q_1}, \widehat{q_2}) = \arg\min_{q_1,q_2} \mathrm{KL}(q_1, q_2||p) = \arg\max_{q_1,q_2} \mathcal{F}(q_1(\theta_1), q_2(\theta_2)). \tag{A.2}$$

Puisque $\mathrm{KL}(q_1, q_2||p)$ n'est pas quadratique en q_1 et q_2, sa dérivée par rapport à ces derniers n'est pas linéaire et une solution explicite n'est pas concevable. Donc la solution de l'équation (A.2) ne peut être obtenue qu'itérativement selon :

$$\widehat{q_1} = \arg\min_{q_1} \mathrm{KL}(q_1, \widehat{q_2}||p) = \arg\max_{q_1} \mathcal{F}(q_1(\theta_1), \widehat{q_2}(\theta_2)), \tag{A.3}$$

$$\widehat{q_2} = \arg\min_{q_2} \mathrm{KL}(\widehat{q_1}, q_2||p) = \arg\max_{q_2} \mathcal{F}(\widehat{q_1}(\theta_1), q_2(\theta_2)). \tag{A.4}$$

Or :

$$\begin{aligned}
\mathcal{F}(q_1(\theta_1), q_2(\theta_2)) &= \int\int q_1(\theta_1)q_2(\theta_2)\log\frac{p(Y,\theta_1,\theta_2)}{q_1(\theta_1)q_2(\theta_2)}d\theta_1 d\theta_2, \\
&= -\int\int q_1(\theta_1)q_2(\theta_2)\log\frac{q_1(\theta_1)q_2(\theta_2)}{p(Y,\theta_1,\theta_2)}d\theta_1 d\theta_2, \\
&= -\int\int q_1(\theta_1)q_2(\theta_2)\Big(\log(q_1(\theta_1)+\log(q_2(\theta_2)) \\
&- \log(p(Y,\theta_1,\theta_2))\Big)d\theta_1 d\theta_2, \\
&= -\int q_1(\theta_1)\log(q_1(\theta_1))d\theta_1\int q_2(\theta_2)d\theta_2 \\
&- \int q_2(\theta_2)\log(q_2(\theta_2))d\theta_2\int q_1(\theta_1)d\theta_1 \\
&+ \int\int q_1(\theta_1)q_2(\theta_2)\log(p(Y,\theta_1,\theta_2))d\theta_1 d\theta_2, \\
&= \int q_1(\theta_1)\log(q_1(\theta_1))d\theta_1 - \int q_2(\theta_2)\log(q_2(\theta_2))d\theta_2 \\
&+ \int\int q_1(\theta_1)q_2(\theta_2)\log(p(Y,\theta_1,\theta_2))d\theta_1 d\theta_2. \qquad \text{(A.5)}
\end{aligned}$$

Ainsi, pour trouver la solution $\widehat{q_1}(\theta_1)$, on calcule la dérivée de l'équation (A.5) par rapport à ce dernier. $\widehat{q_1}(\theta_1)$ est alors la solution de :

$$\frac{\partial\mathcal{F}(q_1(\theta_1), q_2(\theta_2))}{\partial q_1(\theta_1)} = 0, \qquad \text{(A.6)}$$

soit de :

$$1 + \log q_1(\theta_1) - \int q_2(\theta_2)\log p(Y,\theta_1,\theta_2)d\theta_2 = 0, \qquad \text{(A.7)}$$

ou encore de :

$$\log q_1(\theta_1) = \int q_2(\theta_2)\log p(Y,\theta_1,\theta_2)d\theta_2 - 1, \qquad \text{(A.8)}$$

$\widehat{q_1}(\theta_1)$ est donc tel que :

$$\widehat{q_1}(\theta_1) \propto \exp\Big\{\Big\langle\log p(Y,\theta)\Big\rangle_{\widehat{q_2}}\Big\}. \qquad \text{(A.9)}$$

La solution optimale $\widehat{q_2}(\theta_2)$ est obtenue de la même façon. Elle s'écrit :

$$\widehat{q_2}(\theta_2) \propto \exp\Big\{\Big\langle\log p(Y,\theta)\Big\rangle_{\widehat{q_1}}\Big\}. \qquad \text{(A.10)}$$

D'une manière générale, les solutions optimales s'écrivent :

$$\boxed{\widehat{q_i}(\theta_i) = \frac{1}{\kappa_i}\exp\Big\{\Big\langle\log p(Y,\theta)\Big\rangle_{\prod_{i\neq j}\widehat{q_j}(\theta_j)}\Big\},} \qquad \text{(A.11)}$$

où κ_i est une constante de normalisation.

B

Mise à jour des paramètres de forme du G-BV

B.1 Les sources induites w

$$
\begin{aligned}
\log\left(\widetilde{q}^{n}(w(\boldsymbol{r}))\right) &\propto (1-\alpha_w)\log\left(\widehat{q}(w(\boldsymbol{r}))\right) + \alpha_w \left\langle \log\left(p\left(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\boldsymbol{\psi},\boldsymbol{y}\right)\right)\right\rangle_{\tilde{q}(/w(\boldsymbol{r}))} \\
&\propto (1-\alpha_w)\left[\widehat{\boldsymbol{v}}_{w(\boldsymbol{r})}w(\boldsymbol{r})^2 - 2\widehat{\boldsymbol{m}}_{w(\boldsymbol{r})}w(\boldsymbol{r})\right] \\
&+ \alpha_w \left\langle \log\left(p(\boldsymbol{y}|\boldsymbol{w},v_\epsilon)\right) + \log\left(p(\boldsymbol{w}|\boldsymbol{\chi},v_\xi)\right)\right\rangle_{\tilde{q}(/w(\boldsymbol{r}))} \\
&\propto (1-\alpha_w)\left[\widehat{\boldsymbol{v}}_{w(\boldsymbol{r})}w(\boldsymbol{r})^2 - 2\widehat{\boldsymbol{m}}_{w(\boldsymbol{r})}w(\boldsymbol{r})\right] \\
&- \frac{\alpha_w}{2}\left\langle \frac{||\boldsymbol{y}-\boldsymbol{G}^o\boldsymbol{w}||^2_{\mathcal{S}}}{v_\epsilon} + \frac{||\boldsymbol{w}-\boldsymbol{X}\boldsymbol{E}^{inc}-\boldsymbol{X}\boldsymbol{G}^c\boldsymbol{w}||^2_{\mathcal{D}}}{v_\xi}\right\rangle_{\tilde{q}(/w(\boldsymbol{r}))}
\end{aligned}
\tag{B.1}
$$

où $q(/w(\boldsymbol{r})) = \prod_{\boldsymbol{r}'\neq\boldsymbol{r}} q(w(\boldsymbol{r}'))q(\boldsymbol{\chi})q(\boldsymbol{z})q(\boldsymbol{\psi})$. Or

$$
\begin{aligned}
\left\langle ||\boldsymbol{y}-\boldsymbol{G}^o\boldsymbol{w}||^2_{\mathcal{S}}\right\rangle_{\tilde{q}(/w(\boldsymbol{r}))} &\propto \sum_{\boldsymbol{r}'} |G^o(\boldsymbol{r}',\boldsymbol{r})|^2|w(\boldsymbol{r})|^2 \\
&+ 2\,\Re e\left(\sum_{\boldsymbol{r}"} G^{o*}(\boldsymbol{r}",\boldsymbol{r})\left[y(\boldsymbol{r}") - \sum_{\boldsymbol{r}'\neq\boldsymbol{r}} G^o(\boldsymbol{r}",\boldsymbol{r}')\widehat{m}_w(\boldsymbol{r}')\right] w^*(\boldsymbol{r})\right)
\end{aligned}
$$

et

$$
\begin{aligned}
\Big\langle ||\boldsymbol{w} - \boldsymbol{X}\boldsymbol{E}^{inc} - \boldsymbol{X}\boldsymbol{G}^{c}\boldsymbol{w}||^2_{\mathcal{D}} \Big\rangle_{\tilde{q}(/w(\boldsymbol{r}))} \quad &\propto \quad |w(\boldsymbol{r})|^2 - G^{c*}(\boldsymbol{r},\boldsymbol{r})\ \bar{\chi}^*(\boldsymbol{r})\ |w(\boldsymbol{r})|^2 - G^{c}(\boldsymbol{r},\boldsymbol{r})\ \bar{\chi}(\boldsymbol{r})\ |w(\boldsymbol{r})|^2 \\
&+ \quad \sum_{\boldsymbol{r}'} |G^{c}(\boldsymbol{r}',\boldsymbol{r})|\ \overline{|\chi(\boldsymbol{r})|^2}\ |w(\boldsymbol{r})|^2 - 2\ \Re e\left(E^{inc}(\boldsymbol{r})\bar{\chi}(\boldsymbol{r})w^*(\boldsymbol{r})\right) \\
&+ \quad 2\ \Re e\left(\sum_{\boldsymbol{r}'} |G^{c*}(\boldsymbol{r}',\boldsymbol{r})|\ \overline{|\chi(\boldsymbol{r}')|^2}\ |w^*(\boldsymbol{r}')|^2\right) \\
&- \quad 2\ \Re e\left(\sum_{\boldsymbol{r}'\neq\boldsymbol{r}} \bar{\chi}(\boldsymbol{r})\ G^{c}(\boldsymbol{r},\boldsymbol{r}')\ \bar{w}(\boldsymbol{r}')\ w^*(\boldsymbol{r})\right) \\
&- \quad 2\ \Re e\left(\sum_{\boldsymbol{r}'\neq\boldsymbol{r}} \bar{\chi}^*(\boldsymbol{r}')\ G^{c*}(\boldsymbol{r}',\boldsymbol{r})\ \bar{w}(\boldsymbol{r}')\ w^*(\boldsymbol{r})\right) \\
&+ \quad 2\ \Re e\left(\sum_{\boldsymbol{r}'} G^{c*}(\boldsymbol{r}',\boldsymbol{r})\ \overline{|\chi(\boldsymbol{r}')|^2} \sum_{\boldsymbol{r}"\neq\boldsymbol{r}} G^{c}(\boldsymbol{r}',\boldsymbol{r}")\ \bar{w}(\boldsymbol{r}")\ w^*(\boldsymbol{r})\right) \\
&+ \quad cte.
\end{aligned}
$$

où * désigne le complexe conjugué.

Ainsi en combinant tous les termes, à l'itération i la loi approchante devient $q^n(\boldsymbol{w}) = \mathcal{N}(\widehat{\boldsymbol{m}}^n_w, \widehat{\boldsymbol{V}}^n_w)$ où :

$$
\begin{aligned}
\widehat{V}^n_w(\boldsymbol{r}) = \Bigg[&(1-\alpha_w)\widehat{V}_w(\boldsymbol{r}) + \alpha_w \left(\overline{v_\epsilon^{-1}} \sum_{\boldsymbol{r}'} |G^o(\boldsymbol{r}',\boldsymbol{r})|^2 \right. \\
&\left. + \overline{v_\xi^{-1}} \Big[1 - 2\ \Re e\Big(G^c(\boldsymbol{r},\boldsymbol{r})\widehat{m}_\chi(\boldsymbol{r})\Big) + \sum_{\boldsymbol{r}'} |G^c(\boldsymbol{r}',\boldsymbol{r})|^2 \Big(|\widehat{m}_\chi(\boldsymbol{r})|^2 + \widehat{V}_\chi(\boldsymbol{r}) \Big) \Big] \right) \Bigg]^{-1},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\widehat{m}^n_w(\boldsymbol{r})}{\widehat{V}^n_w(\boldsymbol{r})} \quad &= \quad (1-\alpha_w)\,\widehat{m}_w(\boldsymbol{r}) + \alpha_w \left[\overline{v_\epsilon^{-1}} \sum_{\boldsymbol{r}"} G^{o*}(\boldsymbol{r}",\boldsymbol{r}) \left(y(\boldsymbol{r}") - \sum_{\boldsymbol{r}'\neq\boldsymbol{r}} G^o(\boldsymbol{r}",\boldsymbol{r}')\widehat{m}_w(\boldsymbol{r}') \right) \right. \\
&+ \quad \overline{v_\xi^{-1}} \left(E^{inc}(\boldsymbol{r})\widehat{m}_\chi(\boldsymbol{r}) - \sum_{\boldsymbol{r}'} G^{c*}(\boldsymbol{r}',\boldsymbol{r}) \Big(|\widehat{m}_\chi(\boldsymbol{r}')|^2 + \widehat{V}_\chi(\boldsymbol{r}') \Big) E^{inc}(\boldsymbol{r}') \right. \\
&+ \quad \sum_{\boldsymbol{r}'\neq\boldsymbol{r}} \widehat{m}_\chi(\boldsymbol{r}) G^c(\boldsymbol{r},\boldsymbol{r}')\widehat{m}_w(\boldsymbol{r}') + \sum_{\boldsymbol{r}'\neq\boldsymbol{r}} \widehat{m}^*_\chi(\boldsymbol{r}) G^{c*}(\boldsymbol{r}',\boldsymbol{r})\widehat{m}_w(\boldsymbol{r}') \\
&- \quad \left. \left. \sum_{\boldsymbol{r}'} G^{c*}(\boldsymbol{r}',\boldsymbol{r}) \Big(|\widehat{m}_\chi(\boldsymbol{r})|^2 + \widehat{V}_\chi(\boldsymbol{r}) \Big) \sum_{\boldsymbol{r}"\neq\boldsymbol{r}'} G^c(\boldsymbol{r}',\boldsymbol{r}")\overline{w}(\boldsymbol{r}") \right) \right].
\end{aligned}
$$

En rajoutant le terme manquant $\widehat{m}_w$ à droite de l'équation précédente et en écrivant

cette dernière en notation vectorielle, on obtient la forme finale suivante :

$$\left\{\begin{array}{l} \widehat{\boldsymbol{V}}_w^n = \left[(1-\alpha_w)\,\widehat{\boldsymbol{V}}_w^{-1} + \alpha_w\ \mathbf{Diag}\left(\overline{v_\epsilon^{-1}}\Gamma^o + \overline{v_\xi^{-1}}\overline{\Gamma^c}\right)\right]^{-1} \\ \widehat{\boldsymbol{m}}_w^n = \widehat{\boldsymbol{m}}_w + \alpha_w\ \widehat{\boldsymbol{V}}_w^n \Bigg[\Bigg[\overline{v_\epsilon^{-1}}\boldsymbol{G}^{o^H}\left(\boldsymbol{y} - \boldsymbol{G}^o\widehat{\boldsymbol{m}}_w\right) \\ \qquad + \overline{v_\xi^{-1}}\Big(\widehat{\boldsymbol{m}}_\chi \boldsymbol{E}^{inc} - \boldsymbol{G}^{c^H}\left(\widehat{\boldsymbol{m}}_\chi^2 + \widehat{\boldsymbol{V}}_\chi\right)\boldsymbol{E}^{inc} \\ \qquad - \widehat{\boldsymbol{m}}_w + \widehat{\boldsymbol{m}}_\chi \boldsymbol{G}^c\widehat{\boldsymbol{m}}_w + \boldsymbol{G}^{c^H}\widehat{\boldsymbol{m}}_\chi^*\widehat{\boldsymbol{m}}_w - \boldsymbol{G}^{c^H}\left(\widehat{\boldsymbol{m}}_\chi^2 + \widehat{\boldsymbol{V}}_\chi\right)\boldsymbol{G}^c\widehat{\boldsymbol{m}}_w\Big)\Bigg]\Bigg], \end{array}\right. \tag{B.2}$$

où $\overline{u} = \mathrm{E}(u)_q$, l'exposant ^{H} désigne le conjugué transposé et Γ^o et Γ^c sont tels que :

$$\begin{array}{rcl} \Gamma^o(\boldsymbol{r}) & = & \displaystyle\sum_{\boldsymbol{r}'} |G^o(\boldsymbol{r}',\boldsymbol{r})|^2, \\ \overline{\Gamma^c}(\boldsymbol{r}) & = & 1 - 2\Re e(G^c(\boldsymbol{r},\boldsymbol{r})\widehat{m}_\chi(\boldsymbol{r})) \\ & + & (\widehat{m}_\chi^2(\boldsymbol{r}) + \widehat{V}_\chi(\boldsymbol{r}))\displaystyle\sum_{\boldsymbol{r}'} |G^c(\boldsymbol{r}',\boldsymbol{r})|^2. \end{array}$$

B.2 Le contraste $\boldsymbol{\chi}$

$$\begin{array}{rcl} \log\left(\widehat{q}^n(\chi(\boldsymbol{r}))\right) & \propto & (1-\alpha)\log\left(\widehat{q}(\chi(\boldsymbol{r}))\right) + \alpha_\chi\ \left\langle \log\left(p\left(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\boldsymbol{\psi},\boldsymbol{y}\right)\right)\right\rangle_{\tilde{q}(/\chi(\boldsymbol{r}))} \\ & \propto & (1-\alpha_\chi)\left[\widehat{\boldsymbol{v}}_{\chi(\boldsymbol{r})}\chi(\boldsymbol{r})^2 - 2\widehat{\boldsymbol{m}}_{\chi(\boldsymbol{r})}\chi(\boldsymbol{r})\right] \\ & + & \alpha_\chi\ \left\langle \log\left(p(\boldsymbol{w}|\boldsymbol{\chi},v_\epsilon)\right) + \log\left(p(\boldsymbol{\chi}|\boldsymbol{z},\boldsymbol{v},\boldsymbol{m})\right)\right\rangle_{\tilde{q}(/\chi(\boldsymbol{r}))} \\ & \propto & (1-\alpha_\chi)\left[\widehat{\boldsymbol{v}}_{\chi(\boldsymbol{r})}\chi(\boldsymbol{r})^2 - 2\widehat{\boldsymbol{m}}_{\chi(\boldsymbol{r})}\chi(\boldsymbol{r})\right] \\ & + & \alpha_\chi\left\langle \dfrac{||\boldsymbol{w}-\boldsymbol{X}\boldsymbol{E}||_{\mathcal{D}}^2}{v_\xi} + \left(\boldsymbol{\chi}-\boldsymbol{m}_\chi\right)^T\boldsymbol{V}_\chi^{-1}\left(\boldsymbol{\chi}-\boldsymbol{m}_\chi\right)\right\rangle_{\tilde{q}(/\chi(\boldsymbol{r}))}, \end{array} \tag{B.3}$$

où $q(/\chi(\boldsymbol{r})) = \prod_{\boldsymbol{r}'\neq\boldsymbol{r}} q(\chi(\boldsymbol{r}'))q(\boldsymbol{w})q(\boldsymbol{z})q(\boldsymbol{\psi})$. Or

$$\begin{array}{rcl} \left\langle ||\boldsymbol{w}-\boldsymbol{X}\boldsymbol{E}||_{\mathcal{D}}^2\right\rangle_{\tilde{q}(/\chi(\boldsymbol{r}))} & \propto & -2\Re e\left(\overline{w^*(\boldsymbol{r})E(\boldsymbol{r})}\chi(\boldsymbol{r})\right) + \overline{E^2}(\boldsymbol{r})|\chi(\boldsymbol{r})|^2 \\ & + & \left\langle ||\boldsymbol{w}||_{\mathcal{D}}^2 - 2\Re e\left(\displaystyle\sum_{\boldsymbol{r}'\neq\boldsymbol{r}} w^*(\boldsymbol{r}')E(\boldsymbol{r}')\chi(\boldsymbol{r}')\right) + \sum_{\boldsymbol{r}'\neq\boldsymbol{r}}|E(\boldsymbol{r}')\chi(\boldsymbol{r}')|^2\right\rangle_{\tilde{q}(/\chi(\boldsymbol{r}))}, \end{array}$$

et

$$\begin{array}{rcl} \left\langle \left(\boldsymbol{\chi}-\boldsymbol{m}_\chi\right)^T\boldsymbol{V}_\chi^{-1}\left(\boldsymbol{\chi}-\boldsymbol{m}_\chi\right)\right\rangle_{\tilde{q}(/\chi(\boldsymbol{r}))} & \propto & \displaystyle\sum_{\boldsymbol{r}'}\left\langle \frac{||\chi(\boldsymbol{r}')-\boldsymbol{m}_k||_{\mathcal{D}}^2}{v_k}\right\rangle_{\tilde{q}(/\chi(\boldsymbol{r}))} \\ & \propto & \overline{v_k^{-1}}\left(|\chi(\boldsymbol{r})|^2 - 2\Re e\left(\widehat{\boldsymbol{m}}_k\chi^*(\boldsymbol{r})\right)\right). \end{array}$$

Par identification on obtient :

$$\widehat{V}_\chi^n(\boldsymbol{r}) \propto \left[(1-\alpha_\chi)\,\widehat{V}_\chi(\boldsymbol{r})^{-1} + \alpha_\chi\,\left(\overline{v_\xi^{-1}E^2(\boldsymbol{r})} + \overline{\widehat{v}_\chi^{-1}(\boldsymbol{r})}\right)\right]^{-1}$$

et

$$\frac{\widehat{m}_\chi^n(\boldsymbol{r})}{\widehat{V}_\chi^n(\boldsymbol{r})} \propto (1-\alpha_\chi)\,\widehat{m}_\chi(\boldsymbol{r}) + \alpha_\chi\,\left[\overline{\frac{\widehat{m}_\chi(\boldsymbol{r})}{\widehat{V}_\chi(\boldsymbol{r})}} + \overline{v_\xi^{-1}w(\boldsymbol{r})E^*(\boldsymbol{r})}\right],$$

où $\overline{\boldsymbol{wE}^*}$ est la moyenne du vecteur $\boldsymbol{wE}^*$ tel que :

$$\begin{aligned}\overline{wE^*}(\boldsymbol{r}) &= \sum_{N_f N_\vartheta} E^{inc*}(\boldsymbol{r})\widehat{m}_w(\boldsymbol{r}) + \widehat{m}_w(\boldsymbol{r})\sum_{\boldsymbol{r}'} G^{c*}(\boldsymbol{r},\boldsymbol{r}')\widehat{m}_w^*(\boldsymbol{r}') \\ &+ G^{c*}(\boldsymbol{r},\boldsymbol{r})\widehat{V}_w(\boldsymbol{r})\end{aligned}$$

et $\overline{\boldsymbol{E}^2}$ est une matrice diagonale dont les éléments s'écrivent comme suit :

$$\begin{aligned}\overline{E^2}(\boldsymbol{r}) &= \sum_{N_f N_\vartheta} |E^{inc}(\boldsymbol{r})|^2 + 2\Re e\left(E^{inc*}(\boldsymbol{r})G^c\widehat{m}_w(\boldsymbol{r})\right) \\ &+ |\sum_{\boldsymbol{r}'} G^c(\boldsymbol{r},\boldsymbol{r}')\widehat{m}_w(\boldsymbol{r})|^2 + \sum_{\boldsymbol{r}'}|G^c(\boldsymbol{r},\boldsymbol{r}')|^2\widehat{V}_w(\boldsymbol{r}).\end{aligned}$$

Ainsi l'expression sous forme vectorielle est donnée par :

$$\left\{\begin{array}{l}\widehat{\boldsymbol{V}}_\chi^n = \left[(1-\alpha_\chi)\,\widehat{\boldsymbol{V}}_\chi^{-1} + \alpha_\chi\left(\mathbf{Diag}\left(\overline{v_\xi^{-1}\boldsymbol{E}^2} + \overline{\boldsymbol{V}_\chi^{-1}}\right)\right)\right]^{-1} \\ \widehat{\boldsymbol{m}}_\chi^n = \alpha_\chi\,\widehat{\boldsymbol{V}}_\chi^n\left[\left(\sum_k \overline{v_k^{-1}}\widehat{\boldsymbol{\zeta}}_k \circ \widehat{\boldsymbol{m}}_k + \overline{v_\xi^{-1}}\overline{\boldsymbol{w}\circ\boldsymbol{E}^*}\right)\right]\end{array}\right. \tag{B.4}$$

où $\overline{V_\chi^{-1}}(\boldsymbol{r},\boldsymbol{r}) = \sum_k \widehat{\boldsymbol{\zeta}}_k(\boldsymbol{r})\overline{v_k^{-1}}$ et '$\circ$' désigne le produit de Hadamard.

B.3 Le champ caché z

$$\begin{aligned}\log\left(\widehat{q}^n(z(\boldsymbol{r}))\right) &\propto (1-\alpha_z)\log\left(\widehat{q}(z(\boldsymbol{r}))\right) + \alpha\,\left\langle\log\left(p\left(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\boldsymbol{\psi},\boldsymbol{y}\right)\right)\right\rangle_{\tilde{q}(/z(\boldsymbol{r}))} \\ &\propto (1-\alpha_z)\log\left(\widehat{q}(z(\boldsymbol{r}))\right) + \alpha_z\left\langle -\frac{1}{2}\log(v_k) + \frac{||\chi(\boldsymbol{r}) - m_\chi(\boldsymbol{r})||_2^2}{2v_k}\right. \\ &- \left.\lambda\sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})}\delta\left(z(\boldsymbol{r}')-k\right)\right\rangle_{\tilde{q}(/z(\boldsymbol{r}))} \\ &\propto (1-\alpha)\log\left(\widehat{\zeta_k}(\boldsymbol{r})\right) - \frac{\alpha_z}{2}\left[\overline{\log(v_k)} + \overline{v_k^{-1}}\left\langle|\chi(\boldsymbol{r}) - m_\chi(\boldsymbol{r})|\right\rangle_{\tilde{q}(/z(\boldsymbol{r}))}\right. \\ &- \left.2\lambda\sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})}\widehat{\zeta_k}(\boldsymbol{r}')\right] \\ &\propto (1-\alpha_z)\log\left(\widehat{\zeta_k}(\boldsymbol{r})\right) - \frac{\alpha_z}{2}\left[\Psi(\tilde{\eta}_k) - \log\tilde{\phi}_k + \overline{v_k^{-1}}\left[|\widehat{m}_\chi(\boldsymbol{r})|^2\right.\right. \\ &+ \left.\left.\widehat{m}_k^\dagger - 2\Re e\left(\widehat{m}_\chi^\dagger\widehat{m}_\chi^*(\boldsymbol{r})\right)\right] + \lambda\sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})}\widehat{\zeta}_k(\boldsymbol{r}')\right],\end{aligned}$$

où $q(/z(\boldsymbol{r})) = \prod_{\boldsymbol{r}' \neq \boldsymbol{r}} q(z(\boldsymbol{r}'))q(\boldsymbol{\chi})q(\boldsymbol{w})q(\boldsymbol{\psi})$ et $\widehat{m}_\chi^{\dagger 2} = \widehat{m}_k^2 + \widehat{\tau}_k$.

Par identification, nous avons :

$$\begin{aligned}\widehat{\zeta}_k^n &= \widehat{\zeta}_k^{(1-\alpha_z)} \exp\Bigg\{ -\frac{\alpha_z}{2}\Bigg(\Psi(\widehat{\eta}_k) + \log(\widehat{\phi}_k) + \overline{v_k^{-1}}\Big((\widehat{m}_\chi(\boldsymbol{r}) - \widehat{\mu}_k)^2 + \widehat{\tau}_k \\ &+ \widehat{V}_\chi(\boldsymbol{r})\Big) - \lambda \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \widehat{\boldsymbol{\zeta}}_k(\boldsymbol{r}')\Bigg)\Bigg\}, \end{aligned} \tag{B.5}$$

où $\Psi(x) = \dfrac{\partial}{\partial x}\log\Gamma(x)$ est la fonction digamma et $\Gamma(x)$ est la fonction gamma.

B.4 La variance du bruit d'observation v_ϵ

$$\begin{aligned}\log\left(\tilde{q}^n(v_\epsilon)\right) &\propto (1-\alpha_{v_\epsilon})\log\left(\widehat{q}(v_\epsilon)\right) + \alpha_{v_\epsilon}\left\langle \log\left(p\left(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\boldsymbol{\psi},\boldsymbol{y}\right)\right)\right\rangle_{\tilde{q}(/v_\epsilon)} \\ &\propto (1-\alpha_{v_\epsilon})\left(-\frac{v_\epsilon}{\tilde{\eta}_\epsilon} + (\tilde{\phi}_\epsilon \log(v_\epsilon))\right) \\ &+ \alpha_{v_\epsilon}\left\langle p(\boldsymbol{y}|\boldsymbol{w},v_\epsilon)p(v_\epsilon|\eta_\epsilon,\phi_\epsilon)\right\rangle_{\tilde{q}(/v_\epsilon)}, \end{aligned} \tag{B.6}$$

avec $q(/v_\epsilon)) = q(\boldsymbol{\chi})q(\boldsymbol{w}))q(\boldsymbol{z})\prod_{l\neq v_\epsilon} q(\psi_l)$, où :

$$\begin{aligned}\left\langle p(\boldsymbol{y}|\boldsymbol{w},v_\epsilon)p(v_\epsilon|\eta_\epsilon,\phi_\epsilon)\right\rangle_{\tilde{q}(/v_\epsilon)} &\propto \left\langle -\frac{N_{\mathcal{S}}}{2}\log(v_\epsilon) - \frac{||\boldsymbol{y}-\boldsymbol{G}^o\boldsymbol{w}||_{\mathcal{S}}^2}{2v_\epsilon} - \frac{\eta_\epsilon}{v_\epsilon} - (\phi_\epsilon - 1)\log(v_\epsilon)\right\rangle_{\tilde{q}(/v_\epsilon)} \\ &\propto -(1+\phi_\epsilon+\frac{N_{\mathcal{S}}}{2})\log(v_\epsilon) - v_\epsilon^{-1}\left\langle \eta_\epsilon + \frac{||\boldsymbol{y}-\boldsymbol{G}^o\boldsymbol{w}||_{\mathcal{S}}^2}{2}\right\rangle_{\tilde{q}(/v_\epsilon)}, \end{aligned}$$

avec :

$$\begin{aligned}\left\langle ||\boldsymbol{y}-\boldsymbol{G}^o\boldsymbol{w}||_{\mathcal{S}}^2\right\rangle_{\tilde{q}(/v_\epsilon)} &\propto \sum_{\boldsymbol{r}}\Bigg(|y(\boldsymbol{r})|^2 + \sum_{\boldsymbol{r}'}\sum_{\boldsymbol{r}"} G^o(\boldsymbol{r},\boldsymbol{r}")\widehat{m}_w(\boldsymbol{r}')\widehat{m}_w^*(\boldsymbol{r}") + \sum_{\boldsymbol{r}'}|G^o(\boldsymbol{r},\boldsymbol{r}')|\widehat{v}_w(\boldsymbol{r}') \\ &- 2\Re e\Big(y^*\sum_{\boldsymbol{r}'} G^o(\boldsymbol{r},\boldsymbol{r}')\widehat{m}_w^*(\boldsymbol{r}')\Big)\Bigg). \end{aligned}$$

Par identification :

$$\begin{cases} \widehat{\phi}_\epsilon^n = (1-\alpha_{v_\epsilon})\,\widehat{\phi}_\epsilon + \alpha_{v_\epsilon}\left(\phi_\epsilon + \frac{N_{\mathcal{S}}}{2}\right) \\ \widehat{\eta}_\epsilon^n = \frac{1-\alpha_{v_\epsilon}}{\widehat{\eta}_\epsilon} + \alpha_{v_\epsilon}\left(\eta_\epsilon + \frac{1}{2}\Big(||\boldsymbol{y}||_{\mathcal{S}}^2 + ||\boldsymbol{G}^o\widehat{\boldsymbol{m}}_w||_{\mathcal{S}}^2 - 2\Re e(\boldsymbol{y}^H\boldsymbol{G}^o\widehat{\boldsymbol{m}}_w) + ||\boldsymbol{G}^{o2}\widehat{\boldsymbol{v}}_w||_1\Big)\right). \end{cases} \tag{B.7}$$

B.5 La variance du bruit de couplage v_ξ

$$\begin{aligned}\log\left(\tilde{q}^n(v_\xi)\right) &\propto \left(1-\alpha_{v_\xi}\right)\log\left(\widehat{q}(v_\xi)\right) + \alpha_\xi \left\langle \log\left(p\left(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\boldsymbol{\psi},\boldsymbol{y}\right)\right)\right\rangle_{\tilde{q}(/v_\xi)} \\ &\propto \left(1-\alpha_{v_\xi}\right)\left(-\frac{v_\xi}{\tilde{\eta}_\xi} + (\tilde{\phi}_\xi - \log(v_\xi))\right) \\ &+ \alpha_{v_\xi}\ \left\langle p(\boldsymbol{w}|\boldsymbol{\chi},v_\xi)p(v_\xi|\eta_\xi,\phi_\xi)\right\rangle_{\tilde{q}(/v_\xi)}, \end{aligned} \tag{B.8}$$

où $q(/v_\xi)) = q(\boldsymbol{\chi})q(\boldsymbol{w}))q(\boldsymbol{z})\prod_{l\neq v_\xi} q(\psi_l)$ avec :

$$\begin{aligned}\left\langle p(\boldsymbol{w}|\boldsymbol{\chi},v_\xi)p(v_\xi|\eta_\xi,\phi_\xi)\right\rangle_{\tilde{q}(/v_\xi)} &\propto \left\langle \log v_\xi^{-\frac{N_\mathcal{D}}{2}} - \frac{||\boldsymbol{w}-\boldsymbol{X}\boldsymbol{E}||^2_\mathcal{D}}{2v_\xi} - \frac{\eta_\xi}{v_\xi} - (\phi_\xi+1)\log(v_\xi)\right\rangle_{\tilde{q}(/v_\xi)} \\ &\propto \left(\eta_\xi + \frac{N_\mathcal{D}}{2}\right)\log\left(v_\xi\right) - v_\xi^{-1}\left\langle \eta_\xi + \frac{||\boldsymbol{w}-\boldsymbol{X}\boldsymbol{E}||^2_\mathcal{D}}{2}\right\rangle_{\tilde{q}(/v_\xi)}\end{aligned}$$

et

$$\begin{aligned}\left\langle ||\boldsymbol{w}-\boldsymbol{X}\boldsymbol{E}||^2_\mathcal{D}\right\rangle_{\tilde{q}(/v_\xi)} &= \sum_{\boldsymbol{r}}\Bigg(|\widehat{m}_w(\boldsymbol{r})|^2 + \widehat{v}_w(\boldsymbol{r}) + \Big(|\widehat{m}_\chi(\boldsymbol{r})|^2 + \widehat{v}_w(\boldsymbol{r})\Big)\overline{E^2(\boldsymbol{r})} \\ &- 2\Re e\Big(\widehat{m}_\chi(\boldsymbol{r})\overline{w^*(\boldsymbol{r})E(\boldsymbol{r})}\Big)\Bigg).\end{aligned}$$

Par identification nous obtenons :

$$\begin{cases} \widehat{\phi}^n_\xi &= \left(1-\alpha_{v_\xi}\right)\widehat{\phi}_\xi + \alpha_{v_\xi}\left(\phi_\xi + \frac{N_\mathcal{D}}{2}\right)\left(\alpha_{v_\xi}\ \boldsymbol{d}_\chi + \left(1-\alpha_{v_\xi}\right)\widehat{\boldsymbol{m}}_\chi\right) \\ \widehat{\eta}^n_\xi &= \frac{1-\alpha_{v_\xi}}{\widehat{\eta}_\xi} + \alpha_{v_\xi}\Bigg(\eta_\xi + \frac{1}{2}\Big(||\widehat{\boldsymbol{m}}_w||^2_\mathcal{D} + ||\widehat{\boldsymbol{V}}_w||^2_\mathcal{D} + ||\left(\widehat{\boldsymbol{m}}^2_\chi + \widehat{\boldsymbol{V}}_\chi\right)\overline{\boldsymbol{E}^2}||_1 \\ &- 2\Re e\left(\widehat{\boldsymbol{m}}^H_\chi\overline{\boldsymbol{w}\circ\boldsymbol{E}^*}\right)\Big)\Bigg), \end{cases} \tag{B.9}$$

où $\boldsymbol{d}_\chi = \sum_k \overline{v_k^{-1}}\widehat{\boldsymbol{\zeta}}_k \circ \widehat{\boldsymbol{m}}^\dagger_\chi + \overline{v_\xi^{-1}}\overline{\boldsymbol{w}\circ\boldsymbol{E}^*}$.

B.6 La variance des classes v_k

$\forall\kappa \in \{1,...,K\}$ nous avons :

$$
\begin{aligned}
\log\left(\widetilde{q}^n(v_k)\right) &\propto (1-\alpha)\log\left(\widehat{q}(v_k)\right) + \alpha_{v_k}\left\langle \log\left(p\left(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\boldsymbol{\psi},\boldsymbol{y}\right)\right)\right\rangle_{\tilde{q}(/v_k)} \\
&\propto (1-\alpha_{v_k})\left(-\frac{v_k}{\tilde{\eta}_k} + (\tilde{\phi}_k - 1)\log(v_k)\right) + \alpha_{v_k}\left\langle -\frac{1}{2}\sum_{\boldsymbol{r}}\delta(z(\boldsymbol{r})=\kappa)\right. \\
&\times \left.\left(\log(v_k) + \frac{|\chi(\boldsymbol{r})-\boldsymbol{m}_k|^2}{v_k}\right) + \frac{\eta_0}{v_k} + (\phi_0+1)\log(v_k)\right\rangle_{\tilde{q}(/v_k)} \\
&\propto (1-\alpha_{v_k})\left[-\frac{v_k}{\tilde{\eta}_k} + (\tilde{\phi}_k - \log(v_k))\right] + \alpha_{v_k}\left[-\left(\phi_0 + 1 + \frac{\sum \boldsymbol{r}\widehat{\zeta}_k(\boldsymbol{r})}{2}\right)\right. \\
&- \left.\overline{v_k^{-1}}\left\langle \eta_0 + \frac{|\chi(\boldsymbol{r})-\boldsymbol{m}_k|^2}{2}\right\rangle_{\tilde{q}(/v_k)}\right],
\end{aligned}
\tag{B.10}
$$

où $q(/v_k)) = q(\boldsymbol{\chi})q(\boldsymbol{w}))q(\boldsymbol{z})\prod_{l\neq v_k} q(\psi_l)$. Par identification nous obtenons :

$$
\begin{cases}
\widehat{\phi}_k^n = (1-\alpha_{v_k})\,\widehat{\phi}_k + \alpha_{v_k}\left(\phi_0 + \frac{\sum \boldsymbol{r}\widehat{\zeta}_k(\boldsymbol{r})}{2}\right) \\
\widehat{\eta}_k^n = \frac{1-\alpha_{v_k}}{\widehat{\eta}_k} + \alpha_{v_k}\left(\eta_0 + \frac{1}{2}\sum_{\boldsymbol{r}}\widehat{\zeta}_k(\boldsymbol{r})\left(|\widehat{m}_\chi(\boldsymbol{r})|^2 + \widehat{V}_\chi(\boldsymbol{r}) + m_k^2 + \tilde{\tau}_k^2\right)\right).
\end{cases}
\tag{B.11}
$$

B.7 La moyennes des classes m_k

$$
\begin{aligned}
\log\left(\widetilde{q}^n(m_k)\right) &\propto (1-\alpha_{m_k})\log\left(\widehat{q}(m_k)\right) + \alpha_{m_k}\left\langle \log\left(p\left(\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\boldsymbol{\psi},\boldsymbol{y}\right)\right)\right\rangle_{\tilde{q}(/m_k)} \\
&\propto (1-\alpha_{m_k})\left[\widehat{\tau}_k m_k^2 - 2\widehat{\mu}_k m_k\right] + \alpha_{m_k}\left(-\left\langle \frac{1}{2}\sum_{\boldsymbol{r}}\delta\left(z(\boldsymbol{r})=\kappa\right)\right.\right. \\
&\times \left.\left.\left(\frac{|\chi(\boldsymbol{r})-m_\chi(\boldsymbol{r})|^2}{v_k}\right) + \frac{|m_k-\mu_0|}{\tau_0}\right\rangle_{\tilde{q}(/m_k)}\right),
\end{aligned}
\tag{B.12}
$$

avec $q(/m_k)) = q(\boldsymbol{\chi})q(\boldsymbol{w}))q(\boldsymbol{z})\prod_{l\neq m_k} q(\psi_l)$.

Par identification nous avons :

$$
\begin{cases}
\widehat{\tau}_k^n = \left[(1-\alpha_{m_k})\,\widehat{\tau}_k + \alpha_{m_k}\left(\tau_0^{-1} + \overline{v_k^{-1}}\sum_{\boldsymbol{r}}\widehat{\zeta}_k(\boldsymbol{r})\right)\right]^{-1} \\
\widehat{\mu}_k^n = \widehat{\tau}_k^n\left[(1-\alpha_{m_k})\,\widehat{\mu}_k + \alpha_{m_k}\left(\widehat{\tau}_k\left(\frac{\mu_0}{\tau_0} + \overline{v_k^{-1}}\sum_{\boldsymbol{r}}\widehat{\zeta}_k(\boldsymbol{r})\widehat{m}_\chi(\boldsymbol{r})\right)\right)\right].
\end{cases}
\tag{B.13}
$$

C

Détermination du pas optimal de descente du G-BV

C.1 L'énergie libre négative

L'évaluation de l'énergie libre négative durant le processus itératif permet d'observer le comportement de l'algorithme et d'avoir une idée sur sa convergence. L'expression de l'énergie libre négative est donnée par :

$$\mathcal{F}(\boldsymbol{q}) = \langle p(\boldsymbol{y}, \boldsymbol{\chi}, \boldsymbol{w}, \boldsymbol{z}, \boldsymbol{\psi}|\mathcal{M})\rangle_{\boldsymbol{q}} + \mathcal{H}(q(\boldsymbol{\chi}, \boldsymbol{w}, \boldsymbol{z}, \boldsymbol{\psi})), \tag{C.1}$$

où $\mathcal{H}$ est donné par :

$$\begin{aligned}
\mathcal{H}(q(\boldsymbol{\chi}, \boldsymbol{w}, \boldsymbol{z}, \boldsymbol{\psi})) &= \sum_{N_\vartheta}\sum_{N_f}\sum_{\boldsymbol{r}}\sum_{k} \log\left(\sqrt{2\pi e \widehat{v}_w(\boldsymbol{r})}\right) + \sum_{\boldsymbol{r}}\sum_{k} \log\left(\sqrt{2\pi e \widehat{v}_\chi(\boldsymbol{r})}\right) \\
&+ \sum_{\boldsymbol{r}}\sum_{k} \tilde{\zeta}_\kappa(\boldsymbol{r})\left(\lambda \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} \tilde{\zeta}_k(\boldsymbol{r}') - \log\left(\sum_k \exp\left\{\lambda \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} \tilde{\zeta}_k(\boldsymbol{r}')\right\}\right)\right) \\
&+ \sum_{\kappa} \log\left(\sqrt{2\pi e \widehat{\tau}_\kappa}\right) + \sum_k \left(\widehat{\eta}_k + \log\left(\widehat{\phi}_k \Gamma(\widehat{\eta}_k)\right) - (1+\widehat{\eta}_k)\Psi(\widehat{\eta}_k)\right) \\
&+ \widehat{\eta}_\epsilon + \log\left(\widehat{\phi}_\epsilon \Gamma(\widehat{\eta}_\epsilon)\right) - (1+\widehat{\eta}_\epsilon)\Psi(\widehat{\eta}_\epsilon) + \widehat{\eta}_\xi + \log\left(\widehat{\phi}_\xi \Gamma(\widehat{\eta}_\xi)\right) \\
&- (1+\widehat{\eta}_\xi)\Psi(\widehat{\eta}_\xi).
\end{aligned} \tag{C.2}$$

Ainsi :

$$
\begin{aligned}
\langle p(\boldsymbol{y},\boldsymbol{\chi},\boldsymbol{w},\boldsymbol{z},\boldsymbol{\psi})\rangle_{\boldsymbol{q}} &= -\frac{N_{\mathcal{S}}+N_{\mathcal{D}}+K}{2}\log(2\pi)+(\phi_\epsilon-1)\log(\eta_\epsilon)-\Gamma(\phi_\epsilon)\\
&- \frac{N_{\mathcal{S}}N_\vartheta N_f+2\phi_\epsilon+2}{2}\Big(\Psi(\widehat{\eta}_\epsilon)-\log(\widehat{\phi}_\epsilon)\Big)-\eta_\epsilon\overline{v_\epsilon^{-1}}+(\phi_\xi-1)\log(\eta_\xi)-\Gamma(\phi_\xi)\\
&- \frac{N_{\mathcal{D}}N_\vartheta N_f+2\phi_\xi+2}{2}\Big(\Psi(\widehat{\eta}_\xi)-\log(\widehat{\phi}_\xi)\Big)-\eta_\xi\overline{v_\xi^{-1}}+\log(\eta_0^{K(\phi_0-1)})\\
&- \sum_k\frac{\Big(2\phi_0+1+\sum_{\boldsymbol{r}}\widehat{\zeta}_k(\boldsymbol{r})\Big)\Big(\Psi(\widehat{\eta}_k)-\log(\widehat{\phi_k})\Big)-2\eta_0\overline{v_k^{-1}}}{2}-K\,\Gamma(\phi_0)\\
&- \sum_k\frac{\widehat{m}_k^2+\widehat{\tau}_k+\mu_0^2-2\widehat{m}_k\mu_0}{2\tau_0}+\lambda\sum_{\boldsymbol{r}}\sum_k\widehat{\zeta}_k(\boldsymbol{r})\sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})}\widehat{\zeta}_k(\boldsymbol{r}')-\frac{\overline{v_\epsilon^{-1}}}{2}\\
&\times \Big(||\boldsymbol{y}||_{\mathcal{S}}^2+||\boldsymbol{G}^o\widehat{\boldsymbol{m}}_w||_{\mathcal{S}}^2-2\boldsymbol{y}^t\boldsymbol{G}^o\widehat{\boldsymbol{m}}_w+||\widehat{\boldsymbol{V}}_w\Gamma^o||_1\Big)\\
&- \frac{\overline{v_\xi^{-1}}}{2}\Big(||\widehat{\boldsymbol{m}}_w||_{\mathcal{D}}^2+||\widehat{\boldsymbol{V}}_w||_1+\Big|\Big|\Big(\widehat{\boldsymbol{m}}_\chi+\widehat{\boldsymbol{V}}_\chi\Big)\overline{\boldsymbol{E}^2}\Big|\Big|_1-2\widehat{\boldsymbol{m}}_\chi^H\overline{\boldsymbol{w}\circ\boldsymbol{E}^*}\Big)\\
&- \sum_{\boldsymbol{r}}\sum_k\frac{\overline{v_k^{-1}}\widehat{\zeta}_k(\boldsymbol{r})\Big(|\widehat{m}_\chi(\boldsymbol{r})|_{\mathcal{D}}^2+\widehat{v}_\chi(\boldsymbol{r})+\widehat{m}_\chi^{\dagger 2}(\boldsymbol{r})-2\widehat{m}_\chi(\boldsymbol{r})\widehat{m}_\chi^\dagger(\boldsymbol{r})\Big)}{2}. \qquad \text{(C.3)}
\end{aligned}
$$

Nous pouvons noter que le calcul de l'énergie négative libre dépend principalement des termes utilisés pour mettre à jour les valeurs des paramètres et l'évaluation ne nécessite pas de coût de calcul supplémentaire.

C.2 Le calcul du pas optimal de descente

En regardant les équations (B.2), (B.4), (B.5), (B.7), (B.9), (B.11) et (B.13), nous remarquons qu'il n'existe pas de dépendance entre les éléments du même groupe pour les variables χ, z, v_ϵ, v_ξ, v_k et m_k. Par contre, la mise à jour de la valeur moyenne des sources induites $\widehat{\boldsymbol{m}}_w^i$ à l'itération i dépend de la valeur moyenne $\widehat{\boldsymbol{m}}_w$. Nous avons donc décidé de remettre les pas de descente α_ρ, pour $\rho=\chi$, z, v_ϵ, v_ξ, m_k et v_k, à 1 et de chercher le pas optimal α_w qui assure le meilleur comportement de l'algorithme et permet d'atteindre la convergence vers la solution désirée le plus rapidement possible. Pour cela, le pas optimal est calculé par le bais de la méthode de Newton appliquée à l'énergie libre négative, soit :

$$
\alpha_w^{opt}=\nabla\mathcal{F}(\alpha_w)/\nabla^2\mathcal{F}(\alpha_w)\Big|_{\alpha_w=0}, \qquad \text{(C.4)}
$$

où $\nabla\mathcal{F}=\partial\mathcal{F}/\partial\alpha_w$ et $\nabla^2\mathcal{F}=\partial^2\mathcal{F}/\partial^2\alpha_w$.

Afin de déterminer le pas de descente α_w^{opt} pour la technique de gradient comme ap-

proche variationnelle, nous utiliserons les notations suivantes pour la facilité d'écriture :

$$\left\{\begin{array}{l} \widehat{\boldsymbol{V}}_w^n = \left[(1-\alpha_w)\,\widehat{\boldsymbol{V}}_w^{-1} + \alpha\ \boldsymbol{R}_w\right]^{-1} \\ \widehat{\boldsymbol{m}}_w^n = \widehat{\boldsymbol{m}}_w + \alpha_w\ \widehat{\boldsymbol{V}}_w^n\ \boldsymbol{d}_w \end{array}\right. \tag{C.5}$$

avec

$$\begin{aligned} \boldsymbol{R}_w &= \mathbf{Diag}\left(\overline{v_\epsilon^{-1}}\Gamma^o + \overline{v_\xi^{-1}}\overline{\Gamma^{xc}}\right) \\ \boldsymbol{d}_w &= \left[\overline{v_\epsilon^{-1}}\boldsymbol{G}^{o^H}(\boldsymbol{y} - \boldsymbol{G}^o\widehat{\boldsymbol{m}}_w) + \overline{v_\xi^{-1}}\left(\widehat{\boldsymbol{m}}_\chi\boldsymbol{E}^{inc} - \boldsymbol{G}^{c^H}\left(\widehat{\boldsymbol{m}}_\chi^2 + \widehat{\boldsymbol{V}}_\chi\right)\boldsymbol{E}^{inc}\right.\right. \\ &- \left.\left.\widehat{\boldsymbol{m}}_w + \widehat{\boldsymbol{m}}_\chi\boldsymbol{G}^c\widehat{\boldsymbol{m}}_w + \boldsymbol{G}^{c^H}\widehat{\boldsymbol{m}}_\chi^*\widehat{\boldsymbol{m}}_w - \boldsymbol{G}^{c^H}\left(\widehat{\boldsymbol{m}}_\chi^2 + \widehat{\boldsymbol{V}}_\chi\right)\boldsymbol{G}^c\widehat{\boldsymbol{m}}_w\right)\right]. \end{aligned}$$

Ensuite, réécrivons l'expression de l'énergie libre négative en fonction de α_w :

$$\begin{aligned} \mathcal{F}(\alpha_w) = &-\frac{\overline{v_\epsilon^{-1}}}{2}\left(||\boldsymbol{y}||_{\mathcal{S}}^2 + ||\boldsymbol{G}^o\widehat{\boldsymbol{m}}_w||_{\mathcal{S}}^2 - 2\boldsymbol{y}^t\boldsymbol{G}^o\widehat{\boldsymbol{m}}_w + ||\widehat{\boldsymbol{V}}_w\Gamma^o||_1\right) - \frac{\overline{v_\xi^{-1}}}{2}\left(||\widehat{\boldsymbol{m}}_w||_{\mathcal{D}}^2 + ||\widehat{\boldsymbol{V}}_w||_1\right. \\ &\left. + \left|\left|\left(\widehat{\boldsymbol{m}}_\chi + \widehat{\boldsymbol{V}}_\chi\right)\overline{\boldsymbol{E}^2}\right|\right|_1 - 2\widehat{\boldsymbol{m}}_\chi^H\overline{\boldsymbol{w}\circ\boldsymbol{E}^*}\right) + \sum_{\boldsymbol{r}}\log\left(\sqrt{2\pi e\widehat{v}_w(\boldsymbol{r})}\right). \end{aligned} \tag{C.6}$$

Calculons maintenant les dérivées première et seconde de $\mathcal{F}$:

$$\frac{\partial\mathcal{F}(\alpha_w)}{\partial\alpha_w} = \widehat{\boldsymbol{m}}_w'\ \boldsymbol{d}_w^t + \sum_{\boldsymbol{r}}\frac{\widehat{v}_w'(\boldsymbol{r})}{2\ \widehat{v}_w(\boldsymbol{r})} - \widehat{v}_w'(\boldsymbol{r})r_w(\boldsymbol{r}), \tag{C.7}$$

$$\frac{\partial^2\mathcal{F}(\alpha_w)}{\partial^2\alpha_w} = \widehat{\boldsymbol{m}}_w''\ \boldsymbol{d}_w^t + \widehat{\boldsymbol{m}}_w'\ \boldsymbol{d}_w^{t'} + \sum_{\boldsymbol{r}}\frac{\widehat{v}_w''(\boldsymbol{r})\widehat{v}_w(\boldsymbol{r}) - \left(\widehat{v}_w'(\boldsymbol{r})\right)^2}{2\left(\widehat{v}_w(\boldsymbol{r})\right)^2} - \widehat{v}_w''(\boldsymbol{r})r_w(\boldsymbol{r}), \tag{C.8}$$

où $v_w(\boldsymbol{r})$ et $r_w(\boldsymbol{r})$ sont les éléments diagonaux des matrices $\boldsymbol{V}_w$ and $\boldsymbol{R}_w$ et

$$\left\{\begin{array}{l} \widehat{\boldsymbol{V}}_w'^n = \left(\widehat{\boldsymbol{V}}_w^n\right)^2\left(\widehat{\boldsymbol{V}}_w^{-1} - \boldsymbol{R}_w\right) \\ \widehat{\boldsymbol{V}}_w''^n = \left(\widehat{\boldsymbol{V}}_w^n\right)^3\left(\widehat{\boldsymbol{V}}_w^{-1} - \boldsymbol{R}_w\right)^2 \\ \widehat{\boldsymbol{m}}_w'^n = \widehat{\boldsymbol{V}}_w^n\ \boldsymbol{d}_w + \alpha\ \widehat{\boldsymbol{V}}_w'^n\boldsymbol{d}_w \\ \widehat{\boldsymbol{m}}_w''^n = 2\ \widehat{\boldsymbol{V}}_w'^n\boldsymbol{d}_w + \alpha\ \widehat{\boldsymbol{V}}_w''^n\boldsymbol{d}_w. \end{array}\right. \tag{C.9}$$

Le pas optimal s'écrit alors :

$$\alpha_w^{opt} = \frac{\boldsymbol{d}_w^t\widehat{\boldsymbol{m}}_{w0}' + \frac{1}{2}\sum_{\boldsymbol{r}}\boldsymbol{S}_{w\boldsymbol{r}}^2}{\widehat{\boldsymbol{m}}_{w0}''\ \boldsymbol{d}_w^t + \widehat{\boldsymbol{m}}_{w0}'\ \boldsymbol{d}_w^{t'} + \frac{1}{2}\sum_{\boldsymbol{r}}\boldsymbol{S}_{w\boldsymbol{r}}^3}, \tag{C.10}$$

où

$$\begin{cases} \left.\frac{\partial \widehat{\boldsymbol{m}}_w}{\partial \alpha_w}\right|_{\alpha_w=0} = \widehat{\boldsymbol{m}}'_{w0} = \widehat{\boldsymbol{V}}_{w0}\boldsymbol{d}_w, \\ \left.\frac{\partial^2 \widehat{\boldsymbol{m}}_w}{\partial^2 \alpha_w}\right|_{\alpha_w=0} = \widehat{\boldsymbol{m}}''_{w0} = 2\ \widehat{\boldsymbol{V}}'_{w0}\boldsymbol{d}_w, \\ \left.\frac{\partial \widehat{\boldsymbol{V}}_w}{\partial \alpha_w}\right|_{\alpha_w=0} = \widehat{\boldsymbol{V}}'_{w0} = -\widehat{\boldsymbol{V}}_{w0}\boldsymbol{S}_w \end{cases} \tag{C.11}$$

et

$$\begin{aligned} \boldsymbol{d}_w^{t'} &= \widehat{\boldsymbol{m}}_w'^t \left(\overline{v_\epsilon^{-1}}\boldsymbol{G}^o\boldsymbol{G}^H - \overline{v_\xi^{-1}}\Big(1 - \widehat{\boldsymbol{m}}_\chi \boldsymbol{G}^c - \boldsymbol{G}^{c^H}\widehat{\boldsymbol{m}}_\chi^* - \boldsymbol{G}^{c^H}(\widehat{\boldsymbol{m}}_\chi + \widehat{\boldsymbol{V}}_\chi)\boldsymbol{G}^c\Big)\right)^t, \\ \boldsymbol{S}_{w\boldsymbol{r}} &= \boldsymbol{R}_{w\boldsymbol{r}}\widehat{\boldsymbol{V}}_{w0\boldsymbol{r}} - 1. \end{aligned} \tag{C.12}$$

D

Publications

Articles dans des revues internationales avec comité de lecture

— [R1] L. Gharsalli, H. Ayasso, B. Duchêne et A. Mohammad-Djafari (2015). Variational Bayesian inversion for microwave breast imaging. Soumis à special issue of Computer Assisted Methods in Mechanics and Science, 14 pages. **Sollicité**.

— [R2] L. Gharsalli, H. Ayasso, B. Duchêne et A. Mohammad-Djafari (2014). Inverse scattering in a bayesian framework : application to microwave imaging for breast cancer detection. Inverse Problems, 30(11) : 114011. DOI : 10.1088/0266-5611/30/11/114011.

Communications nationales et internationales avec actes

— [C1] L. Gharsalli, H. Ayasso, B. Duchêne et A. Mohammad-Djafari (2014). A Gauss-Markov mixture prior model for a variational Bayesian approach to microwave breast imaging. In 2014 IEEE International Conference on Antenna Measurements and Applications (CAMA), Juan-les-Pins, France, actes : ID SP13.4 – 4 pages.

— [C2] L. Gharsalli, H. Ayasso, B. Duchêne et A. Mohammad-Djafari (2014). A gradient-like variational Bayesian approach : application to microwave imaging for breast tumor detection. In 21st IEEE International Conference on Image Processing, (ICIP), Paris, France, actes : ID TEC-P3.4 – 5 pages.

— [C3] L. Gharsalli, H. Ayasso, B. Duchêne et A. Mohammad-Djafari (2014). Variational Bayesian inversion for microwave imaging applied to breast cancer detection. In 8th International Conference on Inverse Problems in Engineering (ICIPE), Cracovie, Pologne, actes : ID 5-2 – 9 pages.

— [C4] L. Gharsalli, H. Ayasso, B. Duchêne et A. Mohammad-Djafari (2013). Microwave tomography for breast cancer detection within a variational Bayesian approach. In 21st IEEE European Signal Processing Conference, (EUSIPCO), Marrakech, Maroc, actes : ID1569743387 - 5 pages.

— [C5] L. Gharsalli, H. Ayasso, B. Duchêne et A. Mohammad-Djafari (2013). Approche bayésienne variationnelle en tomographie micro-onde appliquée à la détection du cancer du sein. In XXIVe Colloque GRETSI - Traitement du Signal et des Images, Brest France, actes : ID411- 4 pages.

— [C6] L. Gharsalli, A. Mohammad-Djafari, A. Fraysse et T. Rodet (2012). Variational Bayesian approximation with scale mixture prior : a comparison between three algorithms. In 32nd International Workshop on Bayesian Inference and Maximun Entropy Methods in Sciences and Engineering (MaxEnt), actes : volume 1353, pages 130-138, Garching, Allemagne.

Communications nationales et internationales sans actes

— [B1] L. Gharsalli, H. Ayasso, B. Duchêne et A. Mohammad-Djafari (2014). Variational Bayesian Approach with a heavy-tailed prior distribution for solving a nonlinear inverse scattering problem. In 33rd International Workshop on Bayesian Inference and Maximun Entropy Methods in Sciences and Engineering (MaxEnt), Le Clos Lucé, Amboise, France.

— [B2] L. Gharsalli, H. Ayasso, B. Duchêne et A. Mohammad-Djafari (2013). Microwave imaging within a variational Bayesian framework : application to breast tumor detection. Interdisciplinary Symposium on Signals and Systems for Medical Applications (ISSSMA), Paris France.

— [B3] A. Mohammad-Djafari, L. Gharsalli et M. Dumitru (2012). Bayesian variational approximation implementation for linear inverse problem with infinite Gaussian mixture model. Journées Résolution de problèmes inverses : optimisation et parallélisation (GDR ISIS). Paris France.

IOP Publishing Inverse Problems
Inverse Problems **30** (2014) 114011 (26pp) doi:10.1088/0266-5611/30/11/114011

Inverse scattering in a Bayesian framework: application to microwave imaging for breast cancer detection

Leila Gharsalli[1], Hacheme Ayasso[2], Bernard Duchêne[1] and Ali Mohammad-Djafari[1]

[1] Laboratoire des Signaux et Systèmes (L2S), (UMR8506: CNRS—SUPELEC — Univ Paris-Sud), 3 rue Joliot-Curie, 91190 Gif-sur-Yvette, France
[2] GIPSA-LAB, Département Image Signal, (CNRS-Univ de Grenoble), BP 46—38402, Saint Martin d'Hères, France

E-mail: Leila.Gharsalli@lss.supelec.fr, Hacheme.Ayasso@gipsa-lab.grenoble-inp.fr, Bernard.Duchene@lss.supelec.fr and Ali.Mohammad-Djafari@lss.supelec.fr

Received 3 March 2014, revised 16 June 2014
Accepted for publication 1 July 2014
Published 28 October 2014

Abstract

In this paper, we deal with a nonlinear inverse scattering problem where the goal is to detect breast cancer from measurements of the scattered field that results from the interaction between the breast and a known interrogating wave in the microwave frequency range. Modeling of the wave-object (breast) interaction is tackled through a domain integral representation of the electric field in a 2D-TM configuration. The inverse problem is solved in a Bayesian framework where prior information, which consists in the fact that the object is supposed to be composed of compact homogeneous regions made of a restricted number of different materials, is introduced via a Gauss–Markov–Potts model. As an analytic expression for the joint maximum *a posteriori* (MAP) estimators yields an intractable solution, an approximation of the latter is proposed. This is done by means of a variational Bayesian approximation (VBA) technique that is adapted to complex-valued contrast and applied to compute the posterior estimators, and reconstruct maps of both permittivity and conductivity of the sought object. This leads to a joint semi-supervised estimation approach, which allows us to estimate the induced currents, the contrast and all of the parameters introduced in the prior model. The method is tested on two sets of synthetic data generated in different configurations and its performances are compared to that given by a contrast source inversion technique.

 © 2014 IOP Publishing Ltd Printed in the UK

Keywords: inverse scattering, microwave imaging, breast cancer detection, Gauss–Markov–Potts prior, Bayesian framework

(Some figures may appear in colour only in the online journal)

1. Introduction

In last few decades, microwave imaging has received a particular increasing interest for medical applications, such as brain imaging or breast cancer detection [1–8]. Herein, we are concerned with the latter application. In fact, breast cancer is a major health problem because it is the second most common cancer form for women. This is the reason why its diagnosis at an early stage remains a real challenge. The most popular method for breast tumor detection is certainly x-ray mammography, which consists in retrieving images of a compressed breast by means of low doses of ionizing radiations [9]. Nevertheless, the x-ray screening of breast tumors has several drawbacks; for example, radiations are ionizing and the procedure itself is costly, which limits its wide use. In addition, the tissue density and age of the patient can disturb the results of x-ray mammograms [10, 11].

In contrast, although it does not yield a spatial resolution as high as that provided by x-rays, microwave imaging benefits from high sensibility with respect to physical or physiological factors of clinical interest, such as water content, blood-flow rate, and temperature. Furthermore, microwave attenuation in normal breast tissue is low enough to allow wave propagation through relatively large breast volumes. Thus, due to the non negligible dielectric contrast that exists between cancerous and normal healthy breast tissues [12], and to the non-ionizing nature of microwaves, the latter can provide a less costly and less risky imaging alternative than x-rays.

Microwave imaging is taken herein as a nonlinear inverse scattering problem whose resolution requires two major steps. The first step is the construction of a forward model that describes the interaction between the object (the breast) and a known interrogating (or incident) wave. The latter is based upon a domain integral representation of the electric field in a 2D-TM configuration. The second step is the inversion itself, which consists in retrieving a contrast function representative of the dielectric properties (permittivity and conductivity) of the unknown object from measurements of the scattered field that results from its interaction with the incident wave. When the object under test is small compared to the wavelength, or is lowly contrasted with respect to the embedding, the inverse problem can be linearized in the framework of the first order weak scattering Born or Rytov approximations [13, 14]. However, in the present case the sought object is large and it can present high dielectric contrasts, in these conditions the linearizing assumptions fail to provide quantitative information [15, 16].

In the early 1990s, several deterministic inversion algorithms [17–19] were developed to deal with the nonlinear problem at hand through an iterative minimization of a cost functional that expresses the discrepancy between the measured scattered fields and the fields computed by means of the current solution (the sought contrast). However, in addition to being nonlinear, inverse scattering problems are also known to be ill-posed, which means that a regularization is required prior to their resolution. Such a regularization is usually done by introducing *a priori* information on the object, which is not easy with the latter techniques because this information must be introduced in the functional to be minimized [20–22].

This led researchers to turn to stochastic inversion techniques, such as simulated annealing [23] or evolutionary optimization algorithms [24, 25], such as differential evolution

methods [26, 27], genetic algorithms [28, 29] or particle swarm optimizations [30, 31]. Since they are based upon global optimization strategies, these algorithms do not suffer from the drawback of deterministic methods that can be trapped in suboptimal solutions corresponding to local minima. However, this is at the price of a much higher computational cost. Let us emphasize the fact that, over the last few years, leading edge researches have been concerned about sparsity based regularization of inverse scattering problems [32–34] through compressive sensing [35, 36]. However, the latter assumes a linear relationship between unknowns and data, which means a linearization of the inverse problem, and sparsity of the object, which means that it can be represented by a few coefficients in a suitable expansion basis. *A priori* information can be introduced via this expansion basis, which should be wisely chosen.

The method presented in this paper belongs to the class of stochastic inversion techniques and is based upon the probabilistic Bayesian framework, whose main advantage is that it allows us to more easily take into account *a priori* information than the latter methods. *A priori* information consists, herein, in the fact that the object under test is known to be composed of a finite number of different materials, which means that the images to be reconstructed are composed of a finite number of homogeneous areas. This configuration is encountered in many applications where man-made objects are considered, it is also encountered in natural media where inhomogeneities are small compared to the wavelength. The situation is somewhat similar for microwave breast imaging because, although they are very heterogeneous and show a large variability between individuals, the breast tissues may be partitioned into four different groups (one for the tumors [37] and three for the normal tissues, depending upon their fat content [38]) whose dielectric properties mean that the values can be described by one-pole Debye models over the 0.5–6 GHz frequency range considered herein [39].

To account for this prior knowledge, a Gauss–Markov–Potts model [40] is proposed. The latter consists in a Gauss-Markov field with a hidden Potts label field, where the marginal distribution of contrast is sought for as a Gaussian mixture [41], each Gaussian law representing a given material, and the compactness of the regions is accounted for by means of a hidden Markov model.

Taking into account such *a priori* information, the Bayesian computation aims to reconstruct the joint posterior distribution of the unknowns and parameters of the model thanks to the Bayes rule. Generally, the estimators are computed in the maximum *a posteriori* (MAP) sense, where maximization leads to a non-convex optimization issue, or through the posterior mean (PM), which involves an integral whose analytical solution is often hardly tractable. Therefore, an approximation of the posterior law is sought in order to obtain the desired solution at a reasonable computational cost. Two methods are widely used: the Markov Chain Monte Carlo method (MCMC) [42], which estimates the posterior distribution by performing an empirical mean through a stochastic Gibbs sampling, and the deterministic Variational Bayesian Approach (VBA), which is inspired by statistical physics [43]. The latter has been introduced as an alternative to MCMC methods that are rather costly in terms of computation time due to Gibbs sampling [44, 45]. The idea therein is to approximate the posterior distribution by a separable density. Given that the calculations are analytic, the rate of convergence of VBA is expected to be much better than that of MCMC approaches.

This method has already been applied to optical diffraction tomography [46]. The originality, herein, is to apply it to microwave imaging for breast cancer detection, where, contrarily to the case treated in [46], the unknown contrast is complex-valued because biological media are lossy, which means that both permittivity and conductivity images have to be reconstructed. Hence the proposed Bayesian approach and the variational Bayesian

approximation (VBA) method are adapted to the case of complex-valued contrasts by assuming that permittivity and conductivity have the same segmentation (i.e. the same hidden field) but are conditionally independent to this hidden variable.

2. Problem statement

2.1. Domain integral representation

We consider a 2-D configuration in a transverse magnetic polarization case where the object under test is supposed to be cylindrical, of infinite extension along the z axis, and illuminated by a line source whose location can be varied and that operates at several discrete frequencies. This source generates an incident electric field E^{inc} that is polarized along the z axis with an $\exp(-\mathrm{i}\omega t)$ implicit time dependence. The object is supposed to be contained in a test domain $\mathcal{D}$ and the different media are characterized by their propagation constant $k(\boldsymbol{r})$, such that $k(\boldsymbol{r})^2 = \omega^2\epsilon_0\epsilon_r(\boldsymbol{r})\mu_0 + \mathrm{i}\omega\mu_0\sigma(\boldsymbol{r})$, where ω is the angular frequency, ϵ_0 and μ_0 are the permittivity and the permeability of free space, respectively, $\boldsymbol{r} \in \mathcal{D}$ is an observation point and $\epsilon_r(\boldsymbol{r})$ and $\sigma(\boldsymbol{r})$ are the relative permittivity and conductivity of the medium.

We now consider a contrast function χ defined in $\mathcal{D}$ and null outside the object, such that $\chi(\boldsymbol{r}) = (k(\boldsymbol{r})^2 - k_1^2)/k_1^2$, where k_1 is the propagation constant of the embedding medium and we define $w(\boldsymbol{r})$ as the Huygens type sources induced within the object by the incident wave, that is, $w(\boldsymbol{r}) = \chi(\boldsymbol{r})E(\boldsymbol{r})$, where $E(\boldsymbol{r})$ is the total field in the object. By applying Green's theorem to the Helmholtz wave equations satisfied by the fields and by accounting for boundary and radiation conditions, we obtain two coupled contrast source integral equations, whose first one, denoted as *observation* equation, is a first-kind Fredholm integral equation that relates the scattered field $y(\boldsymbol{r})$ observed on a measurement domain $\mathcal{S}$ ($\boldsymbol{r} \in \mathcal{S}$), which results from the interaction between the object and the interrogating wave E^{inc}, to the induced sources $w(\boldsymbol{r})$:

$$y(\boldsymbol{r}) = k_1^2 \int_{\mathcal{D}} G(\boldsymbol{r}, \boldsymbol{r}')w(\boldsymbol{r}')\mathrm{d}\boldsymbol{r}', \tag{1}$$

where $G(\boldsymbol{r}, \boldsymbol{r}')$ is the 2-D Green's function ($G(\boldsymbol{r}, \boldsymbol{r}') = \mathrm{i}H_0^{(1)}(k|\boldsymbol{r} - \boldsymbol{r}'|)/4$, with $H_0^{(1)}$ the first kind Hankel function of order 0).

The second equation, the so-called *coupling* (or *state*) equation, relates the unknown total field E in the object to the induced sources $w(\boldsymbol{r})$:

$$E(\boldsymbol{r}) = E^{inc}(\boldsymbol{r}) + k_1^2 \int_{\mathcal{D}} G(\boldsymbol{r}, \boldsymbol{r}')w(\boldsymbol{r}')\mathrm{d}\boldsymbol{r}'. \tag{2}$$

The forward problem consists in first solving (2) for E, knowing χ, and then solving (1) for y knowing w. This is done from discrete counterparts of the above equations, which are obtained by means of a method of moments [47] with pulse-basis and point matching. The test domain $\mathcal{D}$ is then partitioned into $N_{\mathcal{D}}$ elementary square pixels with side δ that is small enough to consider the field and the contrast as constant over each of them.

2.2. The discrete model

Let us denote as $\boldsymbol{E}$, $\boldsymbol{\chi}$ and $\boldsymbol{w}$ the vectors that contain the values of $E(\boldsymbol{r}')$, $\chi(\boldsymbol{r}')$ and $w(\boldsymbol{r}')$ at the centers $\boldsymbol{r}'$ of the pixels ($\boldsymbol{r}' \in \mathcal{D}$), as $\boldsymbol{y}$ the vector containing the values of the scattered field $y(\boldsymbol{r})$ at the measurement points and as $\boldsymbol{G}^{\mathrm{o}}$ and $\boldsymbol{G}^{\mathrm{c}}$ matrices whose elements result from the integration of the Green's function over the elementary pixels:

$$\boldsymbol{G}^{o,c}(\boldsymbol{r},\boldsymbol{r}') = \begin{cases} \dfrac{i\pi k_1 c}{2} H_0^{(1)}\left(k_1\,|\boldsymbol{r}-\boldsymbol{r}'|\right) J_1\left(k_1 c\right) \text{ if } \boldsymbol{r}\neq\boldsymbol{r}', \boldsymbol{r}\in\mathcal{S} \text{ or } \mathcal{D}, \\ \dfrac{i\pi k_1 c}{2} H_1^{(1)}\left(k_1 c\right) - 1 \text{ if } \boldsymbol{r}=\boldsymbol{r}', \boldsymbol{r}\in\mathcal{D}, \end{cases} \tag{3}$$

where J_1 and H_1^1 are the first order Bessel and Hankel functions of the first kind, respectively, and $c = \delta/\sqrt{\pi}$.

By accounting for the different frequencies (f) and source positions (ν) through a subscript n, and by rewriting (2) in terms of the contrast sources, the discrete counterparts of (1) and (2) now read:

$$\boldsymbol{y}_n = \boldsymbol{G}_f^o\, \boldsymbol{w}_n\,, \tag{4}$$

$$\boldsymbol{w}_n = \boldsymbol{X}_f\, \boldsymbol{E}_n^{inc} + \boldsymbol{X}_f\, \boldsymbol{G}_f^c\, \boldsymbol{w}_n\,, \tag{5}$$

where $\boldsymbol{X}_f = \mathrm{diag}(\boldsymbol{\chi}_f)$. Let us now also introduce two variables, $\boldsymbol{\epsilon}$ and $\boldsymbol{\xi}$, that account for all of the errors: measurement, uncertainties and model errors (discretization and other approximations). In the absence of any information on the noise, except maybe the SNR, they are supposed to be centered and white and to satisfy Gaussian laws ($\boldsymbol{\epsilon}_n \sim \mathcal{N}(\boldsymbol{\epsilon}|\mathbf{0}, v_\epsilon \mathbf{I})$ and $\boldsymbol{\xi}_n \sim \mathcal{N}(\boldsymbol{\xi}|\mathbf{0}, v_\xi \mathbf{I})$) because this is the less informative prior that can be assigned and it is easy to handle concerning the approximating laws and the updating formulas. However, further information could be taken into account at the price of much more involved calculations. Then, we get the following discrete counterparts of the observation and coupling equations:

$$\boldsymbol{y}_n = \boldsymbol{G}_f^o\, \boldsymbol{w}_n + \boldsymbol{\epsilon}_n, \tag{6}$$

$$\boldsymbol{w}_n = \boldsymbol{X}_f\, \boldsymbol{E}_n^{inc} + \boldsymbol{X}_f\, \boldsymbol{G}_f^c\, \boldsymbol{w}_n + \boldsymbol{\xi}_n. \tag{7}$$

The inverse problem consists in estimating the contrast $\boldsymbol{\chi}$ or, more precisely, the relative permittivity ϵ_r and the conductivity σ, from the scattered fields $\boldsymbol{y}$, given the incident fields $\boldsymbol{E}^{\mathrm{inc}}$. It can be noted that since the induced sources $\boldsymbol{w}$ are unknown they have to be estimated at the same time as $\boldsymbol{\chi}$. We will attempt to solve this by means of a technique developed in a Bayesian estimation framework.

3. Bayesian Inversion framework

Inverse scattering problems are known to be ill-posed in the sense of Hadamard [48], which means that they do not admit a unique stable solution. Thus, they must be regularized, which is usually done by introducing *a priori* information on the sought for solution. In a deterministic regularization approach, *a priori* information is usually introduced through a regularization criterion [17–19], which implies several limitations:

- the regularization term, whose choice is limited because it usually introduces smoothness into the solution;
- the determination of a regularization parameter, which is often empirical; and,
- the lack of tools to quantify the remaining uncertainties of the solution.

In fact, Bayesian approaches do not show such limitations and they allow us to introduce suitable *a priori* information in a very easy way.

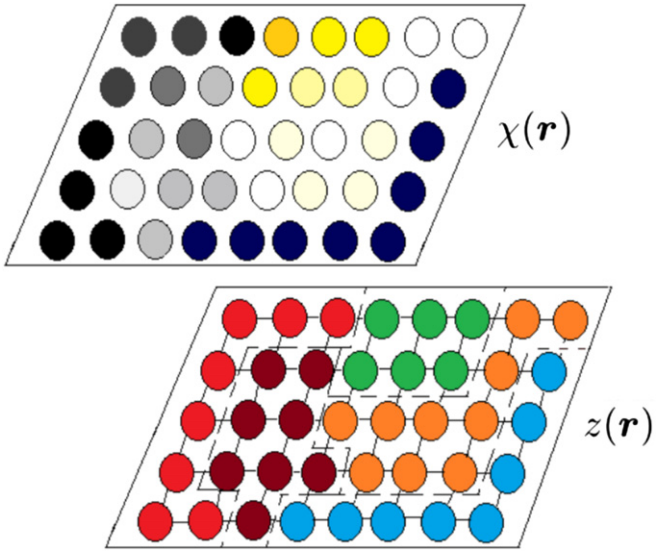


Figure 1. The contrast χ (top) and the associated hidden field z (down).

3.1. The prior modeling

In the present case, the information that we would like to account for is that the object is composed of a finite number K of different materials distributed in compact homogeneous regions. This prior information is introduced by means of a hidden variable $z(\boldsymbol{r})$, which represents the constitutive material of pixel $\boldsymbol{r}$ and that can take discrete values k ($k = 1,..., K$). This label defines different classes of materials and pixels with a given class k, and can be characterized by a contrast that satisfies a Gaussian distribution:

$$p(\chi(\boldsymbol{r})|z(\boldsymbol{r}) = k) = \mathcal{N}\big(\chi(\boldsymbol{r})\big|m_k, v_k\big), \qquad k = 1,..., K, \tag{8}$$

with a mean value m_k that corresponds to the average contrast of the material and a variance v_k that corresponds to the fluctuations around the latter. Hence, the contrast is sought for as a Gaussian mixture, where each Gaussian law represents a given material, and its values belong to a continuous set.

The information that the different materials are distributed in compact homogeneous regions is accounted for by means of a Potts-Markov model on $\boldsymbol{z}$, which expresses the spatial dependence between the neighboring pixels:

$$p(\boldsymbol{z}|\lambda) = \frac{1}{T(\lambda)} \exp\left\{\lambda \sum_{\boldsymbol{r}\in\mathcal{D}} \sum_{\boldsymbol{r}'\in V_{\boldsymbol{r}}} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right\}, \tag{9}$$

where $z = \{z(\boldsymbol{r}), \boldsymbol{r} \in \mathcal{D}\}$ represents the image of the labels (segmentation), λ is a parameter that determines the correlation between neighbors, $T(\lambda)$ is a normalization factor, and $V_{\boldsymbol{r}}$ is a neighborhood of $\boldsymbol{r}$, herein made of the four nearest pixels (figure (1)).

The choice of the correlation parameter λ determines the *a priori* behavior of the hidden field and, hence, the size of the homogeneous areas in the final image: the larger that λ is, the larger the latter are. However, its estimation is difficult since $T(\lambda)$ does not have a closed form. Therefore, λ is set to a value slightly higher than a critical value $\lambda_c = \log(1 + \sqrt{K})$, which yields a good trade-off in the correlation degree [49].

3.2. Posterior laws

It must be noted that a semi-supervised context is considered herein since K is supposed to be known, whereas the contrast $\boldsymbol{\chi}$, the induced currents $\boldsymbol{w}$, the segmentation labels $\boldsymbol{z}$ and the hyper-parameters of the model $\boldsymbol{\Theta} = \{\boldsymbol{m}, \boldsymbol{v}, v_\epsilon, v_\xi\}$ (with $\boldsymbol{m} = \{m_1, m_2, \ldots, m_K\}$ and $\boldsymbol{v} = \{v_1, v_2, \ldots, v_K\}$) are estimated simultaneously. Using the Bayes rule, the expression of the joint posterior distribution of all the unknowns is given by:

$$\begin{aligned} p(\boldsymbol{\chi}, \boldsymbol{w}, \boldsymbol{z}, \boldsymbol{\Theta}|\boldsymbol{y}) \propto\ & p\left(\boldsymbol{y}\middle|\boldsymbol{w}, v_\epsilon\right) p\left(\boldsymbol{w}\middle|\boldsymbol{\chi}, v_\xi\right) p(\boldsymbol{\chi}|\boldsymbol{z}, \boldsymbol{m}, \boldsymbol{v}) \\ & \times p(\boldsymbol{z}|\lambda)\, p\left(\boldsymbol{m}\middle|\mu_0, \tau_0\right) p\left(\boldsymbol{v}\middle|\eta_0, \phi_0\right) \\ & \times p\left(v_\epsilon\middle|\eta_\epsilon, \phi_\epsilon\right) p\left(v_\xi\middle|\eta_\xi, \phi_\xi\right). \end{aligned} \tag{10}$$

The expressions of $p(\boldsymbol{\chi}|\boldsymbol{z}, \boldsymbol{m}, \boldsymbol{v})$ and $p(\boldsymbol{z}|\lambda)$, that appear in the above relationship, are derived from equations (8) and (9), respectively, whereas that of $p(\boldsymbol{y}|\boldsymbol{w}, v_\epsilon)$ and $p(\boldsymbol{w}|\boldsymbol{\chi}, v_\xi)$ are derived from equations (6) and (7). The latter read:

$$\begin{aligned} p\left(\boldsymbol{y}\middle|\boldsymbol{w}, v_\epsilon\right) &= \frac{1}{\left(2\pi v_\epsilon\right)^{N_S/2}} \exp\left\{ -\frac{\|\boldsymbol{y} - \boldsymbol{G}^o\boldsymbol{w}\|_{\mathcal{S}}^2}{2v_\epsilon} \right\}, \\ p\left(\boldsymbol{w}\middle|\boldsymbol{\chi}, v_\xi\right) &= \frac{1}{\left(2\pi v_\xi\right)^{N_{\mathcal{D}} N_f N_v/2}} \exp\left\{ -\frac{\|\boldsymbol{w} - \boldsymbol{X}\boldsymbol{E}^{inc} - \boldsymbol{X}\boldsymbol{G}^c\boldsymbol{w}\|_{\mathcal{D}}^2}{2v_\xi} \right\}, \end{aligned} \tag{11}$$

where N_v, N_f, $N_{\mathcal{S}}$ and $N_{\mathcal{D}}$ are the numbers of source positions, frequencies, measurement points and pixels partitioning the test domain, respectively, and $\|.\|_{\mathcal{A}}$ is the norm associated to the inner product in $L^2(\mathcal{A})$, $\mathcal{A} = \mathcal{D}$ or $\mathcal{S}$.

As for the hyper-parameters, the reconstruction results are less sensitive to their priors than to that of the parameters. Hence, in order to make the computations easier, they are assigned to follow conjugate prior laws because the corresponding posteriors stay in the same family:

$$\begin{aligned} p(m_k) &= \mathcal{N}\left(m_k\middle|\mu_0, \tau_0\right),\ p(v_k) = \mathcal{IG}\left(v_k\middle|\eta_0, \phi_0\right), \\ p(v_\epsilon) &= \mathcal{IG}\left(v_\epsilon\middle|\eta_\epsilon, \phi_\epsilon\right),\ p(v_\xi) = \mathcal{IG}\left(v_\xi\middle|\eta_\xi, \phi_\xi\right), \end{aligned} \tag{12}$$

where $\mathcal{IG}(v|\eta, \phi) \propto v^{-\phi-1} \exp(-\eta/v)$ and $\mathcal{N}(m|\mu, \tau) \propto \exp(-(m-\mu)^2/2\tau)$ stand for inverse-gamma and Gaussian distributions, respectively, and μ_0, τ_0, η_0, ϕ_0, η_ϵ, ϕ_ϵ, η_ξ and ϕ_ξ are meta-hyper-parameters that are appropriately set to obtain almost non-informative flat prior distributions.

From equation (11), different inferences can be done on the unknowns. One of the usual solutions is the Joint Maximum *A Posteriori* (JMAP).

3.3. The JMAP estimators

JMAP solution can be computed *via* an alternate optimization with respect to the different components of the joint MAP criterion:

$$
\begin{aligned}
J_{MAP}(\boldsymbol{\chi}, \boldsymbol{w}, z, \Theta) &\propto -\log p(\boldsymbol{\chi}, \boldsymbol{w}, z, \Theta|\boldsymbol{y}) \\
&\propto -\log p\left(\boldsymbol{y}\middle|\boldsymbol{w}, v_\epsilon\right) - \log p\left(\boldsymbol{w}\middle|\boldsymbol{\chi}, v_\xi\right) - \log p\left(\boldsymbol{\chi}\middle|z, \boldsymbol{m}, \boldsymbol{v}\right) \\
&\quad - \log p\left(z\middle|\lambda\right) - \log p\left(\boldsymbol{m}\middle|\mu_0, \tau_0\right) - \log p\left(\boldsymbol{v}\middle|\eta_0, \phi_0\right) \\
&\quad - \log p\left(v_\epsilon\middle|\eta_\epsilon, \phi_\epsilon\right) - \log p\left(v_\xi\middle|\eta_\xi, \phi_\xi\right) \qquad (13)
\end{aligned}
$$

By accounting for equations (8), (9), (13) and error assignments, this criterion becomes:

$$
\begin{aligned}
J_{MAP}(\boldsymbol{\chi}, \boldsymbol{w}, z, \Theta) \propto{}& \frac{N_S}{2}\log(v_\epsilon) + \frac{1}{2v_\epsilon}\|\boldsymbol{y} - \boldsymbol{G}^o\boldsymbol{w}\|_S^2 \\
&+ \frac{N_D N_f N_\nu}{2}\log(v_\xi) + \frac{1}{2v_\xi}\|\boldsymbol{w} - \boldsymbol{X}\boldsymbol{E}^{inc} - \boldsymbol{X}\boldsymbol{G}^c\boldsymbol{w}\|_D^2 \\
&+ \sum_k\left(\frac{N_k\log(v_k)}{2} + \sum_{\boldsymbol{r}}\frac{\left|\chi(\boldsymbol{r}) - m_k(\boldsymbol{r})\right|^2}{2v_k}\right) \\
&- \lambda\sum_{\boldsymbol{r}}\sum_{\boldsymbol{r}'}\delta(z(\boldsymbol{r}) - z(\boldsymbol{r}')) + \frac{\sum_k\left|m_k - \mu_0\right|^2}{2\tau_0} \\
&+ \sum_k\left(\frac{\eta_0}{v_k} + \left(\phi_0 + 1\right)\log(v_k)\right) \\
&+ \frac{\eta_\epsilon}{v_\epsilon} + \left(\phi_\epsilon + 1\right)\log(v_\epsilon) + \frac{\eta_\xi}{v_\xi} + \left(\phi_\xi + 1\right)\log(v_\xi). \qquad (14)
\end{aligned}
$$

Looking for JMAP solutions $\hat{\mathbf{x}} = \{\widehat{\boldsymbol{\chi}}, \widehat{\boldsymbol{w}}, \hat{z}, \widehat{\Theta}\}$ is done by an alternate maximization of equation (14) with respect to each component; that is, with respect to:

- the induced sources

$$
\begin{aligned}
\nabla_{\boldsymbol{w}} J_{MAP} &= \nabla_{\boldsymbol{w}}\left(\frac{1}{2v_\epsilon}\|\boldsymbol{y} - \boldsymbol{G}^o\boldsymbol{w}\|_S^2 + \frac{1}{2v_\xi}\|\boldsymbol{w} - \boldsymbol{X}\boldsymbol{E}^{inc} - \boldsymbol{X}\boldsymbol{G}^c\boldsymbol{w}\|_D^2\right) \\
&= v_\epsilon^{-1}\boldsymbol{G}^{oH}\left(\boldsymbol{y} - \boldsymbol{G}^o\boldsymbol{w}\right) + v_\xi^{-1}\boldsymbol{Z}^H\left(\boldsymbol{w}\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{E}^{inc}\right) \\
&= v_\epsilon^{-1}\boldsymbol{G}^{oH}\boldsymbol{y} - v_\epsilon^{-1}\boldsymbol{G}^{oH}\boldsymbol{G}^o\boldsymbol{w} + v_\xi^{-1}\boldsymbol{Z}^H\boldsymbol{w}\boldsymbol{Z} - v_\xi^{-1}\boldsymbol{Z}^H\boldsymbol{X}\boldsymbol{G}^c, \qquad (15)
\end{aligned}
$$

where superscript H indicates the adjoint operator and $\boldsymbol{Z} = (\mathbf{I} - \boldsymbol{X}\boldsymbol{G}^c)$. The MAP estimate of the induced sources $\widehat{\boldsymbol{w}}$ is then the solution of $\nabla_{\boldsymbol{w}} J_{MAP} = 0$, which reads:

$$
\widehat{\boldsymbol{w}} = \left(v_\xi^{-1}\boldsymbol{Z}^H\boldsymbol{Z} - v_\epsilon^{-1}\boldsymbol{G}^{oH}\boldsymbol{G}^o\right)^{-1}\left(v_\xi^{-1}\boldsymbol{Z}^H\boldsymbol{X}\boldsymbol{E}^{inc} - v_\epsilon^{-1}\boldsymbol{G}^{oH}\boldsymbol{y}\right). \qquad (16)
$$

- the contrast

$$
\begin{aligned}
\nabla_{\boldsymbol{\chi}} J_{MAP} &= \nabla_{\boldsymbol{\chi}}\left(\frac{1}{2v_\xi}\|\boldsymbol{w} - \boldsymbol{X}\boldsymbol{E}\|_D^2 + \sum_k\sum_{\boldsymbol{r}}\frac{\left|\chi(\boldsymbol{r}) - m_k(\boldsymbol{r})\right|^2}{2v_k}\right) \\
&= v_\xi^{-1}\boldsymbol{X}\boldsymbol{E}^H\boldsymbol{E} + \sum_k v_k^{-1}\boldsymbol{X} + v_\xi^{-1}\boldsymbol{E}^H\boldsymbol{w} - \sum_k v_k^{-1}m_k, \qquad (17)
\end{aligned}
$$

where $\boldsymbol{E} = \boldsymbol{E}^{inc} + \boldsymbol{G}^c\boldsymbol{w}$. The MAP estimate $\widehat{\boldsymbol{\chi}}(\boldsymbol{r})$ is then the solution of $\nabla_{\boldsymbol{\chi}} J_{MAP} = 0$; that

is:

$$\widehat{\chi}(\boldsymbol{r}) = \left(v_\xi^{-1}\boldsymbol{E}^H\boldsymbol{E} + \sum_k v_k^{-1} \right)^{-1} \left(\sum_k v_k^{-1} m_k - v_\xi^{-1}\boldsymbol{w}\boldsymbol{E}^H \right). \tag{18}$$

- the hidden field

$$\begin{aligned} \widehat{z}(\boldsymbol{r}) &= \arg\max_{\{z(\boldsymbol{r})=1,2,\ldots,K\}} J_{MAP}(\boldsymbol{\chi}, \boldsymbol{w}, z(\boldsymbol{r}), \boldsymbol{\Theta}) \\ &= \arg\max_{\{z(\boldsymbol{r})=1,2,\ldots,K\}} \left(\frac{\log(v_k)}{2} + \frac{v_k^{-1}\left|\chi(\boldsymbol{r}) - m_k(\boldsymbol{r})\right|^2}{2} \right. \\ &\left. - \lambda \sum_{\boldsymbol{r}} \sum_{\boldsymbol{r}'} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}')) \right). \end{aligned} \tag{19}$$

- the variance of the observation noise

$$\begin{aligned} \frac{\partial J_{MAP}(\boldsymbol{\chi}, \boldsymbol{w}, z, \boldsymbol{\Theta})}{\partial v_\epsilon} &= \frac{\partial}{\partial v_\epsilon} \left(\frac{N_{\mathcal{S}}}{2} \log(v_\epsilon) + \frac{1}{2v_\epsilon} \|\boldsymbol{y} - \boldsymbol{G}^o\boldsymbol{w}\|_{\mathcal{S}}^2 \right. \\ &\left. + \frac{\eta_\epsilon}{v_\epsilon} + \left(\phi_\epsilon + 1\right) \log(v_\epsilon) \right), \\ &= \frac{1}{v_\epsilon} \left(\frac{N_{\mathcal{S}}}{2} + \left(\phi_\epsilon + 1\right) - \frac{\|\boldsymbol{y} - \boldsymbol{G}^o\boldsymbol{w}\|_{\mathcal{S}}^2}{2v_\epsilon} - \frac{\eta_\epsilon}{v\epsilon} \right). \end{aligned} \tag{20}$$

The MAP estimate $\widehat{v}_\epsilon$ is then the solution of $\frac{\partial J_{MAP}(\boldsymbol{\chi}, \boldsymbol{w}, z, \boldsymbol{\Theta})}{\partial v_\epsilon} = 0$; that is:

$$\widehat{v}_\epsilon = \left(\frac{N_{\mathcal{S}}}{2} + \phi_\epsilon + 1 \right)^{-1} \left(\frac{\|\boldsymbol{y} - \boldsymbol{G}^o\boldsymbol{w}\|_{\mathcal{S}}^2}{2} + \eta_\epsilon \right). \tag{21}$$

- the variance of the coupling noise

$$\begin{aligned} \frac{\partial J_{MAP}(\boldsymbol{\chi}, \boldsymbol{w}, z, \boldsymbol{\Theta})}{\partial v_\xi} &= \frac{\partial}{\partial v_\xi} \left(\frac{N_{\mathcal{D}} N_f N_\nu}{2} \log(v_\xi) + \frac{1}{2v_\xi} \right. \\ &\left. \times \|\boldsymbol{w} - \boldsymbol{X}\boldsymbol{E}\|_{\mathcal{D}}^2 + \frac{\eta_\xi}{v_\xi} + \left(\phi_\xi + 1\right) \log(v_\xi) \right), \\ &= \frac{1}{v_\xi} \left(\frac{N_{\mathcal{D}} N_f N_\nu}{2} + \left(\phi_\xi + 1\right) - \frac{\|\boldsymbol{w} - \boldsymbol{X}\boldsymbol{E}\|_{\mathcal{D}}^2}{2v_\xi} - \frac{\eta_\xi}{v_\xi} \right). \end{aligned} \tag{22}$$

The MAP estimate $\widehat{v}_\xi$ is then the solution of $\frac{\partial J_{MAP}(\boldsymbol{\chi}, \boldsymbol{w}, z, \boldsymbol{\Theta})}{\partial v_\xi} = 0$; that is:

$$\widehat{v_\xi} = \left(\frac{N_{\mathcal{D}} N_f N_v}{2} + \phi_\xi + 1 \right)^{-1} \left(\frac{\|\boldsymbol{w} - \boldsymbol{XE}\|^2_{\mathcal{D}}}{2} + \eta_\xi \right). \tag{23}$$

- the variance of the classes

$$\begin{aligned} \frac{\partial J_{MAP}(\boldsymbol{\chi}, \boldsymbol{w}, z, \Theta)}{\partial v_k} &= \frac{\partial}{\partial v_k} \left(\frac{\sum_k N_k \log(v_k)}{2} + \sum_k \sum_{\boldsymbol{r}} \frac{\left|\chi(\boldsymbol{r}) - m_k(\boldsymbol{r})\right|^2}{2v_k} \right. \\ &\quad \left. + \sum_k \left(\frac{\eta_0}{v_k} + \left(\phi_0 + 1\right) \log(v_k) \right) \right), \\ &= \frac{1}{v_k} \left(\frac{\sum_k N_k}{2} - \frac{\sum_k \sum_r \left|\chi(\boldsymbol{r}) - m_k(\boldsymbol{r})\right|^2}{2v_k} \right. \\ &\quad \left. - \frac{\sum_k \eta_0}{v_k} + \left(\phi_0 + 1\right) \right). \end{aligned} \tag{24}$$

The MAP estimate $\widehat{v_k}$ is then the solution of $\frac{\partial J_{MAP}(\boldsymbol{\chi}, \boldsymbol{w}, z, \Theta)}{\partial v_k} = 0$; that is:

$$\widehat{v_k} = \left(\frac{N_k}{2} + \phi_0 + 1 \right)^{-1} \left(\frac{\sum_k \sum_{\boldsymbol{r}} \left|\chi(\boldsymbol{r}) - m_k(\boldsymbol{r})\right|^2}{2} + k\eta_0 \right). \tag{25}$$

- the mean of the classes

$$\begin{aligned} \frac{\partial J_{MAP}(\boldsymbol{\chi}, \boldsymbol{w}, z, \Theta)}{\partial m_k} &= \frac{\partial}{\partial m_k} \left(\sum_k \left(\sum_{\boldsymbol{r}} \frac{\left|\chi(\boldsymbol{r}) - m_k(\boldsymbol{r})\right|^2}{2v_k} + \frac{\left|m_k - \mu_0\right|^2}{2\tau_0} \right) \right), \\ &= \sum_k v_k^{-1} \sum_{\boldsymbol{r}} \left(m_k(\boldsymbol{r}) - \chi(\boldsymbol{r}) \right) + \tau_0^{-1} \left(\sum_k m_k - \mu_0 \right). \end{aligned} \tag{26}$$

The MAP estimate $\widehat{m_k}$ is then the solution of $\frac{\partial J_{MAP}(\boldsymbol{\chi}, \boldsymbol{w}, z, \Theta)}{\partial m_k} = 0$; that is:

$$\widehat{m_k} = \left(\tau_0^{-1} + \sum_k v_k^{-1} \right)^{-1} \left(\sum_k v_k^{-1} \sum_{\boldsymbol{r}} \chi(\boldsymbol{r}) + \mu_0 \tau_0^{-1} \right). \tag{27}$$

Finally, the following algorithm summarizes the different steps:

$$
\left\{
\begin{array}{l}
\textbf{Alternated optimization algorithm} \\
\textbf{Initialization:} \\
\text{Set initial estimates } (\boldsymbol{\chi}^{(0)}, \boldsymbol{w}^{(0)}, \boldsymbol{z}^{(0)} \text{ and } \Theta^{(0)}). \\
\textbf{Iterations:} \\
\underline{\textbf{Repeat}} \\
\text{Estimate } \widehat{\boldsymbol{z}} \text{ using Equation (19).} \\
\text{Estimate } \widehat{v_\epsilon} \text{ using Equation (21).} \\
\text{Estimate } \widehat{v_\xi} \text{ using Equation (23).} \\
\text{Estimate } \widehat{v_k} \text{ using Equation (25).} \\
\text{Estimate } \widehat{m_k} \text{ using Equation (27).} \\
\text{Estimate } \widehat{\boldsymbol{w}} \text{ using Equation (16).} \\
\text{Estimate } \widehat{\boldsymbol{\chi}} \text{ using Equation (18).} \\
\underline{\textbf{Until convergence}}.
\end{array}
\right. \tag{28}
$$

One of the major difficulties in finding the joint estimator is the mutual dependence between the different variables that renders the computations often complex. Therefore, in the following, a tractable approximation of the joint posterior distribution is preferred, which is expected to give more accurate estimates.

3.4. Variational Bayesian Approximation (VBA)

The variational Bayesian technique (VBA, [50]) consists in finding the best separable distribution $q(\mathbf{x}) = \prod_i q(\mathbf{x}_i)$, where $\mathbf{x} = \{\boldsymbol{\chi}, \boldsymbol{w}, \boldsymbol{z}, \boldsymbol{\Theta}\}$, that approximates the true posterior distribution $p(\mathbf{x}|\mathbf{y})$, in the Kullback-Leibler (KL) sense.

It can be noted that minimization of the KL divergence is equivalent to maximization of the free negative energy derived from statistical physics:

$$
\mathcal{F}(\boldsymbol{q}) = \int_{\mathcal{R}^N} q(\mathbf{x}) \log\left(\frac{p(\mathbf{y}, \mathbf{x})}{q(\mathbf{x})}\right) \mathrm{d}\mathbf{x}. \tag{29}
$$

Hence, the objective of VBA can be summarized as:

$$
\text{find } \boldsymbol{q}^{opt} = \arg\max_{\boldsymbol{q}} \mathcal{F}(\boldsymbol{q}). \tag{30}
$$

Then, assuming separability $\left(q(\mathbf{x}) = \prod_i q(\mathbf{x}_i)\right)$, an analytic form of $\boldsymbol{q}$ can be obtained:

$$
q(\mathbf{x}_i) \propto \exp\left\{ \langle \log(p(\mathbf{x}, \mathbf{y})) \rangle_{\prod_{j \neq i} q(\mathbf{x}_j)} \right\}. \tag{31}
$$

Now, by considering the joint posterior distribution (10), a strong separation form is chosen:

$$q(\mathbf{x}) = q(v_\epsilon)q(v_\xi)\prod_i q(\chi_i)q(w_i)q(z_i)\prod_k q(m_k)q(v_k). \tag{32}$$

Then, the optimal form (equation (31)) is sought for, which leads to the following parametric distributions:

$$\begin{aligned}
q(\boldsymbol{w}) &= \mathcal{N}\left(\widetilde{\boldsymbol{m}}_w, \widetilde{\boldsymbol{V}}_w\right), \qquad q(\boldsymbol{\chi}) = \mathcal{N}\left(\widetilde{\boldsymbol{m}}_\chi, \widetilde{\boldsymbol{V}}_\chi\right),\\
q(m_k) &= \mathcal{N}\left(\widetilde{\mu}_k, \widetilde{\tau}_k\right), \qquad q(v_k) = \mathcal{IG}\left(\widetilde{\eta}_k, \widetilde{\phi}_k\right),\\
q(v_\epsilon) &= \mathcal{IG}\left(\widetilde{\eta}_\epsilon, \widetilde{\phi}_\epsilon\right), \qquad q(v_\xi) = \mathcal{IG}\left(\widetilde{\eta}_\xi, \widetilde{\phi}_\xi\right),\\
q(z) &= \widetilde{\zeta}_k \propto exp\left(\lambda \sum_{\boldsymbol{r}\in\mathcal{D}} \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} \widetilde{\zeta}(\boldsymbol{r}')\right),
\end{aligned} \tag{33}$$

where the expressions of the tilded shaping parameters are given below.

- At iteration step i, the parameters that concern the induced sources are:

$$\begin{aligned}
\widetilde{\boldsymbol{V}}_w^i &= \left[\mathbf{Diag}\left(\overline{v_\epsilon^{-1}}\Gamma^o + \overline{v_\xi^{-1}}\overline{\Gamma^{xc}}\right)\right]^{-1}\\
\widetilde{\boldsymbol{m}}_w^i &= \widetilde{\boldsymbol{m}}_w + \widetilde{\boldsymbol{V}}_w^i\Big[\overline{v_\epsilon^{-1}}\boldsymbol{G}^{o^H}\left(\boldsymbol{y} - \boldsymbol{G}^o\widetilde{\boldsymbol{m}}_w\right)\\
&\quad + \overline{v_\xi^{-1}}\left(\mathbf{I} - \boldsymbol{G}^{c^H}\widetilde{\boldsymbol{m}}_\chi^H\right)\left(\widetilde{\boldsymbol{m}}_\chi\left(\boldsymbol{E}^{inc} + \boldsymbol{G}^c\widetilde{\boldsymbol{m}}_w\right) - \widetilde{\boldsymbol{m}}_w\right)\\
&\quad + \overline{v_\xi^{-1}}\widetilde{\boldsymbol{V}}_\chi\left(\boldsymbol{E}^{inc} + \boldsymbol{G}^c\widetilde{\boldsymbol{m}}_w\right)\Big]
\end{aligned} \tag{34}$$

  where the tilded parameters without superscript are that obtained at iteration step $i-1$, overline denotes the expectation of the variable with respect to q (i.e. $\bar{u} = \mathrm{E}(u)_q$), superscript H indicates the conjugate transpose and Γ^{o} and Γ^{xc} are:

$$\begin{aligned}
\Gamma^o(\boldsymbol{r}) &= \sum_{\boldsymbol{r}'}\left|G^o(\boldsymbol{r}', \boldsymbol{r})\right|^2,\\
\overline{\Gamma^{xc}}(\boldsymbol{r}) &= 1 - 2\Re e\left(G^c(\boldsymbol{r}, \boldsymbol{r})\widetilde{m}_\chi(\boldsymbol{r})\right) + \left(\widetilde{m}_\chi^2(\boldsymbol{r}) + \widetilde{V}_\chi(\boldsymbol{r})\right)\sum_{\boldsymbol{r}'}\left|G^c(\boldsymbol{r}', \boldsymbol{r})\right|^2.
\end{aligned} \tag{35}$$

- The parameters that concern the contrast are:

$$\begin{aligned}
\widetilde{\boldsymbol{V}}_\chi^i &= \left[\mathbf{Diag}\left(\overline{v_\xi^{-1}}\,\overline{\boldsymbol{E}^2} + \overline{\boldsymbol{V}_\chi^{-1}}\right)\right]^{-1}\\
\widetilde{\boldsymbol{m}}_\chi^i &= \widetilde{\boldsymbol{V}}_\chi^i\left(\sum_k \overline{v_k^{-1}}\widetilde{\zeta}_k\widetilde{\boldsymbol{m}}_k + \overline{v_\xi^{-1}}\,\overline{\boldsymbol{w}\boldsymbol{E}^*}\right),
\end{aligned} \tag{36}$$

  where * denotes the complex conjugate and $\overline{\boldsymbol{w}\boldsymbol{E}^*}$ is the mean of the vector $\boldsymbol{w}\boldsymbol{E}^*$:

$$\overline{wE^*}(\boldsymbol{r}) = \sum_{N_f N_\nu} E^{inc*}(\boldsymbol{r})\widetilde{m}_w(\boldsymbol{r}) + \widetilde{m}_w(\boldsymbol{r}) \sum_{N_\mathcal{D}} G^{c*}(\boldsymbol{r}, \boldsymbol{r}')\widetilde{m}_w^*(\boldsymbol{r}') + G^{c*}(\boldsymbol{r}, \boldsymbol{r})\widetilde{V}_w(\boldsymbol{r}), \tag{37}$$

and $\overline{\boldsymbol{V}_\chi^{-1}}$, $\overline{\boldsymbol{E}^2}$ are diagonal matrices whose elements are such that:

$$\overline{V_\chi^{-1}}(\boldsymbol{r}, \boldsymbol{r}) = \sum_k \widetilde{\zeta}_k(\boldsymbol{r})\,\overline{v_k^{-1}}, \tag{38}$$

$$\overline{E^2}(\boldsymbol{r}) = \sum_{N_f N_\nu} \left| E^{inc}(\boldsymbol{r}) \right|^2 + 2\Re e\left(E^{inc*}(\boldsymbol{r}) G^c \widetilde{m}_w(\boldsymbol{r}) \right) + \left| \sum_{\boldsymbol{r}'} G^c(\boldsymbol{r}, \boldsymbol{r}')\widetilde{m}_w(\boldsymbol{r}) \right|^2 + \sum_{\boldsymbol{r}'} \left| G^c(\boldsymbol{r}, \boldsymbol{r}') \right|^2 \widetilde{V}_w(\boldsymbol{r}). \tag{39}$$

- The parameter that concerns the hidden field is:

$$\widetilde{\zeta}_k = \exp\left\{ -\frac{1}{2}\left(\Psi\left(\widetilde{\eta}_k\right) + \log\left(\widetilde{\phi}_k\right) + \overline{v_k^{-1}}\left(\left(\widetilde{m}_\chi(\boldsymbol{r}) - \widetilde{\mu}_k \right)^2 + \tilde{\tau}_k + \widetilde{V}_\chi(\boldsymbol{r}) \right) - \lambda \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \widetilde{\zeta}_k(\boldsymbol{r}') \right) \right\}, \tag{40}$$

where Ψ stands for the digamma function.

- The parameters that concern the variance of the observation noise are:

$$\begin{aligned} \widetilde{\phi}_\epsilon^i &= \phi_\epsilon + \frac{N_\mathcal{S}}{2} \\ \widetilde{\eta}_\epsilon^i &= \eta_\epsilon + \frac{1}{2}\left(\|\boldsymbol{y}\|_\mathcal{S}^2 + \|\boldsymbol{G}^o \widetilde{\boldsymbol{m}}_w\|_\mathcal{S}^2 - 2\Re e\left(\boldsymbol{y}^H \boldsymbol{G}^o \widetilde{\boldsymbol{m}}_w \right) + \|\boldsymbol{G}^{o2} \widetilde{\boldsymbol{v}}_w\|_1 \right). \end{aligned} \tag{41}$$

- The parameters that concern the variance of the coupling noise are:

$$\begin{aligned} \widetilde{\phi}_\xi^i &= \phi_\xi + \frac{N_f N_\nu N_\mathcal{D}}{2} \\ \widetilde{\eta}_\xi^i &= \eta_\xi + \frac{1}{2}\left(\|\widetilde{\boldsymbol{m}}_w\|_\mathcal{D}^2 + \|\widetilde{\boldsymbol{V}}_w\|_\mathcal{D}^2 + \| \left(\widetilde{\boldsymbol{m}}_\chi^2 + \widetilde{\boldsymbol{V}}_\chi \right) \overline{\boldsymbol{E}^2} \|_1 - 2\Re e\left(\widetilde{\boldsymbol{m}}_\chi^H \overline{\boldsymbol{wE}^*} \right) \right), \end{aligned} \tag{42}$$

with $\widetilde{\boldsymbol{m}}_\chi^2 = \widetilde{\boldsymbol{m}}_\chi^H \widetilde{\boldsymbol{m}}_\chi$.

- The parameters that concern the variance of the classes are:

$$\begin{aligned} \widetilde{\phi}_k^i &= \phi_0 + \frac{1}{2}\sum_{\boldsymbol{r}} \widetilde{\zeta}_k(\boldsymbol{r}) \\ \widetilde{\eta}_k^i &= \eta_0 + \frac{1}{2}\sum_{\boldsymbol{r}} \widetilde{\zeta}_k(\boldsymbol{r}) \left(\left| \widetilde{m}_\chi(\boldsymbol{r}) \right|^2 + \widetilde{V}_\chi(\boldsymbol{r}) + m_k^2 + \tilde{\tau}_k^2 \right). \end{aligned} \tag{43}$$

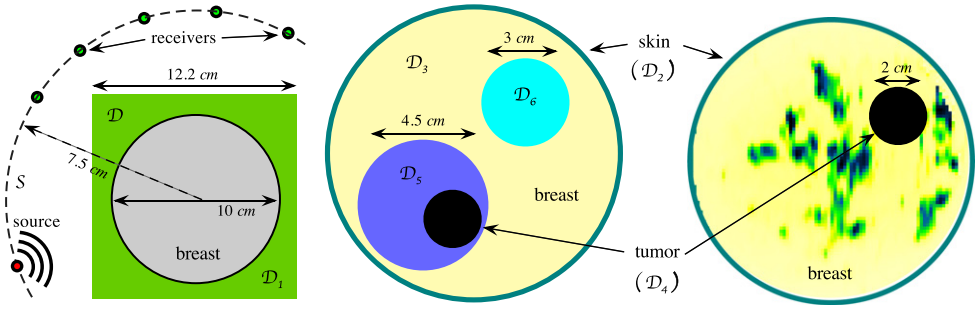


Figure 2. The measurement configuration (left) and the two models of a breast (middle: Model-1, right: Model-2).

- Finally, the parameters that concern the mean of the classes are:

$$\tilde{\tau}_k^i = \left(\tau_0^{-1} + \overline{v_k^{-1}} \sum_{\boldsymbol{r}} \tilde{\zeta}_k(\boldsymbol{r}) \right)^{-1}$$
$$\tilde{\mu}_k^i = \tilde{\tau}_k \left(\frac{\mu_0}{\tau_0} + \overline{v_k^{-1}} \sum_{\boldsymbol{r}} \tilde{\zeta}_k(\boldsymbol{r}) \widetilde{m}_\chi(\boldsymbol{r}) \right). \tag{44}$$

It can be noted that these parameters are mutually dependent and can only be computed in an iterative way.

3.5. The reconstruction algorithm

The initial values of the unknowns $\boldsymbol{w}^{(0)}$ and $\boldsymbol{\chi}^{(0)}$ are obtained by backpropagating the scattered field data from the measurement domain $\mathcal{S}$ onto the test domain $\mathcal{D}$ [51]. Afterwards, knowing $\boldsymbol{w}^{(0)}$ and $\boldsymbol{\chi}^{(0)}$, the initial value of the segmentation $\boldsymbol{z}^{(0)}$ is given by *K-means* clustering [52], with empirical estimators for the hyperparameters $\boldsymbol{\Theta}^{(0)}$. Since the sought contrast is complex-valued, first its real part is segmented and then the same segmentation is used to initialize the imaginary part. It can be noted that in the initial segmentation the number of classes is supposed to be known. However, this is not mandatory. When this number is unknown, it can be initialized at a high level. Then, it has been shown [53] that VBA converges to the optimal number of classes via two processes: i) if the number of pixels in a given class decreases to a low threshold value, then this class ends up vanishing; and, ii) if the means of two different classes are very close from each other, then these classes merge.

From an algorithmic point of view, the tilted parameters of equation (33) are iterated until convergence is reached. The latter is estimated empirically by looking to the evolution of contrast and hyper-parameters in the course of iterations. The VBA algorithm can then be summarized as follows:

Table 1. Electromagnetic parameters of the different media for the two models of breast.

	Medium	D1	D2	D3	D4	D5	D6
Model-1	ϵ_r	21.8, 35	35.7	6.12	55.3	42	21.8
	σ (Sm^{-1})	0.5	0.32	0.11	1.57	1.11	0.5
	Ø (cm)		10	9.6	2	3	4.5
Model-2	ϵ_r	21.8, 35	38	[2.46 − 60.6]	55.3		
	σ (Sm^{-1})	0.5	1.46	[0.01 − 2.28]	1.57		
	Ø (cm)		10		2		

VBA algorithm

Initialization:

Set initial estimates $(\boldsymbol{\chi}^{(0)}, \boldsymbol{w}^{(0)}, \boldsymbol{z}^{(0)}$ and $\Theta^{(0)})$.

Iterations:

Repeat

$$\begin{cases} \text{Estimate } q(\boldsymbol{z}): \text{ Update } \widetilde{\boldsymbol{\zeta}}_k \text{ using Equation (40).} \\ \text{Estimate } q(v_\epsilon): \text{ Update } \widetilde{\eta}_\epsilon \text{ and } \widetilde{\phi}_\epsilon \text{ using Equation (42).} \\ \text{Estimate } q(v_\xi): \text{ Update } \widetilde{\eta}_\xi \text{ and } \widetilde{\phi}_\xi \text{ using Equation (43).} \\ \text{Estimate } q(v_k): \text{ Update } \widetilde{\eta}_k \text{ and } \widetilde{\phi}_k \text{ using Equation (44).} \\ \text{Estimate } q(m_k): \text{ Update } \widetilde{\tau}_k \text{ and } \widetilde{\mu}_k \text{ using Equation (45).} \\ \text{Estimate } q(\boldsymbol{w}): \text{ Update } \widetilde{\boldsymbol{m}}_w \text{ and } \widetilde{\boldsymbol{V}}_w \text{ using Equation (35).} \\ \text{Estimate } q(\boldsymbol{\chi}): \text{ Update } \widetilde{\boldsymbol{m}}_\chi \text{ and } \widetilde{\boldsymbol{V}}_\chi \text{ using Equation (37).} \end{cases} \tag{45}$$

Until convergence.

3.6. VBA versus alternate optimization

A theoretical comparison between VBA and the alternated optimization algorithm shows that expressions of the parameters estimated with both methods are close to each other. For example, by comparing equations (16) and (35), it can be seen that the product that appears in the former equation appears also in the latter between the variance and the mean, with more or less similar components. However, the matrices that appear in the left term of equation (16) are not diagonal and are very large. Their inversion is not so easy. Besides, an efficient optimization for the hidden field appearing in equation (19) is not feasible since this problem is very hard to solve. In contrast, VBA yields more accurate solutions since the precision of the parameters is higher and it also yields an estimation of their variance. Furthermore, at a given iteration step, a variable estimated by means of VBA accounts for the uncertainties on the other variables because the estimation is done by alternate maximization of the joint MAP criterion with respect to the desired component while other variables are assumed to be constant, whereas this is not the case with alternate optimization.

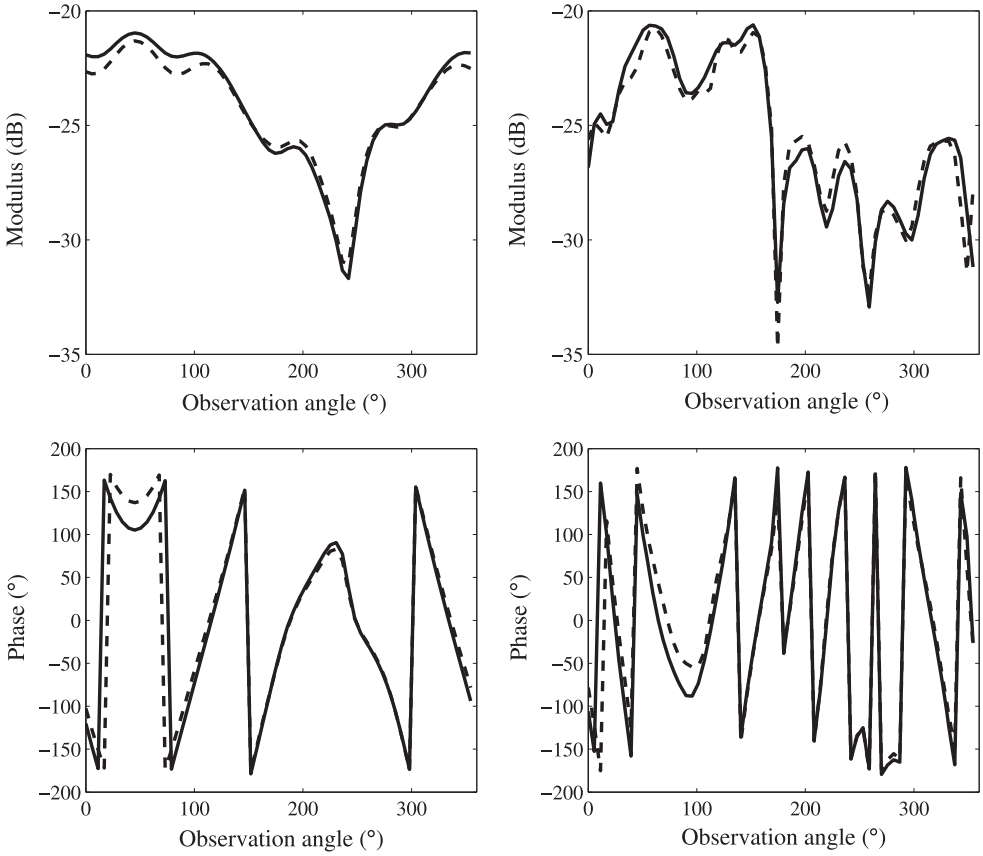


Figure 3. Modulus (up) and phase (down) of the scattered fields computed by means of the data (full line) and forward (dashed line) models for Model-1–1 and for different illumination angles and frequencies: 45° and 1.5 GHz (left), 90° and 3 GHz (right).

Table 2. The relative error between the scattered fields computed by means of the data and forward models for configuration Model-1–1.

Frequency (GHz)	0.5	1	1.5	2	2.5	3
Error %	0.64	1.26	2.31	4.41	7.38	4.18

4. Application and results

In this section, the VBA algorithm (45) is tested on two sets of synthetic data generated in different configurations involving two different numerical breast phantoms: a simple model made of five homogeneous media (denoted as the Model-1) and a more sophisticated one built up from a MRI scan of a real breast (denoted as the Model-2) [54]. Both models are tested with two different backgrounds whose relative permittivity are $\epsilon_r = 21.8$ (background 1) and $\epsilon_r = 35$ (background 2), respectively, and with the same conductivity $\sigma = 0.5$ Sm^{-1}, which leads to four configurations, which are denoted as Model-i-j, where i refers to the breast model ($i = 1, 2$) and j to the background ($j = 1, 2$).

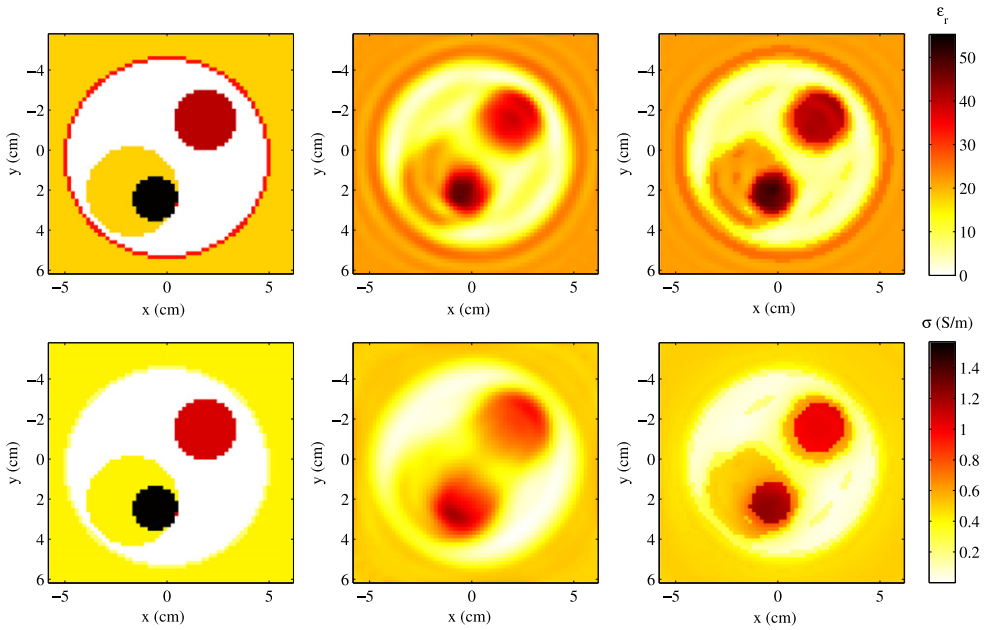


Figure 4. Permittivity (up) and conductivity (down) reconstructed for Model-1–1 by means of CSI (middle) and VBA (right) compared to the real model (left).

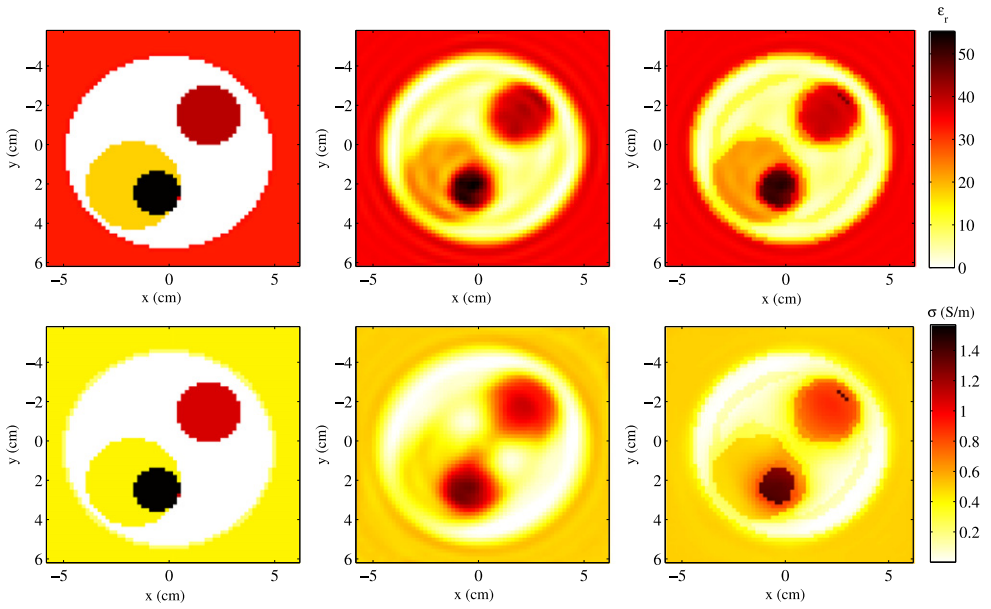


Figure 5. Permittivity (up) and conductivity (down) reconstructed for Model-1–2 by means of CSI (middle) and VBA (right) compared to the real model (left).

The results obtained by means of VBA are compared to those provided by the Contrast Source Inversion approach (CSI, [19]), which was developed in a deterministic framework

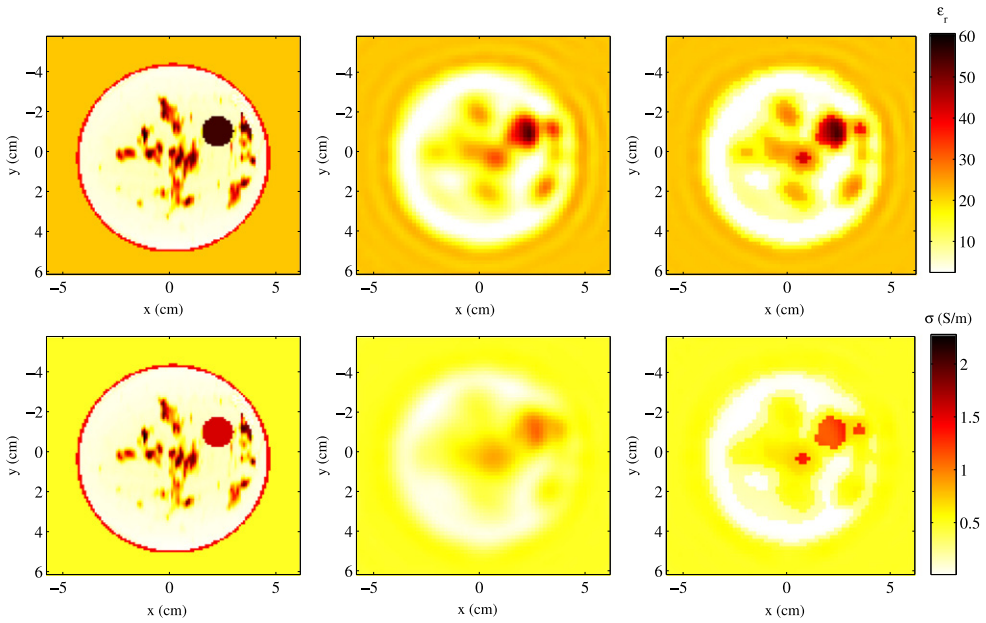


Figure 6. Permittivity (up) and conductivity (down) reconstructed for Model-2–1 by means of CSI (middle) and VBA (right) compared to the real model (left).

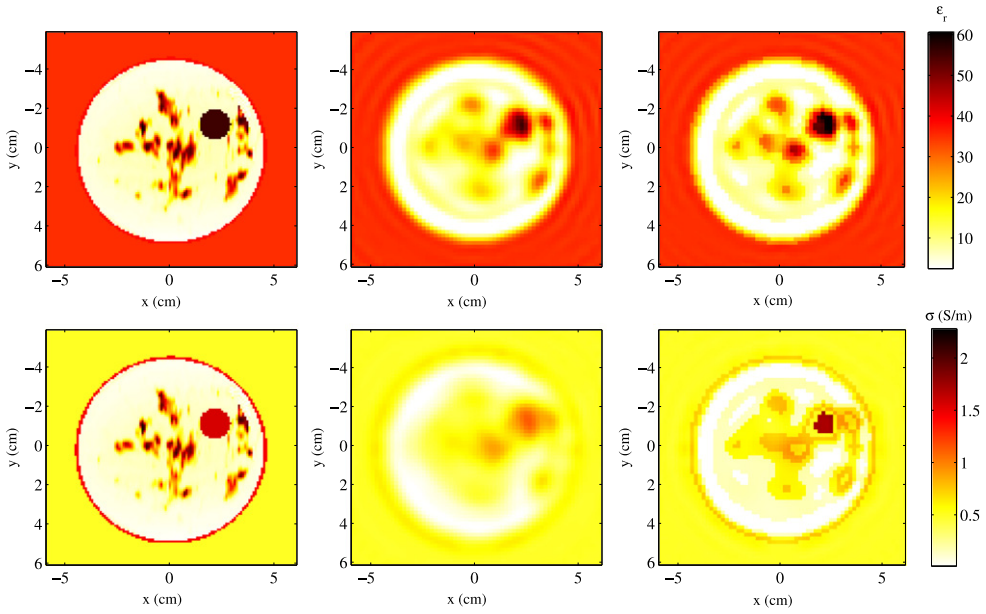


Figure 7. Permittivity (up) and conductivity (down) reconstructed for Model-2–2 by means of CSI (middle) and VBA (right) compared to the real model (left).

that consists of minimizing a cost functional that accounts for both the observation (1) and coupling (2) equations by alternately updating the induced sources $\boldsymbol{w}$ and the contrast χ with

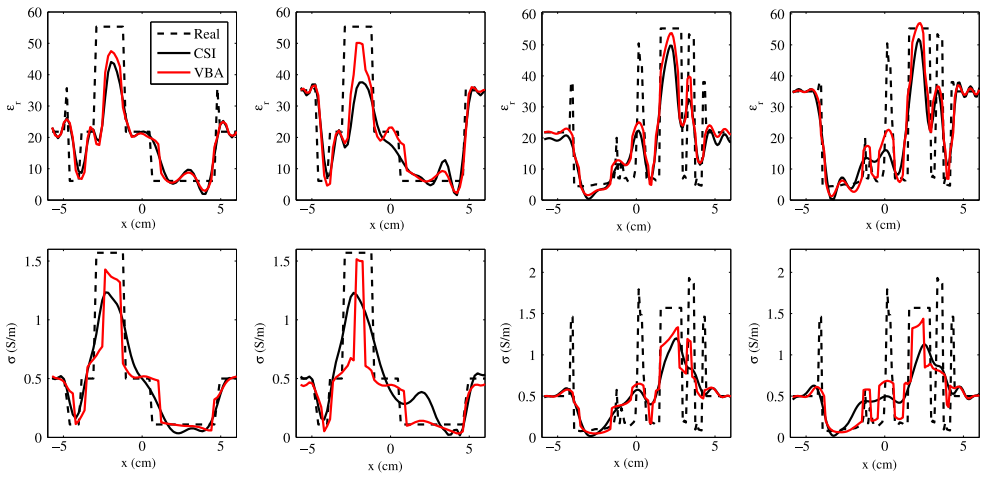


Figure 8. Permittivity (up) and conductivity (down) profiles reconstructed by means of CSI (black full line) and VBA (red full line) compared to real profiles (black dashed line) for Model-1–1 (1st column), Model-1–2 (2nd column), Model-2–1 (3rd column) and Model2–2 (4th column).

Table 3. Comparison of the different approaches in terms of Peak Signal to Noise Ratio (PSNR).

Model		Model-1–1	Model-1–2	Model-2–1	Model-2–2
PSNR (dB)	CSI	44.68	64.17	48.34	46.38
	VBA	69.86	76.19	52.29	51.59

a gradient-based method. The latter has already been used in the framework of breast cancer detection [55, 56] and interesting results have been obtained. Herein, we show the resolution improvement obtained with VBA, due to *a priori* information introduced in the inversion algorithm, as compared to CSI. It can be noted that, in addition to a better resolution, the advantages of VBA when compared to classical deterministic inversion methods is that it yields an estimate of the hyper-parameters of the model figures (11 and 12) and a segmentation of the image that is often even more important than the latter in many applications.

4.1. The experimental configuration

The models of a breast that are considered herein are depicted in (2). Both of them are supposed to be of circular cross-section and are located in a background medium $D1$. They are surrounded by a skin (domain $D2$) and are supposed to be affected by a tumor (domain $D4$). Concerning Model-1, the breast (domain $D3$) includes two glandular domains ($D5$ and $D6$) of circular cross-section. Model-2 is more sophisticated, it is built up from a MRI scan of a real breast and domain $D3$ is now made of a very heterogeneous medium. Table 1 gives the electromagnetic parameters of each medium for both models.

The source illuminates the breast from 64 various angular positions that are uniformly distributed around a 7.5-cm-radius circle that is centered at the origin and at 6 different frequencies in the band 0.5–3 GHz which yields a good compromise between resolution and

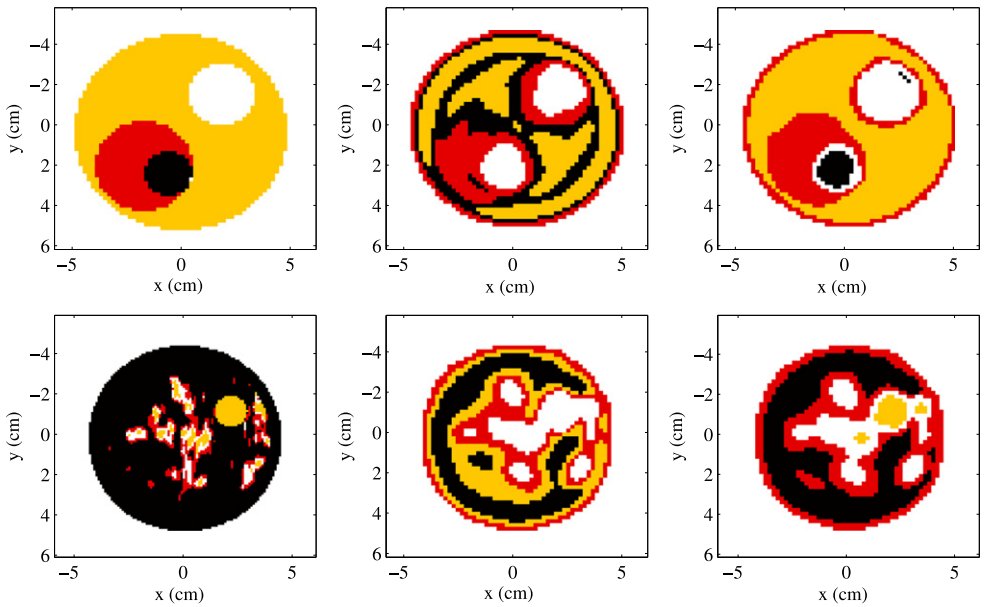


Figure 9. Segmentations obtained with VBA (right) for Model-1–2 (up) and Model-2–2 (down) compared to initial segmentations obtained with *K-means* clustering (middle) and to the actual image (up-left) or its *K-means* segmentation (down-left).

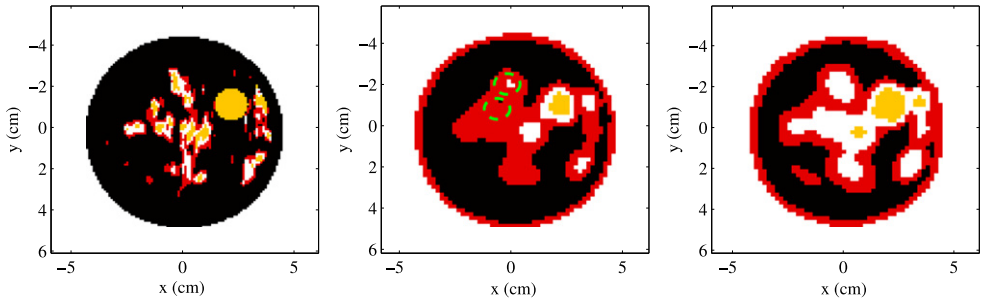


Figure 10. Segmentation of Model-2–2 with different values of the Potts parameter, i.e. $\lambda = 1$ (middle) and $\lambda = 10$ (right), compared to segmentation of the actual image obtained by *K-means* clustering (left).

penetration of the wave in the breast. For each frequency and illumination angle, 64 measurements of the scattered field are performed at angular positions that are also uniformly distributed around the same circle. It can be noted that in the inversion process the data corresponding to the different frequencies and illuminations are processed simultaneously in order to reduce the ill-posedness of the inverse problem.

4.2. Forward model validation

It can be noted that, for Model-1, in order to avoid committing a so-called 'inverse crime' in the sense of [57], which would consist in testing the inversion algorithm on data obtained by means of a model closely related to that used in the inversion, the synthetic data of the inverse

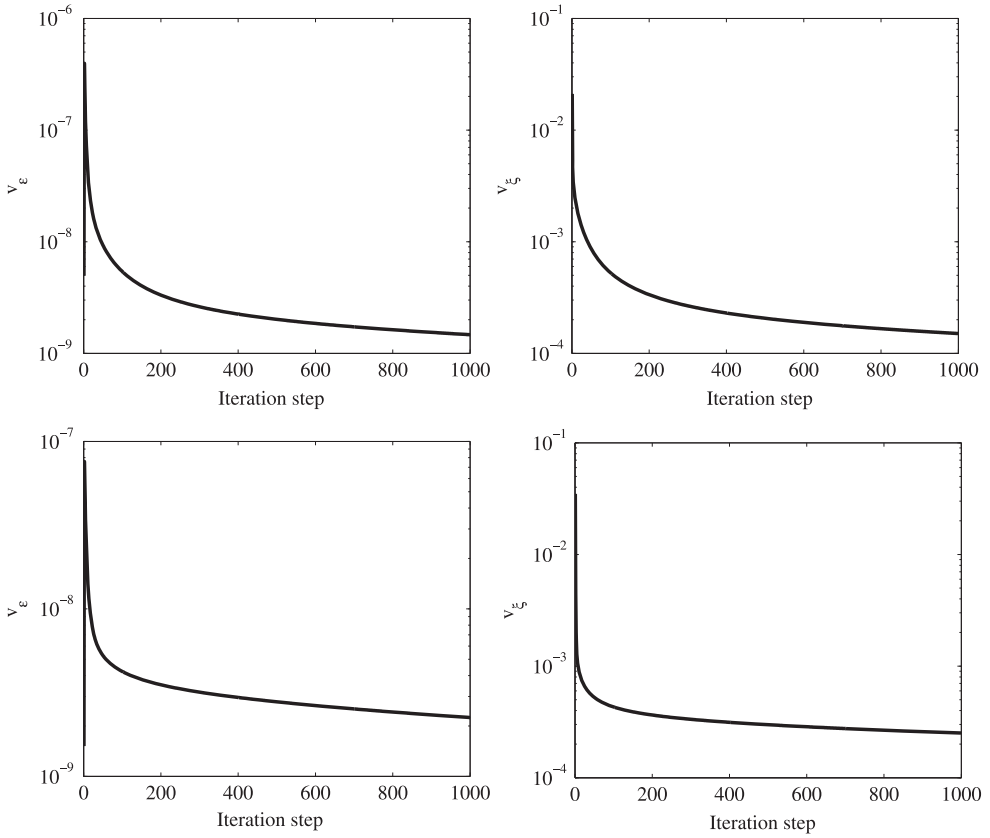


Figure 11. The evolution of the observation v_ε (left) and coupling v_ξ (right) error variances as functions of the iteration step for Model-1–1 (up) and Model-2–2 (down).

problem are computed by means of a forward model (the data model) that is rather different from the one described in section 2 (the forward model). Indeed, whereas in the latter the object under test is the breast as a whole, in the former it consists only in the tumor and glandular medias; the breast and the background are then considered as a cylindrically stratified embedding medium and the Green's function is consequently modified. For inversion, the test domain $\mathcal{D}$ is a 12.16 cm sided square partitioned into 64×64 square pixels with side $\delta = 1.9$ mm. Figure 3 displays the scattered fields obtained by means of the data and forward models for configuration Model-1–1, when the breast is illuminated from two different illumination angles (45° and 90°) and at two different frequencies (1.5 GHz and 3 GHz), whereas table 2 displays the relative error between these fields at different frequencies. It can be observed that the scattered fields fit relatively well, which validates the forward model.

4.3. Reconstruction

(4), (5), (6) and (7) display maps of permittivity and conductivity reconstructed by means of CSI (after 500 iteration steps) and VBA (after 1,000 iteration steps) for Model-1–1, Model-

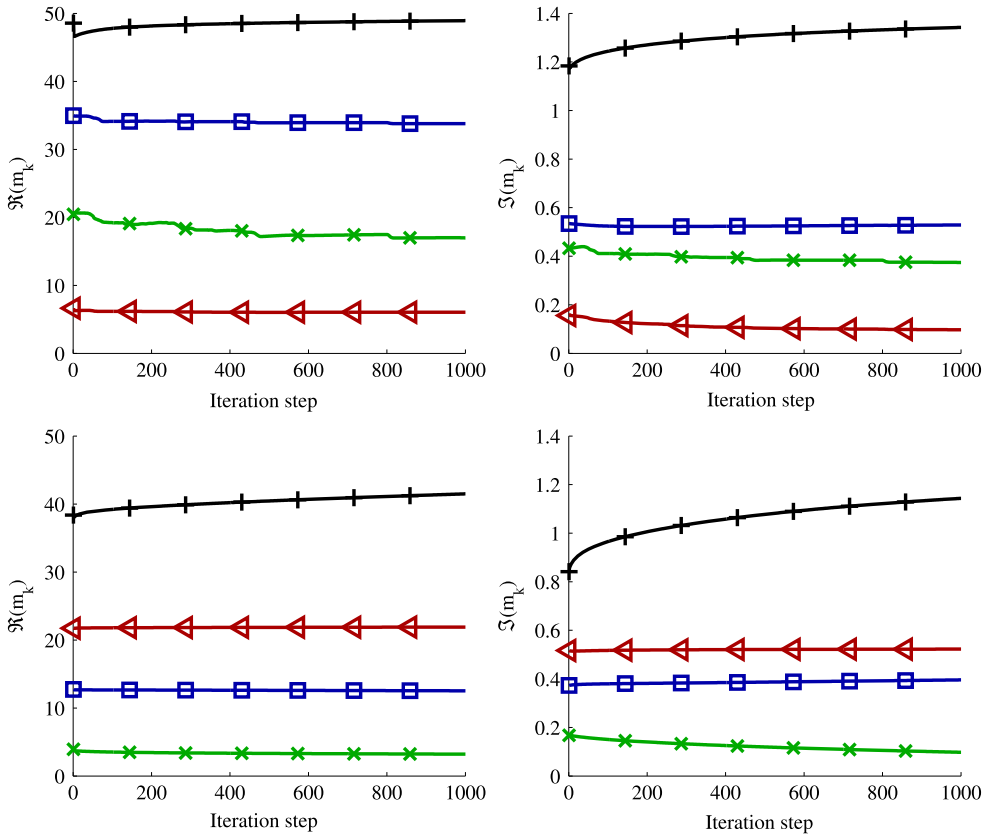


Figure 12. The real (left) and imaginary (right) parts of the means of the four classes for Model-1–2 (up) and Model-2–1 (down) as functions of the iteration step.

1–2, Model-2–1 and Model-2–2, respectively. Due to the prior information introduced in the inversion algorithm, the quality of reconstruction is significantly improved with VBA, especially for the conductivity, as compared to CSI, which gives good results but with an insufficient resolution, particularly for the tumor. In addition, the algorithm succeeds in general in retrieving homogeneous regions that correspond to the skin, the breast, the glandular media, and the tumor with accurate values of contrast, as can be observed in figure 8, which depicts the profiles reconstructed with both methods along an horizontal line crossing the center of the tumor and evidences the fact that VBA outperforms CSI, particularly with respect to the conductivity profile. This is confirmed by table 3, which displays the Peak Signal to Noise Ratio (PSNR) obtained with both methods for the four configurations. Concerning their computational performances, both methods are equivalent: indeed, they have almost the same memory requirements (VBA needs a little bit more space as there are more variables such as the hidden field and the hyper-parameters), whereas CSI needs less iteration steps to reach convergence than VBA, but consumes much more time per iteration than the latter, that is, 14.79 s for CSI (convergence time $\simeq$20 min) instead of 4.16 s for VBA (convergence time $\simeq$26 min). Figure 9 displays the segmentation of images obtained by means of VBA for Model-1–2 and Model-2–2, compared to the initial segmentation obtained with a *K-means* clustering algorithm. It can be noted that for Model-1–2 and Model-2–2, respectively, 496 and 531 pixels have undergone a change of class during the iterative

process, that is, between the middle and right images, which emphasizes the pros of VBA. In particular, with the latter, the tumor stands out in the segmentation, whereas it does not in a *K-means* segmentation of the image obtained with CSI. Finally, (10) emphasizes the influence of the Potts parameter on the segmentation results, as discussed in the subsection (3.1). It can be noted that with $\lambda = 1$, which is close to the critical value $\lambda_c = 1.1$ but smaller than the latter, the segmented regions are not completely homogeneous and may show some granulations, such as those circled in green, whereas with $\lambda = 10$, which is greater than the critical value, they are more uniform and homogeneous. This illustrates the properties of Potts fields: high values of λ yield larger and more homogeneous regions than low ones, but they can lead us to miss small regions of interest.

5. Conclusion

In this paper, a Bayesian inversion method is applied to microwave imaging for breast cancer detection. Microwave imaging is taken here as a non-linear inverse scattering problem, which is known to be ill-posed. Ill-posedness is overcome by introducing *a priori* information, which consists in the fact that the sought object is composed of a restricted known number of different materials distributed in compact regions. This information is taken into account by means of a Gauss–Markov–Potts prior modeling of the contrast distribution developed in a Bayesian estimation framework.

First, an alternate optimization scheme with respect to the different components of the joint maximum a posteriori criterion is investigated. However, in this framework optimization with respect to induced currents or contrast is not easy, not only because of their complex expressions but also because the estimation of a given variable at a given iteration does not account for the uncertainties on the other variables, which leads to a lack of precision and to high errors in the inversion results. Hence, an approximation of the joint posterior distribution is sought. This approximation is based upon an appropriate variational Bayesian technique (VBA) and consists in the best separable distribution that approximates the true posterior law in the Kullback–Leibler sense. This leads to an implicit parametric optimization scheme, which is solved iteratively with respect to the mutual dependency between the different variables. By applying VBA to two different configurations of a breast affected by a tumor, good results have been obtained concerning the retrieved complex-valued contrast and the geometry of the object. It has also been shown that this technique performs better than deterministic iterative techniques, such as the contrast source inversion method. Moreover, VBA has the advantage of providing not only an estimate of the contrast distribution but also its segmentation in regions and contrast parameters (means and variances) in each of the latter; in addition to the variances of the measurement and model errors, and the different hyper-parameters involved in the model. In the application considered herein, the latter results are even more important than the reconstruction itself.

Acknowledgments

The authors would like to acknowledge E Zastrow, S K Davis, M Lazebnik, F Kelcz, B D Van Veem and S C Hagness for the numerical breast phantom.

References

[1] Larsen L E and Jacobi J H 1985 *Medical Applications of Microwave Imaging* (New York: IEEE)
[2] Souvorov A E, Bulyshev A E, Semenov S Y, Svenson R H and Tatsis G P 2000 Two-dimensional computer analysis of a microwave flat antenna array for breast cancer tomography *IEEE Trans. Microwave Theory Tech.* **MTT-48** 1413–5
[3] Meaney P M, Fanning M W, Li D, Poplack S P and Paulsen K D 2000 A clinical prototype for active microwave imaging of the breast *IEEE Trans. Microwave Theory Tech.* **MTT-48** 1841–53
[4] Fear E C *et al* 2002 Conformal microwave imaging for breast cancer detection: localization of tumors in three dimensions *IEEE Trans. Biomed. Eng.* **BME-49** 812–22
[5] Li X, Davis S K, Hagness S C, van der Weide D W and van Veen B D 2004 Microwave imaging via spacetime beamforming: experimental investigation of tumor detection in multilayer breast phantoms *IEEE Trans. Microwave Theory Tech.* **MTT-52** 1856–65
[6] Klemm M, Craddock I J, Leendertz J A, Preece A and Benjamin R 2009 Radar-based breast cancer detection using a hemispherical antenna array: experimental results *IEEE Trans. Antennas Propag.* **AP-57** 1692–704
[7] Henriksson T, Joachimowicz N, Conessa C and Bolomey J-C 2010 Quantitative microwave imaging for breast cancer detection using a planar 2.45 GHz system, *IEEE Trans. Instrum. Meas.* **IM-59** 2691–9
[8] Hagness S C, Fear E C and Massa A 2012 Guest Editorial: Special Cluster on Microwave Medical Imaging *IEEE Antennas Wirel. Propag. Lett.* **11** 1592–7
[9] Brown M L, Houn F, Sickles E A and Kessler L G 1995 Screening mammography in community practice: positive predictive value of abnormal findings and yield of follow-up diagnostic procedures *Amer. J. Roentgen.* **165** 1373–7
[10] Huynh P T, Jarolimek A M and Daye S 1998 The false-negative mammogram *Radiographics* **18** 1137–54
[11] Elmore J G, Barton M B, Moceri V M, Plok S and Arena P J 1998 Ten year risk of false positive screening mammograms and clinical breast examinations *N. Engl. J. Med.* **338** 1089–96
[12] Joines W T, Jirtle R L, Rafal M D and Schaefer D J 1980 Microwave power absorption differences between normal and malignant tissue *Int. J. Radiation Oncology Biol. Phys.* **6** 681–7
[13] Devaney A J 1983 A computer simulation study of diffraction tomography *IEEE Trans. Biomed. Eng.* **BME-30** 377–86
[14] Paoloni F J 1987 Implementation of microwave diffraction tomography for measurement of dielectric constant distribution *IEE Proc. H Microwaves Antennas Propag.* **134** 25–29
[15] Azimi M and Kak A C 1983 Distorsion in diffraction tomography caused by multiple scattering *IEEE Trans. Med. Imaging* **MI-2** 176–95
[16] Slaney M, Kak A C and Larsen L E 1984 Limitations of imaging with first-order diffraction tomography *IEEE Trans. Microwave Theory Tech.* **MTT-32** 860–74
[17] Joachimowicz N, Pichot C and Hugonin J P 1991 Inverse scattering: an iterative numerical method for electromagnetic imaging *IEEE Trans. Antennas Propag.* **AP-39** 1742–52
[18] Kleinman R E and van den Berg P M 1992 A modified gradient method for two-dimensional problems in tomography *J. Comput. Appl. Math.* **42** 17–35
[19] van den Berg P M and Kleinman R E 1997 A contrast source inversion method *Inverse Problems* **13** 1607–20
[20] Souriau L, Duchêne B, Lesselier D and Kleinman R E 1996 Modified gradient approach to inverse scattering for binary objects in stratified media *Inverse Prob.* **12** 463–81
[21] Abubakar A, van den Berg P M and Mallorqui J J 2002 Imaging of biomedical data using a multiplicative regularized contrast source inversion method *IEEE Trans. Microwave Theory Tech.* **MTT-50** 1761–71
[22] Gilmore C, Mojabi P and LoVetri J 2009 Imaging of biomedical data using a multiplicative regularized contrast source inversion method *IEEE Trans. Antennas Propag.* **AP-57** 2341–51
[23] Garnero L, Franchois A, Hugonin J-P, Pichot C and Joachimowicz N 1991 Microwave imaging-complex permittivity reconstruction-by simulated annealing *IEEE Trans. Microwave Theory Tech.* **MTT-39** 1801–7
[24] Pastorino M 2007 Stochastic optimization methods applied to microwave imaging: A review *IEEE Trans. Antennas Propag.* **AP-55** 538–48

[25] Rocca P, Benedetti M, Donelli M, Franceschini D and Massa A 2009 Evolutionary optimization as applied to inverse scattering problems *Inverse Problems* **25** 123003
[26] Qing A 2003 Electromagnetic inverse scattering of multiple two-dimensional perfectly conducting objects by the differential evolution strategy *IEEE Trans. Antennas Propag.* **AP-51** 1251–62
[27] Rocca P, Oliveri G and Massa A 2011 Differential evolution as applied to electromagnetics *IEEE Antennas Propag. Mag.* **53** 38–49
[28] Rahmat-Samii Y and Michielssen E 1999 *Electromagnetic Optimization by Genetic Algorithms* (New York: Wiley)
[29] Caorsi S, Massa A, Pastorino M and Rosani A 2004 Microwave medical imaging: Potentialities and limitations of a stochastic optimization technique *IEEE Trans. Microwave Theory Tech.* **MTT-52** 1909–16
[30] Randazzo A 2012 Swarm optimization methods in microwave imaging *International Journal of Microwave Science and Technology* **2012**
[31] Donelli M and Massa A 2005 Computational approach based on a particle swarm optimizer for microwave imaging of two-dimensional dielectric scatterers *IEEE Trans. Microwave Theory Tech.* **MTT-53** 1761–76
[32] Oliveri G, Rocca P and Massa A 2011 Bayesian-compressive-sampling-based inversion for imaging sparse scatterers *IEEE Trans. Geosci. Remote Sens.* **GRS-49** 3993–4006
[33] Poli L, Oliveri G and Massa A 2012 Microwave imaging within the first-order Born approximation by means of the contrast-field Bayesian compressive sensing *IEEE Trans. Antennas Propag.* **AP-60** 2865–792
[34] Poli L, Oliveri G, Rocca P and Massa A 2013 Bayesian compressive sensing approaches for the reconstruction of two-dimensional sparse scatterers under TE illuminations *IEEE Trans Geosci. Remote Sens.* **GRS-51** 2920–36
[35] Candès E J and Wakin M B 2008 An introduction to compressive sampling *IEEE Signal Process. Mag.* **25** 21–30
[36] Eldar Y C and Kutyniok G (ed) 2012 *Compressed Sensing: Theory and Applications* (Cambridge: Cambridge University Press)
[37] Lazebnik M *et al* 2007 A large-scale study of the ultrawideband microwave dielectric properties of normal, benign and malignant breast tissues obtained from cancer surgeries *Phys. Med. Biol.* **52** 6093–115
[38] Lazebnik M *et al* 2007 A large-scale study of the ultrawideband microwave dielectric properties of normal breast tissue obtained from reduction surgeries *Phys. Med. Biol.* **52** 2637–56
[39] Lazebnik M *et al* 2007 Highly accurate Debye models for normal and malignant breast tissue dielectric properties at microwave frequencies *IEEE Microwave Wireless Comp. Lett.* **17** 822–4
[40] Mohammad-djafari A 2008 Gauss–Markov–Potts priors for images in computer tomography resulting to joint optimal reconstruction and segmentation *Int. J. Tomography Statistics* **11** 76–92
[41] Féron O and Mohammad-Djafari A 2005 Image fusion and unsupervised joint segmentation using a HMM and MCMC algorithms *J. Electron. Imaging* **14** 023014
[42] Robert C and Casella G 2000 *Monte Carlo Statistical Methods* (New York: Springer)
[43] MacKay D J C 1995 Ensemble learning and evidence maximization *Proc. Nips*
[44] Féron O, Duchêne B and Mohammad-Djafari A 2005 Microwave imaging of inhomogeneous objects made of a finite number of dielectric and conductive materials from experimental data *Inverse Problems* **21** 95–115
[45] Ayasso H, Duchêne B and Mohammad-Djafari A 2010 Bayesian inversion for optical diffraction tomography *J. Mod. Opt.* **57** 765–76
[46] Ayasso H, Duchene B and Mohammad-Djafari A 2012 Optical diffraction tomography within a variational Bayesian framework *Inverse Prob. Sci. Eng* **20** 59–73
[47] Gibson W C 2008 *The Method of Moments in Electromagnetics* (Boca Raton: Chapman & Hall/CRC)
[48] Hadamard J 2003 *Lectures on Cauchy's Problem in Linear Partial Differential Equations* (Mineola: Dover Publications)
[49] Beffara V and Duminil-Copin H 2012 The self-dual point of the two-dimensional random-cluster model is critical for $q \geqslant 1$ *Probab. Theory Relat. Fields* **153** 511–42
[50] Smídl V and Quinn A 2006 *The Variational Bayes Method in Signal Processing* (New York: Springer)

Bibliographie

[1] A. Maréchal : *Optique Géométrique Générale*, volume 5/24. Springer Berlin Heidelberg, 1956.

[2] K. Fuwa et B. L. Valle : The physical basis of analytical atomic absorption spectrometry. The pertinence of the Beer-Lambert law. *Analytical Chemistry*, 35(8):942–946, 1963.

[3] J.A. Richards : *Remote Sensing with Imaging Radar*. Springer, 2009.

[4] R. Lencrerot : *Outils de modélisation et d'imagerie pour un scanner micro-onde : Application au contrôle de la teneur en eau d'une colonne de sol*. Thèse de doctorat, Institut Fresnel, 2009.

[5] S. Caorsi, A. Massa, M. Pastorino et M. Donelli : Improved microwave imaging procedure for nondestructive evaluations of two-dimensional structures. *IEEE Transactions on Antennas and Propagation*, AP-52(6):1386–1397, 2004.

[6] G. Acciani, G. Brunetti, G. Fornarelli et C. Guaragnella : Model based microwave non destructive testing of pipes. *In 5th WSEAS/IASME International Conference on Systems Theory and Scientific Computation*, pages 120–125, Stevens Point, Wisconsin, USA, 2005. World Scientific and Engineering Academy and Society.

[7] L. Chommeloux : *Application de l'imagerie micro-onde active à la détection d'objets enfouis dans un milieu homogène ou stratifié*. Thèse de doctorat, Université Paris-Sud 11, 1987.

[8] S. Kidera, T. Sakamoto et T. Sato : High-resolution 3-D imaging algorithm with an envelope of modified spheres for UWB through-the-wall radars. *IEEE Transactions on Antennas and Propagation*, AP-57(11):3520–3529, 2009.

[9] N. Maaref, P. Millot, C. Pichot et O. Picon : FMCW ultra-wideband radar for through-the-wall detection of human beings. *IEEE Transactions on Geoscience and Remote Sensing*, 47(5):1297–1300, 2009.

[10] B. Boudamouz, P. Millot et C. Pichot : Through the wall radar imaging with MIMO beamforming processing-Simulation and experimental results. *American Journal of Remote Sensing*, 1(1):7–12, 2013.

[11] S. Caorsi, A. Massa et M. Pastorino : A crack identification microwave procedure based on a genetic algorithm for nondestructive testing. *IEEE Transactions on Antennas and Propagation*, AP-49(12):1812–1820, 2001.

[12] M. Benedetti, M. Donelli, D. Lesselier et A. Massa : A two-step inverse scattering procedure for the qualitative imaging of homogeneous cracks in known host media - Preliminary results. *IEEE Antennas and Wireless Propagation Letters*, 6:592–595, 2007.

[13] R. Zoughi : *Microwave Non-Destructive Testing and Evaluation Principles*. Springer, 2000.

[14] W. T. Joines, R. L. Jirtle, M. D. Rafal et D. J. Schaefer : Microwave power absorption differences between normal and malignant tissue. *International Journal of Radiation Oncology* Biology* Physics*, 6(6):681–687, 1980.

[15] J. Lewiner : *L'Imagerie du Corps Humain*. Physique et Médecine. Editions de physique, 1984.

[16] J. M. Elissalt : *Imagerie micoonde : reconstruction quantitative bidimensionnelle. Applications au domaine biomedical*. Thèse de doctorat, Université Paris-Sud 11, 1995.

[17] L. Jofre, M. S. Hawley, A. Broquetas, E. De los Reyes, M. Ferrando et A. R. Elias-Fuste : Medical imaging with a microwave tomographic scanner. *IEEE Transactions on Biomedical Engineering*, BE-37(3):303–312, 2002.

[18] Y. S. Semenov et D. R. Corfield : Microwave tomography for brain imaging : Feasibility assessment for stroke detection. *International Journal of Antennas and Propagation*, page ID 254830, 2008.

[19] T. Henriksson : *Contribution to quantitative microwave imaging techniques for biomedical applications*. Thèse de doctorat, Université Paris-Sud 11 - Mälardalen University, 2009.

[20] S. C Hagness, E. C. Fear et A. Massa : Guest editorial : special cluster on microwave medical imaging. *IEEE Antennas and Wireless Propagation Letters*, 11:1592–1597, 2012.

[21] E. C. Fear, P. M. Meaney et M. A. Stuchly : Microwaves for breast cancer detection ? *IEEE Potentials*, 22(1):12–18, 2003.

[22] P. M. Meaney, Q. Fang, M. W. Fanning, S. A Pendergrass, T. Raynolds, C. J. Fox et K. D. Paulsen : Microwave breast imaging with an under-determined reconstruction parameter mesh. *In IEEE International Symposium on Biomedical Imaging : Nano to Macro*, volume 2, pages 1369–1372, 2004.

[23] A. E. Souvorov, A. E. Bulyshev, S. Y. Semenov, R. H. Svenson et G. P. Tatsis : Two-dimensional computer analysis of a microwave flat antenna array for breast cancer tomography. *IEEE Transactions on Microwave Theory and Techniques*, MTT-48(8):1413–1415, 2000.

[24] P. M. Meaney, M. W. Fanning, D. Li, S. P. Poplack et K. D. Paulsen : A clinical prototype for active microwave imaging of the breast. *IEEE Transactions on Microwave Theory and Techniques*, MTT-48(11):1841–1853, 2000.

[25] E. C. Fear, X. Li, S. C. Hagness et M. A. Stuchly : Confocal microwave imaging for breast cancer detection : Localization of tumors in three dimensions. *IEEE Transactions on Biomedical Engineering*, BE-49(8):812–822, 2002.

[26] X. Li, S. K. Davis, S.C. Hagness, D. W. van der Weide et B.D. van Veen : Microwave imaging via spacetime beamforming : experimental investigation of tumor detection in multilayer breast phantoms. *IEEE Transactions on Microwave Theory and Techniques*, MTT-52(8):1856–1865, 2004.

[27] M. Klemm, I. J. Craddock, J. A. Leendertz, A. Preece et R. Benjamin : Radar-based breast cancer detection using a hemispherical antenna array : experimental results. *IEEE Transactions on Antennas and Propagation*, AP-57(6):1692–1704, 2009.

[28] T. Henriksson, N. Joachimowicz, C. Conessa et J.-Ch. Bolomey : Quantitative microwave imaging for breast cancer detection using a planar 2.45 GHz system. *IEEE Transactions on Instrumentation and Measurements*, IM-59(10):2691–2699, 2010.

[29] M. Persson, Z. Xuezhi et A. Fhager : Microwave imaging for medical applications. *In 5th European Conference on Antennas and Propagation (EUCAP)*, pages 3070–3072, 2011.

[30] L. E. Larsen et J. H. Jacobi : Microwave scattering parameter imagery of an isolated canine kidney. *Medical Physics*, 6(5):394–403, 1979.

[31] K. L. Carr, P. Cevasco, P. Dunlea et J. Shaeffer : Radiometric sensing : an adjuvant to mammography to determine breast biopsy. *In IEEE MTT-S International Microwave Symposium Digest*, volume 2, pages 929–932, 2000.

[32] S. Iudicello : *Microwave radiometry for breast cancer detection.* Thèse de doctorat, Universita' degli studi di Roma, 2009.

[33] R. A. Kruger, W. L. Kiser, D. R. Reinecke, G. A. Kruger et R. L. Eisenhart : Thermoacoustic computed tomography of the breast at 434 MHz. *In IEEE MTT-S International Microwave Symposium Digest*, volume 2, pages 591–594, 1999.

[34] S. Gabriel, R. W. Lau et C. Gabriel : The dielectric properties of biological tissues : II. Measurements in the frequency range 10 Hz to 20 GHz. *Physics in Medicine and Biology*, 41(11):2251, 1996.

[35] Xu. Li, E. J. Bond, B. D. van Veen et S. C. Hagness : An overview of ultra-wideband microwave imaging via space-time beamforming for early-stage breast-cancer detection. *IEEE Antennas and Propagation Magazine*, 47(1):19–34, 2005.

[36] X. Yao, G. Bin, X. Luzhou, L. Jian et P. Stoica : Multistatic adaptive microwave imaging for early breast cancer detection. *IEEE Transactions on Biomedical Engineering*, BE-53(8):1647–1657, 2006.

[37] T. Takenaka, H. Jia et T. Tanaka : Microwave imaging of electrical property distributions by a forward-backward time-stepping method. *Journal of Electromagnetic Waves and Applications*, 14(12):1609–1626, 2000.

[38] A. Fhager et M. Persson : Comparison of two image reconstruction algorithms for microwave tomography. *Radio Science*, 40(3), 2005.

[39] W. C. Chew : *Waves and Fields in Inhomogeneous Media (Electromagnetic Waves).* IEEE Computer Society Press, 1995.

[40] R.F. Harrington : *Field Computation by Moment Methods.* Wiley-IEEE Press, 1993.

[41] W. C. Gibson : *The Method of Moments in Electromagnetics.* Chapman & Hall/CRC, Boca Raton, 2008.

[42] M. Bertero et P. Boccacci : *Introduction to Inverse Problems in Imaging.* CRC press, 2010.

[43] J. J. Sakurai : *Modern Quantum Mechanics (Revised Edition).* Addison Wesley, 1993.

[44] R. Carminati et J. Greffet : Influence of dielectric contrast and topography on the near field scattered by an inhomogeneous surface. *Journal of the Optical Society of America A*, 12(12):2716–2725, 1995.

[45] M. F. Pascual, W. Zierau, T. A. Leskova et A. A. Maradudin : Surface electromagnetic waves in near-field optical scanning microscopy. *Optics Communications*, 155(4–6):351–360, 1998.

[46] M. Slaney, A. C. Kak et L. E. Larsen : Limitations of imaging with first-order diffraction tomography. *IEEE Transactions on Microwave Theory and Techniques*, MTT-32(8):860–874, 1984.

[47] A. J. Devaney et M. L. Oristaglio : Inversion procedure for inverse scattering within the distorted-wave Born approximation. *Physical Review Letters*, 51:237–240, 1983.

[48] R. J. Wombell et M. A. Fiddy : Inverse scattering within the distorted-wave Born approximation. *Inverse Problems*, 4(3):L23, 1988.

[49] A. J. Devaney : A filtered backpropagation algorithm for diffraction tomography. *Ultrasonic Imaging*, 4(4):336–350, 1982.

[50] F. Natterer et F. Wübbeling : *Mathematical Methods in Image Reconstruction*. Society for Industrial and Applied Mathematics, 2001.

[51] R. Pierri, R. Persico et R. Bernini : Information content of the Born field scattered by an embedded slab : multifrequency, multiview, and multifrequency–multiview cases. *Journal of the Optical Society of America A*, 16(10):2392–2399, 1999.

[52] A. Schatzberg et A. J. Devaney : Super-resolution in diffraction tomography. *Inverse Problems*, 8(1):149, 1992.

[53] G. Oliveri, P. Rocca et A. Massa : A Bayesian-compressive-sampling-based inversion for imaging sparse scatterers. *IEEE Transactions on Geoscience and Remote Sensing*, GRS-49(10):3993–4006, 2011.

[54] L. Poli, G. Oliveri et A. Massa : Microwave imaging within the first-order Born approximation by means of the contrast-field Bayesian compressive sensing. *IEEE Transactions on Antennas and Propagation*, AP-60(6):2865–2879, 2012.

[55] L. Poli, G. Oliveri, P. Rocca et A. Massa : Bayesian compressive sensing approaches for the reconstruction of two-dimensional sparse scatterers under TE illuminations. *IEEE Transactions on Geoscience and Remote Sensing*, GRS-51(5): 2920–2936, 2013.

[56] A. Massa, G. Oliveri, P. Ding et N. Anselmi : A CS-based approach for physical imaging. *In Inverse Problems from Theory to Applications Conference (IPTA 2014)*, pages 85–89, 2014.

[57] E. J. Candès et M. B. Wakin : An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008.

[58] Y. C. Eldar et G. Kutyniok : *Compressed Sensing : Theory and Applications*. Cambridge University Press, 2012.

[59] N. Joachimowicz, C. Pichot et J.-P. Hugonin : Inverse scattering : an iterative numerical method for electromagnetic imaging. *IEEE Transactions on Antennas and Propagation*, AP-39(12):1742–1752, 1991.

[60] W. C. Chew et Y. M. Wang : Reconstruction of two-dimensional permittivity distribution using the distorted Born iterative method. *IEEE Transactions on Medical Imaging*, MI-9(2):218–225, 1990.

[61] Y. Chun, Y. Mengqing, J. Stang, E. Bresslour, R. T. George, G. A. Ybarra, W. T. Joines et L. Qing-Huo : Active microwave imaging II : 3-D system prototype and image reconstruction from experimental data. *IEEE Transactions on Microwave Theory and Techniques*, MTT-56(4):991–1000, 2008.

[62] J. D. Shea, P. Kosmas, S. C. Hagness et B. D. van Veen : Three-dimensional microwave imaging of realistic numerical breast phantoms via a multiple-frequency inverse scattering technique. *Medical Physics*, 37(8):4210–4226, 2010.

[63] S. Y. Semenov, A. E. Bulyshev, A. E. Souvorov, A. G. Nazarov, Y. E. Sizov, R. H. Svenson, V. G. Posukh, A. Pavlovsky, P. N. Repin et G. P. Tatsis : Three-dimensional microwave tomography : experimental imaging of phantoms and biological objects. *IEEE Transactions on Microwave Theory and Techniques*, MTT-48(6):1071–1074, 2000.

[64] C. Estatico, G. Bozza, A. Massa, M. Pastorino et A. Randazzo : A two-step iterative inexact-Newton method for electromagnetic imaging of dielectric structures from real data. *Inverse problems*, 21(6):S81, 2005.

[65] P. N. Mojabi et J. LoVetri : Microwave biomedical imaging using the multiplicative regularized Gauss–Newton inversion. *IEEE Antennas and Wireless Propagation Letters*, 8:645–648, 2009.

[66] A. Franchois et Ch. Pichot : Microwave imaging-complex permittivity reconstruction with a Levenberg-Marquardt method. *IEEE Transactions on Antennas and Propagation*, AP-45(2):203–215, 1997.

[67] R. E. Kleinman et P. M. van den Berg : A modified gradient method for two-dimensional problems in tomography. *Journal of Computational and Applied Mathematics*, 42(1):17–35, 1992.

[68] P. M. van den Berg et R. E. Kleinman : A contrast source inversion method. *Inverse Problems*, 13(6):1607–1620, 1997.

[69] P. A. Barrière : *Développement d'algorithmes d'inversion rapides et propositions relatives à la configuration du montage de mesures dans un contexte de tomographie micro-onde appliquée à la détection du cancer du sein.* Thèse de doctorat, Ecole Centrale de Nantes, 2008.

[70] N. Ozmen-Eryilmaz et K. W. van Dongen : A contrast source inversion method for breast cancer detection. *Journal of the Acoustical Society of America*, 19(1): 075003, 2013.

[71] P. M. van den Berg, K. F. Haak et R. E. Kleinman : Contrast source inversion method using multi-frequency data. *IEEE Antennas and Propagation Society International Symposium*, 2:710–713, 1998.

[72] P. M. van den Berg, A. van Broekhoven et A. Abubakar : Extented contrast source inversion. *Inverse problems*, 15(5):1325–1344, 1999.

[73] P. M. van den Berg et A. Abubakar : Contrast source inversion method : state of art. *Journal of Electromagnetic Waves and Applications*, 15(11):1503–1505, 2001.

[74] H. Carfantan : *Approche bayésienne pour un problème inverse non linéaire en imagerie à ondes diffractées.* Thèse de doctorat, Université Paris-Sud 11, Orsay, 1996.

[75] A. Baussard, K. Belkebir et D. Prémel : A Markovian regularization approach of modified gradient method for solving a two dimensional inverse scattering problem. *Journal of Electromagnetic Waves and Applications*, 17(7):989–1008, 2003.

[76] L. Garnero, A. Franchois, J.-P. Hugonin, C. Pichot et N. Joachimowicz : Microwave imaging-complex permittivity reconstruction by simulated annealing. *IEEE Transactions on Microwave Theory and Techniques*, MTT-39(11):1801–1807, 1991.

[77] P. Caorsi, G. L. Gragnani, S. Medicina, M. Pastorino et G. Zunino : Microwave imaging method using a simulated annealing approach. *IEEE Microwave and Guided Wave Letters*, 1(11):331–333, 1991.

[78] H.-P. P. Schwefel : *Evolution and Optimum Seeking : the Sixth Generation.* John Wiley & Sons, 1993.

[79] P. Rocca, M. Benedetti, M. Donelli, D. Franceschini et A. Massa : Evolutionary optimization as applied to inverse scattering problems. *Inverse Problems*, 25(12):123003, 2009.

[80] M. Pastorino, A. Massa et S. Caorsi : A microwave inverse scattering technique for image reconstruction based on a genetic algorithm. *IEEE Transactions on Instrumentation and Measurement*, IM-49(3):573–578, 2000.

[81] S. Caorsi, A. Massa, M. Pastorino et A. Rosani : Microwave medical imaging : potentialities and limitations of a stochastic optimization technique. *IEEE Transactions on Microwave Theory and Techniques*, MTT-52(8):1909–1916, 2004.

[82] A. Massa, M. Pastorino et A. Randazzo : Reconstruction of two-dimensional buried objects by a differential evolution method. *Inverse Problems*, 20(6):S135, 2004.

[83] P. Rocca, G. Oliveri et A. Massa : Differential evolution as applied to electromagnetics. *IEEE Antennas and Propagation Magazine*, 53(1):38–49, 2011.

[84] M. Donelli et A. Massa : Computational approach based on a particle swarm optimizer for microwave imaging of two-dimensional dielectric scatterers. *IEEE Transactions on Microwave Theory and Techniques*, MTT-53(5):1761–1776, 2005.

[85] A. Randazzo : Swarm optimization methods in microwave imaging. *International Journal of Microwave Science and Technology*, 2012:ID 491713, 2012.

[86] M. Dorigo, M. Birattari et T. Stutzle : Ant colony optimization. *IEEE Computational Intelligence Magazine*, 1(4):28–39, 2006.

[87] C. Solnon : *Optimisation par Colonies de Fourmis.* Hermes Science Publications, 2008.

[88] V. Pascazio et G. Ferraiuolo : Statistical regularization in linearized microwave imaging through MRF-based MAP estimation : hyperparameter estimation and image computation. *IEEE Transactions on Image Processing*, IP-12(5):572–582, 2003.

[89] O. Féron : *Champs de Markov cachés pour les problèmes inverses. Application à la fusion de données et à la reconstruction d'images en tomographie micro-onde.* Thèse de doctorat, Université Paris-Sud 11, 2006.

[90] H. Ayasso : *Une approche bayésienne de l'inversion. Application à l'imagerie de diffraction dans les domaines micro-onde et optique.* Thèse de doctorat, Université Paris-Sud 11, 2010.

[91] M. Lazebnik et *al.* : A large-scale study of the ultrawideband microwave dielectric properties of normal, benign and malignant breast tissues obtained from cancer surgeries. *Physics in Medicine and Biology*, 52(20):6093–115, 2007.

[92] M. Lazebnik et *al.* : A large-scale study of the ultrawideband microwave dielectric properties of normal breast tissue obtained from reduction surgeries. *Physics in Medicine and Biology*, 52(20):2637–56, 2007.

[93] M. Lazebnik, M. Okoniewski, J. H. Booske et S. C. Hagness : Highly accurate Debye models for normal and malignant breast tissue dielectric properties at microwave frequencies. *IEEE Microwave and Wireless Components Letters*, 17(12):822–824, 2007.

[94] A. Litman : Reconstruction by level sets of n-ary scattering obstacles. *Inverse Problems*, 21(6):S131, 2005.

[95] C. P. Robert et G. Casella : *Monte-Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag, New York, 2005.

[96] D. J. C. Mackay : *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[97] E. Zastrow, S. K. Davis, M. Lazebnik, F. Kelcz, B. D. van Veem et S. C. Hagness : Database of 3D grid-based numerical breast phantoms for use in computational electromagnetics simulations. University of Wisconsin-Madison, Online, 2008.

[98] D. Colton et R. Kress : *Inverse Acoustic and Electromagnetic Scattering Theory*. Springer, New York, 1992.

[99] M. J. Hagmann, O. P. Gandhi et C. H. Durney : Upper bound on cell size for moment-method solutions. *IEEE Transactions on Microwave Theory and Techniques*, MTT-25(10):831–832, 1977.

[100] J. Richmond : Scattering by a dielectric cylinder of arbitrary cross section shape. *IEEE Transactions on Antennas and Propagation*, AP-13(3):334–341, 1965.

[101] J. Richmond : TE-wave scattering by a dielectric cylinder of arbitrary cross-section shape. *IEEE Transactions on Antennas and Propagation*, AP-14(4):460–464, 1966.

[102] X. Millard : *The BCGS-FFT method for electromagnetic scattering from inhomogeneous objects of arbitrary shape embedded in a planarly layered medium*. Thèse de doctorat, Duke University, 2002.

[103] J. Hadamard : *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. Yale University Press, New Haven, 1923.

[104] Ali Mohammad-Djafari : Gauss-Markov-Potts priors for images in computer tomography resulting to joint optimal reconstruction and segmentation. *International Journal of Tomography and Statistics (IJTS)*, 11:76–92, 2008.

[105] P. Wojciech : Modèles de Markov en traitement d'images. *Traitement du signal*, 20(3):255–278, 2003.

[106] V. Beffara et H. Duminil-Copin : The self-dual point of the two-dimensional random-cluster model is critical for $q \geq 1$. *Probability Theory and Related Fields*, 153(3-4):511–542, 2012.

[107] H. Ayasso, B. Duchêne et A. Mohammad-Djafari : Approches MCMC et variationnelle de l'inversion bayésienne en imagerie de diffraction. *In* J.-F. Giovannelli et J. Idier, éditeurs : *Méthodes d'inversion appliquées au traitement du signal et de l'image*, pages 223–247. Hermes Science Publications, 2013.

[108] C. Volkan et G. Ryan : Laplace approximation. Rapport technique, STAT 631 – ELEC 639 : Graphical Models, Rice University, 2008.

[109] F. Gersende : *Méthodes de Monte-Carlo et chaînes de Markov pour la simulation.* Thèse de doctorat, Habilitation à Diriger les Recherches, 2009.

[110] H. Ayasso, B. Duchêne et A. Mohammad-Djafari : Optical diffraction tomography within a variational Bayesian framework. *Inverse Problems in Science and Engineering*, 20(1):59–73, 2012.

[111] H. Ayasso, B. Duchêne et A. Mohammad-Djafari : A variational Bayesian approach for frequency diverse nonlinear microwave imaging. *In 19th IEEE International Conference on Image Processing (ICIP)*, pages 2069–2072, 2012.

[112] V. Smídl et A. Quinn : *The Variational Bayes Method in Signal Processing (Signals and Communication Technology).* Springer-Verlag, New York, 2005.

[113] M. J. Wainwright : *Graphical Models, Exponential Families, and Variational Inference.* New publishers, 2008.

[114] A. C. Likas et N. P. Galatsanos : A variational approach for Bayesian blind image deconvolution. *IEEE Transactions on Signal Processing*, SP-52(8):2222–2233, 2004.

[115] R. A. Choudrey : *Variational Methods for Bayesian Independent Component Analysis.* Thèse de doctorat, University of Oxford, 2002.

[116] N. Bali et A. Mohammad-Djafari : Bayesian approach with hidden Markov modeling and mean field approximation for hyperspectral data analysis. *IEEE Transactions on Signal Processing*, SP-17(2):217–225, 2005.

[117] J. Rustaji : *Variational Methods in Statistics.* Mathematics in Science and Engineering. Elsevier Science, 1976.

[118] Z. Ghahramani et M. I. Jordan : Factorial hidden Markov models. *Machine Learning*, 29(2-3):245–273, 1997.

[119] H. Attias : Independent factor analysis. *Neural Computation*, 11:803–851, 1999.

[120] J. Idier : *Approche bayésienne pour les problèmes inverses.* Traité IC2, Série traitement du signal et de l'image, Hermès, Paris, 2001.

[121] H. B. Callen : *Thermodynamics and an Introduction to Thermostatistics.* John Wiley and Sons (WIE), 1985.

[122] R. P. Feynman : *Statistical Mechanics : a Set of Lectures.* Frontiers in physics, 1972.

[123] N. Nasios et A. Bors : A variational approach for Bayesian blind image deconvolution. *IEEE transactions on Signal Processing*, SP-52(8):2222–2233, 2004.

[124] B. Duchêne, A. Joisel et M. Lambert : Nonlinear inversions of immersed objects using laboratory-controlled data. *Inverse Problems*, 20(6):81–98, 2001.

[125] J. B. MacQueen : Some methods for classification and analysis of multivariate observations. *In* L. M. Le Cam et J. Neyman, éditeurs : *5th Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967.

[126] A. Fraysse et T. Rodet : A measure-theoretic variational Bayesian algorithm for large dimensional problems. *SIAM Journal on Imaging Sciences*, 7(4):2591–2622, 2014.

[127] H. Ayasso, T. Rodet et A. Abergel : A variational Bayesian approach for unsupervised super-resolution using mixture models of point and smooth sources applied to astrophysical map-making. *Inverse Problems*, 28(12):125005, 2012.

[128] I. Molchanov et S. Zuyev : Steepest descent algorithms in a space of measures. *Statistics and Computing*, 12(2):115–123, 2002.

[129] J. Kivinen et M. K. Warmuth : Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 132(1):1–63, 1997.

[130] Y. Zheng : *Algorithmes bayésiens variationnels accélérés et applications aux problèmes inverses de grande taille*. Thèse de doctorat, Université Paris-Sud 11, 2014.

[131] A. Mohammad-Djafari : Bayesian approach with prior models which enforce sparsity in signal and image processing. *EURASIP Journal on Advances in Signal Processing*, Special issue on Sparse Signal Processing(1):1–19, 2012.

[132] C. Eyraud, A. Litman, A. Hérique et W. Kofman : Microwave imaging from experimental data within a Bayesian framework with realistic random noise. *Inverse problems*, 25(2):024005, 2009.

[133] S. Blanchard, D. Rousseau, D. Gindre et F. Chapeau-Blondeau : Constructive action of the speckle noise in a coherent imaging system. *Optics Letters*, 32(14):1983–1985, 2007.

[134] A. Levada, N. Mascarenhas et A. Tannus : Improving Potts MRF model parameter estimation in image analysis. *In 11th IEEE International Conference on Computational Science and Engineering (CSE '08)*, pages 211–218, 2008.

[135] J.-F. Giovannelli : Ising field parameter estimation from incomplete and noisy data. *In 18th IEEE International Conference on Image Processing (ICIP)*, pages 1853–1856, 2011.

[136] Q. Fang, P. M. Meaney et K. D. Paulsen : Microwave image reconstruction of tissue property dispersion characteristics utilizing multiple-frequency information. *IEEE Transactions on Microwave Theory and Techniques*, 52(8):1866–1875, 2004.

Table des figures

1.1 Propagation d'une onde éléctromagnétique. 14

1.2 Spectre éléctromagnétique. 14

1.3 Configuration géométrique en imagerie micro-onde bi-dimensionnelle (2D) 16

1.4 a) Schéma anatomique d'un sein (source : Société Canadienne du Cancer) et b) représentation 3D d'un sein atteint par une tumeur (source : Vivant Santé, Médecine et Sciences du Vivant). 18

1.5 Schéma de la réflexion sur les approches de l'imagerie micro-onde et les méthodes de résolution du problème inverse. 24

2.1 La configuration en imagerie micro-onde bi-dimensionnelle en polarisation transverse magnétique (2D-TM). 28

2.2 Discrétisation du domaine test en pixels élémentaires. 31

2.3 Configuration de mesure 36

2.4 Les différents fantômes de sein considérés : fantôme 1 (en haut à gauche), fantôme 2 (en haut à droite), fantôme 3 (en bas à gauche) et fantôme 4 (en bas à droite). 37

2.5 Validation du modèle direct pour le fantôme 1-0 pour deux fréquences et deux positions de source différentes : 45° et 1,5 GHz (haut) et 90° et 2,5 GHz (bas). La sortie du modèle (en noir discontinu) est comparée aux données (en rouge continu). 39

2.6 Validation du modèle direct pour le fantôme 1-3 pour deux fréquences et deux positions de source différentes : 45° et 1,5 GHz (haut) et 90° et 3 GHz (bas). La sortie du modèle (en noir discontinu) est comparée aux données (en rouge continu). 40

2.7 Validation du modèle direct pour le fantôme 2-2 pour une position de source de 135° et deux fréquences différentes : 0,5 GHz (haut) et 2,5 GHz (bas). La sortie du modèle (en noir discontinu) est comparée aux données (en rouge continu). 40

2.8 Validation du modèle direct pour le fantôme 3-2 pour deux fréquences et deux positions de source différentes : 45° et 1 GHz (haut) et 90° et 3 GHz (bas). La sortie du modèle (en noir discontinu) est comparée aux données (en rouge continu). 41

3.1 Thomas Bayes (1701-1761) (gauche, source : Wikipédia) et Pierre-Simon de Laplace (1749-1827) (droite, source : site de l'Université de Caen http://www.math.unicaen.fr/ reyssat/laplace/). 44

3.2 Représentation de l'association de la variable discrète cachée $\boldsymbol{z}$ à $\boldsymbol{\chi}$: a) les $\boldsymbol{\chi}$ sont indépendants et les $\boldsymbol{z}$ sont corrélés (processus markovien), b) les $\boldsymbol{\chi}$ et les $\boldsymbol{z}$ suivent un processus markovien. 46

3.3 Notion de voisinage dans une structure : les quatre pixels rouges forment le voisinage $\mathcal{V}$ du pixel gris. 47

3.4 Représentations 2D des modèles *a priori* proposés pour les images $\chi(\boldsymbol{r})$ segmentées en K classes. $z(\boldsymbol{r})$ représente ces classes (segmentation). A gauche, le modèle MGI où les pixels de chaque classe sont indépendants et, à droite, le modèle MGM où les pixels de chaque classe sont dépendants. 49

3.5 Représentation hiérarchique du modèle *a priori* : la distribution du contraste $\boldsymbol{\chi}$, sachant le champ caché $\boldsymbol{z}$, peut suivre un modèle de gaussiennes indépendantes ou un modèle de Gauss-Markov de paramètres $(\boldsymbol{m}, \boldsymbol{v})$. Le champ caché $\boldsymbol{z}$ suit une loi de Potts de paramètres (k, λ). Les hyper-paramètres ($\boldsymbol{m}$, $\boldsymbol{v}$, v_ϵ, v_ξ) sont choisis dans des familles de lois conjuguées et les meta-hyper-paramètres (μ_0, τ_0, η_0, ϕ_0, η_ϵ, ϕ_ϵ, η_ξ, ϕ_ξ) le sont de façon à obtenir des lois plates "non-informatives". 51

5.1 Exemples de comportement des parties réelles et imaginaires des moyennes des classes pour 4 fantômes différents. 76

5.2 Exemples de comportement des variances des bruits d'observation et de couplage pour 2 fantômes différents. 77

5.3 La permittivité ϵ_r et la conductivité σ reconstruites à l'aide de CSI (2^e colonne), BV (3^e colonne) et G-BV (4^e colonne) comparées aux valeurs réelles (1^{re} colonne). 78

5.4 La permittivité ϵ_r et la conductivité σ reconstruites à l'aide de CSI (2^e colonne), BV (3^e colonne) et G-BV (4^e colonne) comparées aux valeurs réelles (1^{re} colonne) pour les fantômes 2-1, 2-2 et 2-3. 79

5.5 La permittivité ϵ_r et la conductivité σ reconstruites à l'aide de CSI (2^e colonne), BV (3^e colonne) et G-BV (4^e colonne) comparées aux valeurs réelles (1^{re} colonne) pour les fantômes 3-1, 3-2 et 3-3. 80

5.6 La permittivité ϵ_r et la conductivité σ reconstruites à l'aide de CSI (2^e colonne), BV (3^e colonne) et G-BV (4^e colonne) comparées aux valeurs réelles (1^{re} colonne) pour les fantômes 4-1, 4-2 et 4-3. 81

5.7 Profils de permittivité reconstruits le long d'une ligne qui passe par le centre de la tumeur pour les différents fantômes (gauche : milieu d'adaptation 1, centre : milieu d'adaptation 2, droite : milieu d'adaptation 3). . 83

5.8 Profils de conductivité reconstruits le long d'une ligne qui passe par le centre de la tumeur pour les différents fantômes (gauche : milieu d'adaptation 1, centre : milieu d'adaptation 2, droite : milieu d'adaptation 3). . 84

5.9 Champs cachés obtenus par le biais des approches BV et G-BV comparés aux segmentations des images réelles et de celles obtenues par CSI par le biais de la méthode K-means (K=4) pour 4 fantômes différents. 85

5.10 Résultats de reconstruction du fantôme 3-3 après 1000 itérations par le biais de la méthode BV utilisant trois nombres de classes différents. . . . 86

5.11 Erreur relative de reconstruction des fantômes 3-1 et 4-3 en fonction du rapport signal sur bruit. 88

5.12 Résultats de reconstruction du fantôme 3-1 à l'aide de la méthode BV pour des données de champ diffracté bruitées avec des rapports signal à bruit de 5 dB, 10 dB et 20 dB. 89

5.13 Résultats de reconstruction du fantôme 4-3 à l'aide de la méthode BV pour des données de champ diffracté bruitées avec des rapports signal à bruit de 5 dB, 10 dB et 20 dB. 90

5.14 Champs cachés correspondant aux fantômes 3-3 (haut), 4-2 (milieu) et 4-3 (bas) obtenus avec deux valeurs différentes du paramètre de Potts : $\lambda = 1$ (milieu), $\lambda = 10$ (droite) comparés aux segmentations des objets réels (gauche) par la méthode K-means ($K = 4$). 91

5.15 Résultats de reconstruction du fantôme 4-2 en absence de tumeur par le biais des méthodes CSI, BV-MGI et BV-MGM comparés au fantôme d'origine. 92

6.1 Variation en fréquence de la permittivité relative et de la conductivité décrite par un modèle de Debye. 100

Liste des tableaux

1.1 Permittivité diélectrique et conductivité de divers organes humains à la fréquence de 3 GHz. 17

2.1 Les paramètres électromagnétiques des différents milieux et des différents fantômes 38

2.2 L'erreur relative entre les données et la sortie du modèle pour le fantôme 1-0 en fonction de la fréquence. 40

2.3 L'erreur relative entre les données et la sortie du modèle pour le fantôme 3-2 en fonction de la fréquence. 41

5.1 Récapitulatif des paramètres des configurations pour les 4 fantômes. . . . 77

5.2 Récapitulatif des paramètres des algorithmes BV et G-BV. 77

5.3 Comparaison des différentes approches en termes de temps de calcul et de PSNR pour 4 fantômes différents. 82

5.4 Le nombre de pixels mal classés par rapport à la segmentation de référence pour 4 fantômes différents. 82

5.5 Tableau résumant le nombre d'itérations et le temps de calcul nécessaires pour assurer la convergence des 3 algorithmes pour le cas du fantôme 1-1. 87

5.6 Tableau résumant le nombre d'itérations et le temps de calcul nécessaires pour assurer la convergence des 3 algorithmes pour le cas du fantôme 4-2. 87

5.7 Comparaison des approches CSI, BV-MGI et BV-MGM en termes de PSNR pour la reconstruction du fantôme 4-2 en absence de tumeur. . . . 92