A systematic approach to robust preconditioning for gradient based inverse scattering algorithms

Nordebo, Sven; Fhager, Andreas; Gustafsson, Mats; Persson, Mikael

2008

CODEN:LUTEDX/(TEAT-7164)/1-23/(2008)

A Systematic Approach to Robust Preconditioning for Gradient Based Inverse Scattering Algorithms

Sven Nordebo, Andreas Fhager, Mats Gustafsson, and Mikael Persson

Electromagnetic Theory
Department of Electrical and Information Technology
Lund University
Sweden

Sven Nordebo
sven.nordebo@msi.vxu.se

School of Mathematics and Systems Engineering
Växjö University
SE-351 95 Växjö
Sweden

Andreas Fhager
andreas.fhager@chalmers.se

Department of Signals and Systems
Chalmers University of Technology
SE-412 96 Göteborg
Sweden

Mats Gustafsson
Mats.Gustafsson@eit.lth.se

Department of Electrical and Information Technology
Electromagnetic Theory
P.O. Box 118
SE-221 00 Lund
Sweden

Mikael Persson
mikael.persson@chalmers.se

Department of Signals and Systems
Chalmers University of Technology
SE-412 96 Göteborg
Sweden

Editor: Gerhard Kristensson

Abstract

This paper presents a systematic approach to robust preconditioning for gradient based non-linear inverse scattering algorithms. In particular, one- and two-dimensional inverse problems are considered where the permittivity and conductivity profiles are unknown and the input data consists of the scattered field over a certain bandwidth. A time-domain least-squares formulation is employed and the inversion algorithm is based on a conjugate gradient, or quasi-Newton algorithm together with an FDTD-electromagnetic solver. A Fisher information analysis is used to estimate the Hessian of the error functional. A robust preconditioner is then obtained by incorporating a parameter scaling such that the scaled Fisher information has a unit diagonal. By improving the conditioning of the Hessian, the convergence rate of the conjugate gradient or quasi-Newton methods are improved. The preconditioner is robust in the sense that the scaling, *i.e.*, the diagonal Fisher information is virtually invariant to the numerical resolution and the discretization model that is employed. Numerical examples of image reconstruction are included to illustrate the efficiency of the proposed technique.

1 Introduction

Microwave tomography is an inverse problem with several promising applications in *e.g.*, non-invasive medical imaging, early detection of breast tumors [7–9, 16, 20, 21, 27], and non-destructive testing [2, 13, 14]. However, there still remains many important analytical and computational challenges related to these highly non-linear and often ill-conditioned imaging/estimation problems [17, 30]. In particular, computational cost and preconditioning of numerical algorithms are very important issues.

To accelerate the convergence of iterative methods, preconditioning is often used [11, 19]. There are very few theoretical results on preconditioning, and finding a good preconditioner is sometimes considered as a combination of art and science [26]. The gradient based algorithms are often based on a least-squares formulation and are hence justified in the case of Gaussian noise [15, 30]. The major drawback with the gradient based method is the high computational cost and the risk of getting trapped in a local minimum. It has been observed that a preconditioner in the form of a simple parameter scaling can increase the convergence rate and reduce the problem with local minima [14].

The idea of employing sensitivity analysis as a means to improve the tradeoff between the quality of images and the complexity of the reconstruction is a well known issue in inverse problems. In [3], a sensitivity analysis for a linearized electrical impedance tomography problem is used to obtain a regularization scheme based on parameter variance uniformization. In this way, the inherent ill-posedness of the reconstruction due to strong attenuation phenomena, is greatly alleviated.

A Fisher information analysis and the Cramér-Rao lower bound provides a very useful instrument for sensitivity analysis of various wave propagation phenomena, and which facilitates valuable physical interpretations, see *e.g.*, [4–6, 12, 15, 23, 24, 28, 31]. Cramér-Rao bounds for the location, size and orientation of a known object

has been studied in the context of diffraction tomography and Maximum Likelihood (ML) estimation in [5, 23, 31]. Previously, the Cramér-Rao bound has been employed as an analytical tool to investigate the one-dimensional inverse scattering problem of multilayer structures [12], and a canonical two-dimensional microwave tomography set-up is analyzed in [24]. In [12, 24], the Cramér-Rao bound is employed as an analytical tool to quantify the ill-posedness of the reconstruction and to explicitely describe the inherent trade-off between the accuracy and the resolution.

In this paper, the Fisher information analysis is introduced as a systematic approach to obtain a robust preconditioner for gradient based non-linear inverse scattering algorithms. One- and two-dimensional inverse problems are considered where the permittivity and conductivity profiles are unknown and the input data consists of the scattered field over a certain bandwidth. A time-domain least-squares formulation is employed [13] and the inversion algorithm is based on a conjugate gradient, or quasi-Newton algorithm [10] together with an FDTD-electromagnetic solver [29].

In the first step of the preconditioning, the Fisher information analysis is performed to estimate the Hessian of the error functional corresponding to some known background. A robust preconditioner is then obtained by incorporating a parameter scaling such that the scaled Fisher information has a unit diagonal, *cf.*, the Jacobi preconditioner in numerical analysis [11, 19]. By improving the conditioning of the Hessian in the corresponding Maximum Likelihood estimation problem, the convergence rate of the conjugate gradient or quasi-Newton methods are improved. The preconditioner is robust in the sense that the scaling, *i.e.*, the diagonal Fisher information is virtually invariant to the numerical resolution and the discretization model that is employed.

One- and two-dimensional model problems are considered in parallel as these offer a natural introduction to electromagnetic inverse scattering, wave splitting and numerical algorithms, see *e.g.*, [13]. In particular, the numerical scheme for two-dimensional microwave tomography as described in [8, 9, 16] has been modified by employing the Fisher information analysis developed in [24]. Further, the one-dimensional model problem [12] is employed to illustrate the use of wave splitting techniques. It should be noted that even though the detailed analysis given here is with fairly simple one- and two-dimensional model problems which have been chosen for simplicity and clarity of exposition, it is straightforward to extend the proposed Fisher information analysis and preconditioning strategy to more elaborate inverse problems including full three-dimensional modelling, etc.

The preconditioning strategy described in this paper bears several similarities with the ideas presented in [3], eventhough [3] addresses the regularization parameters. In fact, the sensitivity matrix for the linearized problem used in [3] has a direct correspondence to the Fisher information matrix used in the present context. In particular, both techniques are efficiently capturing the impact of attenuation phenomena. In this respect, this paper presents a systematic and unifying approach to preconditioning which does not require prior linearization of the forward problem. Numerical examples of image reconstruction are included to illustrate the performance of the proposed preconditioner.

The rest of the paper is outlined as follows. In section 2 is presented the model

problems in one and two dimensions, including the statistical signal modelling. In section 3 is presented the Maximum Likelihood (ML) estimation interpretation of the time-domain least-squares inverse problem formulation. In section 4 is presented the Fisher information analysis for the one- and two-dimensional model problems. In section 5 is presented the general Fisher information based preconditioning strategy and in section 6 the numerical examples.

2 Model Problems in One and Two Dimensions

In this paper, two distinct model problems are considered related to the least-squares optimization approach to time-domain electromagnetic inverse problems, *cf.*, [8, 9, 12–14, 16, 24]. These model problems are: a two-dimensional microwave tomography set-up and a one-dimensional inverse scattering problem, as described in detail below.

Throughout the paper, let $\mathrm{e}^{\mathrm{i}\omega t}$ be the time-convention where $\omega = 2\pi f$ is the angular frequency. Let k_0, c_0, ϵ_0, μ_0 and η_0 denote the wave number, the speed of light, the permittivity, the permeability and the wave impedance of free space, respectively.

2.1 Two-Dimensional Microwave Tomography Set-up

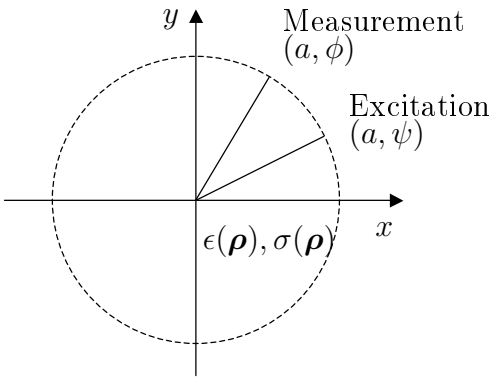

Figure 1: Two-dimensional microwave tomography set-up for an isotropic cylinder with relative permittivity $\epsilon(\boldsymbol{\rho})$ and conductivity $\sigma(\boldsymbol{\rho})$. Measurement cylinder of radius a with excitation at (a, ψ) and measurement at (a, ϕ). The background space is homogenous and isotropic.

Let (ρ, ϕ, z) denote the cylindrical coordinates, $(\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{z}})$ the corresponding unit vectors and $\boldsymbol{\rho} = \rho\hat{\boldsymbol{\rho}}$ the two-dimensional radius vector with coordinates (ρ, ϕ). It is assumed that all fields depend on the two-dimensional spatial domain $\boldsymbol{\rho} \in \mathbb{R}^2$ only and that the electric field $\boldsymbol{E}$ is vertically polarized with $\boldsymbol{E} = E(\boldsymbol{\rho})\hat{\boldsymbol{z}}$.

Consider the electromagnetic inverse problem of imaging an isotropic two-dimensional circular cylinder of radius a in a homogenous and isotropic background space,

cf., figure 1. The measurement is performed on a cylinder of radius a with excitation at (a, ψ) and measurement at (a, ϕ). The inverse problem consists of estimating the relative permittivity $\epsilon(\boldsymbol{\rho})$ and conductivity $\sigma(\boldsymbol{\rho})$ within the cylindrical object (for $\rho < a$), based on measurements (or observations) of the electric field $E(\phi, \psi, t)$ for $(\phi, \psi, t) \in [0, 2\pi] \times [0, 2\pi] \times [0, T]$ where T is the length of the observation interval.

The electric and magnetic fields $\boldsymbol{u} = (\boldsymbol{E}, \boldsymbol{H})$ satisfy Maxwell's equations[1]

$$\mathcal{P}\boldsymbol{u} = \begin{cases} \epsilon \partial_t \boldsymbol{E} - \nabla \times \boldsymbol{H} + \sigma \boldsymbol{E} = -\boldsymbol{J}_{\mathrm{s}} \\ \partial_t \boldsymbol{H} + \nabla \times \boldsymbol{E} = \boldsymbol{0} \end{cases} \tag{2.1}$$

for $t \in [0, T]$, together with the initial conditions $\boldsymbol{E}(\boldsymbol{\rho}, t) = \boldsymbol{0}$ for $t = 0$ and $\boldsymbol{\rho} \in \mathbb{R}^2$. It is assumed that the excitation is a line source $\boldsymbol{J}_{\mathrm{s}} = \hat{\boldsymbol{z}} s(t) \delta(\boldsymbol{\rho} - \boldsymbol{\rho}')$ at $\boldsymbol{\rho}' = (a, \psi)$ where $s(t)$ is the broad band excitation signal and $\delta(\cdot)$ denotes a spatial impulse function.

In the frequency domain, Maxwell's equations (2.1) yield the following wave equation for the scalar field E, *i.e.*, the Helmholtz equation

$$\mathcal{L}E = \left\{ \frac{1}{\rho} \frac{\partial}{\partial \rho} \rho \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2}{\partial \phi^2} + k^2 \right\} E = \mathrm{i}\omega J, \tag{2.2}$$

where $k = k_0 \sqrt{\epsilon_{\mathrm{c}}} = \omega \sqrt{\epsilon_{\mathrm{c}}}$, $\epsilon_{\mathrm{c}}(\boldsymbol{\rho}) = \epsilon(\boldsymbol{\rho}) - \mathrm{i}\sigma(\boldsymbol{\rho})/\omega$ and the source is vertically polarized with $\boldsymbol{J}_{\mathrm{s}} = J\hat{\boldsymbol{z}}$. The corresponding Green's function $G(\boldsymbol{\rho}, \boldsymbol{\rho}') = G(\rho, \phi, \rho', \phi')$ for a line source at $\boldsymbol{\rho}' = (\rho', \phi')$ satisfies $\mathcal{L}G(\boldsymbol{\rho}, \boldsymbol{\rho}') = -\delta(\boldsymbol{\rho} - \boldsymbol{\rho}')$.

For a homogenous background and cylindrical coordinates, the background Green's function is given by

$$G(\rho, \phi, \rho', \phi') = -\frac{\mathrm{i}}{4} \sum_{m=-\infty}^{\infty} \mathrm{J}_m(k\rho_<) \mathrm{H}_m^{(2)}(k\rho_>) \mathrm{e}^{\mathrm{i}m(\phi - \phi')} \tag{2.3}$$

where $\mathrm{J}_m(\cdot)$ and $\mathrm{H}_m^{(2)}(\cdot)$ are the Bessel function and the Hankel function of the second kind, respectively, both of order m, see *e.g.*, [1]. Here, $\rho_< = \min\{\rho, \rho'\}$ and $\rho_> = \max\{\rho, \rho'\}$.

Assuming that the source is a line source at $\boldsymbol{\rho}' = (a, \psi)$ with $J = S(f)\delta(\boldsymbol{\rho} - \boldsymbol{\rho}')$, and that the measurement is performed at $\boldsymbol{\rho} = (a, \phi)$, the observed quantity is then given by

$$E(\phi, \psi, f) = \int_{-\infty}^{\infty} E(\phi, \psi, t) \mathrm{e}^{-\mathrm{i}\omega t} \, \mathrm{d}t = -\mathrm{i}\omega S(f) G(a, \phi, a, \psi), \tag{2.4}$$

where $S(f)$ is the frequency domain excitation signal.

The two-dimensional Fourier coefficients of the $2\pi \times 2\pi$ periodic function $E(\phi, \psi, f)$ are defined by

$$E_{mn}(f) = \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} E(\phi, \psi, f) \mathrm{e}^{-\mathrm{i}m\phi - \mathrm{i}n\psi} \, \mathrm{d}\phi \, \mathrm{d}\psi. \tag{2.5}$$

[1]Here, the common SI-unit quantities are normalized as $(t, \omega, \epsilon, \sigma, \boldsymbol{E}, \boldsymbol{H}, \boldsymbol{J})^{\mathrm{norm}} = (c_0 t, \omega/c_0, \epsilon, \eta_0 \sigma, \sqrt{\epsilon_0}\boldsymbol{E}, \sqrt{\mu_0}\boldsymbol{H}, \sqrt{\mu_0}\boldsymbol{J})$ so that the speed of wave propagation is normalized to unity and all fields are measured in the same energy unit (Energy/Volyme)$^{1/2}$.

Further, let $G_m(\rho,\rho',\phi') = -\frac{\mathrm{i}}{4}\mathrm{J}_m(k\rho_<)\mathrm{H}_m^{(2)}(k\rho_>)\mathrm{e}^{-\mathrm{i}m\phi'}$ be the one-dimensional Fourier coefficients of $G(\rho,\phi,\rho',\phi')$ with respect to ϕ, as defined by (2.3). Due to the circular symmetry, it is readily seen that $G_m(\rho,\rho',\phi') = G_m(\rho,\rho',0)\mathrm{e}^{-\mathrm{i}m\phi'}$ and hence

$$E_{mn}(f) = -\mathrm{i}\omega S(f)G_m(a,a,0)\delta_{-m,n} = -\frac{\omega}{4}S(f)\mathrm{J}_m(ka)\mathrm{H}_m^{(2)}(ka)\delta_{-m,n} \tag{2.6}$$

where δ_{mn} denotes the Kronecker delta.

Consider now a finite time interval $[-\tau/2,\tau/2]$ and let E_{mnq} denote the three-dimensional Fourier coefficients

$$E_{mnq} = \frac{1}{(2\pi)^2\tau}\int_0^{2\pi}\int_0^{2\pi}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} E(\phi,\psi,t)\mathrm{e}^{-\mathrm{i}m\phi-\mathrm{i}n\psi-\mathrm{i}q\frac{2\pi}{\tau}t}\,\mathrm{d}t\,\mathrm{d}\phi\,\mathrm{d}\psi \tag{2.7}$$

corresponding to a τ-periodic extension of the time-domain pulses. Assuming that $E(\phi,\psi,t)$ is a time-limited pulse with support in $[0,T]$, it is seen that $\tau E_{mnq} \to E_{mn}(f)|_{f=\frac{q}{\tau}}$ as $\tau\to\infty$.

Consider the following statistical measurement model

$$E^{(\mathrm{m})}(\phi,\psi,t) = E(\phi,\psi,t) + N(\phi,\psi,t), \tag{2.8}$$

where $E^{(\mathrm{m})}(\phi,\psi,t)$ denotes the measured field, $N(\phi,\psi,t)$ is additive measurement noise and $(\phi,\psi,t)\in[0,2\pi]\times[0,2\pi]\times[-\tau/2,\tau/2]$. Here, $N(\phi,\psi,t)$ is modeled as a spatially uncorrelated zero mean Gaussian random process [18] with correlation function

$$\mathcal{E}\left\{N(\phi+\Delta\phi,\psi+\Delta\psi,t+\Delta t)N(\phi,\psi,t)\right\} = (2\pi)^2\delta(\Delta\phi)\delta(\Delta\psi)r_{\mathrm{N}}(\Delta t), \tag{2.9}$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation operator and $\delta(\cdot)$ an impulse function with period 2π. Here, $r_{\mathrm{N}}(\Delta t)$ denotes the temporal correlation function with power spectral density $R_{\mathrm{N}}(f) = \int_{-\infty}^{\infty} r_{\mathrm{N}}(t)\mathrm{e}^{-\mathrm{i}2\pi ft}\,\mathrm{d}t$.

A discrete measurement model is now obtained by first creating the complex (time-domain) Hilbert pair [25] corresponding to (2.8), and then considering the corresponding three-dimensional Fourier series representation. Note that the complex Hilbert pair corresponding to a real Gaussian random process is complex Gaussian, and the Fourier transform of a complex Gaussian process is complex Gaussian [22]. Hence, (2.8) yields

$$2E_{mnq}^{(\mathrm{m})} = 2E_{mnq} + N_{mnq},\ q\geq 0, \tag{2.10}$$

where N_{mnq} is discrete zero mean complex Gaussian noise with correlation function given by

$$\mathcal{E}\left\{N_{mnq}^* N_{m'n'q'}\right\} = \delta_{mm'}\delta_{nn'}\frac{1}{\tau}4R_{\mathrm{N}}(\frac{q}{\tau})\delta_{qq'},\ q\geq 0 \tag{2.11}$$

where $(\cdot)^*$ denotes the complex conjugate and where it has been assumed that the support time t_{N} for the correlation function $r_{\mathrm{N}}(\Delta t)$ is much less then the period, $t_{\mathrm{N}} \ll \tau$.

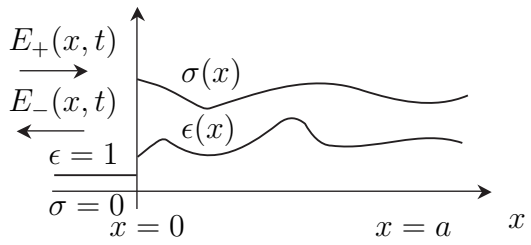

Figure 2: One-dimensional inverse problem for an isotropic half space with relative permittivity $\epsilon(x)$ and conductivity $\sigma(x)$.

2.2 One-Dimensional Inverse Scattering Problem

Let $(x, y, z) \in \mathbb{R}^3$ denote the cartesian coordinates and $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ the corresponding unit vectors. It is assumed that all fields depend on the x-coordinate only and that the electric and magnetic fields $\boldsymbol{E}$ and $\boldsymbol{H}$ are linearly polarized with $\boldsymbol{E} = E(x)\hat{\boldsymbol{z}}$ and $\boldsymbol{H} = H(x)\hat{\boldsymbol{y}}$.

Consider the electromagnetic inverse problem of imaging a one-dimensional isotropic half space $x \geq 0$, with relative permittivity $\epsilon(x)$ and conductivity $\sigma(x)$, see figure 2. The imaging is based on a measurement of the incident field $E_+(x,t)$ and the scattered field $E_-(x,t)$ at the boundary $x = 0$ for $t \in [0, T]$ where T is the length of the observation interval. The left half space $x < 0$ is free space with $\epsilon = 1$ and $\sigma = 0$.

The electric and magnetic fields $\boldsymbol{u} = (E, H)$ satisfy Maxwell's equations

$$\mathcal{P}\boldsymbol{u} = \begin{cases} \epsilon\partial_t E - \partial_x H + \sigma E = 0 \\ \partial_t H - \partial_x E = 0 \end{cases} \tag{2.12}$$

for $x \in \mathbb{R}$ and $t \in [0, T]$, together with the initial conditions $E(x, 0) = 0$ for $x \geq 0$ and $t = 0$, and the boundary conditions $E_+(0,t) = E_+^{(\mathrm{m})}(0,t)$ for $x = 0$ and $t \in [0, T]$, where the superscript $(\cdot)^{(\mathrm{m})}$ denotes measured field quantities.

In the frequency domain, Maxwell's equations (2.12) yield the following wave equation for the scalar field E

$$\mathcal{L}E = \left\{\partial_x^2 + k^2\right\} E = 0, \tag{2.13}$$

where $k = k_0\sqrt{\epsilon_{\mathrm{c}}} = \omega\sqrt{\epsilon_{\mathrm{c}}}$, $\epsilon_{\mathrm{c}}(x) = \epsilon(x) - \mathrm{i}\sigma(x)/\omega$ and the boundary conditions are $E_+(0, f) = E_+^{(\mathrm{m})}(0, f)$. Here, the Fourier transform is given by $E(x, f) = \int_{-\infty}^{\infty} E(x,t)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}t$.

Assuming that the right half space $x \geq 0$ is homogenous, the transmitted field is given by the corresponding transmission coefficient. Hence, for $x \geq 0$

$$E(x, f) = E_+(0, f)\frac{2\eta}{\eta + 1}\mathrm{e}^{-\mathrm{i}kx}, \tag{2.14}$$

where $\eta = 1/\sqrt{\epsilon_{\rm c}}$. The corresponding Green's function $G(x,x')$ for a point source at x' satisfies $\mathcal{L}G(x,x') = -\delta(x-x')$ where $\delta(\cdot)$ is an impulse function. For a homogenous half space and $x'=0$, the Green's function is given by

$$G(x,0) = \frac{1}{{\rm i}\omega + {\rm i}k}{\rm e}^{-{\rm i}kx} \tag{2.15}$$

for $x \geq 0$.

Consider now the finite time interval $[-\tau/2, \tau/2]$ and let $E(x,q)$ denote the one-dimensional Fourier coefficients

$$E(x,q) = \frac{1}{\tau}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} E(x,t){\rm e}^{-{\rm i}q\frac{2\pi}{\tau}t}\,{\rm d}t \tag{2.16}$$

corresponding to a τ-periodic extension of the time-domain pulses. Assuming that $E(x,t)$ is a time-limited pulse with support in $[0,T]$, it is seen that $\tau E(x,q) \rightarrow E(x,f)|_{f=\frac{q}{\tau}}$ as $\tau \rightarrow \infty$.

Consider the following statistical measurement model

$$E_-^{(\rm m)}(0,t) = E_-(0,t) + N(t) \tag{2.17}$$

where $t \in [-\tau/2, \tau/2]$ and $N(t)$ is additive measurement noise. Here, $N(t)$ is modeled as a zero mean Gaussian random process with correlation function $r_{\rm N}(\Delta t) = \mathcal{E}\left\{N(t+\Delta t)N(t)\right\}$ and power spectral density $R_{\rm N}(f)$.

As before, a discrete measurement model is now obtained by first creating the complex Hilbert pair corresponding to (2.17), and then considering the corresponding Fourier series representation. Hence,

$$2E_-^{(\rm m)}(0,q) = 2E_-(0,q) + N_q,\ q \geq 0 \tag{2.18}$$

where N_q is discrete zero mean complex Gaussian noise with correlation function given by

$$\mathcal{E}\left\{N_q^* N_{q'}\right\} = \frac{1}{\tau}4R_{\rm N}(\frac{q}{\tau})\delta_{qq'},\ q \geq 0 \tag{2.19}$$

where it has been assumed that the support time $t_{\rm N}$ for the correlation function $r_{\rm N}(\Delta t)$ is much less then the period, $t_{\rm N} \ll \tau$.

3 Maximum Likelihood Estimation

3.1 The two-dimensional model problem

Consider first the two-dimensional model problem described above. It is assumed that the cylinder region $\mathcal{S} = \{\boldsymbol{\rho} | \rho \leq a_0 < a\} = \cup_{i=1}^I \mathcal{S}_i$ is decomposed into a finite set of disjoint image cells or pixels $\mathcal{S}_i$ corresponding to some specific scale of resolution. The relative permittivity and conductivity within the cylinder is hence discretized according to the finite expansions

$$\begin{cases} \epsilon(\boldsymbol{\rho}) = \sum_{i=1}^I \epsilon_i \chi_i(\boldsymbol{\rho}) \\ \sigma(\boldsymbol{\rho}) = \sum_{i=1}^I \sigma_i \chi_i(\boldsymbol{\rho}) \end{cases} \tag{3.1}$$

where ϵ_i and σ_i are the parameters to be estimated and $\chi_i(\boldsymbol{\rho})$ is the characteristic function for pixel $\mathcal{S}_i$, *i.e.*, $\chi_i(\boldsymbol{\rho}) = 1$ if $\boldsymbol{\rho} \in \mathcal{S}_i$ and $\chi_i(\boldsymbol{\rho}) = 0$ if $\boldsymbol{\rho} \notin \mathcal{S}_i$.

Let $\boldsymbol{\nu} = [\boldsymbol{\epsilon}^{\mathrm{T}} \boldsymbol{\sigma}^{\mathrm{T}}]^{\mathrm{T}}$ denote the parameter vector with elements ϵ_i and σ_i as defined in (3.1), and let $\boldsymbol{x}$ denote the measurement vector $\boldsymbol{x} = \{2E^{(\mathrm{m})}_{mnq}\}$ with probability density function $p(\boldsymbol{x}|\boldsymbol{\nu})$. Since the measurement noise N_{mnq} is an uncorrelated complex Gaussian random process, the negative loglikelihood function [15, 18] is given by

$$-\log p(\boldsymbol{x}|\boldsymbol{\nu}) = b + \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \sum_{q=0}^{\infty} \frac{\tau}{R_{\mathrm{N}}(\frac{q}{\tau})} |E_{mnq} - E^{(\mathrm{m})}_{mnq}|^2 \tag{3.2}$$

which becomes, in the limit as $\tau \to \infty$,

$$-\log p(\boldsymbol{x}|\boldsymbol{\nu}) = b + \frac{1}{(2\pi)^2 2N_0} \mathcal{J}(\boldsymbol{\nu}) \tag{3.3}$$

where the power spectral density is assumed to be a constant $R_{\mathrm{N}}(f) = N_0$ over the relevant bandwidth, b is a constant, and $\mathcal{J}(\boldsymbol{\nu})$ is the misfit functional

$$\mathcal{J}(\boldsymbol{\nu}) = \int_0^{2\pi} \int_0^{2\pi} \int_0^{T} |E(\phi, \psi, t) - E^{(\mathrm{m})}(\phi, \psi, t)|^2 \, \mathrm{d}t \, \mathrm{d}\phi \, \mathrm{d}\psi. \tag{3.4}$$

Hence, with the Gaussian noise model adopted here, the optimization problem stated in two-dimensional microwave tomography [9, 14] is equivalent to the classical Maximum Likelihood (ML) criterion [15, 18].

3.2 The one-dimensional model problem

The ML criterion for the one-dimensional model problem is merely a straightforward modification and a change of notation in the analysis above. The spatial region $\mathcal{S} = \{x | 0 \le x \le a\} = \cup_{i=1}^{I} \mathcal{S}_i$ is decomposed into a finite set of disjoint intervals $\mathcal{S}_i$ corresponding to some specific scale of resolution. The incident wave is assumed to be attenuated such that it is negligible at the computational boundary $x = a$. The relative permittivity and conductivity within the material is discretized according to the finite expansions

$$\begin{cases} \epsilon(x) = \sum_{i=1}^{I} \epsilon_i \chi_i(x) \\ \sigma(x) = \sum_{i=1}^{I} \sigma_i \chi_i(x) \end{cases}, \tag{3.5}$$

where $\chi_i(x)$ is the characteristic function for the interval (pixel) $\mathcal{S}_i$.

Here $\boldsymbol{\nu} = [\boldsymbol{\epsilon}^{\mathrm{T}} \boldsymbol{\sigma}^{\mathrm{T}}]^{\mathrm{T}}$ denotes the parameter vector with elements ϵ_i and σ_i as defined in (3.5), and $\boldsymbol{x}$ denotes the measurement vector $\boldsymbol{x} = \{2E^{(\mathrm{m})}_{-}(0, q)\}$ with probability density function $p(\boldsymbol{x}|\boldsymbol{\nu})$. The measurement noise N_q is an uncorrelated complex Gaussian random process, and the negative loglikelihood function is given by

$$-\log p(\boldsymbol{x}|\boldsymbol{\nu}) = b + \lim_{\tau \to \infty} \sum_{q=0}^{\infty} \frac{\tau}{R_{\mathrm{N}}(\frac{q}{\tau})} |E_{-}(0, q) - E^{(\mathrm{m})}_{-}(0, q)|^2 = b + \frac{1}{2N_0} \mathcal{J}(\boldsymbol{\nu}), \tag{3.6}$$

where $\mathcal{J}(\boldsymbol{\nu})$ is the misfit functional

$$\mathcal{J}(\boldsymbol{\nu}) = \int_0^T |E_-(0,t) - E_-^{(\mathrm{m})}(0,t)|^2 \,\mathrm{d}t \tag{3.7}$$

which is employed with one-dimensional inverse scattering problems, see *e.g.*, [13].

4 Fisher Information Analysis

4.1 The two-dimensional model problem

The Fisher information matrix [18] for the parameters ϵ_i and σ_i based on the statistical measurement model (2.10) is given by

$$[\boldsymbol{\mathcal{I}}_{\nu\zeta}]_{ij} = 2\,\mathrm{Re} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \sum_{q=0}^{\infty} \frac{\tau}{R_{\mathrm{N}}(\frac{q}{\tau})} \frac{\partial E^*_{mnq}}{\partial \nu_i} \frac{\partial E_{mnq}}{\partial \zeta_j}, \tag{4.1}$$

where ν and ζ are either ϵ or σ, and $i, j = 1, \ldots, I$. In the limit as $\tau \to \infty$, the expression (4.1) becomes

$$[\boldsymbol{\mathcal{I}}_{\nu\zeta}]_{ij} = \int_{-\infty}^{\infty} \frac{1}{R_{\mathrm{N}}(f)} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{\partial E^*_{mn}}{\partial \nu_i} \frac{\partial E_{mn}}{\partial \zeta_j} \,\mathrm{d}f. \tag{4.2}$$

The differentiated field, or sensitivity field, satisfy the wave equation (2.2)

$$\mathcal{L}\frac{\partial E}{\partial \nu_i} = \left\{ \frac{1}{\rho}\frac{\partial}{\partial \rho}\rho\frac{\partial}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2}{\partial \phi^2} + k^2 \right\} \frac{\partial E}{\partial \nu_i} = \mathrm{i}\omega g_\nu \chi_i E, \tag{4.3}$$

where $g_\nu = 1$ if $\nu = \sigma$ and $g_\nu = \mathrm{i}\omega$ if $\nu = \epsilon$. Note that the solution E of (2.2) now appears in the source term of (4.3). The sensitivity field can hence be expressed as

$$\begin{aligned} \frac{\partial E(\phi,\psi,f)}{\partial \nu_i} &= -\mathrm{i}\omega g_\nu \int_{\mathcal{S}} G(\boldsymbol{\rho},\boldsymbol{\rho}')\chi_i(\boldsymbol{\rho}')E(\boldsymbol{\rho}')\,\mathrm{d}S' \\ &= -\omega^2 g_\nu S(f) \int_{\mathcal{S}_i} G(a,\phi,\rho',\phi')G(a,\psi,\rho',\phi')\,\mathrm{d}S', \end{aligned} \tag{4.4}$$

where the symmetry of the Green's function has been employed. Considering the two-dimensional Fourier series representation of (4.4), the differentiated Fourier coefficients become

$$\begin{aligned} \frac{\partial E_{mn}}{\partial \nu_i} &= -\omega^2 g_\nu S(f) \int_{\mathcal{S}_i} G_m(a,\rho',\phi')G_n(a,\rho',\phi')\,\mathrm{d}S' \\ &= \frac{\omega^2}{16} g_\nu S(f) \mathrm{H}_m^{(2)}(ka)\mathrm{H}_n^{(2)}(ka) \int_{\mathcal{S}_i} \mathrm{J}_m(k\rho')\mathrm{J}_n(k\rho')\mathrm{e}^{-\mathrm{i}(m+n)\phi'}\,\mathrm{d}S' \end{aligned} \tag{4.5}$$

where the last line is valid under the assumption that the background space is homogenous with $G_m(\rho,\rho',\phi') = -\frac{\mathrm{i}}{4}\mathrm{J}_m(k\rho_<)\mathrm{H}_m^{(2)}(k\rho_>)\mathrm{e}^{-\mathrm{i}m\phi'}$ as defined in (2.3).

After evaluation of (4.2), the total Fisher information matrix is assembled as

$$\boldsymbol{\mathcal{I}} = \begin{pmatrix} \boldsymbol{\mathcal{I}}_{\epsilon\epsilon} & \boldsymbol{\mathcal{I}}_{\epsilon\sigma} \\ \boldsymbol{\mathcal{I}}_{\sigma\epsilon} & \boldsymbol{\mathcal{I}}_{\sigma\sigma} \end{pmatrix}. \tag{4.6}$$

The Cramér-Rao bound (CRB) [18] for estimating the parameter ν_i is finally given by

$$\mathcal{E}\left\{|\widehat{\nu}_i - \nu_i|^2\right\} \geq \left[\left[\boldsymbol{\mathcal{I}}^{-1}\right]_{\nu\nu}\right]_{ii} \tag{4.7}$$

where $\widehat{\nu}_i$ is any unbiased estimate of ν_i. For further details about the calculation of (4.2) and (4.5) for circularly symmetrical pixels $\mathcal{S}_i$, see [24].

In order to define the signal to noise ratio, it is assumed that the noise power spectral density is a constant $R_{\mathrm{N}}(f) = N_0$ over the relevant bandwidth, and that the incident signal is a normalized Gaussian pulse with

$$\begin{cases} s^2(t) = \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{t}}} \mathrm{e}^{-t^2/2\sigma_{\mathrm{t}}^2} \\ S^2(f) = \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{f}}} \mathrm{e}^{-f^2/2\sigma_{\mathrm{f}}^2} \end{cases} \tag{4.8}$$

which has been centered at the origin (base-band) for simplicity. Here, $\int s^2(t)\,\mathrm{d}t = \int S^2(f)\,\mathrm{d}f = 1$ and $\sigma_{\mathrm{t}}\sigma_{\mathrm{f}} = 1/4\pi$. Defining the time and frequency bandwidths as $t_{\mathrm{B}} = 2 \cdot 1.96\sigma_{\mathrm{t}}$ and $f_{\mathrm{B}} = 2 \cdot 1.96\sigma_{\mathrm{f}}$ for 95 % of the pulse energy, the time-frequency bandwidth product is given by $t_{\mathrm{B}} f_{\mathrm{B}} = \frac{1.96^2}{\pi} \approx 1$. Hence,

$$\frac{\frac{1}{T}\int s^2(t)\,\mathrm{d}t}{\int R_{\mathrm{N}}(f)\,\mathrm{d}f} \approx \frac{\int s^2(t)\,\mathrm{d}t}{t_{\mathrm{B}} 4N_0 f_{\mathrm{B}}} \approx \frac{1}{4N_0} = \mathrm{SNR} \tag{4.9}$$

where the factor $4N_0$ is due to the base-band translation.

4.2 The one-dimensional model problem

The Fisher information matrix based on the statistical measurement model (2.18) is given by

$$[\boldsymbol{\mathcal{I}}_{\nu\zeta}]_{ij} = 2\,\mathrm{Re}\sum_{q=0}^{\infty} \frac{\tau}{R_{\mathrm{N}}(\frac{q}{\tau})} \frac{\partial E_-^*(0,q)}{\partial \nu_i} \frac{\partial E_-(0,q)}{\partial \zeta_j} \tag{4.10}$$

where ν and ζ are either ϵ or σ, and $i, j = 1, \ldots, I$. In the limit as $\tau \to \infty$, the expression (4.10) becomes

$$[\boldsymbol{\mathcal{I}}_{\nu\zeta}]_{ij} = \int_{-\infty}^{\infty} \frac{1}{R_{\mathrm{N}}(f)} \frac{\partial E_-^*(0,f)}{\partial \nu_i} \frac{\partial E_-(0,f)}{\partial \zeta_j}\,\mathrm{d}f. \tag{4.11}$$

The sensitivity field satisfy the wave equation (2.13)

$$\mathcal{L}\frac{\partial E}{\partial \nu_i} = \left\{\partial_x^2 + k^2\right\}\frac{\partial E}{\partial \nu_i} = \mathrm{i}\omega g_\nu \chi_i E \tag{4.12}$$

where $g_\nu = 1$ if $\nu = \sigma$ and $g_\nu = \mathrm{i}\omega$ if $\nu = \epsilon$, and the solution E of (2.13) appears in the source term of (4.12). The sensitivity field can hence be represented as

$$\frac{\partial E(x,f)}{\partial \nu_i} = -\mathrm{i}\omega g_\nu \int_{\mathcal{S}} G(x,x')\chi_i(x')E(x',f)\,\mathrm{d}x', \tag{4.13}$$

or

$$\frac{\partial E_-(0,f)}{\partial \nu_i} = -\mathrm{i}\omega g_\nu \int_{\mathcal{S}_i} G(x',0)E(x',f)\,\mathrm{d}x' \tag{4.14}$$

where the symmetry of the Green's function $G(x,x') = G(x',x)$ and $\frac{\partial E_+(0,f)}{\partial \nu_i} = 0$ have been employed. By inserting (2.14) and (2.15) in (4.14) and evaluating the integral, the sensitivity field for a homogenous background is given by

$$\frac{\partial E_-(0,f)}{\partial \nu_i} = -2g_\nu \frac{E_+(0,f)}{(1+\sqrt{\epsilon_\mathrm{c}})^2} \frac{\sin(k\Delta x)}{k\Delta x} \Delta x \mathrm{e}^{-\mathrm{i}k(2i-1)\Delta x} \tag{4.15}$$

where $\mathcal{S}_i = [(i-1)\Delta x, i\Delta x]$, Δx the spatial sampling interval and $i = 1,\ldots,I$.

As can be clearly seen in (4.15), the magnitude of the sensitivity field is almost linear in Δx (for small Δx). A similar property can be deduced for the two-dimensional case (4.5), *i.e.*, the sensitivity field is almost proportional to the pixel area ΔS. In this particular sense, we may say that the Fisher information is only weakly dependent on, or *virtually invariant* to the assumed discretization model. Note however, that the *inverse* Fisher information and the Cramér-Rao bound are strongly dependent on, or extremely sensitive to the assumed discretization model.

The second essential property here is that the magnitude of the Fisher information is closely related to the attenuation of the wave field. In fact, for the one-dimensional problem above the magnitude of the Fisher information decays inside the material exactly according to the (round-trip) attenuation of a plane wave at a single frequency, as can be clearly seen in (4.15).

Finally, the signal to noise ratio for the one-dimensional problem is defined in the same manner as in (4.9).

5 Preconditioning by Gradient Scaling

5.1 Calculation of the gradient

To obtain the gradient of $\mathcal{J}(\boldsymbol{\nu})$ defined in (3.4) for the two-dimensional model problem, a first order perturbation analysis is considered as in [9, 14] where the Fréchet derivative is used to define the gradient. In the case when the relative permittivity and conductivity within the cylinder is discretized according to a finite expansion as described in (3.1), the finite gradients are given by

$$\begin{cases} \dfrac{\partial \mathcal{J}}{\partial \epsilon_i} = -\displaystyle\int_0^{2\pi}\int_{\mathcal{S}_i}\int_0^T \widetilde{\boldsymbol{E}}\cdot\partial_t\boldsymbol{E}\,\mathrm{d}t\,\mathrm{d}S\,\mathrm{d}\psi \\ \dfrac{\partial \mathcal{J}}{\partial \sigma_i} = -\displaystyle\int_0^{2\pi}\int_{\mathcal{S}_i}\int_0^T \widetilde{\boldsymbol{E}}\cdot\boldsymbol{E}\,\mathrm{d}t\,\mathrm{d}S\,\mathrm{d}\psi. \end{cases} \tag{5.1}$$

Here, the adjoint electric and magnetic fields $\widetilde{\boldsymbol{u}} = (\widetilde{\boldsymbol{E}}, \widetilde{\boldsymbol{H}})$ satisfy the adjoint Maxwell's equations

$$\mathcal{P}^\dagger\widetilde{\boldsymbol{u}} = \begin{cases} -\epsilon\partial_t\widetilde{\boldsymbol{E}} + \nabla\times\widetilde{\boldsymbol{H}} + \sigma\widetilde{\boldsymbol{E}} = \widetilde{\boldsymbol{J}} \\ -\partial_t\widetilde{\boldsymbol{H}} - \nabla\times\widetilde{\boldsymbol{E}} = \boldsymbol{0} \end{cases} \tag{5.2}$$

where $\widetilde{\boldsymbol{J}} = 2\hat{\boldsymbol{z}}\frac{\delta(\rho-a)}{\rho}(E(\phi,\psi,t) - E^{(\mathrm{m})}(\phi,\psi,t))$ is the source term corresponding to the solution of (2.1). Note that (5.2) is solved backwards in time for $t \in [0,T]$ and the "initial" conditions are $\widetilde{\boldsymbol{E}}(\boldsymbol{\rho},t) = \boldsymbol{0}$ for $t = T$ and $\boldsymbol{\rho} \in \mathbb{R}^2$.

In a practical situation where a finite number of measurement points (sensors) are used, the integrals over ϕ and ψ in (3.4) and (5.1) are replaced by finite sums and the source term in (5.2) is replaced by $\widetilde{\boldsymbol{J}} \to \widetilde{\boldsymbol{J}} \sum_j \delta(\phi - \phi_j)$.

The calculation of the gradient in one dimension is similar to the two-dimensional case above. However, since this is the wave splitting case the excitation is given by boundary conditions instead of sources. For the one-dimensional model problem the misfit functional is given by (3.7) and the discretization by (3.5). The finite gradients are given by

$$\begin{cases} \dfrac{\partial \mathcal{J}}{\partial \epsilon_i} = -\displaystyle\int_{\mathcal{S}_i}\int_0^T \widetilde{E}\partial_t E \,\mathrm{d}t\,\mathrm{d}x \\ \dfrac{\partial \mathcal{J}}{\partial \sigma_i} = -\displaystyle\int_{\mathcal{S}_i}\int_0^T \widetilde{E} E \,\mathrm{d}t\,\mathrm{d}x. \end{cases} \tag{5.3}$$

Here, the adjoint electric and magnetic fields $\widetilde{\boldsymbol{u}} = (\widetilde{E}, \widetilde{H})$ satisfy the adjoint Maxwell's equations

$$\mathcal{P}^\dagger \widetilde{\boldsymbol{u}} = \begin{cases} -\epsilon\partial_t \widetilde{E} + \partial_x \widetilde{H} + \sigma \widetilde{E} = 0 \\ -\partial_t \widetilde{H} + \partial_x \widetilde{E} = 0 \end{cases} \tag{5.4}$$

for $x \in \mathbb{R}$ and $t \in [0,T]$, together with the boundary conditions $\widetilde{E}_-(0,t) = E_-(0,t) - E_-^{(\mathrm{m})}(0,t)$ corresponding to the solution of (2.12). Note that (5.4) is solved backwards in time and the "initial" conditions are $\widetilde{E}(x,T) = 0$ for $x \geq 0$.

5.2 Preconditioning based on the Fisher information analysis

The optimization problem aims at minimizing the mistfit functionals (3.4) or (3.7), or equivalently, to minimize the negative loglikelihood functions (3.3) or (3.6). The Hessian of the negative loglikelihood function is given by

$$\boldsymbol{\mathcal{H}}(\boldsymbol{x}|\boldsymbol{\nu}) = -\frac{\partial^2 \log p(\boldsymbol{x}|\boldsymbol{\nu})}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}^{\mathrm{T}}} \tag{5.5}$$

and the Fisher information matrix is defined by

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\nu}) = \mathcal{E}\{\boldsymbol{\mathcal{H}}(\boldsymbol{x}|\boldsymbol{\nu})\} = -\mathcal{E}\{\frac{\partial^2 \log p(\boldsymbol{x}|\boldsymbol{\nu})}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}^{\mathrm{T}}}\} \tag{5.6}$$

where $[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ij} = -\mathcal{E}\{\frac{\partial^2 \log p(\boldsymbol{x}|\boldsymbol{\nu})}{\partial \nu_i \partial \nu_j}\}$, see *e.g.*, [18]. The preconditioning strategy is now to chose a parameter scaling such that the resulting Fisher information matrix becomes well-conditioned at some known background parameter value $\boldsymbol{\nu}$.

Since the Fisher information matrix is the mean value of the Hessian, it is expected that such a strategy will stabilize any gradient based numerical inversion

algorithm. In particular, by improving the conditioning of the Hessian and hence decreasing its eigenvalue spread, the convergence rate of the conjugate gradient and quasi-Newton methods are expected to be significantly improved since these optimization algorithms are usually initiated as "steepest-descent" algorithms.

Assume that a linear prescaling is employed such that $\boldsymbol{\nu} = \mathbf{K}\boldsymbol{\xi}$ where $\mathbf{K}$ is a positive definite matrix and $\boldsymbol{\xi}$ denotes the scaled parameters. By using $\frac{\partial}{\partial \boldsymbol{\xi}} = \mathbf{K}^{\mathrm{T}} \frac{\partial}{\partial \boldsymbol{\nu}}$ it is readily found that $\boldsymbol{\mathcal{I}}(\boldsymbol{\xi}) = \mathbf{K}^{\mathrm{T}} \boldsymbol{\mathcal{I}}(\boldsymbol{\nu}) \mathbf{K}$ and a natural choice of prescaler is $\mathbf{K} = \boldsymbol{\mathcal{I}}^{-1/2}(\boldsymbol{\nu})$ so that the scaled Fisher information matrix becomes a unit matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\xi}) = \mathbf{I}$. The corresponding gradient scaling is given by

$$\frac{\partial}{\partial \boldsymbol{\xi}} = \boldsymbol{\mathcal{I}}^{-1/2}(\boldsymbol{\nu}) \frac{\partial}{\partial \boldsymbol{\nu}}. \tag{5.7}$$

However, the inverse matrix operation above becomes highly ill-conditioned for high resolution images, and the scaling (5.7) is thus extremely sensitive to modelling errors. A better and more robust strategy is therefore to employ a diagonal scaling matrix $\mathbf{K}$ and requiring that the scaled Fisher information matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\xi})$ obtains a unit diagonal, $\mathrm{diag}\{\boldsymbol{\mathcal{I}}(\boldsymbol{\xi})\} = \mathbf{I}$, *cf.*, the Jacobi preconditioner in numerical analysis [11, 19]. The appropriate scaling matrix is then

$$\mathbf{K} = (\mathrm{diag}\{\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})\})^{-1/2} \tag{5.8}$$

and the resulting scaled Fisher information matrix is given by

$$[\boldsymbol{\mathcal{I}}(\boldsymbol{\xi})]_{ij} = \frac{1}{\sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}} \sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{jj}}} [\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ij}. \tag{5.9}$$

The corresponding gradient scaling is given by

$$\frac{\partial}{\partial \boldsymbol{\xi}} = \mathbf{K} \frac{\partial}{\partial \boldsymbol{\nu}} = (\mathrm{diag}\{\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})\})^{-1/2} \frac{\partial}{\partial \boldsymbol{\nu}}. \tag{5.10}$$

An exponential transformation is sometimes advantageous since it efficiently incorporates the a priori information inherently given by modelling a non-negative parameter [15, 30]. Assume that the non-negative parameter ν_i is given by

$$\nu_i = \alpha_i \mathrm{e}^{\xi_i/\beta_i} + \nu_{0i} \tag{5.11}$$

where α_i and ν_{0i} are a priori known constants and β_i is the scaling parameter. The gradient scaling is then given by $\frac{\partial}{\partial \xi_i} = \frac{\alpha_i}{\beta_i} \mathrm{e}^{\xi_i/\beta_i} \frac{\partial}{\partial \nu_i}$, and the diagonal elements of the scaled Fisher information is given by $[\boldsymbol{\mathcal{I}}(\boldsymbol{\xi})]_{ii} = \frac{\alpha_i^2}{\beta_i^2} \mathrm{e}^{2\xi_i/\beta_i} [\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}$. Hence, a robust Fisher information based Jacobi preconditioner with $[\boldsymbol{\mathcal{I}}(\boldsymbol{\xi})]_{ii} = 1$, is given by

$$\beta_i = \alpha_i \mathrm{e}^{\xi_i/\beta_i} \sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}} \tag{5.12}$$

at some known background parameter value ν_i.

Assuming that the known background corresponds to $\xi_i = 0$ ($\nu_i = \alpha_i + \nu_{0i}$), the appropriate scaling is then given by

$$\begin{cases} \beta_i = \alpha_i \sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}} \\ \dfrac{\partial}{\partial \xi_i} = \dfrac{1}{\sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}}} \mathrm{e}^{\xi_i/\beta_i} \dfrac{\partial}{\partial \nu_i}. \end{cases} \tag{5.13}$$

The resulting scaled Fisher information matrix at the background level $\boldsymbol{\xi} = \mathbf{0}$ is again given by (5.9).

6 Numerical Examples

6.1 The one-dimensional model problem

Consider the one-dimensional model problem as depicted in figure 2. An exponential transformation was used and a gradient scaling (5.13) was calculated based on the Fisher information analysis described in section 4.2. The Fisher information matrix is defined in (4.11). The calculation was performed for a homogeneous background with $\epsilon = 15$ ($\alpha = 1, \nu_0 = 14$) and $\sigma = 0.4\,\mathrm{S/m}$ ($\alpha = 0.4\eta_0, \nu_0 = 0$). The center frequency was $f_0 = 6\,\mathrm{GHz}$, the bandwidth $f_\mathrm{B} = 8\,\mathrm{GHz}$ and the signal to noise ratio $\mathrm{SNR} = 50\,\mathrm{dB}$. Note that $\epsilon_\mathrm{c} = 15 - \mathrm{i}1.2\omega_0/\omega$ gives a loss tangent of 0.08 at the center frequency. The computational boundary was $a = 15\lambda$ where $\lambda = 7.7\,\mathrm{mm}$ denotes the wavelength in the background medium at 10 GHz.

In figure 3 is shown the diagonal Fisher information $[\boldsymbol{\mathcal{I}}_{\epsilon\epsilon}]_{ii}, [\boldsymbol{\mathcal{I}}_{\sigma\sigma}]_{ii}$ and the Cramér-Rao bound[2] $[\boldsymbol{\mathcal{I}}_{\epsilon\epsilon}^{-1}]_{ii}, [\boldsymbol{\mathcal{I}}_{\sigma\sigma}^{-1}]_{ii}$ for the parameters ϵ and σ/ω_0 versus x/λ, plotted for different resolutions $d = \Delta x/\lambda = \{0.4, 0.5, 0.6\}$. Here, the conductivity parameter has been scaled as σ/ω_0 (where ω_0 is the center frequency) in order for the parameters ϵ and σ/ω_0 to obtain similar sensitivity, *cf.*, [12]. The graphs in figure 3 illustrate the fact that the diagonal Fisher information is virtually invariant to the resolution d, whereas the Cramér-Rao bound is extremely sensitive to the assumed discretization model. Hence, the diagonal Fisher information is a robust sensitivity measure which can be used as a basis for parameter scaling and preconditioning. Note that the reduced sensitivity is due to attenuation of the wave field. At the highest frequency $1.3\,\mathrm{dB}/\lambda$, resembling the results in figure 3 (yielding approximately 40 dB attenuation of the Fisher information at $x = 15\lambda$).

Next, we consider a numerical implementation of the one-dimensional inverse problem. An inversion algorithm was implemented based on a quasi-Newton algorithm using the BFGS formula and Golden section line search, see *e.g.*, [10], together with the gradient calculations and preconditioning that are described in section 5 above. The solutions to the related direct and adjoint electromagnetic problems were based on an implementation of the FDTD algorithm, see *e.g.*, [29], where the spatial resolution was 10 points per wave length. The signal to noise ratio was $\mathrm{SNR} = 50$ dB, and artificial noise was added correspondingly prior to the inversion.

[2]Here, $\left[\boldsymbol{\mathcal{I}}_{\nu\nu}^{-1}\right]_{ii}$ is used as a simplified notation for $\left[\left[\boldsymbol{\mathcal{I}}^{-1}\right]_{\nu\nu}\right]_{ii}$.

In figure 4 is shown the reconstruction for the one-dimensional model problem. The graphs show the parameters ϵ and σ/ω_0 versus x/λ, with and without preconditioning. The true parameter values are shown as a dashed line. The 4 columns illustrate the first 5 iterations in the reconstruction of ϵ with preconditioning (40 dB), ϵ without preconditioning (0 dB), σ/ω_0 with preconditioning (40 dB) and σ/ω_0 without preconditioning (0 dB), respectively. Here, an iteration is referred to as a calculation of a new search direction and a completed line search. As can be seen in these reconstructions, the scaled version of the algorithm (with preconditioning) has improved capabilities of finding objects in the interior of the material early in the iteration process. This is due to the fact that the gradient scaling takes into account the effect of losses (attenuation) and amplifies the gradient further inside the material.

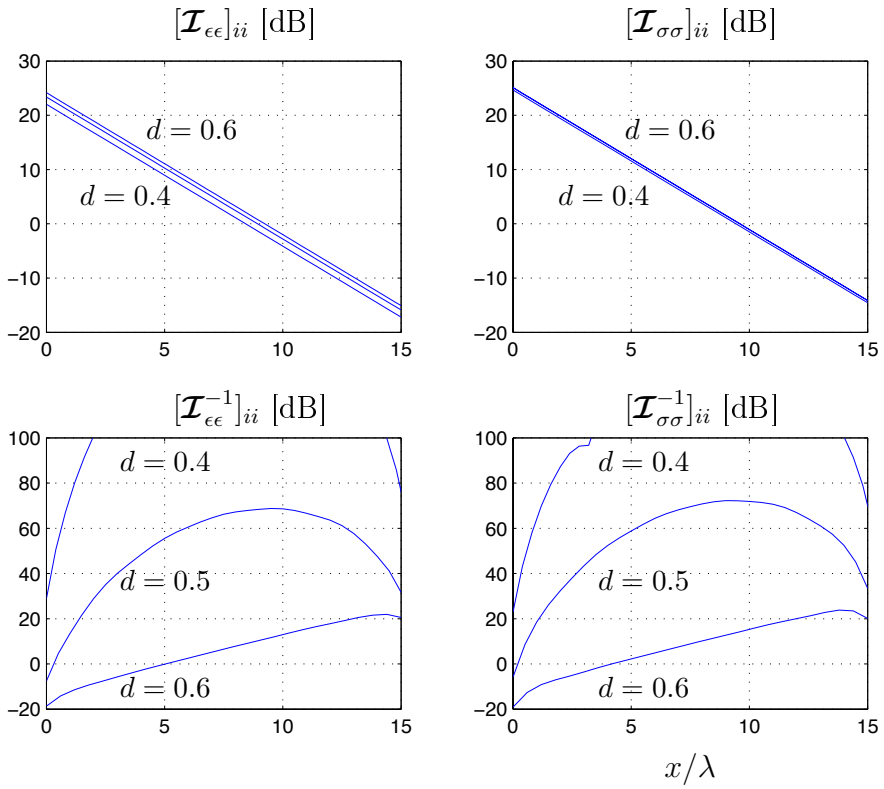

Figure 3: Fisher information and Cramér-Rao bound in the one-dimensional model problem. The graphs show $[\boldsymbol{\mathcal{I}}_{\epsilon\epsilon}]_{ii}$, $[\boldsymbol{\mathcal{I}}_{\sigma\sigma}]_{ii}$ and $[\boldsymbol{\mathcal{I}}^{-1}_{\epsilon\epsilon}]_{ii}$, $[\boldsymbol{\mathcal{I}}^{-1}_{\sigma\sigma}]_{ii}$ for the parameters ϵ and σ/ω_0 versus x/λ. The resolution is $d = \Delta x/\lambda = \{0.4, 0.5, 0.6\}$.

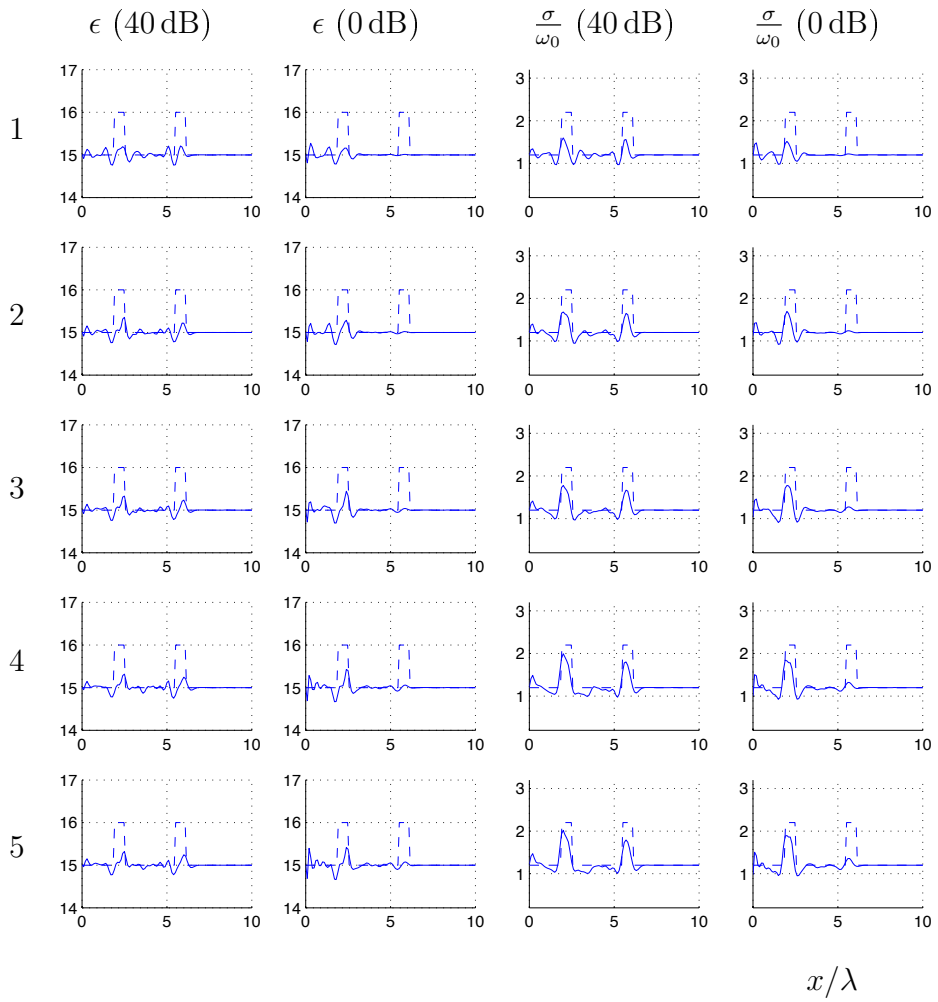

Figure 4: Reconstruction for the one-dimensional model problem. The graphs show the parameters ϵ and σ/ω_0 versus x/λ, respectively, with preconditioning (40 dB) and without preconditioning (0 dB), and iteration 1–5. The true parameter values are shown as a dashed line.

6.2 The two-dimensional model problem

6.2.1 Calculation of gradient scaling based on the Fisher Information

Consider the two-dimensional microwave tomography set-up depicted in figure 1. A gradient scaling (5.10) was calculated based on the Fisher information analysis described in section 4.1. The calculation was performed for a measurement cylinder radius $a = 0.1\,\mathrm{m}$, center frequency $f_0 = 1\,\mathrm{GHz}$, bandwidth $f_\mathrm{B} = 1\,\mathrm{GHz}$, and signal to noise ratio $\mathrm{SNR} = 80\,\mathrm{dB}$. The homogeneous background was modeled with $\epsilon = 78$ and $\sigma = 0.2\,\mathrm{S/m}$, corresponding to $\epsilon_\mathrm{c} = 78 - \mathrm{i}3.6\omega_0/\omega$ and a loss tangent of 0.046 at the center frequency. Note that the radius is $a = 4.4\lambda$ where $\lambda = 23\,\mathrm{mm}$ denotes the wavelength in the background medium at 1.5 GHz. The Fisher information matrix is defined in (4.2) and is circularly symmetrical for the circularly symmetrical background under consideration.

In figure 5 is shown the Cramér-Rao bound $[\boldsymbol{\mathcal{I}}_{\epsilon\epsilon}^{-1}]_{ii}$ and $[\boldsymbol{\mathcal{I}}_{\sigma\sigma}^{-1}]_{ii}$ as a function of radius for different scales of resolution $d = \sqrt{\Delta A_0}/\lambda$, where A_0 is the area of each circularly symmetrical pixel and λ the shortest wave length, *cf.*, [24]. Again, the conductivity parameter has been scaled as σ/ω_0 in order for the parameters ϵ and σ/ω_0 to obtain similar sensitivity. The figure 5 also shows the Cramér-Rao bound for the scaled parameters ξ_ϵ and ξ_σ defined by a linear scaling and (5.8). As can be seen in figure 5, the behaviour of the Cramér-Rao bound for the parameters ϵ and σ/ω_0 is very similar as for the parameters ξ_ϵ and ξ_σ. Hence, the simple scaling σ/ω_0 makes the parameters ϵ and σ/ω_0 be essentially "equal" in terms of sensitivity and estimation performance.

A good estimation accuracy requires that the Cramér-Rao bound is well below the required contrast level with respect to the background permittivity and conductivity. Note, however, that regularization of the inverse problem implies adding a priori information [17, 30], which in turn alleviates the resolution limit indicated by the Cramér-Rao bound analysis employed here (which does not take a priori information or regularization into account). In practice, when employing a particular regularization scheme, it is quite difficult to say what the "effective" resolution and estimation accuracy is. Figure 5 illustrates that the scaling (5.7) based on the Cramér-Rao bound is very sensitive to the assumed scale of resolution, and is hence very sensitive to the assumed pixel model or discretization. Experiments have also shown that the scaling (5.7) is very sensitive to a small shift of the discretization grid. In fact, the only situation when the scaling (5.7) can be used reliably, is when the assumed pixel model (discretization) is known to be correct. However, since this assumption is an "inverse crime" [17], the inverse Fisher information in (5.7) is not a suitable basis for the gradient scaling.

In figure 6 is shown the Fisher information diagonal elements $[\boldsymbol{\mathcal{I}}_{\epsilon\epsilon}]_{ii}$ and $[\boldsymbol{\mathcal{I}}_{\sigma\sigma}]_{ii}$ as a function of radius for different scales of resolution. Again, the figure 6 has been calculated and plotted for the parameters ϵ and σ/ω_0 showing that these parameters have similar sensitivity properties. The figure 6 also illustrates that the radial slope of the diagonal Fisher information is virtually invariant to the scale of resolution. Hence, the diagonal Fisher information employed in (5.10) is a suitable basis for scaling the gradient with respect to the radially varying loss (attenuation). In this particular example, it is concluded from figure 6 that a suitable approximation of the radial scaling is a straight line (in the dB scale) ranging approximately 12 dB from the interior to the boundary of the measurement cylinder. Note that the corresponding attenuation of a plane wave is $0.84\,\mathrm{dB}/\lambda$ giving an approximate radial scaling of 7.4 dB.

6.2.2 Microwave tomography simulation

Consider the two-dimensional microwave tomography set-up depicted in figure 7. Here, we have employed the numerical algorithm described in [8, 9, 16] which has

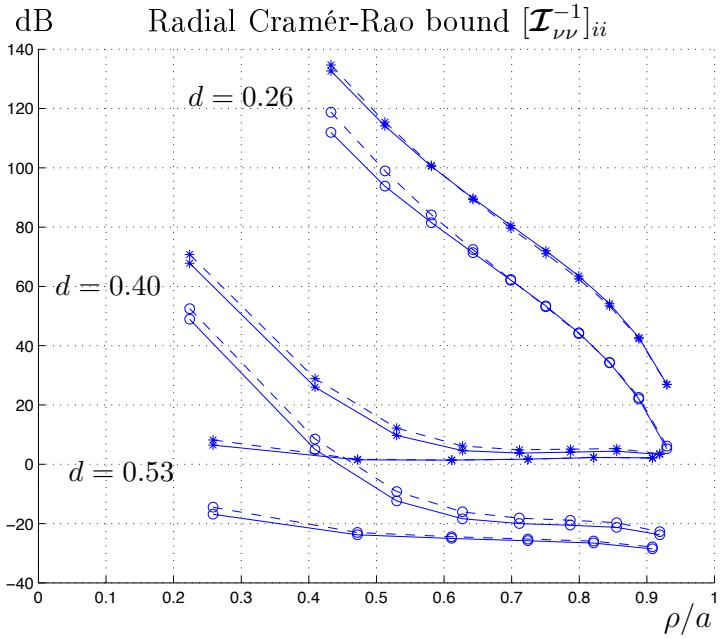

Figure 5: Cramér-Rao bound $[\mathcal{I}_{\nu\nu}^{-1}]_{ii}$ as a function of radius ρ for different scales of resolution, $d = \sqrt{\Delta A_0}/\lambda$. The solid "o"-lines indicate $[\mathcal{I}_{\epsilon\epsilon}^{-1}]_{ii}$ and the dashed "o"-lines $[\mathcal{I}_{\sigma\sigma}^{-1}]_{ii}$, corresponding to the parameters ϵ and σ/ω_0. The solid "$\star$"-lines indicate $[\mathcal{I}_{\epsilon\epsilon}^{-1}]_{ii}$ and the dashed "$\star$"-lines $[\mathcal{I}_{\sigma\sigma}^{-1}]_{ii}$, corresponding to the scaled parameters ξ_ϵ and ξ_σ.

been modified by using the Fisher information analysis and preconditioning technique as described above. The inversion algorithm was implemented based on a least-squares optimization approach [9, 13, 14] and a conjugate-gradient algorithm, see *e.g.*, [10], together with the gradient calculations and preconditioning that is described in section 5 above. The solutions to the related direct and adjoint electromagnetic problems were based on an implementation of the FDTD algorithm, see *e.g.*, [29], where the spatial resolution was 10 points per wave length. The signal to noise ratio was SNR = 10 dB, and artificial Gaussian noise was added correspondingly prior to the inversion.

In the numerical simulation, 17 antennas were placed on a circle with the radius 10 cm. The centre frequency of the electromagnetic Gaussian pulse was set to 1 GHz, and the bandwidth to 1 GHz. The theoretical expression (5.8) was approximated and used for the spatial gradient scaling (5.10) such that the scaling was ranging from 0 dB at the outer edge of the reconstruction domain and 12 dB at the centre (approximate linear radial function in dB).

Figure 7 a and b show the simulated true values of the relative permittivity ϵ and conductivity σ (in S/m), respectively, where the background is $\epsilon = 78$ and $\sigma = 0.2$ S/m. The scatterer consists of 4 cylindrical objects, where the cylindrical object with radius 16 mm $\approx 0.71\lambda$ is modeled by $\epsilon = 50$ and $\sigma = 1.8$ S/m and the 3 cylindrical objects with radius 8 mm $\approx 0.35\lambda$ are modeled by $\epsilon = 20$ and

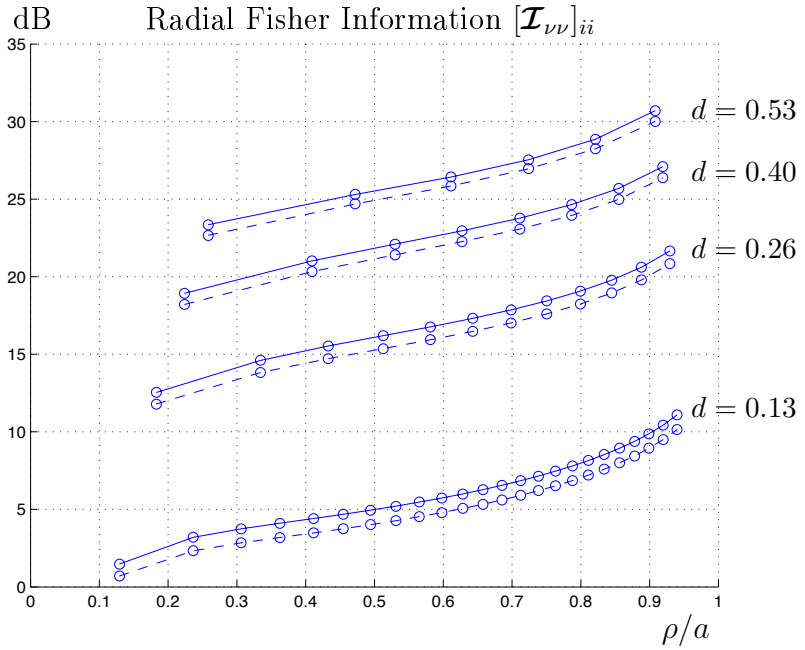

Figure 6: Fisher Information $[\boldsymbol{\mathcal{I}}_{\nu\nu}]_{ii}$ as a function of radius ρ for different scales of resolution, $d = \sqrt{\Delta A_0}/\lambda$. The solid lines indicate $[\boldsymbol{\mathcal{I}}_{\epsilon\epsilon}]_{ii}$ and the dashed lines $[\boldsymbol{\mathcal{I}}_{\sigma\sigma}]_{ii}$, corresponding to the parameters ϵ and σ/ω_0.

$\sigma = 3.4\,\mathrm{S/m}$. Here, $\lambda = 23\,\mathrm{mm}$ denotes the wavelength in the background medium at 1.5 GHz. In figure 7 c is shown the cost function $\sqrt{\mathcal{J}_n/\mathcal{J}_0}$ for the first 25 iterations. The solid line indicates the scaled version of the algorithm (12 dB) and the dashed line the algorithm without preconditioning (0 dB). Here, an iteration is referred to as a calculation of a new search direction and a completed line search.

The 4 columns in figure 7 d and e illustrate the first 25 iterations in the numerical reconstruction of the permittivity ϵ and the conductivity σ, respectively, with preconditioning (12 dB) and without preconditioning (0 dB). Similar to the one-dimensional numerical example above, it is clearly visible that the scaled version of the algorithm (with preconditioning) has improved capabilities of finding objects in the interior of the cylinder early in the iteration process. Again, this is due to the fact that the gradient scaling takes into account the effect of losses (attenuation) and amplifies the gradient in the vicinity of the cylinder center. The example also shows that the scaled version of the algorithm has a capability to avoid spurious local minima by decreasing the eigenvalue spread of the Hessian as discussed in section 5.2 above. In particular, see figure 7 d (reconstruction of ϵ) at iteration 10 and 25.

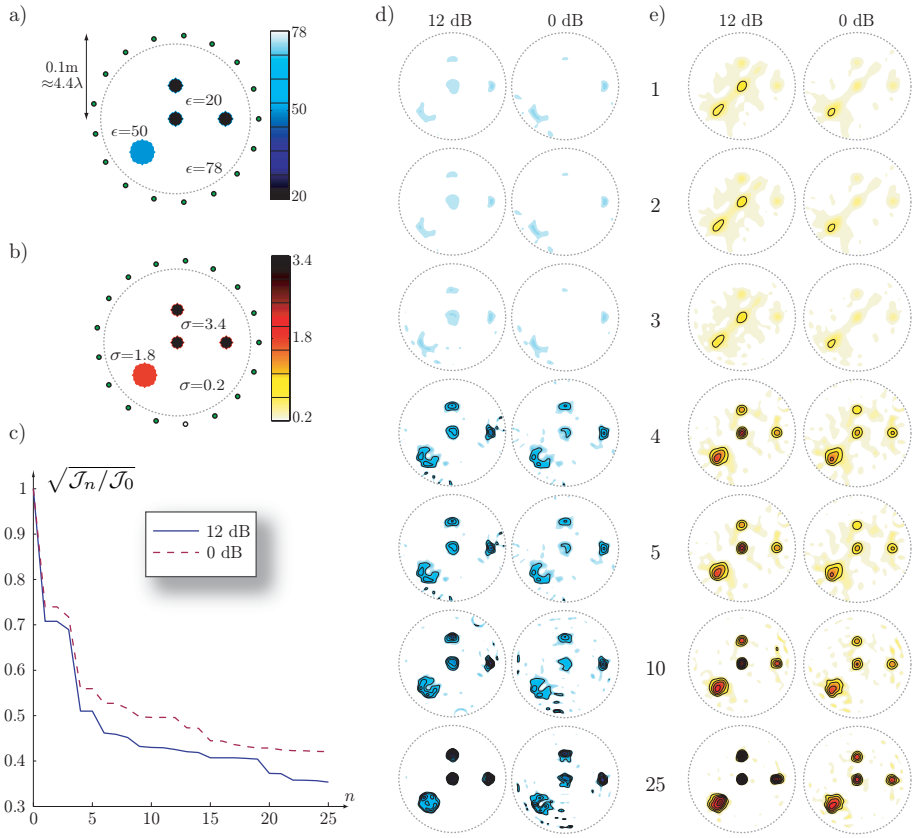

Figure 7: a) and b) Microwave tomography measurement set-up and simulated true values of the relative permittivity ϵ and conductivity σ, respectively. c) Cost function $\sqrt{\mathcal{J}_n/\mathcal{J}_0}$ for the first 25 iterations. d) and e) Contour plots depicting reconstructions of the permittivity ϵ and the conductivity σ, respectively, with preconditioning (12 dB) and without preconditioning (0 dB), and iterations 1–5, 10, 25.

7 Summary and conclusions

In this paper, the Fisher information analysis is introduced as a systematic approach to obtain a robust preconditioner for gradient based non-linear inverse scattering algorithms. One and two-dimensional inverse problems are considered where the permittivity and conductivity profiles are unknown and the input data consists of the scattered field over a certain bandwidth. A time-domain least-squares formulation is employed and the inversion algorithm is based on a conjugate gradient, or quasi-Newton algorithm together with an FDTD-electromagnetic solver.

In the first step of the preconditioning, the Fisher information analysis is performed to estimate the Hessian of the error functional corresponding to some known background. A robust preconditioner is then obtained by incorporating a parameter

scaling such that the scaled Fisher information has a unit diagonal. The preconditioner is robust in the sense that the scaling, *i.e.*, the diagonal Fisher information is virtually invariant to the numerical resolution and the discretization model that is employed. On the other hand, a full inversion of the Fisher information matrix (a calculation of the Cramér-Rao bound) is extremely sensitive in this respect and can not be used as a basis for the preconditioning.

Since the Fisher information matrix is the mean value of the Hessian, the proposed preconditioner has a capability to stabilize any gradient based numerical inversion algorithm. In particular, by improving the conditioning of the Hessian and hence decreasing its eigenvalue spread, the convergence rate of the conjugate gradient or quasi-Newton methods are improved since these optimization algorithms are usually initiated as "steepest-descent" algorithms.

Numerical examples of image reconstruction are included to illustrate the efficiency of the proposed technique. As can be seen in these reconstructions, the scaled version of the algorithm has improved capabilities of finding objects in the interior of the material as well as avoiding the problem with local minima.

Acknowledgement

The authors gratefully acknowledge the financial support by the Swedish Research Council and the Swedish Foundation for Strategic Research.

References

[1] A. Boström, G. Kristensson, and S. Ström. Transformation properties of plane, spherical and cylindrical scalar and vector wave functions. In V. V. Varadan, A. Lakhtakia, and V. K. Varadan, editors, *Field Representations and Introduction to Scattering*, Acoustic, Electromagnetic and Elastic Wave Scattering, chapter 4, pages 165–210. Elsevier Science Publishers, Amsterdam, 1991.

[2] M. Cheney and D. Isaacson. Issues in electrical impedance imaging. *IEEE Computational Science & Engineering*, pages 53–62, 1995.

[3] C. Cohen-Bacrice, Y. Goussard, and R. Guardo. Regularized reconstruction in electrical impedance tomography using a variance uniformization constraint. *IEEE Trans. Med. Imaging*, **16**(5), 562–571, oct 1997.

[4] S. L. Collier. Fisher information for a complex Gaussian random variable: Beamforming applications for wave propagation in a random medium. *IEEE Trans. Signal Process.*, **53**(11), 4236–4248, November 2005.

[5] A. J. Devaney and G. A. Tsihirintzis. Maximum likelihood estimation of object location in diffraction tomography. *IEEE Trans. Signal Process.*, **39**(3), 672–682, March 1991.

[6] A. Dogandzic and A. Nehorai. Cramér–Rao bounds for estimating range, velocity, and direction with an active array. *IEEE Trans. Signal Process.*, **49**(6), 1122–1137, June 2001.

[7] E. Fear and M. A. Stuchly. Microwave detection of breast cancer. *IEEE Trans. Microwave Theory Tech.*, **48**, 1854–1863, nov 2000.

[8] A. Fhager, P. Hashemzadeh, and M. Persson. Reconstruction quality and spectral content of an electromagnetic time-domain inversion algorithm. *IEEE Trans. Biomed. Eng.*, **53**(8), 1594–1604, August 2006.

[9] A. Fhager and M. Persson. Comparison of two image reconstruction algorithms for microwave tomography. *Radio Sci.*, **40**(RS3017), June 2005.

[10] R. Fletcher. *Practical Methods of Optimization.* John Wiley & Sons, Ltd., Chichester, 1987.

[11] A. Greenbaum. *Iterative Methods for Solving Linear Systems.* SIAM Press, Philadelphia, 1997.

[12] M. Gustafsson and S. Nordebo. Bandwidth, Q-factor, and resonance models of antennas. *Progress in Electromagnetics Research*, **62**, 1–20, 2006.

[13] M. Gustafsson. *Wave Splitting in Direct and Inverse Scattering Problems.* PhD thesis, Lund Institute of Technology, Department of Electromagnetic Theory, P.O. Box 118, S-221 00 Lund, Sweden, 2000. `http://www.eit.lth.se`.

[14] M. Gustafsson and S. He. An optimization approach to two-dimensional time domain electromagnetic inverse problems. *Radio Sci.*, **35**(2), 525–536, 2000.

[15] T. M. Habashy and A. Abubakar. A general framework for constraint minimization for the inversion of electromagnetic measurements. *Progress in Electromagnetics Research*, **46**, 265–312, 2004.

[16] P. Hashemzadeh, A. Fhager, and M. Persson. Experimental investigation of an optimization approach to microwave tomography. *Electromagnetic Biology and Medicine*, **25**(1), 1–12, 2006.

[17] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems.* Springer-Verlag, New York, 2005.

[18] S. M. Kay. *Fundamentals of Statistical Signal Processing, Estimation Theory.* Prentice-Hall, Inc., NJ, 1993.

[19] C. T. Kelley. *Iterative Methods for Linear and Nonlinear Equations.* SIAM Press, Philadelphia, 1995.

[20] P. Kosmas and C. Rappaport. A FDTD-based time reversal for microwave breast cancer detection – localization in three dimensions. *IEEE Trans. Microwave Theory Tech.*, **54**, 1921–1927, apr 2006.

[21] P. M. Meaney, M. W. Fanning, T. Raynolds, C. J. Fox, Q. Q. Fang, C. A. Kogel, S. P. Poplack, and K. D. Paulsen. Initial clinical experience with microwave breast imaging in women with normal mammography. *Academic Radiology*, **14**(2), 207–218, feb 2007.

[22] K. S. Miller. *Complex Stochastic Processes.* Addison–Wesley Publishing Company, Inc., 1974.

[23] P. S. Naidu and A. Buvaneswari. A study of Cramér–Rao bounds on object shape parameters from scattered field. *IEEE Trans. Signal Process.*, **47**(5), 1478–1481, May 1999.

[24] S. Nordebo, M. Gustafsson, and B. Nilsson. Fisher information analysis for two-dimensional microwave tomography. *Inverse Problems*, **23**, 859–877, 2007.

[25] J. G. Proakis. *Digital Communications.* McGraw-Hill, third edition, 1995.

[26] Y. Saad. *Iterative Methods for Sparse Linear Systems.* PWS Publishing Company, Boston, 1996.

[27] J. M. Sill and E. C. Fear. Tissue sensing adaptive radar for breast cancer detection – experimental investigation of simple tumor models. *IEEE Trans. Microwave Theory Tech.*, **53**(11), 3312–3319, nov 2005.

[28] S. T. Smith. Statistical resolution limits and the complexified Cramér–Rao bound. *IEEE Trans. Signal Process.*, **53**(5), 1597–1609, May 2005.

[29] A. Taflove and S. C. Hagness. *Computational electrodynamics: The Finite-Difference Time-Domain Method.* Artech House, Boston, London, 2000.

[30] A. Tarantola. *Inverse problem theory and methods for model parameter estimation.* Society for Industrial and Applied Mathematics, Philadelphia, 2005.

[31] G. A. Tsihirintzis and A. J. Devaney. Maximum likelihood estimation of object location in diffraction tomography, Part II; strongly scattering objects. *IEEE Trans. Signal Process.*, **39**(6), 1466–1470, June 1991.