A priori modelling for gradient based inverse scattering algorithms

Nordebo, Sven; Gustafsson, Mats

2008

CODEN:LUTEDX/(TEAT-7165)/1-21/(2008)

A priori modelling for gradient based inverse scattering algorithms

Sven Nordebo and Mats Gustafsson

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

Mats Gustafsson
Mats.Gustafsson@eit.lth.se

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

Editor: Gerhard Kristensson

Abstract

This paper presents a Fisher information based Bayesian approach to analysis and design of the regularization and preconditioning parameters used with gradient based inverse scattering algorithms. In particular, a one-dimensional inverse problem is considered where the permittivity and conductivity profiles are unknown and the input data consist of the scattered field over a certain bandwidth. A priori parameter modelling is considered by nonlinear exponential and arctangential parameter scalings and robust preconditioners are obtained by choosing the related scaling parameters based on a Fisher information analysis of the known background. The Bayesian approach and a principal parameter (singular value) analysis of the stochastic Cramér-Rao bound provide a natural interpretation of the regularization that is necessary to achieve stable inversion, as well as an indicator to predict the feasibility of achieving successful reconstruction in a given problem set-up. In particular, the Tikhonov regularization scheme is put into a Bayesian estimation framework. A time-domain least-squares inversion algorithm is employed which is based on a quasi-Newton algorithm together with an FDTD-electromagnetic solver. Numerical examples are included to illustrate and verify the analysis.

1 Introduction

Inverse scattering problems offer a variety of applications in *e.g.*, medicine, nondestructive testing, surveillance, quantum mechanics, acoustics and optics. These problems are in general ill-posed, *i.e.*, they are not well-posed in the sense of existence, uniqueness, and the solution being a continuous function of the data [1, 4, 12, 15, 16, 19, 21, 33]. Since the solution does not generally depend continuously on the data, it is almost always necessary to employ some kind of *regularization* to control the imaging error, see *e.g.*, [1, 4, 19]. Typically, the number of degrees of freedom (NDF) pertaining the number of significant singular values of a linear operator is a very useful tool, see *e.g.*, [1, 2, 27, 28]. The NDF, which in many cases is virtually independent of the noise level, can be used to estimate the number of retrievable parameters of an object, and hence the resolution. Another approach is the Tikhonov regularization [16], which controls the modelling error as well as a suitable norm of the image itself. However, these approaches do not quantify the amount of a priori information that is inherent with the regularization scheme.

A Fisher information analysis and the Cramér-Rao bound provides a very useful instrument for sensitivity analysis of various wave propagation phenomena, and which facilitates valuable physical interpretations, see *e.g.*, [3, 5, 6, 11, 14, 23–26, 31, 34]. 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, 34]. Previously, the Cramér-Rao bound has been employed as an analytical tool to investigate the one-dimensional inverse scattering problem of multilayer structures [11], and a canonical two-dimensional microwave tomography set-up is analyzed in [24]. The Fisher information analysis has also been used with electromagnetic inverse source problems, see *e.g.*, [25, 26]. In [11, 24–26], 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.

The resolution limit of an ill-posed imaging inverse problem is often in practice restricted to about a half wave length, see *e.g.*, [11, 24–26]. In this view, the regularization scheme supplies a priori information to an extent that the resolution limit may be achieved, while at the same time stability is maintained. On the other hand, if stability requires that too much a priori information is supplied, the only thing that can be retrieved is the a priori information itself and the image can not be resolved. In order for an inverse problem to be feasible, there must be a balance between these two contradictory requirements. In this paper, the Bayesian approach with MAP estimation [14, 16], and a principal parameter (singular value) analysis of the stochastic Cramér-Rao bound [35] is employed to provide a natural interpretation of the regularization that is necessary to achieve stable inversion, as well as an indicator to predict the feasibility of achieving successful reconstruction in a given problem set-up. In particular, the Tikhonov regularization scheme is put into the Bayesian estimation framework.

Closely related to the regularization is the concept of preconditioning, which is often used to accelerate the convergence of iterative methods, see *e.g.*, [10, 18]. There are very few theoretical results on preconditioning and finding a good preconditioner is sometimes considered as a combination of art and science [30]. In this paper, a robust preconditioner is considered which is obtained by incorporating a parameter scaling (or transformation) such that the scaled Fisher information has a unit diagonal at some known background parameter value, *cf.*, the Jacobi preconditioner in numerical analysis [10, 18]. This 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, see [7] for a detailed study about this issue. Here, an a priori parameter modelling is considered by nonlinear exponential [33] and arctangential parameter scalings.

The purpose of this paper is to present a Fisher information based Bayesian approach to analysis and design of the regularization and preconditioning parameters used with gradient based inverse scattering algorithms. A one-dimensional inverse scattering problem is considered as it offers a natural introduction to parameter identification and wave splitting techniques *cf.*, [12, 20]. A time-domain least-squares inversion algorithm [12] based on a quasi-Newton algorithm [9] together with an FDTD-electromagnetic solver [32] has been employed in order to generate the numerical examples.

The rest of the paper is outlined as follows. In section 2 is presented the one-dimensional inverse scattering problem with basic gradient expressions. In section 3 is presented a conditional statistical analysis containing Maximum Likelihood (ML) estimation and Fisher information analysis. In section 4 is presented the robust Fisher information based preconditioning strategy covering linear, exponential and arctangential parameter scalings. In section 5 is given the Bayesian, or Maximum A Posteriori (MAP) framework for Tikhonov regularization and a principal (SVD) parameter analysis. Section 5 also covers a discussion about nonlinear a priori

models in general, and the lognormal distribution in particular. Section 6 contains the numerical examples and section 7 the summary and conclusions.

2 The one-dimensional inverse scattering problem

Throughout the paper, 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 $\boldsymbol{H} = H(x)\hat{\boldsymbol{y}}$. Further, 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 Problem formulation

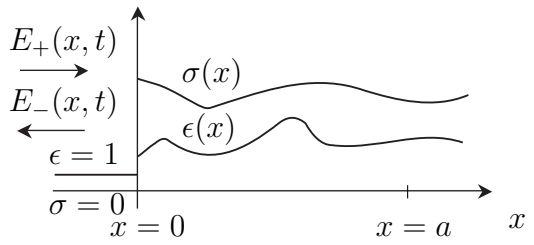

Figure 1: One-dimensional inverse problem for an isotropic half space with relative permittivity $\epsilon(x)$ and conductivity $\sigma(x)$.

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 Fig.1. 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[1]

$$\mathcal{P}\boldsymbol{u} = \begin{cases} \epsilon\partial_t E - \partial_x H + \sigma E = 0 \\ \partial_t H - \partial_x E = 0 \end{cases} \tag{2.1}$$

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 $E_+(0, t) = E_+^{(\mathrm{m})}(t)$ for $x = 0$ and $t \in [0, T]$, where the superscript $(\cdot)^{(\mathrm{m})}$ denotes measured field quantities.

[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)$^{1/2}$.

In inverse problems and microwave tomography, see *e.g.*, [8, 12, 13], it is common to employ the following least squares misfit functional

$$\mathcal{J} = \frac{1}{2}\int_0^T (|E(0,t) - E^{(\mathrm{m})}(t)|^2 + |H(0,t) - H^{(\mathrm{m})}(t)|^2)\,\mathrm{d}t \tag{2.2}$$

where the aim is to minimize $\mathcal{J}$ with respect to the material parameters ϵ and σ. Using the split fields $E_\pm = (E \mp H)/2$, the misfit functional (2.2) becomes

$$\mathcal{J}(\epsilon,\sigma) = \int_0^T |E_-(0,t) - E_-^{(\mathrm{m})}(t)|^2\,\mathrm{d}t \tag{2.3}$$

where the boundary conditions of (2.1) have been used.

It is assumed that 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 attenuation is assumed to be very high at the computational boundary $x = a$. The relative permittivity and conductivity within the material is discretized according to the finite expansions

$$\begin{cases} \epsilon(x) = \displaystyle\sum_{i=1}^I \epsilon_i \chi_i(x) \\ \sigma(x) = \displaystyle\sum_{i=1}^I \sigma_i \chi_i(x) \end{cases} \tag{2.4}$$

where ϵ_i and σ_i are the optimization variables and $\chi_i(x)$ the characteristic function for pixel $\mathcal{S}_i$, *i.e.*, $\chi_i(x) = 1$ if $x \in \mathcal{S}_i$ and $\chi_i(x) = 0$ if $x \notin \mathcal{S}_i$.

2.2 Derivation of the gradient

To obtain the gradient of $\mathcal{J}(\epsilon,\sigma)$ defined in (2.3), a first order perturbation analysis is considered as in [8, 12, 13]. Here, let $(\epsilon,\sigma) \to (\epsilon,\sigma) + \delta(\epsilon',\sigma')$ and $(E,H) \to (E,H) + \delta(E',H') + \mathcal{O}(\delta^2)$ where $\|\mathcal{O}(\delta^2)\| \le C\delta^2$ as $\delta \to 0$. The incremental fields $\boldsymbol{u}' = (E',H')$ satisfy Maxwell's equations

$$\mathcal{P}\boldsymbol{u}' = \begin{cases} \epsilon\partial_t E' - \partial_x H' + \sigma E' = -\epsilon'\partial_t E - \sigma' E \\ \partial_t H' - \partial_x E' = 0 \end{cases} \tag{2.5}$$

where the boundary conditions are $E'_+(0,t) = 0$ and E the solution to (2.1). The first variation $\delta\mathcal{J}$ of (2.3) is given by

$$\delta\mathcal{J}(\epsilon,\sigma) = 2\int_0^T (E_-(0,t) - E_-^{(\mathrm{m})}(t))E'_-(0,t)\,\mathrm{d}t \tag{2.6}$$

where $E'_-(0,t)$ is the solution to (2.5).

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

$$\mathcal{P}^\dagger\widetilde{\boldsymbol{u}} = \begin{cases} -\epsilon\partial_t\widetilde{E} + \partial_x\widetilde{H} + \sigma\widetilde{E} = 0 \\ -\partial_t\widetilde{H} + \partial_x\widetilde{E} = 0 \end{cases} \tag{2.7}$$

for $x \in \mathbb{R}$ and $t \in [0,T]$, together with the boundary conditions $\widetilde{E}_-(0,t) = E_-(0,t) - E_-^{(\mathrm{m})}(t)$ yielded from the solution of (2.1). Note that (2.7) is solved backwards in time and the "initial" conditions are $\widetilde{E}(x,T) = 0$ for $x \geq 0$.

By defining the inner product

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \int_{\mathbb{R}} \int_0^T \boldsymbol{u} \cdot \boldsymbol{v} \, \mathrm{d}t \, \mathrm{d}x \tag{2.8}$$

and by employing (2.1) and (2.7), it can be readily verified that

$$\langle \widetilde{\boldsymbol{u}}, \mathcal{P}\boldsymbol{u} \rangle - \langle \mathcal{P}^\dagger \widetilde{\boldsymbol{u}}, \boldsymbol{u} \rangle = \int_0^T (\widetilde{H}(0,t)E(0,t) + \widetilde{E}(0,t)H(0,t)) \, \mathrm{d}t \tag{2.9}$$

where all fields are assumed to vanish for $x \to \infty$. Using the split fields $E_\pm = (E \mp H)/2$ as well as the boundary conditions $E'_+(0,t) = 0$ and $\widetilde{E}_-(0,t) = E_-(0,t) - E_-^{(\mathrm{m})}(t)$, the relation (2.9) yields

$$\langle \widetilde{\boldsymbol{u}}, \mathcal{P}\boldsymbol{u}' \rangle - \langle \mathcal{P}^\dagger \widetilde{\boldsymbol{u}}, \boldsymbol{u}' \rangle = 2 \int_0^T (E_-(0,t) - E_-^{(\mathrm{m})}(t)) E'_-(0,t)) \, \mathrm{d}t. \tag{2.10}$$

Since $\langle \mathcal{P}^\dagger \widetilde{\boldsymbol{u}}, \boldsymbol{u}' \rangle = 0$, it follows from (2.10) and (2.6) that $\langle \widetilde{\boldsymbol{u}}, \mathcal{P}\boldsymbol{u}' \rangle = \delta\mathcal{J}$, and hence by (2.5)

$$\delta\mathcal{J}(\epsilon, \sigma) = \langle \widetilde{\boldsymbol{u}}, \mathcal{P}\boldsymbol{u}' \rangle = -\int_{\mathbb{R}} \int_0^T (\epsilon' \partial_t E + \sigma' E) \widetilde{E} \, \mathrm{d}t \, \mathrm{d}x. \tag{2.11}$$

The gradients $G_\epsilon(x)$ and $G_\sigma(x)$ are defined by the Fréchet derivative $\delta\mathcal{J} = \langle G_\epsilon(x), \epsilon'(x) \rangle + \langle G_\sigma(x), \sigma'(x) \rangle$ where $\langle u, v \rangle = \int_{\mathbb{R}} uv \, \mathrm{d}x$. Hence,

$$\begin{cases} G_\epsilon(x) = -\displaystyle\int_0^T \widetilde{E} \partial_t E \, \mathrm{d}t \\ G_\sigma(x) = -\displaystyle\int_0^T \widetilde{E} E \, \mathrm{d}t. \end{cases} \tag{2.12}$$

Assume now that the relative permittivity and conductivity are discretized as in (2.4), where ϵ_i and σ_i are the optimization variables and $\mathcal{S}_i$ the disjoint intervals. By inserting $\epsilon'(x) = \chi_i(x) \, \mathrm{d}\epsilon_i$ and $\sigma'(x) = \chi_i(x) \, \mathrm{d}\sigma_i$ in (2.11), the finite gradients are given by

$$\begin{cases} \dfrac{\partial \mathcal{J}}{\partial \epsilon_i} = -\displaystyle\int_{\mathcal{S}_i} \int_0^T \widetilde{E} \partial_t E \, \mathrm{d}t \, \mathrm{d}x \\ \dfrac{\partial \mathcal{J}}{\partial \sigma_i} = -\displaystyle\int_{\mathcal{S}_i} \int_0^T \widetilde{E} E \, \mathrm{d}t \, \mathrm{d}x. \end{cases} \tag{2.13}$$

3 Conditional Statistical Analysis

3.1 Maximum Likelihood Estimation

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

$$\mathcal{L}E = \left\{ \partial_x^2 + k^2 \right\} E = 0 \tag{3.1}$$

where $k = k_0\sqrt{\epsilon_{\rm c}} = \omega\sqrt{\epsilon_{\rm c}}$, $\epsilon_{\rm c}(x) = \epsilon(x) - {\rm i}\sigma(x)/\omega$ and the boundary conditions are $E_+(0,f) = E_+^{(\rm m)}(f)$. Here, the Fourier transform is given by $E(x,f) = \int_{-\infty}^{\infty} E(x,t){\rm e}^{-{\rm i}2\pi ft}\,{\rm d}t$.

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

$$E(x,p) = \frac{1}{\tau}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} E(x,t){\rm e}^{-{\rm i}2\pi\frac{p}{\tau}t}\,{\rm d}t \tag{3.2}$$

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 $\tau E(x,p) \to E(x,f)|_{f=\frac{p}{\tau}}$ as $\tau \to \infty$.

Consider the following statistical measurement model for the time-domain

$$E_-^{(\rm m)}(t) = E_-(0,t) + N(t),\ t \in [-\tau/2, \tau/2] \tag{3.3}$$

where $N(t)$ is zero mean Gaussian noise with correlation function $r_{\rm N}(\Delta t) = \mathcal{E}\left\{N(t+\Delta t)N(t)\right\}$ and power spectral density $R_{\rm N}(f) = \int_{-\infty}^{\infty} r_{\rm N}(t){\rm e}^{-{\rm i}2\pi ft}\,{\rm d}t$. Here, $\mathcal{E}\left\{\cdot\right\}$ denotes the expectation operator.

A discrete measurement model is now obtained by first creating the complex (time-domain) Hilbert pair [29] corresponding to (3.3), and then considering the corresponding 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, (3.3) yields

$$2E_-^{(\rm m)}(p) = 2E_-(0,p) + N_p,\ p \geq 0 \tag{3.4}$$

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

$$\mathcal{E}\left\{N_p^* N_{p'}\right\} = \frac{1}{\tau}4R_{\rm N}(\frac{p}{\tau})\delta_{pp'},\ p \geq 0 \tag{3.5}$$

where $(\cdot)^*$ denotes the complex conjugate and 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, $t_{\rm N} \ll \tau$.

Let $\boldsymbol{\nu} = [\boldsymbol{\epsilon}^{\rm T}\boldsymbol{\sigma}^{\rm T}]^{\rm T}$ denote the parameter vector with elements ϵ_i and σ_i as defined in (2.4) and let $\boldsymbol{x}$ denote the measurement vector $\boldsymbol{x} = \{2E_-^{(\rm m)}(p)\}$ with probability density function $p(\boldsymbol{x}|\boldsymbol{\nu})$. Since the measurement noise N_p is an uncorrelated complex Gaussian random process, the negative loglikelihood function [14, 16, 17] is given by

$$\begin{aligned} -\log p(\boldsymbol{x}|\boldsymbol{\nu}) &= b + \lim_{\tau\to\infty}\sum_{p=0}^{\infty}\frac{\tau}{R_{\rm N}(\frac{p}{\tau})}|E_-(0,p) - E_-^{(\rm m)}(p)|^2 \\ &= b + \frac{1}{2}\int_{-\infty}^{\infty}\frac{1}{R_{\rm N}(f)}|E_-(0,f) - E_-^{(\rm m)}(f)|^2\,{\rm d}f = b + \frac{1}{2N_0}\mathcal{J}(\boldsymbol{\nu}) \end{aligned} \tag{3.6}$$

where the power spectral density is assumed to be a constant $R_{\rm N}(f) = N_0$ over the relevant bandwidth, b a constant, and $\mathcal{J}(\boldsymbol{\nu})$ the misfit functional defined in (2.3).

Hence, with the Gaussian noise model adopted here, the optimization problem stated in section 2 is equivalent to the classical Maximum Likelihood (ML) criterion [17].

In order to define the signal to noise ratio, it is assumed that the incident signal is a normalized Gaussian pulse with

$$\begin{cases} E_+^2(0,t) = \frac{1}{\sqrt{2\pi}\sigma_\mathrm{t}} \mathrm{e}^{-t^2/2\sigma_\mathrm{t}^2} \\ E_+^2(0,f) = \frac{1}{\sqrt{2\pi}\sigma_\mathrm{f}} \mathrm{e}^{-f^2/2\sigma_\mathrm{f}^2} \end{cases} \tag{3.7}$$

which has been centered at the origin (base-band) for simplicity. Note that $\sigma_\mathrm{t}\sigma_\mathrm{f} = 1/4\pi$. 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,

$$\frac{\frac{1}{T}\int E_+^2(0,t)\,\mathrm{d}t}{\int R_\mathrm{N}(f)\,\mathrm{d}f} \approx \frac{\int E_+^2(0,t)\,\mathrm{d}t}{t_\mathrm{B} 4N_0 f_\mathrm{B}} \approx \frac{1}{4N_0} = \mathrm{SNR} \tag{3.8}$$

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

3.2 Fisher information analysis

The Fisher information matrix [17] for the parameters ϵ_i and σ_i based on the statistical measurement model (3.4) is given by

$$[\boldsymbol{\mathcal{I}}_{\nu\zeta}]_{ij} = 2\,\mathrm{Re}\sum_{p=0}^{\infty} \frac{\tau}{R_\mathrm{N}(\frac{p}{\tau})} \frac{\partial E_-^*(0,p)}{\partial \nu_i} \frac{\partial E_-(0,p)}{\partial \zeta_j} \tag{3.9}$$

where ν and ζ are either ϵ or σ, and $i,j = 1,\ldots,I$. In the limit as $\tau \to \infty$, the expression (3.9) becomes

$$[\boldsymbol{\mathcal{I}}_{\nu\zeta}]_{ij} = \int_{-\infty}^{\infty} \frac{1}{R_\mathrm{N}(f)} \frac{\partial E_-^*(0,f)}{\partial \nu_i} \frac{\partial E_-(0,f)}{\partial \zeta_j}\,\mathrm{d}f. \tag{3.10}$$

The differentiated field, or sensitivity field, satisfy the wave equation (3.1)

$$\mathcal{L}\frac{\partial E}{\partial \nu_i} = \left\{\partial_x^2 + k^2\right\}\frac{\partial E}{\partial \nu_i} = \mathrm{i}\omega g_\nu \chi_i E \tag{3.11}$$

where $g_\nu = 1$ if $\nu = \sigma$ and $g_\nu = \mathrm{i}\omega$ if $\nu = \epsilon$. Note that the solution E of (3.1) now appears in the source term of (3.11).

After solving (3.11) and evaluating (3.10), the total Fisher information matrix is assembled as

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\nu}) = \begin{pmatrix} \boldsymbol{\mathcal{I}}_{\epsilon\epsilon} & \boldsymbol{\mathcal{I}}_{\epsilon\sigma} \\ \boldsymbol{\mathcal{I}}_{\sigma\epsilon} & \boldsymbol{\mathcal{I}}_{\sigma\sigma} \end{pmatrix}. \tag{3.12}$$

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

$$E(x,f) = E_+(0,f)\frac{2\eta}{\eta+1}\mathrm{e}^{-\mathrm{i}kx} \tag{3.13}$$

where $\eta = 1/\sqrt{\epsilon_c}$. The corresponding background 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

$$G(x, 0) = \frac{1}{\mathrm{i}\omega + \mathrm{i}k} \mathrm{e}^{-\mathrm{i}kx} \tag{3.14}$$

for $x \geq 0$. The sensitivity field can now be expressed as

$$\frac{\partial E(x, f)}{\partial \nu_i} = -\mathrm{i}\omega g_\nu \int_{\mathcal{S}} G(x, x')\chi_i(x')E(x', f)\,\mathrm{d}x', \tag{3.15}$$

or

$$\frac{\partial E_-(0, f)}{\partial \nu_i} = -\mathrm{i}\omega g_\nu \int_{\mathcal{S}_i} G(x', 0)E(x', f)\,\mathrm{d}x' \tag{3.16}$$

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 (3.13) and (3.14) in (3.16) and evaluating the integral, the sensitivity field for a homogenous background is given by

$$\frac{\partial E_-(0, f)}{\partial \nu_i} = -2g_\nu \frac{E_+(0, f)}{(1 + \sqrt{\epsilon_c})^2} \frac{\sin(k\Delta x)}{k\Delta x} \Delta x \mathrm{e}^{-\mathrm{i}k(2i-1)\Delta x} \tag{3.17}$$

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

4 Parameter scaling

A robust preconditioning as well as an efficient a priori modelling of the inverse problem may be obtained by appropriate parameter scaling. The basic ideas are briefly outlined below for the present context, see also [7, 14, 33].

The optimization problem aims at minimizing the mistfit functional (2.3), or equivalently, to minimize the negative loglikelihood function (3.6). The Hessian of the negative loglikelihood function is given by

$$\boldsymbol{\mathcal{H}}(\boldsymbol{x}|\boldsymbol{\nu}) = -\frac{\partial^2 \log p(\boldsymbol{x}|\boldsymbol{\nu})}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}^{\mathrm{T}}} \tag{4.1}$$

and the Fisher information matrix is defined by

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\nu}) = \mathcal{E}\{\boldsymbol{\mathcal{H}}(\boldsymbol{x}|\boldsymbol{\nu})\} = -\mathcal{E}\{\frac{\partial^2 \log p(\boldsymbol{x}|\boldsymbol{\nu})}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}^{\mathrm{T}}}\}, \tag{4.2}$$

see *e.g.*, [17].

A robust preconditioner is obtained by incorporating a parameter scaling (or transformation) such that the scaled Fisher information has a unit diagonal at some known background parameter value $\boldsymbol{\nu}$, *cf.*, the Jacobi preconditioner in numerical analysis [10, 18]. Since the Fisher information matrix is the mean value of the Hessian in the corresponding Maximum Likelihood estimation problem, it is expected that such a strategy will stabilize any gradient based numerical inversion algorithm and

that the problem with local minima should be alleviated [7]. 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, see [7] for a detailed study about this issue.

Consider the following linear and nonlinear parameter models

$$\begin{array}{lcll} \nu_i & = & \xi_i/\beta_i & \text{Linear} \\ \nu_i & = & \alpha_i \mathrm{e}^{\xi_i/\beta_i} + \nu_{0i} & \text{Exponential} \\ \nu_i & = & \alpha_i \arctan(\xi_i/\beta_i + \bar{\xi}_i) + \nu_{0i} & \text{Arctangential} \end{array} \tag{4.3}$$

where ξ_i is the new optimization variable and β_i the scaling constant. Here, α_i, ν_{0i}, and $\bar{\xi}_i$ are a priori known model constants.

With the linear scaling, the gradient is given by $\frac{\partial}{\partial \xi_i} = G_i \frac{\partial}{\partial \nu_i}$ and the scaled Fisher information is $[\boldsymbol{\mathcal{I}}(\boldsymbol{\xi})]_{ij} = G_i G_j [\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ij}$ where $G_i = \frac{1}{\beta_i}$. Hence, a robust Fisher information based Jacobi preconditioner with $[\boldsymbol{\mathcal{I}}(\boldsymbol{\xi})]_{ii} = 1$, is given by

$$\begin{cases} \beta_i = \sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}} \\ \dfrac{\partial}{\partial \xi_i} = \dfrac{1}{\sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}}} \dfrac{\partial}{\partial \nu_i} \end{cases} \tag{4.4}$$

and the resulting scaled Fisher information matrix is given by

$$[\boldsymbol{\mathcal{I}}(\boldsymbol{\xi})]_{ij} = \frac{1}{\sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}}\sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{jj}}} [\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ij}. \tag{4.5}$$

With the exponential scaling, it is assumed that the known background corresponds to $\xi_i = 0$, or $\nu_i = \alpha_i + \nu_{0i}$ where ν_{0i} represents a lower parameter bound. The gradient is given by $\frac{\partial}{\partial \xi_i} = G_i \frac{\partial}{\partial \nu_i}$ and the scaled Fisher information is $[\boldsymbol{\mathcal{I}}(\boldsymbol{\xi})]_{ij} = G_i G_j [\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ij}$ where $G_i = \frac{\alpha_i}{\beta_i} \mathrm{e}^{\xi_i/\beta_i}$. The appropriate scaling is then given by

$$\begin{cases} \beta_i = \alpha_i \sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}} \\ \dfrac{\partial}{\partial \xi_i} = \dfrac{1}{\sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}}} \mathrm{e}^{\xi_i/\beta_i} \dfrac{\partial}{\partial \nu_i} \end{cases} \tag{4.6}$$

and the resulting scaled Fisher information matrix at the background level $\boldsymbol{\xi} = \mathbf{0}$ is again given by (4.5).

With the arctangential scaling, it is assumed that the known background corresponds to $\xi_i = 0$, or $\nu_i = \alpha_i \arctan(\bar{\xi}_i) + \nu_{0i}$. Further, if v_{ui} and v_{li} denotes upper and lower parameter bounds, respectively, it is seen that $2v_{0i} = v_{ui} + v_{li}$ and $\pi\alpha_i = v_{ui} - v_{li}$. The gradient is given by $\frac{\partial}{\partial \xi_i} = G_i \frac{\partial}{\partial \nu_i}$ and the scaled Fisher information is $[\boldsymbol{\mathcal{I}}(\boldsymbol{\xi})]_{ij} = G_i G_j [\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ij}$ where $G_i = \frac{\alpha_i}{\beta_i}(1 + (\xi_i/\beta_i + \bar{\xi}_i)^2)^{-1}$. The appropriate scaling is then given by

$$\begin{cases} \beta_i = \dfrac{\alpha_i}{1 + \bar{\xi}_i^{\,2}} \sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}} \\ \dfrac{\partial}{\partial \xi_i} = \dfrac{1 + \bar{\xi}_i^{\,2}}{1 + (\xi_i/\beta_i + \bar{\xi}_i)^2} \dfrac{1}{\sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}}} \dfrac{\partial}{\partial \nu_i} \end{cases} \tag{4.7}$$

and the resulting scaled Fisher information matrix at the background level $\boldsymbol{\xi} = \mathbf{0}$ is again given by (4.5).

The contrast in the scaled parameter $\Delta\xi_i$ corresponding to a deviation $\Delta\nu_i$ with respect to the known background, is given for the three parameter models in (4.3) as follows

$$\Delta\xi_i = \sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}}\Delta\nu_i \tag{4.8}$$

$$\Delta\xi_i = \alpha_i\sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}}\log(1+\frac{\Delta\nu_i}{\alpha_i}) \tag{4.9}$$

$$\Delta\xi_i = \frac{\alpha_i\sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}}}{1+\bar{\xi}_i^{\,2}} \tag{4.10}$$

$$\times\left[\tan\left(\frac{\Delta\nu_i}{\alpha_i}+\arctan(\bar{\xi}_i)\right)-\bar{\xi}_i\right]. \tag{4.11}$$

Obviously, the nonlinear models above yield $\Delta\xi_i \approx \sqrt{[\boldsymbol{\mathcal{I}}(\boldsymbol{\nu})]_{ii}}\Delta\nu_i$ when $\frac{\Delta\nu_i}{\alpha_i}$ is small.

5 A priori modelling and regularization

5.1 The a priori statistics of Tikhonov regularization

In inverse problem theory and applications, it is common to employ a Tikhonov type of regularization that punish rapid spatial variations in the medium parameters [12]. Hence, the following Tikhonov regularization scheme may be considered

$$\min_{\xi}\left\{\mathcal{J}(\xi)+\gamma\int_0^a\left(\frac{\partial\xi}{\partial x}\right)^2\mathrm{d}x\right\} \tag{5.1}$$

where γ is the regularization constant. By employing the boundary conditions $\xi(0) = \xi(a) = 0$ and integrating by parts, the integral above can be approximated in discrete form as

$$\int_0^a\left(\frac{\partial\xi}{\partial x}\right)^2\mathrm{d}x = -\int_0^a\xi(x)\frac{\partial^2\xi}{\partial x^2}\,\mathrm{d}x \approx \frac{1}{\Delta x}\boldsymbol{\xi}^{\mathrm{T}}\bar{\mathbf{C}}^{-1}\boldsymbol{\xi} \tag{5.2}$$

where Δx is the discretization interval, $\boldsymbol{\xi}$ an $N\times 1$ sample vector and

$$\bar{\mathbf{C}}^{-1} = \begin{pmatrix} 2 & -1 & 0 & & \\ -1 & 2 & -1 & & \\ & & \ddots & & \\ & & -1 & 2 & -1 \\ & & 0 & -1 & 2 \end{pmatrix} \tag{5.3}$$

an $N\times N$ symmetric Toeplitz matrix. It is noted that the matrix $\bar{\mathbf{C}}^{-1}$ is the inverse of the symmetric matrix $\bar{\mathbf{C}}$ with elements $[\bar{\mathbf{C}}]_{ij} = i(N-j+1)/(N+1)$ for $j \geq i$ and $[\bar{\mathbf{C}}]_{ij} = [\bar{\mathbf{C}}]_{ji}$.

Now, the Tikhonov regularization scheme can be given an unconditional statistical (or Bayesian estimation) interpretation as follows. Assume that the parameter

vector $\boldsymbol{\xi}$ has zero mean and known prior Gaussian distribution with probability density function $p(\boldsymbol{\xi})$ where $\log p(\boldsymbol{\xi}) = d - \frac{1}{2}\boldsymbol{\xi}^{\mathrm{T}}\mathbf{C}^{-1}\boldsymbol{\xi}$, d is a constant and the correlation matrix $\mathbf{C}$ is given by

$$\mathbf{C} = \mathcal{E}\left\{\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}\right\} = \frac{\Delta x N_0}{\gamma}\bar{\mathbf{C}} \tag{5.4}$$

where N_0 is the spectral density of the measurement noise. The spatial variance and correlation coefficient corresponding to the correlation matrix $\bar{\mathbf{C}}$ is depicted in Fig. 2 below.

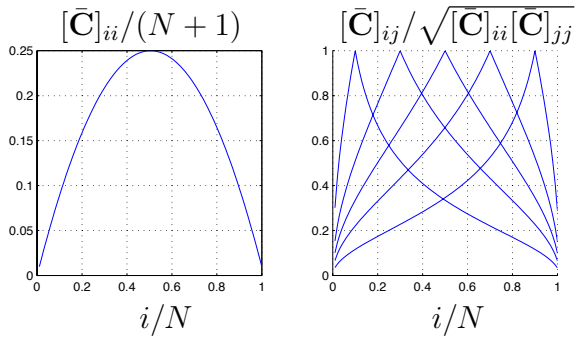

Figure 2: Spatial variance $[\bar{\mathbf{C}}]_{ii}/(N+1)$ and correlation coefficient $[\bar{\mathbf{C}}]_{ij}/\sqrt{[\bar{\mathbf{C}}]_{ii}[\bar{\mathbf{C}}]_{jj}}$ plotted for large N ($N = 100$). In the right figure, $j/N = \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

The Maximum A Posteriori (MAP) criterion [14, 16, 17] is to maximize the posterior conditional density function $p(\boldsymbol{\xi}|\boldsymbol{x})$ with respect to $\boldsymbol{\xi}$ where $\boldsymbol{x}$ is the measurement vector, or equivalently, to minimize the function $-\log p(\boldsymbol{x}|\boldsymbol{\xi}) - \log p(\boldsymbol{\xi})$ where $-\log p(\boldsymbol{x}|\boldsymbol{\xi})$ is the negative loglikelihood function defined in (3.6). Hence, the MAP criterion can be stated as

$$\min_{\boldsymbol{\xi}}\left\{\frac{1}{2N_0}\mathcal{J}(\boldsymbol{\xi}) + \frac{1}{2}\boldsymbol{\xi}^{\mathrm{T}}\mathbf{C}^{-1}\boldsymbol{\xi}\right\} \tag{5.5}$$

which is equivalent to (5.1) when $\mathbf{C}^{-1} = \frac{\gamma}{\Delta x N_0}\bar{\mathbf{C}}^{-1}$ according to the definition (5.4). Note that the gradient corresponding to (5.5) is given by

$$\frac{1}{2N_0}\frac{\partial}{\partial\boldsymbol{\xi}}\mathcal{J}(\boldsymbol{\xi}) + \mathbf{C}^{-1}\boldsymbol{\xi} \tag{5.6}$$

where $\frac{\partial}{\partial\boldsymbol{\xi}}$ is defined in section 4 and $\frac{\partial}{\partial\boldsymbol{\nu}}\mathcal{J}(\boldsymbol{\nu})$ in section 2.2.

According to the Bayesian statistical analogue given above, the use of the Tikhonov regularization scheme is equivalent to the assumption of a certain prior Gaussian parameter distribution together with an application of the MAP criterion. As will be demonstrated by using numerical experiments below, even if this parameter model is not physically justified, the statistical analogue is useful for characterizing the balance between the estimation error (the Cramér-Rao bound), the regularization constant, the signal to noise ratio and the spatial resolution in a given measurement situation. The aim is to employ the Fisher information analysis as a tool to predict the feasibility of successful reconstruction as well as to choose a proper regularization constant.

5.2 Principal parameter analysis

The Fisher information matrix for a stochastic vector parameter $\boldsymbol{\xi}$ is given by

$$\begin{aligned}\boldsymbol{\mathcal{I}} &= -\mathcal{E}_{\boldsymbol{x},\boldsymbol{\xi}}\left\{\frac{\partial^2 \log p(\boldsymbol{\xi}|\boldsymbol{x})}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^{\mathrm{T}}}\right\} \\ &= -\mathcal{E}_{\boldsymbol{\xi}}\mathcal{E}_{\boldsymbol{x}|\boldsymbol{\xi}}\left\{\frac{\partial^2 \log p(\boldsymbol{x}|\boldsymbol{\xi})}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^{\mathrm{T}}}\right\} - \mathcal{E}_{\boldsymbol{\xi}}\left\{\frac{\partial^2 \log p(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^{\mathrm{T}}}\right\} \\ &= \mathcal{E}_{\boldsymbol{\xi}}\left\{\boldsymbol{\mathcal{I}}(\boldsymbol{\xi})\right\} + \mathbf{C}^{-1} \quad (5.7)\end{aligned}$$

where $\mathcal{E}_{\boldsymbol{x},\boldsymbol{\xi}}$ and $\mathcal{E}_{\boldsymbol{x}|\boldsymbol{\xi}}$ denote the unconditional and the conditional expectation operator, respectively, see *e.g.*, [35]. In (5.7), $\boldsymbol{\mathcal{I}}(\boldsymbol{\xi})$ denotes the conditional Fisher information defined in *e.g.*, (4.5), (4.2) and (3.10), and $\mathbf{C}$ the correlation matrix defined in (5.4). The stochastic Cramér-Rao bound is given by the inverse of $\boldsymbol{\mathcal{I}}$, see *e.g.*, [35].

In practice, it will be extremely costly to compute $\mathcal{E}_{\boldsymbol{\xi}}\{\boldsymbol{\mathcal{I}}(\xi)\}$ in the expression (5.7), and hence the approximation $\boldsymbol{\mathcal{I}} \approx \boldsymbol{\mathcal{I}}(\boldsymbol{\xi})|_{\boldsymbol{\xi}=0} + \mathbf{C}^{-1}$ will be used where the conditional Fisher information is calculated at the known background. Note that this approximation is assymptotically unbiased and converges in probability as $\gamma \to \infty$.

The Fisher information matrix $\boldsymbol{\mathcal{I}}$ is extremely ill-conditioned and a calculation of the Cramér-Rao bound for individual pixels ξ_i is virtually impossible (and irrelevant) if the pixel resolution Δx is far below the resolution limit $\Delta x \ll \lambda/2$, *cf.*, [11]. However, a principal parameter analysis using the singular value decomposition (SVD) may be carried out to identify the significant number of retrievable parameters, and hence the resolution. The following notation will be employed

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\xi})|_{\boldsymbol{\xi}=0} + \mathbf{C}^{-1} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^{\mathrm{T}} \quad (5.8)$$

where $\mathbf{U}$ contains the singular vectors and $\boldsymbol{\Sigma}$ the singular values. The principal parameters are defined by $\mathbf{U}^{\mathrm{T}}\boldsymbol{\xi}$ and the corresponding Cramér-Rao bounds are given by the diagonal elements of $\boldsymbol{\Sigma}^{-1}$.

5.3 On the lognormal a priori model

When considering an a priori model such as the multivariate lognormal distribution it may be appealing to approach the problem from the point of view of estimating the parameter $\boldsymbol{\nu}$ directly. Suppose that the parameter is given by $\boldsymbol{\nu} = \mathrm{e}^{\boldsymbol{\xi}}$ where $\boldsymbol{\xi}$ is Gaussian distributed with mean $\boldsymbol{\eta}$ and correlation matrix $\mathbf{C}$. In this case, the MAP criterion (expressed in the variable $\boldsymbol{\xi}$) becomes

$$\min_{\boldsymbol{\xi}}\left\{\frac{1}{2N_0}J(\boldsymbol{\xi}) + \mathbf{1}^{\mathrm{T}}\boldsymbol{\xi} + \frac{1}{2}(\boldsymbol{\xi}-\boldsymbol{\eta})^{\mathrm{T}}\mathbf{C}^{-1}(\boldsymbol{\xi}-\boldsymbol{\eta})\right\} \quad (5.9)$$

where $\mathbf{1}$ is a vector of ones. The high-noise solution when $N_0 \to \infty$ is given by the a priori distribution only and is obtained here as $\boldsymbol{\xi} = \boldsymbol{\eta} - \mathbf{C}\mathbf{1}$.

It is instructive to consider the a priori statistics of the multivariate lognormal distribution. It is straightforward to derive the first and second order statistics

in terms of the a priori parameters $\boldsymbol{\eta}$ and $\mathbf{C}$ by completing the squares in the corresponding exponentials. Of particular interest is the mean $\mathcal{E}\{\nu_i\} = \mathrm{e}^{\eta_i+\sigma_i^2/2}$, the correlation function

$$[\mathbf{C}_\nu]_{ij} = \mathcal{E}\left\{(\nu_i - \mathcal{E}\{\nu_i\})(\nu_j - \mathcal{E}\{\nu_j\})\right\} = \mathrm{e}^{\eta_i+\eta_j+\sigma_i^2/2+\sigma_j^2/2}\left(\mathrm{e}^{C_{ij}} - 1\right) \tag{5.10}$$

and the a priori Fisher information

$$[\boldsymbol{\mathcal{I}}_\nu]_{ij} = -\mathcal{E}\left\{\frac{\partial^2 \log p(\boldsymbol{\nu})}{\partial\nu_i\partial\nu_j}\right\} = \delta_{ij}\mathrm{e}^{-2\eta_i+2\sigma_i^2} + \mathrm{e}^{-\eta_i-\eta_j+\sigma_i^2/2+\sigma_j^2/2}\mathrm{e}^{C_{ij}}C_{ij}^{-1} \tag{5.11}$$

where C_{ij} and C_{ij}^{-1} denote the elements of $\mathbf{C}$ and $\mathbf{C}^{-1}$, respectively, and $\sigma_i^2 = C_{ii}$.

The assymptotics of (5.10) and (5.11) is readily investigated for large and small variances by putting $C_{ij} = \frac{1}{\gamma}\bar{C}_{ij}$ where $\bar{C}_{ij}$ and η_i are fixed. Hence, by letting $\gamma \to \infty$ the correlation tends to zero and the Fisher information tends to infinity. In particular $[\mathbf{C}_\nu]_{ij} \to \mathrm{e}^{\eta_i+\eta_j}C_{ij}$, $[\mathbf{C}_\nu^{-1}]_{ij} \to \mathrm{e}^{-\eta_i-\eta_j}C_{ij}^{-1}$ and $[\boldsymbol{\mathcal{I}}_\nu]_{ij} \to \delta_{ij}\mathrm{e}^{-2\eta_i} + \mathrm{e}^{-\eta_i-\eta_j}C_{ij}^{-1} \approx [\mathbf{C}_\nu^{-1}]_{ij}$. This makes perfect sense and corresponds to a collapse of the lognormal distribution at $\boldsymbol{\nu} = \mathrm{e}^{\boldsymbol{\eta}}$.

On the other hand, when $\gamma \to 0$ the correlation tends to infinity $[\mathbf{C}_\nu]_{ij} \to \infty$, but the Fisher information does not tend to zero, instead $[\boldsymbol{\mathcal{I}}_\nu]_{ij} \to \infty$. This is obviously a reflection of the fact that the lognormal distribution collapses at the origin when $\boldsymbol{\eta}$ is fixed and $\mathbf{C} \to \infty$. Hence, if $\boldsymbol{\eta}$ is a given constant (as in section 5.1), the high-noise solution $\boldsymbol{\nu} \to \mathbf{0}$ as $\gamma \to 0$. To avoid this undesirable situation, it is necessary to invoke a condition such as $\boldsymbol{\eta}-\mathbf{C}\mathbf{1} = \log\boldsymbol{\nu}_\mathrm{b}$ where $\boldsymbol{\nu}_\mathrm{b}$ is the (fixed) known background. However, it is not obvious how the choose a proper preconditioning and parameter scaling ξ_i/β_i such that the required high-noise solution $\boldsymbol{\nu}_\mathrm{b}$ is maintained at the same time.

In conclusion, it is generally rather difficult to choose appropriate model parameters in the multivariate lognormal distribution in order to control its assymptotic behaviour with respect to the regularization and preconditioning (scaling) parameters. It is analytically and numerically much more tractable to consider the estimation of the Gaussian a priori parameters $\boldsymbol{\xi}$ as outlined in section 5.1 and 5.2.

6 numerical examples

Consider the one-dimensional inverse scattering problem as described in section 2 and depicted in Fig. 1. A conditional Fisher information analysis and related preconditioning was carried out for the exponential and arctangential parameter scalings as described in section 3 and 4, and an a priori model principal parameter analysis as described in section 5. The calculations were performed for a homogenous background with $\epsilon = 15$ and $\sigma = 0.4\,\mathrm{S/m}$ ($\sigma = 0.4\eta_0$). The center frequency was $f_0 = 6$ GHz and the bandwidth $f_\mathrm{B} = 8$ GHz. The computational boundary was $a = 15\lambda$ where $\lambda = 7.7\,\mathrm{mm}$ denotes the wavelength in the background medium at 10 GHz. The pixel resolution was $\Delta x = 0.1\lambda$. The conditional Fisher information employed is defined in (3.10). Further, 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.*, [11]. Note that $\epsilon_c = 15 - \mathrm{i}1.2\omega_0/\omega$ gives a loss tangent of 0.08 at the center frequency.

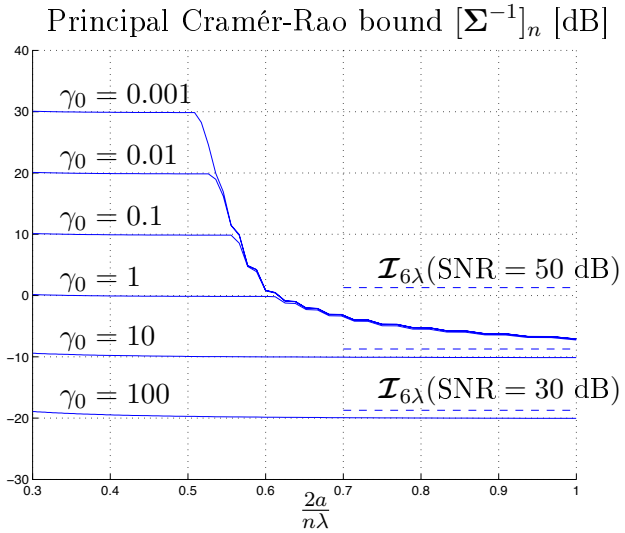

Figure 3: Principal Cramér-Rao bound $[\mathbf{\Sigma}^{-1}]_n$ as a function of resolution $2a/n\lambda$. Here γ_0 denotes the regularization parameter and $\boldsymbol{\mathcal{I}}_{6\lambda}$ the required contrast level at $x = 6\lambda$ for SNR = 30, 40 and 50 dB.

In Fig. 3 is shown the principal Cramér-Rao bound (singular values) $[\mathbf{\Sigma}^{-1}]_n$ as defined in (5.8), plotted as a function of resolution $2a/n\lambda$. The resolution is here defined as the size of the spatial domain in wave lengths divided by the number $(n/2)$ of retrieved parameters for ϵ and σ, respectively. In Fig. 3, the regularization parameter defined in (5.1) and (5.4) is given by $\gamma = \gamma_0 2N_0 10^{-4}$ where γ_0 ranges from 10^{-3} to 10^2, and $\boldsymbol{\mathcal{I}}_{6\lambda}$ is the required contrast level (4.8) at $x = 6\lambda$ for $\Delta\nu_i = 1$ and SNR = 30, 40 and 50 dB. Note that the scaled Fisher information (4.5) used in (5.8) is independent of the noise strength N_0, and hence motivates the scaling of γ with N_0 as above.

In Fig. 3 can be clearly seen the sharp resolution limit at $\lambda/2$, beyond which reconstruction (estimation) becomes virtually unfeasible, *cf.*, also [11]. For a resolution limit not below about 0.6λ and a signal to noise ratio of 50 dB, the analysis predicts that reconstruction is feasible and virtually independent of the regularization constant if $\gamma_0 \leq 1$ (the Cramér-Rao bound is significantly below the required contrast level $\boldsymbol{\mathcal{I}}_{6\lambda}$ with $\Delta\nu_i = 1$). On the other hand, with a signal to noise ratio of 30 dB, a regularization constant of $\gamma_0 = 100$ is required to force the Cramér-Rao bound below the required contrast level. In this case, the singular values as a function of resolution has become totally flat, indicating that the regularization has saturated, *i.e.*, the regularization has become so dominating that the only thing that can be retrieved is the a priori information itself, *i.e.*, the background parameters. Hence, reconstruction is unfeasible for SNR = 30 dB.

Next, a numerical implementation of the one-dimensional inverse problem is con-

sidered with the exponential and arctangential scaling and preconditioning defined in (4.3), (4.6) and (4.7). A priori model parameters for the exponential scaling was $\alpha_\epsilon = 1$, $\epsilon_0 = 14$, $\alpha_\sigma = 1.2$ and $\sigma_0 = 0$ and for the arctangential scaling $\pi\alpha_\epsilon = 3$, $\epsilon_0 = 15.5$, $\bar{\epsilon} = \tan\frac{-0.5}{3/\pi}$, $\pi\alpha_\sigma = 3$, $\sigma_0 = 1.5$ and $\bar{\sigma} = \tan\frac{-0.3}{3/\pi}$. The contrast level was $\Delta\epsilon = \Delta\sigma = 1$.

An inversion algorithm was implemented based on a quasi-Newton algorithm using the BFGS formula and Golden section line search, see *e.g.*, [9], together with the gradient calculations that are described in section 2.2 and (5.6) above. The solution to the related direct and adjoint electromagnetic problems were based on an implementation of the FDTD algorithm, see *e.g.*, [32], where the spatial resolution was 10 points per wavelength. A different spatial grid was used for the generation of input data in order to avoid the "inverse crime" [16, 33]. The signal to noise ratio was SNR = 30-50 dB, and artificial noise was added correspondingly prior to the reconstruction.

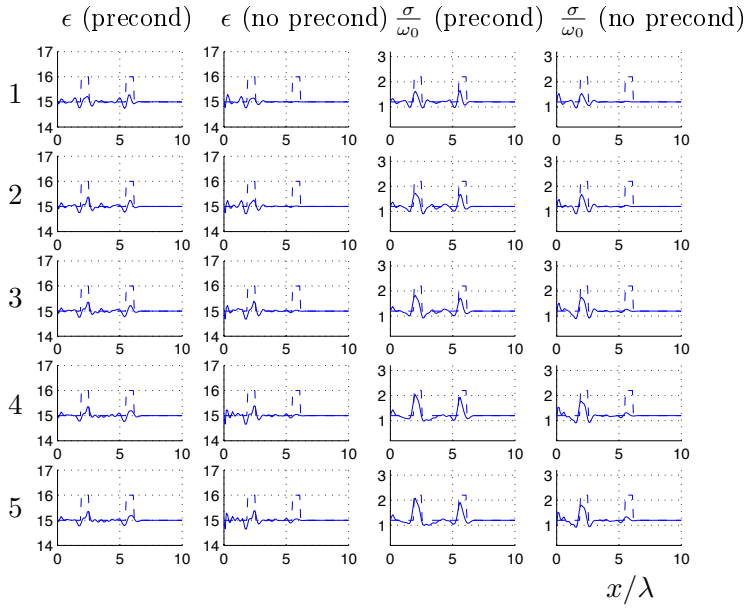

Figure 4: Reconstruction for the one-dimensional inverse problem with exponential parameter scaling. The graphs show the parameters ϵ and σ/ω_0 versus x/λ, with and without preconditioning, and iteration 1–5. The true parameter values are shown as a dashed line. The signal to noise ratio is SNR = 50 dB and the regularization constant $\gamma_0 = 1$.

In Figs. 4 and 5 are shown the reconstruction for the one-dimensional inverse problem with exponential and arctangential parameter scalings, respectively. The signal to noise ratio is SNR = 50 dB and the regularization constant $\gamma_0 = 1$. 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), ϵ (without

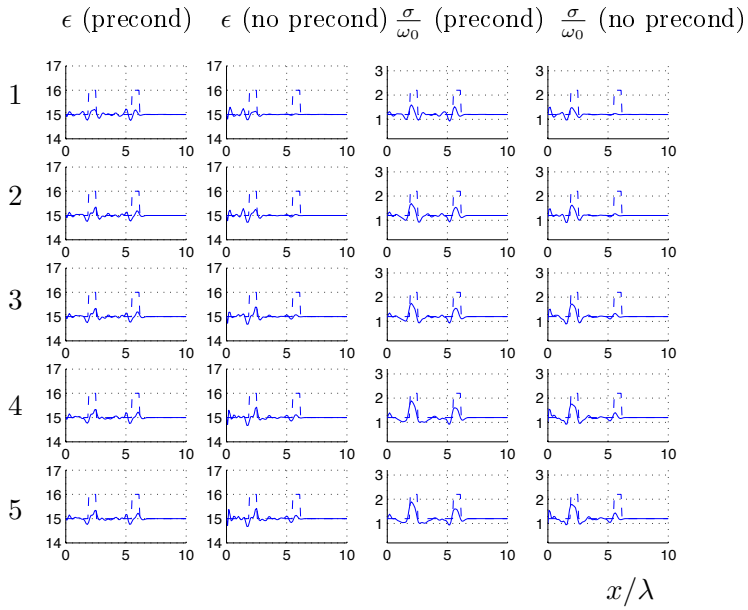

Figure 5: Reconstruction for the one-dimensional inverse problem with arctangential parameter scaling. The graphs show the parameters ϵ and σ/ω_0 versus x/λ, with and without preconditioning, and iteration 1–5. The true parameter values are shown as a dashed line. The signal to noise ratio is SNR = 50 dB and the regularization constant $\gamma_0 = 1$.

preconditioning), σ/ω_0 (with preconditioning) and σ/ω_0 (without preconditioning), 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.

In Figs. 6 and 7 are shown similarly as above, the reconstructions with exponential and arctangential parameter scalings in a comparison where the signal to noise ratio is SNR = 30 dB and the regularization constant is either $\gamma_0 = 1$ or $\gamma_0 = 100$. The case with SNR = 30 dB and $\gamma_0 = 1$ represents a situation where the regularization is insufficient and the parameter values tend to diverge as with the exponential scaling shown in Fig. 6. Note however, that the arctangential parameter scaling still regularizes the problem by limiting the parameter values by the upper and lower bounds as shown in Fig. 7. By increasing the regularization constant to $\gamma_0 = 100$, the regularization saturates and the parameters are forced to the known background level. Hence, with SNR = 30 dB, it does not seem to be feasible to retrieve the parameters accurately.

The reconstruction results in Figs. 4 through 7 can be compared and evaluated against the principal parameter analysis which was discussed above and illustrated in

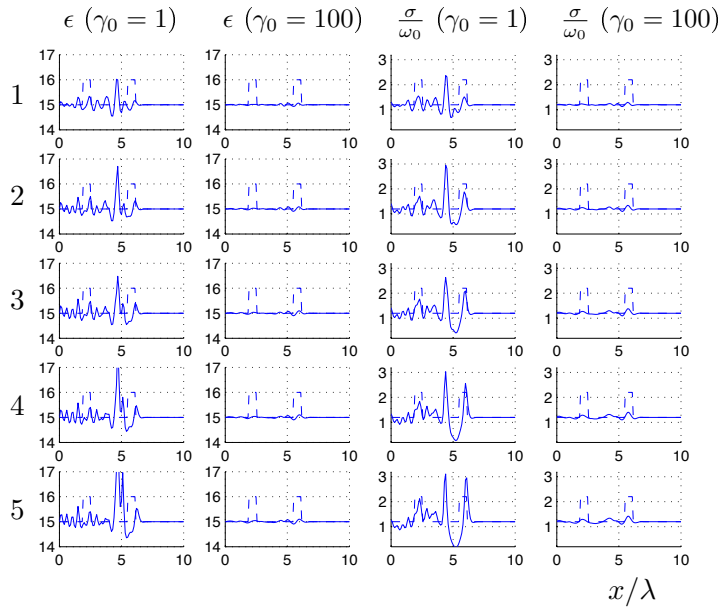

Figure 6: Reconstruction for the one-dimensional inverse problem with exponential parameter scaling. The graphs show the parameters ϵ and σ/ω_0 versus x/λ, with regularization constant $\gamma_0 = 1$ and $\gamma_0 = 100$, and iteration 1–5. The true parameter values are shown as a dashed line. The signal to noise ratio is SNR = 30 dB.

Fig. 3. As was predicted by the principal Fisher information analysis, the inversion problem is feasible (reconstruction works reasonably well) when the signal to noise ratio is SNR = 50 dB and is unfeasible (inversion is either unstable or saturated) when SNR = 30 dB. Furthermore, with the higher signal to noise ratio SNR = 50 dB, the behaviour of the inversion algorithm was (as predicted) rather independent of the regularization constant for $\gamma_0 \leq 1$. In conclusion, the principal parameter analysis worked well as an indicator to whether the inverse problem was feasible or not, as well as an indicator to a proper choice of regularization constant.

7 Summary and conclusions

In this paper, a Fisher information based Bayesian approach is presented for analysis and design of the regularization and preconditioning parameters used with gradient based inverse scattering algorithms. In particular, a one-dimensional inverse problem is considered where the permittivity and conductivity profiles are unknown and the input data consist of the scattered field over a certain bandwidth. A priori parameter modelling with nonlinear exponential and arctangential parameter scalings is treated and robust preconditioners are obtained by choosing the related scaling parameters based on a Fisher information analysis of the known background.

The Bayesian approach and a principal parameter (singular value) analysis of

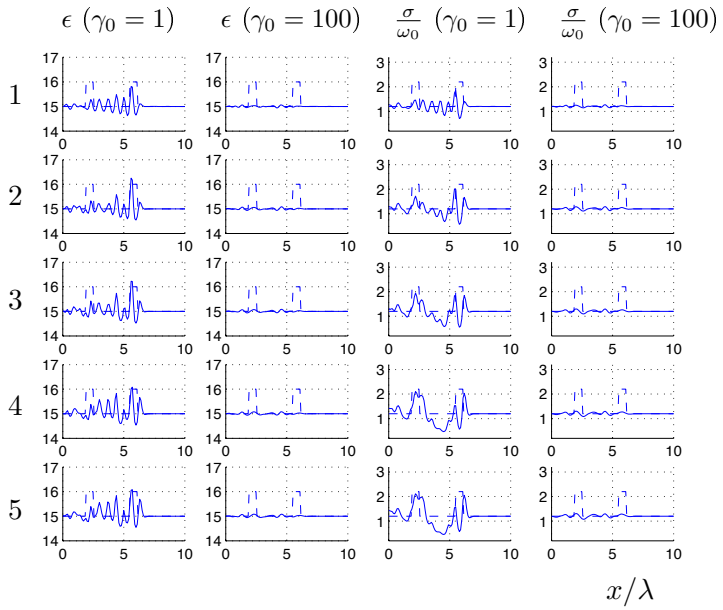

Figure 7: Reconstruction for the one-dimensional inverse problem with arctangential parameter scaling. The graphs show the parameters ϵ and σ/ω_0 versus x/λ, with regularization constant $\gamma_0 = 1$ and $\gamma_0 = 100$, and iteration 1–5. The true parameter values are shown as a dashed line. The signal to noise ratio is SNR = 30 dB.

the stochastic Cramér-Rao bound is used to investigate the regularization that is necessary to achieve stable inversion, as well as to predict the feasibility of achieving successful reconstruction in a given problem set-up. In particular, the Tikhonov regularization scheme is put into a Bayesian estimation framework.

A time-domain least-squares inversion algorithm based on a quasi-Newton algorithm together with an FDTD-electromagnetic solver has been employed in order to generate the numerical examples. The numerical examples verify the principal parameter analysis by considering low and high noise situations corresponding to feasible and unfeasible inverse problem set-ups, respectively. In a low noise situation, the behaviour of the inversion algorithm is typically independent of the regularization constant if the constant is below a certain limit which is predicted by the principal parameter analysis. In a high noise situation, the inverse problem is typically unfeasible (if the noise is high enough) and the regularization will saturate the reconstruction, yielding the a priori known background as an output.

Acknowledgement

The authors gratefully acknowledge the financial support by the Swedish Research Council.

References

[1] M. Bertero. Linear inverse and ill–posed problems. *Advances in electronics and electron physics*, **75**, 1–120, 1989.

[2] O. M. Bucci, L. Crocco, T. Isernia, and V. Pascazio. Subsurface inverse scattering problems: Quantifying qualifying and achieving the available information. *IEEE Trans. Geoscience and Remote Sensing*, **39**(11), 2527–2538, November 2001.

[3] 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.

[4] D. Colton and R. Kress. *Inverse Acoustic and Electromagnetic Scattering Theory*. Springer-Verlag, Berlin, 1992.

[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] A. Fhager, M. Gustafsson, S. Nordebo, and M. Persson. A statistically based preconditioner for two-dimensional microwave tomography. In *Proceedings of The Second International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP2007*, pages 173–176, October 2007.

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

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

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

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

[12] 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`.

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

[14] 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.

[15] V. Isakov. *Inverse Problems for Partial Differential Equations.* Springer-Verlag, Berlin, 1998.

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

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

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

[19] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems.* Springer-Verlag, New York, 1996.

[20] G. Kristensson and R. J. Krueger. Direct and inverse scattering in the time domain for a dissipative wave equation. Part 1: Scattering operators. *J. Math. Phys.*, **27**(6), 1667–1682, 1986.

[21] E. A. Marengo and R. W. Ziolkowski. Nonradiating and minimum energy sources and their fields: Generalized source inversion theory and applications. *IEEE Trans. Antennas Propagat.*, **48**(10), 1553–1562, October 2000.

[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] S. Nordebo, M. Gustafsson, and K. Persson. Sensitivity analysis for antenna near-field imaging. *IEEE Trans. Signal Process.*, **55**(1), 94–101, January 2007.

[26] S. Nordebo and M. Gustafsson. Statistical signal analysis for the inverse source problem of electromagnetics. *IEEE Trans. Signal Process.*, **54**(6), 2357–2361, June 2006.

[27] R. Pierri, A. Liseno, and F. Soldovieri. Shape reconstruction from PO multifrequency scattered fields via the singular value decomposition approach. *IEEE Trans. Antennas Propagat.*, **49**(9), 1333–1343, September 2001.

[28] R. Pierri and F. Soldovieri. On the information content of the radiated fields in the near zone over bounded domains. *Inverse Problems*, **14**(2), 321–337, 1998.

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

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

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

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

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

[34] 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.

[35] H. L. Van Trees. *Detection, Estimation and Modulation Theory, part I.* John Wiley & Sons, Inc., New York, 1968.