An Overview of Nonlinear Diffraction Tomography within Bayesian Estimation Framework

Hervé Carfantan and Ali Mohammad-Djafari

Laboratoire des Signaux et Systèmes (CNRS/Supélec/UPS)
Plateau de Moulon, 91 192 Gif-sur-Yvette Cedex, France

Abstract. The Bayesian approach has been proven to give a common estimation structure to existing image reconstruction and restoration methods, in spite of their apparent diversity (Demoment 1989). The goal of this paper is to investigate diffraction tomography within the Bayesian estimation framework. A regularized solution to this ill-posed nonlinear inverse problem is defined as the maximum a posteriori estimate, introducing prior information on the object to reconstruct. Two equivalent formulations of this definition are available which lead to solution of a constrained or an unconstrained optimization problem to compute this solution. Different existing methods for solving this problem – such as *Born Iterative Method* (Wang and Chew 1989), *Newton-Kantorovitch method* (Joachimovicz et al. 1991), *Distorted Born Iterative method* (Chew and Wang 1990) and *Modified Gradient method* (Kleinman and van den Berg 1992) – are interpreted as algorithms to compute the defined solution. This common point of view allows an objective comparison between these methods, from the standpoint of their convergence properties and the solution they provide.

Introduction

Diffraction tomography consists in constructing an image representing the spatial variation of some physical properties of an inhomogeneous object (such as dielectric permittivity and conductivity for electro-magnetic waves), from a finite set of field data scattered by this object. This problem is intrinsically ill-posed and a satisfactory solution cannot be obtained from imperfect data without any introduction of a priori information on the object. The objectives of this paper are to define a regularized solution to this nonlinear inverse problem within the Bayesian estimation framework and to interpret some of the existing methods to solve this problem as algorithms to compute the defined solution.

First, we briefly present the direct model in a functional and in an algebraic framework. The algebraic framework allows a compact presentation and notably allows us to perceive strong similarities between some classical methods, which cannot be distinguished in the functional framework in which they have been proposed.

Then, we define a regularized solution within the Bayesian estimation framework. Bayes rule is a consistent way to combine information provided by

the data and prior information on the solution. In this paper, we use Markov Random Fields to model this a priori information. We define the solution as the maximum a posteriori estimate; so the solution's computation requires resolution of an optimization problem.

Then some of the existing methods to solve the diffraction tomography problem are interpreted and analyzed as algorithms to compute the defined regularized solution. Among these methods are *Born Iterative Method* (Wang and Chew 1989), *Newton-Kantorovitch method* (Joachimovicz et al. 1991), *Distorted Born Iterative method* (Chew and Wang 1990) and *Modified Gradient method* (Kleinman and van den Berg 1992). Three classes of methods are distinguished: the first one considers successive linearizations of the forward model, the second define the solution as the minimum of a joint criterion depending on the object and the field on the object, while methods of the third class minimize a criterion which only depends on the object.

Finally, an objective comparison between these different classes of methods and the solution they provide is proposed.

1   Problem Statement

We consider an inhomogeneous 2-D object, embedded in a known homogeneous medium, illuminated with a pure harmonic Transverse Magnetic (TM) plane wave. The object is characterized by its complex contrast function $x(\boldsymbol{r}) = k^2(\boldsymbol{r}) - k_0^2$, which is related to the dielectric permittivity $\epsilon(\boldsymbol{r})$ and the conductivity $\sigma(\boldsymbol{r})$ of the object by $k^2(\boldsymbol{r}) = \omega^2 \mu_0 \left(\epsilon(\boldsymbol{r}) + j\sigma(\boldsymbol{r})/\omega \right)$, k_0 is the wave number of the background homogeneous medium and $\boldsymbol{r}$ denotes a position in $\mathbb{R}^2$. The direct scattering problem is modeled by the coupled integral equations:

$$y(\boldsymbol{r}_i) = \iint_{D_O} \mathcal{G}(\boldsymbol{r}_i, \boldsymbol{r}')x(\boldsymbol{r}')\phi(\boldsymbol{r}')\mathrm{d}\boldsymbol{r}', \ \boldsymbol{r}_i \in D_M \ , \tag{1}$$

$$\phi(\boldsymbol{r}) = \phi_0(\boldsymbol{r}) + \iint_{D_O} \mathcal{G}(\boldsymbol{r}, \boldsymbol{r}')x(\boldsymbol{r}')\phi(\boldsymbol{r}')\mathrm{d}\boldsymbol{r}', \ \boldsymbol{r} \in D_O \ , \tag{2}$$

where $y(\boldsymbol{r}_i), \boldsymbol{r}_i \in D_M$ is the scattered field on a sensor located at $\boldsymbol{r}_i$ in the measurement area D_M, $\phi(\boldsymbol{r}), \boldsymbol{r} \in D_O$ and $\phi_0, \boldsymbol{r} \in D_O$ are the total and the incident field on the object area D_O, and $\mathcal{G}$ is the Green function for the homogeneous background medium.

From an algebraic viewpoint, discretization of (1–2) with a moment method (Howard and Kretzschmar 1986), leads to:

$$\boldsymbol{y} = \boldsymbol{G}_M \boldsymbol{X} \boldsymbol{\phi} \ , \tag{3}$$

$$\boldsymbol{\phi} = \boldsymbol{\phi}_0 + \boldsymbol{G}_O \boldsymbol{X} \boldsymbol{\phi} \ , \tag{4}$$

where $\boldsymbol{y} \in \mathbb{C}^{n_M}, \boldsymbol{\phi} \in \mathbb{C}^{n_O}, \boldsymbol{\phi}_0 \in \mathbb{C}^{n_O}$, $\boldsymbol{X}$ is a diagonal matrix $(n_O \times n_O)$ with the components of the vector $\boldsymbol{x} \in \mathbb{C}^{n_O}$ as diagonal elements, n_O is the number

of pixels of the discrete object and n_{M} is the number of measurement sensors. Note that these notations can be extended for emission from n_{S} different positions.

Formally, the total field ϕ on the object can be expressed from (4) and introduced in (3). It gives an explicit relation between the contrast and the data $y = \mathcal{A}(x)$ with:

$$\mathcal{A}(x) = G_{\mathrm{M}} X \left(I - G_{\mathrm{O}} X \right)^{-1} \phi_0 \ . \tag{5}$$

The inverse problem, which we are concerned with consists in determining the contrast x from a given finite set of noisy data y. Moreover, note that one can have $n_{\mathrm{O}} \gg n_{\mathrm{M}} \times n_{\mathrm{S}}$ (number of unknowns larger than number of data) so that the system of algebraic equations can be highly under-determined.

2   A Bayesian Approach for the Inverse Problem

The Bayesian inference is now a common way to handle ill-posed inverse problems in signal and image processing (Demoment 1989). We recall the main basis of the Bayesian framework before considering its application to nonlinear diffraction tomography.

2.1   General Framework

In a general Bayesian framework of parameter estimation from experimental data, the relation between the unknown parameters $x \in \mathbb{R}^n$ or $\mathbb{C}^n$ and the data $y \in \mathbb{R}^m$ or $\mathbb{C}^m$ can be written:

$$y = \mathcal{A}(x) + n \ ,$$

where $\mathcal{A}$ models the observation mechanism (direct problem) and n models errors on the measurements (measurement noise as well as modeling and discretization errors, which can often be considered additive on the data). Without particular knowledge on the errors, they are usually modeled by zero mean white Gaussian random variables, circular in case of complex quantities, with known variance σ_n^2 and independent of x. These assumptions are considered hereafter.

From this modeling, the likelihood function of the object for given data can be deduced:

$$p(y|x) = \left(\frac{1}{\pi \sigma_n^2} \right)^m \exp \left(-\frac{1}{\sigma_n^2} \| y - \mathcal{A}(x) \|^2 \right) \ .$$

The a priori state of knowledge, that is before any measurement is carried on, is taken into account through a probability law:

$$p(x) \propto \exp \left\{ -\mu \mathcal{U}(x) \right\} \ ,$$

where $\mathcal{U}$ has to be chosen to enforce desired properties on the solution.

Bayes rule allows to combine information supplied by the data and prior model in the a posteriori probability law of the parameters:

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{y})} \ ,$$

where, $p(\boldsymbol{y})$ is a normalizing coefficient.

From a strictly Bayesian viewpoint, the posterior law is the solution to the problem as it sums up all information available on the object. However, it is necessary to decide on a value to give to $\boldsymbol{x}$. Different estimators can be exhibited following the chosen decision rule, such as Maximum a posteriori (MAP), Maximum Marginal a posteriori (MMAP) or Posterior Mean (PM) estimators. Parameters which maximize the a posteriori law (MAP) are frequently chosen and this leads to an optimization problem. Indeed, the MAP estimate corresponds to the minimizer of the criterion:

$$\mathcal{J}(\boldsymbol{x}) = \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{x})\|^2 + \lambda\mathcal{U}(\boldsymbol{x}) \ ,$$

where $\lambda = \sigma_n^2\mu$ can be considered as a regularization parameter which balances between fidelity to the data and prior information.

2.2   Application to Nonlinear Diffraction Tomography

This general framework can be applied on many ways to the considered problem. We propose herafter two distinct formulations, depending on whether the contrast $\boldsymbol{x}$ has to be estimated from the data $\boldsymbol{y}$ or both the contrast $\boldsymbol{x}$ and the field $\boldsymbol{\phi}$ on the object have to be estimated.

First Formulation : Estimation of $\boldsymbol{x}$.   This formulation is straightforward. The solution is defined as the MAP estimate of $\boldsymbol{x}$:

$$\boldsymbol{x}_{\text{MAP}} = \arg\max_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{y}) \ .$$

From the explicit relation (5) it corresponds to the global minimizer of the criterion

$$\mathcal{J}^{\text{MAP}}(\boldsymbol{x}) = \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{x})\|^2 + \lambda\mathcal{U}(\boldsymbol{x}) \ , \tag{6}$$

with

$$\mathcal{A}(\boldsymbol{x}) = \boldsymbol{G}_{\text{M}}\boldsymbol{X}(\boldsymbol{I} - \boldsymbol{G}_{\text{O}}\boldsymbol{X})^{-1}\boldsymbol{\phi}_0 \ .$$

Second Formulation : Joint Estimation of x and ϕ. The solution is defined as the joint MAP estimate of x and ϕ:

$$(x, \phi)_{\text{MAP}} = \arg \max_{(x, \phi)} p(x, \phi | y) \ .$$

Thanks to Bayes rule, the a posteriori law can be written:

$$p(x, \phi | y) = \frac{p(y | x, \phi) p(\phi | x) p(x)}{p(y)} \ . \tag{7}$$

In this relation, $p(y)$ is a constant with respect to x and ϕ, so only the three numerator terms intervene in the MAP criterion:

– Using (3), with the considered error model, the first term can be written:

$$p(y | x, \phi) \propto \exp \left\{ -\frac{1}{\sigma_b^2} \|y - G_{\text{M}} X \phi\|^2 \right\} \ ;$$

– The second term corresponds to the probability law of ϕ for a known x. As ϕ is the total field on the object, it is uniquely determined for a given x by (4). Thus, if δ denotes the Dirac distribution:

$$p(\phi | x) = \delta(\phi - \phi_0 - G_{\text{O}} X \phi) \ ;$$

– $p(x)$ corresponds to the prior model on the object: $p(x) \propto \exp \left\{ -\mu \mathcal{U}(x) \right\}.$

Using these expressions, the posterior probability law can be written:

$$p(x, \phi | y) \propto \exp \left\{ -\frac{1}{\sigma_b^2} \|y - G_{\text{M}} X \phi\|^2 - \mu \mathcal{U}(x) \right\} \delta(\phi - \phi_0 - G_{\text{O}} X \phi) \ .$$

The MAP estimate of (x, ϕ) corresponds to the maximum of $p(x, \phi | y)$, i.e. it minimizes the criterion:

$$\mathcal{J}_c^{\text{MAP}}(x, \phi) = \|y - G_{\text{M}} X \phi\|^2 + \lambda \mathcal{U}(x) \ , \tag{8}$$

subject to the constraint:

$$\phi - \phi_0 - G_{\text{O}} X \phi = 0 \ . \tag{9}$$

2.3 Prior Models

Introduction of a priori information on the object is the basis of regularization. In the Bayesian framework, this information is modeled by a probability law $p(x)$ or equivalently by an energy function $\mathcal{U}(x)$.

We consider the class of Markov Random Fields (MRF) models, which is frequently used in image processing (Geman 1990) and allows the introduction of local correlations between the elements of an image. The energy function of a MRF can generally be written:

$$\mathcal{U}(x) = \sum_i \sum_{i \sim j} \rho(x_i - x_j) \ ,$$

with $\rho(t)$ a *potential* function, and $i \sim j$ stands for neighbors pixels i and j. Note that for complex fields, ρ operate separately on real and imaginary parts of x if they are considered independent.

A large choice of such potential functions has been proposed in the literature and certain of them are summarized in Table 1 and represented Fig. 1.

Table 1. Some potential functions and their characteristics

Name	Potential function	Characteristics
L_2 norm, Gaussian	$\rho(t) = t^2$	strictly convex, scale invariant
L_1 norm, Laplacian	$\rho(t) = \lvert t \rvert$	convex, scale invariant
L_p norm	$\rho(t) = \lvert t \rvert^p, 1 < p < 2$	strictly convex, scale invariant
Hubert function	$\rho(t) = \begin{cases} \lvert t \rvert^2 & \text{if } \lvert t \rvert \leq 1 \\ 2\lvert t \rvert - 1 & \text{if } \lvert t \rvert \geq 1 \end{cases}$	convex
Truncated Quadratic	$\rho(t) = \begin{cases} \lvert t \rvert^2 & \text{if } \lvert t \rvert \leq 1 \\ 1 & \text{if } \lvert t \rvert \geq 1 \end{cases}$	non convex, implicit line process

The L_2 norm corresponds to a first order Tikhonov regularization. This kind of regularization is of significant interest when the relation between the unknown and the data is linear because a linear explicit relation between the MAP estimate and the data is then available:

$$x_{\text{MAP}} = (A^\dagger A + \lambda W)^{-1} A^\dagger y \ ,$$

with W^{-1} the correlation matrix of the Gaussian process. However, such an interest decrease for nonlinear direct models, unless linear approximations are considered.

Nonconvex potential functions, like the truncated quadratic or other models including implicit or explicit line processes, can improve considerably the reconstruction of piecewise continuous images (Künsch 1994). However, as

local minima may appear in the energy function, choosing such a model generally largely increases the difficulty of computing the solution.

Convex potential functions, such as L_p norms or Hubert function, seem to be a reasonable choice for nonlinear inverse problems. They corresponds to a compromise between L_2 norm and nonconvex functions, as large variations of the field are less penalized than for the L_2 norm, but more than nonconvex functions. Using such models, allows better reconstruction of piecewise continuous images than Tikhonov regularization with no difficulty increase of the solution computation.
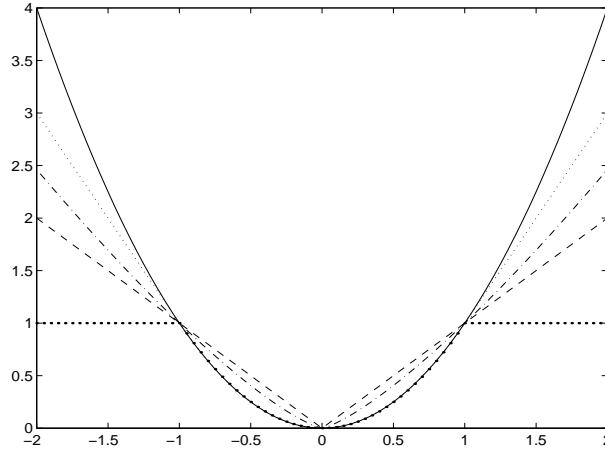

Fig. 1. 1D representation of some potential functions: L_2 $(-)$; L_1 $(--)$; L_p, $p = 1.3$ $(\cdots)$; Huber function, $T = 1$ $(-\cdot)$; Truncated Quadratic $T = 1(\cdot \cdot)$.

2.4   A Computational Challenge

The regularized solution has been defined as the contrast x which minimizes criterion (6) or as the contrast x and the total field ϕ that jointly minimize criterion (8) under constraint (9). These two distinct formulations are equivalent in the sense that they define the same solution (for x), but one may consider using different techniques to solve them.

Note that the Bayesian framework is not indispensable for defining the solution as the minimum of the criterion (6). Indeed, this criterion can be considered as a penalized least square criterion within a deterministic framework. However, the definition of the joint solution as the minimum of (8) under constraint (9) is not straightforward from deterministic arguments and other joint criteria are often proposed, which will be studied in § 4. Anyway, the Bayesian framework is not only useful to define a regularized solution to

an inverse problem but also offers probabilistic tools e.g. to characterize the solution (Tarantola 1987) and to estimate some additional parameters such as the regularization parameter (Idier et al. 1996).

Due to the non-linearity of the direct problem, it is easy to show that the criteria (6) and (8) are not convex functions. Thus, even if the prior information is modeled with a convex energy function, the criteria may have local minima. From simulation experiments, appearance of local minima is closely linked to a high contrast value, a limited number of measurements and a low signal-to-noise ratio. Thus computation of the solution may be a cumbersome task, especially in these *difficult* configurations. However, the problem seems to be less difficult in more favorable configurations.

In the multiplicity of methods proposed for solving nonlinear diffraction tomography problems, we tried to establish a classification, even if not exhaustive. Three types of methods has been emphasized which can be interpreted and analyzed in terms of algorithms to compute the defined regularized solution.

3 Successive Linearizations

Methods of the *first type* consider iteratively linear approximations of the direct model, which leads to solve successively linear inverse problems. Different methods of this type have been proposed in the literature to solve the nonlinear inverse problem of diffraction tomography. As the nonlinear inverse problem is ill-posed, each of the linear inverse problem is ill-posed and regularization has often been introduced to stabilize the solution of each linear problem.

Before comparing these different methods, we propose a successive linearizations algorithm specifically designed to minimize the criterion (6). Finally, we study the convergence properties of such techniques.

3.1 A Successive Linearizations Algorithm to Minimize $\mathcal{J}^{\mathrm{MAP}}$

At a given iteration n, a linear approximation of $\mathcal{A}$ has to be taken into account for x near x_n. The theoretical most coherent linear approximation of $\mathcal{A}$ near x_n is given by its first order Taylor series expansion:

$$\mathcal{A}(x) = \mathcal{A}(x_n) + \nabla_x \mathcal{A}(x_n)(x - x_n) + \mathcal{O}\left((x - x_n)^2\right)$$

(strictly speaking, one has to account for the Taylor series expansion of the real and imaginary parts of $\mathcal{A}$ to define such a relation for complex values functions). Let $A_n^{\mathrm{SLMAP}} = \nabla_x \mathcal{A}(x_n)$, calculus of A_n^{SLMAP} can be done easily. If $\phi_n = (I - G_0 X_n)^{-1}\phi_0$ denotes the field on the object for contrast x_n, and Φ_n its corresponding diagonal matrix, A_n^{SLMAP} can be written:

$$A_n^{\mathrm{SLMAP}} = G_{\mathrm{M}}\left[I + X_n(I - G_0 X_n)^{-1}G_0\right]\Phi_n \ .$$

Thus minimization of $\mathcal{J}^{\mathrm{MAP}}$ can be performed with successive linearizations of $\mathcal{A}$:

Initialize $n = 0, \boldsymbol{x}_0$.
Iterate for $n = 1, 2 \ldots$ until convergence towards a stationary point:
1. Compute the field on the object $\boldsymbol{\phi}_n$ and the matrix $\boldsymbol{A}_n^{\mathrm{SLMAP}}$ corresponding to the linear approximation of $\mathcal{A}$ near current solution $\boldsymbol{x}_n$.
2. Compute $\boldsymbol{x}_{n+1} = \arg\min_{\boldsymbol{x}} \mathcal{J}_n^{\mathrm{SLMAP}}(\boldsymbol{x})$ with
$$\mathcal{J}_n^{\mathrm{SLMAP}}(\boldsymbol{x}) = \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{x}_n) - \boldsymbol{A}_n^{\mathrm{SLMAP}}(\boldsymbol{x} - \boldsymbol{x}_n)\|^2 + \lambda \mathcal{U}(\boldsymbol{x}) \ .$$

Note that in such a scheme, for convex energy functions $\mathcal{U}$, all these criteria are convex functions and consequently have a unique global minimum.

3.2   The Born Iterative Method

The Born Iterative Method (BIM) has been introduced to circumvent the non-linearity, solving iteratively each of the coupled equations (1–2) (Wang and Chew 1989). Indeed, both integral equations are bilinear with respect to x and ϕ and solving each equation with respect to one of these variables leads to solve linear equations. Using algebraic notations, the BIM scheme can be summarized:

Initialize $\boldsymbol{\phi}_n = \boldsymbol{\phi}_0$ (Born approximation).
Iterate for $n = 1, 2 \ldots$ until convergence towards a stationary point:
1. Compute $\boldsymbol{x}_{n+1}$ for field $\boldsymbol{\phi}_n$ on the object, i.e. solve the linear inverse problem: $\boldsymbol{y} = \boldsymbol{G}_{\mathrm{M}} \boldsymbol{\Phi}_n \boldsymbol{x}$.
2. Compute the total field on the object $\boldsymbol{\phi}_{n+1}$, corresponding to contrast $\boldsymbol{x}_{n+1}$.

The linear approximation of the direct model which is accounted for in step 1. can be written:

$$\mathcal{A}(\boldsymbol{x}) \approx \mathcal{A}(\boldsymbol{x}_n) + \boldsymbol{A}_n^{\mathrm{BIM}}(\boldsymbol{x} - \boldsymbol{x}_n), \qquad \text{with} \qquad \boldsymbol{A}_n^{\mathrm{BIM}} = \boldsymbol{G}_{\mathrm{M}} \boldsymbol{\Phi}_n \ .$$

It appears in calculus of $\boldsymbol{A}_n^{\mathrm{SLMAP}}$, that $\boldsymbol{A}_n^{\mathrm{BIM}}$ corresponds to take a zeroth order approximation, with respect to $\delta\boldsymbol{x}$, of $[\boldsymbol{I} - \boldsymbol{G}_0(\boldsymbol{X}_n + \delta\boldsymbol{X})]^{-1}$. This term is approximated by $[\boldsymbol{I} - \boldsymbol{G}_0 \boldsymbol{X}_n]^{-1}$ so that the approximation of the BIM is coarser than the approximation of SLMAP.

In (Wang and Chew 1989), the linear inverse problem of step 1. is solved using a zero order Tikhonov regularization on $\boldsymbol{x}$. Hence, the original BIM is equivalent to SLMAP where $\boldsymbol{A}_n^{\mathrm{SLMAP}}$ is replaced by $\boldsymbol{A}_n^{\mathrm{BIM}}$, with $\mathcal{U}(\boldsymbol{x}) = \|\boldsymbol{x}\|^2$ and $\boldsymbol{x} = \boldsymbol{0}$ is taken as initial solution.

3.3   The Distorted Born Iterative Method

The Distorted Born Iterative Method (DBIM) (Chew and Wang 1990) is based on a scheme similar to the BIM, using distorted wave Born approximations. At each iteration, a known inhomogeneous background medium with contrast x_n is considered, with corresponding Green function $\mathcal{G}^n$ and incident field ϕ_n, and an additional inhomogeneity δx has to be computed.

Hereafter, $\phi_{\mathrm{O}}^n, \phi_{\mathrm{M}}^n$ denote the field ϕ_n on the object and on the measurement points respectively, G_{M}^n denotes a matrix corresponding to discretization of the Green function for inhomogeneous medium x_n. The DBIM scheme can then be summarized:

Initialize $x_0 = 0, G_{\mathrm{M}}^n = G_{\mathrm{M}}$ and the incident fields $\phi_{\mathrm{O}}^n = \phi_{\mathrm{O}}^0, \phi_{\mathrm{M}}^n = \phi_{\mathrm{M}}^0$ (Born approximation).
Iterate for $n = 1, 2 \ldots$ until convergence towards a stationary point:
1. Compute contrast $x_{n+1} = x_n + \delta x$ for the distorted wave Born approximation (field on the object ϕ_{O}^n and matrix G_{M}^n), i.e. solve the linear inverse problem: $y + \phi_{\mathrm{M}}^0 = \phi_{\mathrm{O}}^n + G_{\mathrm{M}}^n \Phi_{\mathrm{O}}^n \delta x$.
2. Compute incident fields $\phi_{\mathrm{O}}^{n+1}, \phi_{\mathrm{M}}^{n+1}$ and matrix G_{M}^{n+1} corresponding to the new inhomogeneous background x_{n+1}.

If discretization is performed with a moment method with pulse *basis* and *test functions*, as suggested in (Chew and Wang 1990), the update of G_{M}^n can be written:

$$G_{\mathrm{M}}^n = G_{\mathrm{M}} + G_{\mathrm{M}}^n X_n G_{\mathrm{O}} \ .$$

Note that this algebraic relation is not valid for other basis and test functions such as piecewise continuous ones, in which case the study of DBIM in an algebraic framework is not as easy. Using algebraic notations, it can be shown that, at each iteration, the first step accounts for a linear approximation of $\mathcal{A}$ which can be written:

$$\mathcal{A}(x) \approx \mathcal{A}(x_n) + A_n^{\mathrm{DBIM}}(x - x_n) \ , \quad \text{with} \quad A_n^{\mathrm{DBIM}} = G_{\mathrm{M}}(I - X_n G_{\mathrm{O}})^{-1} \Phi_n \ .$$

It can be shown that the approximation of the DBIM is identical to that of the SLMAP. Indeed,

$$(I - X_n G_{\mathrm{O}})^{-1} = I + X_n (I - G_{\mathrm{O}} X_n)^{-1} G_{\mathrm{O}} \ ,$$

which can be verified by calculating the product of these matrices.

In (Chew and Wang 1990), zero order Tikhonov regularization on δx has been introduced to solve the linear inverse problem of step 1, i.e. it accounts for an energy function $\mathcal{U}(x - x_n)$ instead of $\mathcal{U}(x)$ in the SLMAP scheme. Thus the solution given by this method does not correspond to a minimum of the MAP criterion (6).

3.4   The Newton Kantorovitch Method

The BIM and the DBIM are specific to the modeling of the forward problem with coupled equations such as (2-1). The Newton-Kantorovitch method is a more general method to solve nonlinear functional equations $y = \mathcal{A}(x)$ (Roger 1981). An iterative scheme is introduced, whose iteration consists in calculating variation δx added to x_n so that $y - \mathcal{A}(x_n) = \mathcal{A}(x_n + \delta x)$. As δx is assumed to be small, $\mathcal{A}(x_n + \delta x)$ is linearized for each iteration.

A Newton-Kantorovitch Method (NKM) has been proposed to solve the problem concerned (Joachimovicz et al. 1991). The linear approximation taken into account in (Joachimovicz et al. 1991) can be written:

$$\mathcal{A}(\boldsymbol{x}) \approx \mathcal{A}(\boldsymbol{x}_n) + \boldsymbol{A}_n^{\text{NKM}}(\boldsymbol{x} - \boldsymbol{x}_n) \ , \quad \text{with} \quad \boldsymbol{A}_n^{\text{NKM}} = \boldsymbol{G}_{\text{M}}(\boldsymbol{I} - \boldsymbol{X}_n \boldsymbol{G}_{\text{O}})^{-1} \boldsymbol{\Phi}_n \ ,$$

which is strictly equivalent to that of the DBIM. Let us recall that this relation can be established for the DBIM when the discretization is performed with a moment method with pulse basis and test functions, while it is still valid for other functions for the NKM.

In (Joachimovicz et al. 1991), the solution of each linear inverse problem is computed using zero order Tikhonov regularization on $\delta \boldsymbol{x}$. Thus the DBIM and the NKM are strictly equivalent.

3.5   Interpretation and Analysis of the Solutions

In terms of linear approximations, the SLMAP, the DBIM and the NKM are strictly equivalent, while the BIM accounts for a coarser approximation of $\mathcal{A}$ at each iteration.

The DBIM and the NKM are identical and only differ the from the SLMAP on the way according to which the regularization is introduced. Indeed, in the DBIM and the NKM, regularization is introduced to stabilize the solution of each linear inverse problem and not to regularize the whole nonlinear inverse problem. Regularization is performed on $\delta \boldsymbol{x}$ and does not take into account any prior model on $\boldsymbol{x}$; so the provided solution does not correspond to a minimum of $\mathcal{J}^{\text{MAP}}$. Note that for such a regularization, the algorithms seem to be very sensitive to the regularization parameter. In (Joachimovicz et al. 1991) a specific adjusting method has been proposed for this parameter in the NKM. In (Chew and Wang 1990) it has been observed that the DBIM can diverge more easily than the BIM. It seems to be contradictory with the fact that the BIM accounts for a coarser approximation of $\mathcal{A}$ than the DBIM, but it can be easily understood from the fact that the DBIM does not regularize the nonlinear inverse problem satisfactorily but each linear inverse problem independently.

On the other hand, the SLMAP is a successive linearizations algorithm designed to compute a regularized solution to the nonlinear inverse problem, defined as the minimum of $\mathcal{J}^{\text{MAP}}$.

The properties of the SLMAP in terms of minimization of $\mathcal{J}^{\mathrm{MAP}}$ can be studied. At each step, $\mathcal{J}^{\mathrm{MAP}}$ is approximated by a convex criterion $\mathcal{J}_n^{\mathrm{SLMAP}}$ with same value at $\boldsymbol{x}_n$ and – due to the first order Taylor series expansion – same slope at this point:

$$\mathcal{J}_n^{\mathrm{SLMAP}}(\boldsymbol{x}_n) = \mathcal{J}^{\mathrm{MAP}}(\boldsymbol{x}_n) \quad \text{and} \quad \nabla_{\boldsymbol{x}} \mathcal{J}_n^{\mathrm{SLMAP}}(\boldsymbol{x}_n) = \nabla_{\boldsymbol{x}} \mathcal{J}^{\mathrm{MAP}}(\boldsymbol{x}_n) \ .$$

The properties of such an algorithm are:

1. There exists no convergence guarantee and algorithm could diverge.
2. If it converges, a stationary point $\boldsymbol{x}_\infty$ is reached and

$$\nabla_{\boldsymbol{x}} \mathcal{J}_n^{\mathrm{SLMAP}}(\boldsymbol{x}_\infty) = \nabla_{\boldsymbol{x}} \mathcal{J}^{\mathrm{MAP}}(\boldsymbol{x}_\infty) = \boldsymbol{0} \ ,$$

   so this point corresponds to a local minimum of the criterion $\mathcal{J}^{\mathrm{MAP}}$.
3. Possible convergence and reached stationary point depend upon the initialization of the algorithm.

Note that second property is not valid for the linear approximation taken into account in the BIM. If the BIM converges towards a stationary point, this point is not guaranteed to be a minimum of $\mathcal{J}^{\mathrm{MAP}}$ because $\nabla_{\boldsymbol{x}} \mathcal{J}_n^{\mathrm{BIM}}(\boldsymbol{x}_n) \neq \nabla_{\boldsymbol{x}} \mathcal{J}^{\mathrm{MAP}}(\boldsymbol{x}_n)$. In this sense, the BIM is sub-optimal compared to the SLMAP (moreover, the SLMAP has been shown to converge more rapidly than the BIM (Carfantan and Djafari 1996)).

4   Minimization of a Joint Criterion

Some recently proposed methods – methods of the *second type* – define the solution as the minimum of criteria which account for errors on both coupled equations (3-4) with possible additional terms (Kleinman and van den Berg 1992), (Sabbagh and Lautzenheiser 1993), (Caorsi et al. 1993). In these methods, the solution is defined as the minimizer of a criterion, jointly on the contrast $\boldsymbol{x}$ and the field on the object $\boldsymbol{\phi}$, with the following generic form:

$$\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{\phi}) = \alpha_{\mathrm{M}} \|\boldsymbol{y} - \boldsymbol{G}_{\mathrm{M}} \boldsymbol{X} \boldsymbol{\phi}\|^2 + \alpha_{\mathrm{o}} \|\boldsymbol{\phi} - \boldsymbol{\phi}_0 - \boldsymbol{G}_{\mathrm{O}} \boldsymbol{X} \boldsymbol{\phi}\|^2 + \lambda \mathcal{U}(\boldsymbol{x}, \boldsymbol{\phi}) \ . \quad (10)$$

Such a criterion is very easy to understand intuitively: it corresponds to minimize jointly the errors on (3) and (4) and, as the problem is ill-posed, a penalization term on the unknowns is added to regularize it.

The proposed methods differ on several points:

− Criteria differ from value of parameters α_{M} and α_{o}. For example, these parameters are fixed to normalize the errors on both equations for $\boldsymbol{\phi} = \boldsymbol{0}$: $\alpha_{\mathrm{o}} = 1/\|\boldsymbol{\phi}_0\|^2$ and $\alpha_{\mathrm{M}} = 1/\|\boldsymbol{y}\|^2$ in (Kleinman and van den Berg 1992), while they are fixed to $1/2$ in (Sabbagh and Lautzenheiser 1993).

- There are differences on the regularization term. Originally, no regularization was introduced (Kleinman and van den Berg 1992), (Sabbagh and Lautzenheiser 1993). Then, it has been proposed to regularize both on x and ϕ, with $\mathcal{U}(x, \phi) = \lambda_x \|x\|^2 + \lambda_\phi \|\phi\|^2$ in (Barkeshli and Lautzenheizer 1994) and with $\mathcal{U}(x, \phi) = \lambda_\phi \|\Delta_1 \phi + (X + k_0^2 I)\phi\|^2 + \lambda_x \|D_1 x\|^2$ in (Caorsi et al. 1993), where Δ_1 and D_1 corresponds to discretization of Laplacian and gradient operators. Finally, it has been proposed to introduce a single regularization term on x, corresponding to a Markov random field with a line process in (Caorsi et al. 1995), and to a total variation penalization in (van den Berg and Kleinman 1995), which is equivalent to a L_1 regularization term.
- The methods also differ from the techniques used to compute the solution. Usual gradient type local minimization techniques has been used (Sabbagh and Lautzenheiser 1993), (Barkeshli and Lautzenheizer 1994) as well as local techniques specially designed for such a criterion (Kleinman and van den Berg 1992) and global minimization techniques such as Simulated Annealing (Caorsi et al. 1995).

4.1   Bayesian Interpretation

Recall that joint estimation of x and ϕ leads to problem $\mathcal{P}_c$: minimization of criterion (8) subject to constraint (9). The constraint can be equivalently written $\|\phi - \phi_0 - G_O X \phi\|^2 = 0$, so that the Lagrangian of $\mathcal{P}_c$ can be written:

$$L(x, \phi_0, \mu) = \|y - G_M X \phi\|^2 + \mu \|\phi - \phi_0 - G_O X \phi\|^2 + \lambda \mathcal{U}(x), \qquad (11)$$

with the scalar Lagrange multiplier μ. This Lagrangian looks like generic criterion (10), so that the adopted Bayesian framework gives a new way to look at it. It corresponds to the Lagrangian of the constraint optimization problems $\mathcal{P}_c$ in which the Lagrange multiplier is fixed intuitively ($\mu = \|y\|^2/\|\phi_0\|^2$ (Kleinman and van den Berg 1992) or $\mu = 1$ (Sabbagh and Lautzenheiser 1993)).

Moreover, this viewpoint gives indications for regularizing such a criterion with an energy function $\mathcal{U}(x)$. Using Bayes rule for the considered model of errors on measurements, we can see on (7) that there is no need to introduce prior model on ϕ.

Note that in (Caorsi et al. 1995) another Bayesian interpretation has been given for this criterion. It is shown that if additive gaussian error are assumed on both coupled equations (3–4), the joint MAP estimate of x and ϕ minimizes a criterion of form (10). However, it can be shown that such a criterion is obtained introducing zero mean circular Gaussian (conditionally to x) noise with covariance matrix $C_M = \sigma_2^2 I + \sigma_2^2 \left[(G_M X)^\dagger (G_M X) \right]^{-1}$ on the measurements. It seems to be a very strong and unjustified hypothesis as it considers a particular correlation between the measurement errors and the unknown object.

4.2 Analysis of the Solutions

The Lagrangian theory provides a link between solutions of constrained optimization problems and saddle-point of the corresponding Lagrangian. However, in the considered case where neither the criterion nor the constraint are convex, the only available property is the following:
If $((x, \phi), \mu)$ is a saddle-point of Lagrangian (11), (x, ϕ) is a solution to constrained optimization problem $\mathcal{P}_c$.
However, there is no guarantee that such a saddle-point exists.

Among the different methods proposed to minimize a criterion of form (10), none tries to reach a possible saddle-point of Lagrangian (11), but only a minimum of it for a fixed value of Lagrange multiplier. The given solution is then not necessarily a solution of $\mathcal{P}_c$. Moreover, note that criterion (10) is not convex, so it can have local minima. The solution computed with local minimization techniques will then possibly corresponds to a local minimum of the Lagrangian, for fixed Lagrange parameters.

It is possible that the given solution corresponds to a saddle-point of the Lagrangian. If the fixed Lagrange multiplier corresponds to a maximum of the Lagrangian, the solution is solution of $\mathcal{P}_c$. It can be shown that if (x, ϕ) is a local minimum of L, for μ fixed and that constraint is verified, then x corresponds to a local extremum of the unconstrained criterion $\mathcal{J}^{\text{MAP}}$. But if the constraint (9) is not verified, the solution cannot be characterized as easily.

Note that from this definition of the joint solution as the solution of $\mathcal{P}_c$, specific algorithms can be designed to compute this solution (Carfantan 1996).

5 Minimization of the MAP Criterion

From presentation of § 2, a natural idea to compute the defined solution – which corresponds to methods of the *third type* – is to minimize directly the MAP criterion (6).

Different methods have been proposed in the literature which defined the solution as the minimizer of the mean square error (MSE) between experimental and simulated data, possibly taking into account a regularization penalty term (*e.g.* (Garnero et al. 1991), (Xia et al. 1994)). However, an explicit formulation of criterion (6) using the explicit algebraic relation (5) of $\mathcal{A}$ has only been proposed recently (Carfantan and Djafari 1995). Note that it is not necessary to express such a relation to try to minimize the MSE and it is sufficient to be able to simulate the forward problem. However, one can take advantage of such an expression to design specific algorithms to minimize this MSE and the criterion (6).

Different optimization techniques have been used to compute the minimum of this criterion and will not be detailed hereafter: local techniques such

as conjugate gradient (Xia et al. 1994), (Lobel et al. 1996) or global ones such as Simulated Annealing (SA) (Garnero et al. 1991), Graduated non Convexity (GNC) (Carfantan and Djafari 1995), or a cheaper Iterative Conditional Mode (ICM) (Carfantan et al 1996).

Using any minimization technique to minimize $\mathcal{J}^{\mathrm{MAP}}$ guarantees the solution to be a minimum, possibly local, of this criterion. However, in *difficult configurations*, when this criterion has local minima, a global minimization technique may be used to compute the solution.

6   A Comparative Study

The presented classification of existing methods allows to better compare them. One can study the number of considered unknowns, the computation cost, the convergence properties and the robustness with respect to some parameters for each type of methods. No simulation results are shown in this paper and the reader can refer to (Carfantan 1996),(Carfantan and Djafari 1996) for more details. The main conclusions of this study is presented in the following.

6.1   Number of Unknown

In methods of both first and third types, the unknown is the contrast $x \in \mathbb{C}^{n_O}$ while in methods of second type, the contrast and the total field on the object for each incident field $\phi \in \mathbb{C}^{n_O \times n_S}$ have to be reconstructed. So, if the number of data is increased, considering more source positions, the number of unknown is increased as well in methods of the second type.

6.2   Computation Cost

In methods of the first type, evaluation of the criterion requires an order of $\mathcal{O}(n_O * n_M * n_S)$ complex operations. However, these methods require the update of some matrices between each iteration which includes resolution of the direct problem. For the BIM, the cost of these updates is of order $\mathcal{O}(n_O{}^3 + n_O{}^2 * n_S + n_O n_M n_S)$ while it is of order $\mathcal{O}(n_O{}^3 + n_O{}^2 * (n_S + n_M))$ for the SLMAP, the DBIM and the NKM.

The methods of the second type do not need any updates and evaluation of the criterion has a cost of order $\mathcal{O}(n_O{}^2 n_S + n_O * n_S * n_M)$ complex operations.

On the other hand, the third type methods require evaluations of criterion (6) whose computation order is $\mathcal{O}(n_O{}^3 + n_O{}^2 n_S + n_O * n_M * n_S)$. Indeed, for each evaluation of the criterion, the direct problem has to be solved. Fortunately, algorithm such as SA (Garnero et al. 1991) and ICM (Carfantan et al 1996), which update the contrast image pixel by pixel, can perform these updates without computing the whole criterion for each pixel but only once for the sweep of the whole image.

6.3   Convergence Properties

We already studied the convergence properties of each type of methods which can be summarized as follows. Methods of the first type, such as the SLMAP, can diverge while the other are guaranteed to converge towards a stationary point. On the other hand when the SLMAP converges, the provided solution corresponds to a local minimum of the MAP criterion (6) (which is not true for the BIM, the DBIM and the NKM). Methods of the second type are guaranteed to converge towards a minimum, maybe local, of the criterion (10). This solution corresponds to the MAP estimate only if the constraint (9) is verified, which is not guaranteed by these methods. The third type methods are guaranteed to converge towards a minimum, possibly local, of $\mathcal{J}^{\mathrm{MAP}}$.

6.4   Robustness with Respect to the Regularization Parameter

It has been experimentally established that methods of the second and third types are more robust with respect to the value of the regularization parameter λ (Carfantan 1996) than first type methods. For example, method of the first type can give good results for a value of λ and diverge for a nearby value, while the second and third types methods are in general not very sensitive to a change of a factor ten of this parameter, on the same configuration. This is an important point to consider as no automatic adjustment of this parameter is available, only the user's experience.

7   Conclusion

In this paper, we have studied diffraction tomography within the Bayesian estimation framework. It allows to consistently introduce prior information on the solution of this nonlinear ill-posed inverse problem and to define a regularized solution, the MAP estimate, with reasonable assumptions.

Different existing methods have been classified in terms of algorithms to compute the MAP estimate. Three types of methods have been distinguished. Methods of the first class correspond to successively approximating the nonlinear object/data relation with a linear one. Methods of the second type define the solution as the joint minimizer of a criterion depending on the object and on the total field on the object. Third type methods directly minimize the MAP criterion depending on the object. These methods have been compared on their convergence properties and on the solution they provide.

Three major key ideas can be emphasized:

- As regularization consists in introducing prior information on the solution, one can get benefits from introducing more advanced models than a simple L_2 (Tikhonov) one.

- A successive linearizations algorithm has been proposed to compute a
  regularized solution to this nonlinear inverse problem. It is both more
  efficient than the BIM for its linear approximation and than the NKM
  and the DBIM from a regularization viewpoint.
- The solution given by the minimization of a (penalized) joint criterion
  does not correspond to the minimum of the (penalized) mean square error
  on the measurement.

References

Barkeshli S., Lautzenheizer R. G. (1994): An iterative method for inverse scattering
problems based on an exact gradient search. *Radio Science*, vol. 29, no. 4, pp.
1119–1130

Caorsi S., Gragnani G. L., Pastorino M., Peraso A. (1993): Electromagnetic in-
verse scattering numerical method for non invasive diagnostic of dielectric mate-
rials. in *3rd International Conference on Electromagnetics in Aerospace Applica-
tions*, Torino, Italy

Caorsi S., Gragnani G. L., Medicina S., Pastorino M., Pinto A. (1995): A Gibbs
random fields-based active electromagnetic method for noninvasive diagnostics
in biomedical applications. *Radio Science*, vol. 30, no. 1, pp. 291–301

Carfantan H. (1996): Approche bayésienne pour un problème inverse non linéaire
en imagerie à ondes diffractées. *PhD Thesis* Université de Paris-Sud Orsay

Carfantan H., Mohammad-Djafari A. (1995): A Bayesian approach for nonlinear
inverse scattering tomographic imaging. *Proc. IEEE ICASSP*, Detroit, U.S.A.,
vol. IV, pp. 2311–2314

Carfantan H., Mohammad-Djafari A. (1996): Beyond the Born approximation in
inverse scattering with a Bayesian approach *2nd Int. Conf. on Inverse Problems
in Engineering*, Le Croisic, France

Carfantan H., Idier J., Mohammad-Djafari A. (1996): A single site update al-
gorithm for nonlinear diffraction tomography. accepted to *IEEE ICASSP* 1997,
Munich, Germany

Chew W. C., Wang Y. M. (1990): Reconstruction of two-dimensional permittivity
distribution using the distorted Born iterative method. *IEEE Trans. Medical
Imaging*, vol. MI-9, pp. 218–225

Demoment G. (1989): Image reconstruction and restoration : Overview of common
estimation structure and problems. *IEEE Trans. Acoust. Speech, Signal Process-
ing*, vol ASSP-37, no. 12, pp. 2024–2036

Garnero L., Franchois A., Hugonin J.-P., Pichot C., Joachimowicz N. (1991): Mi-
crowave imaging – complex permittivity reconstruction by simulated annealing.
IEEE Trans. Microwave Theory and Technology, vol. 39, no. 11, pp. 1801–1807

Geman D. (1990): Random fields and inverse problems in imaging. *École d'Été de
Probabilités de Saint-Flour XVIII - 1988*, vol. 1427, pp. 117–193, Springer-Verlag,
lecture notes in mathematics

Howard A. Q. J., Kretzschmar J. L. (1986): Synthesis of EM geophysical tomo-
graphic data. *Proc. IEEE*, vol. 74, no. 2, pp. 353–360

Idier J., Mohammad-Djafari A., Demoment G. (1996): Regularization methods
and inverse problems: an information theory standpoint. *2nd Int. Conf. on In-
verse Problems in Engineering*, Le Croisic, France

Joachimovicz N., Pichot C., Hugonin J.-P. (1991): Inverse scattering: An iterative numerical method for electromagnetic imaging. *IEEE Trans. Ant. Propag.*, vol. AP-39, no. 12, pp. 1742–1752

Kleinman R. E., van den Berg P. M. (1992): A modified gradient method for two-dimensional problems in tomography. *J. Computational and Applied Mathematics*, vol. 42, pp. 17–35

Künsch H. R. (1994): Robust priors for smoothing and image restoration. *Annals Institute Statistical Mathematics*, vol. 46, no. 1, pp. 1–19

Lobel P., Kleinman R.E., Pichot C., Blanc-Féraud L., Barlaud M. (1996): Conjugate gradient method for solving inverse scattering with experimental data. *IEEE Trans. Ant. Propag. Magazine*, vol. 38, no. 3, pp. 48–51

Roger A. (1981): Newton-Kantorovitch Algorithm Applied to an Electromagnetic Inverse problem. *IEEE Trans. Ant. Propag.*, vol. AP-29, pp. 232–238

Sabbagh H. A., Lautzenheiser R. G. (1993): Inverse problems in electromagnetic nondestructive evaluation. *International Journal of Applied Electromagnetics in Materials*, vol. 3, pp. 235–261

Tarantola A. (1987): Inverse problem theory: Methods for data fitting and model parameter estimation. *Elsevier Science Publisher*

van den Berg P. M., Kleinman R. E. (1995): A total variation enhanced modified gradient algorithm for profile reconstruction. *Inverse Problems*, vol. 11, pp. L5–L10

Wang Y. M., Chew W. C. (1989): An iterative solution of the two-dimensional electromagnetic inverse scattering problem. *Int. J. Imaging Systems and Technology*, vol. 1, pp. 100–108

Xia J. J., Habashy M., Kong J. A. (1994): Profile inversion in a cylindrically stratified lossy medium. *Radio Science*, vol. 29, no. 4, pp. 1131–1141