Surv Geophys (2016) 37:109–147
DOI 10.1007/s10712-015-9340-4

Applied Mathematics in EM Studies with Special Emphasis on an Uncertainty Quantification and 3-D Integral Equation Modelling

Oleg Pankratov[1] · **Alexey Kuvshinov**[2]

Received: 7 December 2014 / Accepted: 8 September 2015 / Published online: 27 October 2015

Abstract Despite impressive progress in the development and application of electromagnetic (EM) deterministic inverse schemes to map the 3-D distribution of electrical conductivity within the Earth, there is one question which remains poorly addressed—uncertainty quantification of the recovered conductivity models. Apparently, only an inversion based on a statistical approach provides a systematic framework to quantify such uncertainties. The Metropolis–Hastings (M–H) algorithm is the most popular technique for sampling the posterior probability distribution that describes the solution of the statistical inverse problem. However, all statistical inverse schemes require an enormous amount of forward simulations and thus appear to be extremely demanding computationally, if not prohibitive, if a 3-D set up is invoked. This urges development of fast and scalable 3-D modelling codes which can run large-scale 3-D models of practical interest for fractions of a second on high-performance multi-core platforms. But, even with these codes, the challenge for M–H methods is to construct proposal functions that simultaneously provide a good approximation of the target density function while being inexpensive to be sampled. In this paper we address both of these issues. First we introduce a variant of the M–H method which uses information about the local gradient and Hessian of the penalty function. This, in particular, allows us to exploit adjoint-based machinery that has been instrumental for the fast solution of deterministic inverse problems. We explain why this modification of M–H significantly accelerates sampling of the posterior probability distribution. In addition we show how Hessian handling (inverse, square root) can be made practicable by a low-rank approximation using the Lanczos algorithm. Ultimately we discuss uncertainty analysis based on stochastic inversion results. In addition, we

✉ Alexey Kuvshinov
kuvshinov@erdw.ethz.ch

Oleg Pankratov
oleg.pankratov@gmail.com

[1] Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radiowave Propagation, Russian Academy of Sciences, Kaluzhskoe Hwy 4, Troitsk, Moscow, Russia 142190

[2] Institute of Geophysics, ETH Zurich, Sonneggstrasse 5, 8092 Zurich, Switzerland

demonstrate how this analysis can be performed within a deterministic approach. In the second part, we summarize modern trends in the development of efficient 3-D EM forward modelling schemes with special emphasis on recent advances in the integral equation approach.

Keywords Quantification of uncertainties · Bayesian approach · Metropolis–Hastings · Gradient and Hessian · Adjoint sources approach · Finite elements · Contracting integral equations

1 Introduction

Electromagnetic (EM) three-dimensional (3-D) studies of the Earth have advanced significantly over the past decade. However, one of the key questions in the theory and practice of EM studies—uncertainty quantification of the recovered electrical conductivity models—remains practically unexplored. It is rather clear that only a Bayesian (stochastic) approach (Tarantola 2005) provides a systematic framework to quantify such uncertainties. Note that the number of works devoted to implementation of a stochastic approach in EM studies is rather limited (cf. Grandis et al. 1999; Chen et al. 2007, 2011; Rosas-Carbajal et al. 2013; Grandis et al. 2012; Minsley 2011; Brown et al. 2012) and moreover, the works are confined mostly to 1-D problems.

In contrast to a deterministic approach, the solution of the stochastic inverse problem is not a single model, but a family of models that samples the *posterior* probability density function (PDF). This family (ideally) delivers a complete statistical characterization for any nonlinear inverse problem (Bodin et al. 2012). The Metropolis–Hastings (M–H) algorithm (Metropolis et al. 1953)—which is one of the variants of the more general Markov Chain Monte Carlo (MCMC) concept—is the most popular approach for sampling posterior PDF (Gilks et al. 1996). However all stochastic methods, including M–H, require a huge number of forward modellings and thus appear to be extremely demanding computationally, if not prohibitive, at least for 3-D set-ups. This prompts development of highly efficient, robust and scalable modelling codes which can run large-scale 3-D models (say, models with millions of degrees of freedom) within seconds, in order to make M–H analysis feasible. Bearing in mind that nowadays high-performance multi-core platforms are readily available, the scalability of the codes becomes especially important. Another important challenge in M–H is a construction of proposal density functions that simultaneously provide a good approximation of the target density function while being inexpensive to be sampled. In this paper we address both of these challenges by confining ourselves, however, to the frequency-domain formulation.

The paper is organized as follows. In Sect. 2 we discuss the Bayesian formulation of the inverse problem. In Sect. 3 we explain the classical Metropolis–Hastings (M–H) algorithm as applied to an inverse problem. In Sect. 4 we introduce a proposal density based on local quadratic approximation of the penalty function (via gradient and Hessian). In Sect. 5 we discuss low-rank approximation of the Hessian-based proposal density. In Sect. 6 we present M–H algorithm with a proposal density based on low-rank approximation of the Hessian. In Sect. 7 we discuss uncertainty analysis based on stochastic results and in Sect. 8 we show how uncertainty analysis can be performed within a deterministic framework. In Sects. 9 and 10 we briefly discuss efficient calculation of gradient and Hessian-vector products which are needed to implement efficiently the extended M–H and deterministic

approaches introduced in previous sections. In the second part of the paper (Sect. 11) we summarize modern trends in the development of efficient 3-D EM forward modelling schemes with special emphasis on the integral equation approach. The paper also includes a number of appendices which detail results presented in the main body of the paper.

2 Bayesian Formulation of the Inverse Problem

The main challenge in solving inverse problems lies in the fact that they are usually ill-posed: many very different conductivity models may be consistent with the data. In other words there always exists a set of inverse problem solutions (models), $\mathbf{m}$, for which the following inequality holds

$$\beta_d(\mathbf{m}) \leq \min \beta_d + \Delta\beta_d, \tag{1}$$

where β_d stands for the data misfit function, and $\Delta\beta_d$ relates to the noise in the data. We refer to

$$\mathcal{E}(\Delta\beta_d) = \{\mathbf{m} \,|\, \beta_d(\mathbf{m}) \leq \min \beta_d + \Delta\beta_d\}, \tag{2}$$

as an *equivalence domain* (ED). Usually this domain is *narrow* but has a large size (see Fig. 1). A more illustrative explanation is presented in Fig. 2. Note that in the latter figure we assume that the number of model parameters is equal to 2. Terms *well resolved* and *poorly resolved* will be discussed later in the paper (in Sects. 7.2, 8).

The geometry of ED determines the uncertainty of the inverse solution which we are interested in. A conventional (deterministic) approach—which is a mainstream of EM studies—provides only a single solution without comprehensive survey of ED. However, the study of ED is possible in the framework of the Bayesian approach (Kaipio and Somersalo 2005; Tarantola 2005), namely by reformulating the inverse problem as a problem in statistical inference.

A Bayesian approach deals with *posterior*, *likelihood*, *prior* and *evidence* as follows

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}. \tag{3}$$

These terms will be explained later in the text. In the geophysical context a Bayesian approach assumes that there exists some joint probability distribution of the model vector $\mathbf{m} \in \mathbb{R}^{N_\mathcal{M}}$ and the vector of observed responses $\mathbf{\Phi} \in \mathbb{R}^{N_d}$, where $N_\mathcal{M}$ is the number of (real-valued) model parameters, and N_d is the number of (real-valued) responses. This distribution which can be written as

$$P(\mathbf{m}, \mathbf{\Phi})\; d^{N_\mathcal{M}}\mathbf{m}\; d^{N_d}\mathbf{\Phi}, \tag{4}$$

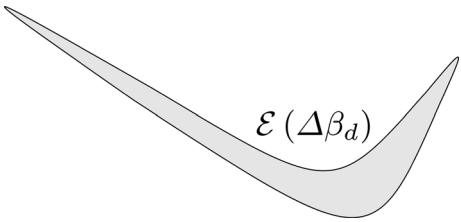

Fig. 1 Equivalence domain is *narrow* but has a large size

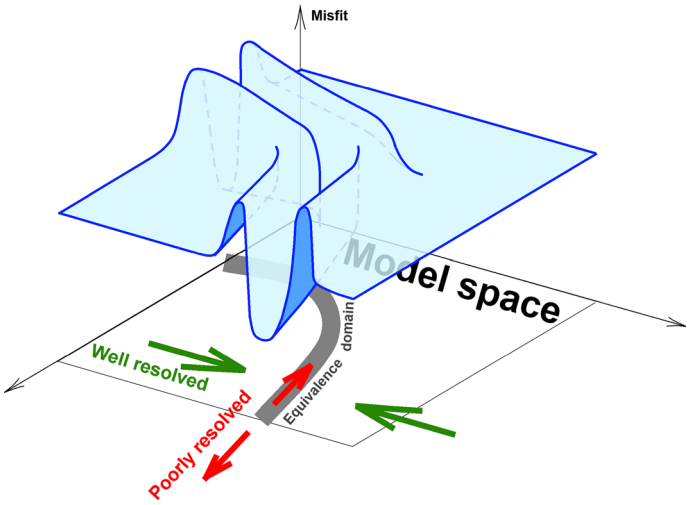

Fig. 2 An illustration of equivalence domain concept for an example where the number of model parameters is equal to 2. Terms *well resolved* and *poorly resolved* will be discussed later in the paper (in Sects. 7.2, 8).

specifies that a model vector belongs to volume $\mathbf{m} + d^{N_{\mathcal{M}}}\mathbf{m}$ and simultaneously a response vector belongs to volume $\mathbf{\Phi} + d^{N_d}\mathbf{\Phi}$ with probability $P(\mathbf{m}, \mathbf{\Phi})\, d^{N_{\mathcal{M}}}\mathbf{m}\, d^{N_d}\mathbf{\Phi}$. Here $P(\mathbf{m}, \mathbf{\Phi})$ stands for probability density of the joint distribution, $d^{N_{\mathcal{M}}}\mathbf{m}$ is a standard measure in the model space (or a small rectangular parallelepiped) and $d^{N_d}\mathbf{\Phi}$ is a standard measure in the data space (or, again, a small rectangular parallelepiped).

Marginal densities

$$P(\mathbf{m}) = \int P(\mathbf{m}, \mathbf{\Phi})\, d^{N_d}\mathbf{\Phi}, \tag{5}$$

$$P(\mathbf{\Phi}) = \int P(\mathbf{m}, \mathbf{\Phi})\, d^{N_{\mathcal{M}}}\mathbf{m}, \tag{6}$$

define marginal probability distributions for model

$$P(\mathbf{m})\, d^{N_{\mathcal{M}}}\mathbf{m}, \tag{7}$$

and responses

$$P(\mathbf{\Phi})\, d^{N_d}\mathbf{\Phi}, \tag{8}$$

respectively. Conditional probability densities $P(\mathbf{m}|\mathbf{\Phi})$ and $P(\mathbf{\Phi}|\mathbf{m})$ are defined as follows

$$P(\mathbf{m}|\mathbf{\Phi}) = P(\mathbf{m}, \mathbf{\Phi})/P(\mathbf{m}), \tag{9}$$

$$P(\mathbf{\Phi}|\mathbf{m}) = P(\mathbf{m}, \mathbf{\Phi})/P(\mathbf{\Phi}). \tag{10}$$

Bayes's theorem for the densities states that

$$P(\mathbf{m}|\mathbf{\Phi}) = \frac{P(\mathbf{\Phi}|\mathbf{m}) \cdot P(\mathbf{m})}{P(\mathbf{\Phi})}. \tag{11}$$

In this form the theorem looks trivial, since it immediately follows from Eqs. (9)–(10) but in 1763 when it was published (Bayes 1763), it shed a light on the "Doctrine of Chances". Since that time the doctrine significantly changed, the theory progressed (e.g., the Lebesgue measure theory was developed), and new nomenclature (e.g., *statistics*, *probability theory*, and *probability spaces*) was introduced; see, e.g., Kolmogorov (1956). Nowadays the concept of probability space includes a *sample space*, a set of *events* and a measure that assigns probabilities to events. A standard Lebesgue measure in $\mathbb{R}^n$ is denoted by $\mathrm{d}x_1, \ldots, \mathrm{d}x_n$, and any continuous measure can be represented as the product of the *measure density* and the standard Lebesgue measure, $f(x_1, \ldots, x_n)\mathrm{d}x_1, \ldots, \mathrm{d}x_n$. In the context of stochastic inverse problems the above densities and probability space concept are used as follows

- $P(\mathbf{m}|\mathbf{\Phi})$ is called *posterior density*, and it is the main object to be explored. Later in the paper we will discuss approaches to build a family of models that samples this distribution;
- $P(\mathbf{\Phi}|\mathbf{m})$ is called *a likelihood density*,

$$P(\mathbf{\Phi}|\mathbf{m}) = C_l \ \exp\left\{-\tfrac{1}{2}\beta_d(\mathbf{m})\right\}, \tag{12}$$

where

$$C_l = \left(\int \exp\left\{-\tfrac{1}{2}\beta_d(\mathbf{m})\right\} d^{N_{\mathcal{M}}}\mathbf{m}\right)^{-1} \tag{13}$$

is a normalizing constant, and $\beta_d(\mathbf{m})$ is the data misfit, which is a quadratic function of the residuals usually defined as

$$\beta_d(\mathbf{m}) = \sum_{g\in\mathcal{G}} \left|\frac{\theta_g(\mathbf{m}) - \Phi_g}{\Delta\Phi_g}\right|^2. \tag{14}$$

Note that a more general form reads

$$\beta_d(\mathbf{m}) = (\mathbf{\Theta}(\mathbf{m}) - \mathbf{\Phi})^T \mathbf{\Gamma}_{\text{noise}}^{-1} (\mathbf{\Theta}(\mathbf{m}) - \mathbf{\Phi}), \tag{15}$$

where $\mathbf{\Phi}$ is the vector of the experimental responses, $\mathbf{\Theta}(\mathbf{m})$ is the vector of the predicted responses for trial model $\mathbf{m}$, and $\mathbf{\Gamma}_{\text{noise}}$ is the data covariance matrix. If this matrix is diagonal, then Eq. (15) degenerates to Eq. (14).

- $P(\mathbf{m})$ is called *prior density*,

$$P(\mathbf{m}) = C_r \ \exp\left\{-\tfrac{1}{2}\beta_{\text{reg}}(\mathbf{m})\right\}, \tag{16}$$

where

$$C_r = \left(\int \exp\left\{-\tfrac{1}{2}\beta_{\text{reg}}(\mathbf{m})\right\} d^{N_{\mathcal{M}}}\mathbf{m}\right)^{-1}, \tag{17}$$

is a normalizing constant, and $\beta_{\text{reg}}(\mathbf{m})$ is a regularization term, which is a convex downward function of the model vector aimed to make the inverse problem well posed. This function is usually a quadratic function defined as

$$\beta_{\text{reg}}(\mathbf{m}) = (\mathbf{m} - \mathbf{m}_{\text{prior}})^T \mathbf{\Gamma}_{\text{prior}}^{-1} (\mathbf{m} - \mathbf{m}_{\text{prior}}), \tag{18}$$

where $\mathbf{\Gamma}_{\text{prior}}$ is a symmetric positive-definite matrix.

- $P(\mathbf{\Phi})$ is called *an evidence density*; it is a constant in Eq. (11) because $\mathbf{\Phi}$ is a fixed dataset that we are going to invert; thus the evidence density can be ignored in the further consideration.

Following Bayes's theorem (11) and Eqs. (12) and (16) posterior density reads

$$\pi(\mathbf{m}) := P(\mathbf{m}|\mathbf{\Phi}) = C_p \ \exp\left\{-\tfrac{1}{2}\beta_d(\mathbf{m}) - \tfrac{1}{2}\beta_{\text{reg}}(\mathbf{m})\right\}, \tag{19}$$

where normalizing constant C_p is as follows

$$C_p = \left(\int \exp\left\{-\tfrac{1}{2}\beta_d(\mathbf{m}) - \tfrac{1}{2}\beta_{\text{reg}}(\mathbf{m})\right\} d^{N_{\mathcal{M}}}\mathbf{m}\right)^{-1}. \tag{20}$$

Stochastic inversion is the generation of a Markov Chain Monte Carlo (MCMC) sequence of models which samples the Bayesian posterior density (19). Note that an MCMC sequence is a random chain in which the current model depends only on the previous model and not on the models that preceded it. In contrast to deterministic inversion which in fact relies on a minimization of the negative argument of the exponential function in Eq. (19) and usually delivers a single model, the stochastic inversion allows for a statistical characterization of the inverse problem (e.g., mean value, covariance, etc.), and ultimately is capable specifying the shape of the equivalence domain. In the next section we discuss the most popular MCMC algorithm—Metropolis–Hastings (M–H).

3 Metropolis–Hastings Sampling Algorithm

The Metropolis–Hastings (M–H) algorithm generates a sequence of models iteratively. At each iteration M–H: (1) randomly generates a new model $\mathbf{m}'$ from the previous model $\mathbf{m}$ using *a proposal density* $q(\mathbf{m}, \mathbf{m}')$; (2) rejects/accepts the new model $\mathbf{m}'$ using the M–H criterion. The proposal density means that new model $\mathbf{m}'$ belongs to volume $\mathbf{m}' + d^{N_{\mathcal{M}}}\mathbf{m}'$ with probability

$$q(\mathbf{m}, \mathbf{m}')\, d^{N_{\mathcal{M}}}\mathbf{m}', \tag{21}$$

provided that the previous model is $\mathbf{m}$. One of the simplest choices for $q(\mathbf{m}, \mathbf{m}')$ is a symmetric (Gaussian) proposal density (see illustration in Fig. 3)

$$q(\mathbf{m}, \mathbf{m}') = \frac{1}{(2\pi)^{N_{\mathcal{M}}}} \exp\left(-\tfrac{1}{2}|\mathbf{m} - \mathbf{m}'|^2\right), \tag{22}$$

which is easy to sample. This sampling scheme was introduced by Metropolis et al. (1953) who also suggested a criterion to reject/accept the new model. Namely, $\mathbf{m}'$ is accepted with the probability

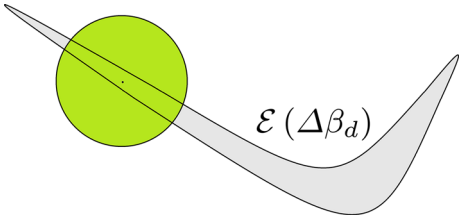

Fig. 3 Equivalence domain and a symmetric (Gaussian) proposal density (depicted as a *circle*)

$$\min\left(1, \frac{\pi(\mathbf{m}')}{\pi(\mathbf{m})}\right), \tag{23}$$

which means that a model with a smaller penalty function has greater chance to be accepted. Later (Hastings 1970) generalized the concept to work with an arbitrary proposal density $q(\mathbf{m}, \mathbf{m}')$ and modified the criterion in such a way that $\mathbf{m}'$ is accepted with the probability

$$\min\left(1, \frac{\pi(\mathbf{m}')q(\mathbf{m}', \mathbf{m})}{\pi(\mathbf{m})q(\mathbf{m}, \mathbf{m}')}\right). \tag{24}$$

This generalization underlies the Metropolis–Hastings sampling algorithm which is summarized below as a pseudo-code.

Algorithm 1 The standard Metropolis-Hastings sampling algorithm

- Choose$^{(A)}$ initial model $\mathbf{m}^{(0)} = \left(m_1^{(0)}, ..., m_{N_\mathcal{M}}^{(0)}\right)$
- Calculate$^{(B)}$ regularization term $\lambda\beta_{\text{reg}}(\mathbf{m}^{(0)})$
- Perform forward modelling to calculate $\beta_d(\mathbf{m}^{(0)})$
- Calculate posterior probability $\pi(\mathbf{m}^{(0)})/C_p = \exp\left(-\frac{1}{2}\beta(\mathbf{m}^{(0)})\right)$
- **for** $k = 0, ..., N_{it} - 1^{(C)}$ **do**
  - Draw$^{(D)}$ a new model $\mathbf{m}'$ from the proposal density $q(\mathbf{m}^{(k)}, \cdot)$
  - Calculate regularization term $\lambda\beta_{\text{reg}}(\mathbf{m}')$
  - Perform forward modelling to calculate $\beta_d(\mathbf{m}')$
  - Calculate posterior probability $\pi(\mathbf{m}')/C_p = \exp\left(-\frac{1}{2}\beta(\mathbf{m}')\right)$
  - Calculate proposal density function values $q(\mathbf{m}', \mathbf{m}^{(k)})$ and $q(\mathbf{m}^{(k)}, \mathbf{m}')$
  - Calculate criterion $\alpha(\mathbf{m}^{(k)}, \mathbf{m}') = \min\left(1, \frac{\pi(\mathbf{m}')q(\mathbf{m}', \mathbf{m}^{(k)})}{\pi(\mathbf{m}^{(k)})q(\mathbf{m}^{(k)}, \mathbf{m}')}\right)$
  - $\mathbf{m}^{(k+1)} = \begin{cases} \mathbf{m}' \text{ with probability } \alpha(\mathbf{m}^{(k)}, \mathbf{m}')^{(E)} \\ \mathbf{m}^{(k)} \text{ with probability } 1 - \alpha(\mathbf{m}^{(k)}, \mathbf{m}') \end{cases}$
- **end for**

Below are the explanatory notes on the code.

(A) It is very reasonable to take a (regularized) solution of the deterministic inverse problem as an initial model $\mathbf{m}^{(0)}$. Making this choice there is a hope that $\mathbf{m}^{(0)}$ are inside, or in the vicinity of, the equivalence domain.

(B) λ is specified during deterministic inversion.

(C) Deciding when to stop the chain is an important practical issue. The aim is to run the chain long enough to sample adequately posterior density. One of the most obvious informal methods for determining N_{it} is to run several chains in parallel, with different starting models, and compare marginal densities, where the estimate of marginal density, $p_i(m_i)$, of the i-th model parameter is a histogram computed from $m_i^{(k)}, k = 0, \ldots, N_{it}$ (see illustration in Fig. 4). More advanced approaches on convergence diagnostics, integrated autocorrelation, chain mixing and ergodicity can be found in the literature on MCMC [see Gilks et al. (1996) and Martin et al.

(2012), among others]. Note that in any case N_{it} is large (from thousands to ten thousand or more) but is much less than the number of samples in straightforward grid-based sampling.

(D) A specific algorithm to draw a new model $\mathbf{m}'$ is determined by the nature of the proposal density $q(\mathbf{m}, \mathbf{m}')$ chosen. For example, if symmetric $q(\mathbf{m}, \mathbf{m}')$ in the form of Eq. (22) is invoked then a Gaussian sampler should be used. Note that in this case the M–H criterion degenerates to the criterion specified by Eq. (23).

(E) The last step in the loop of the pseudo-code stands for the following action

- Draw a random number u uniformly distributed in the interval (0, 1)

if $u < \alpha(\mathbf{m}^{(k)}, \mathbf{m}')$ **then**
  "Accept": set $\mathbf{m}^{(k+1)} = \mathbf{m}'$
else
  "Reject": set $\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)}$
end if

Concluding this section we have to stress that in most applications of the M–H concept the Gaussian proposal density [see Eq. (22)] is implemented, i.e. all model parameters have equal importance (see illustration in Fig. 3). This means that specifics of the forward problem are not accounted for in the choice of the proposal density. An important and undesirable side effect of this is that during the M–H process an excessively large number of models are rejected. Since each step of the M–H algorithm involves forward modelling, it may lead to prohibitive computational loads. In the next section, following Martin et al. (2012), we discuss construction of the proposal density based on local quadratic approximation of the penalty function.

4 A Proposal Density Based on Local Quadratic Approximation of the Penalty Function

Following Martin et al. (2012) we introduce a proposal density $q(\mathbf{m}^{(k)}, \mathbf{m}')$ to be an approximation of the posterior density via a second-order (Taylor's) expansion of the penalty function β

$$q(\mathbf{m}^{(k)}, \mathbf{m}') = \tilde{\pi}(\mathbf{m}^{(k)}, \mathbf{s}) \approx \pi(\mathbf{m}^{(k)} + \mathbf{s}), \tag{25}$$

with

$$\tilde{\pi}(\mathbf{m}^{(k)}, \mathbf{s}) = C_1 \exp\left(-\tfrac{1}{2}\left(2\mathbf{g}_k^T\mathbf{s} + \mathbf{s}^T\mathrm{H}_k\mathbf{s}\right)\right). \tag{26}$$

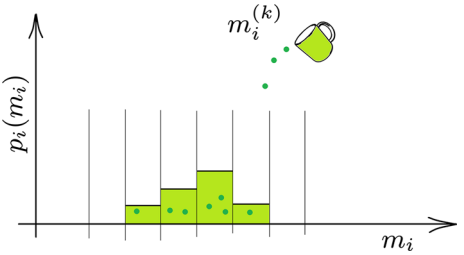

Fig. 4 A marginal density, $p_i(m_i)$, of the i-th model parameter is a histogram computed from $m_i^{(k)}, k = 0, \ldots, N_{it}$

Here

$$\mathbf{s} = \mathbf{m}' - \mathbf{m}^{(k)}, \tag{27}$$

stands for the jump from the current model $\mathbf{m}^{(k)}$, $\mathbf{g}_k = \nabla\beta(\mathbf{m}^{(k)})$ is the gradient of the penalty function, and $\mathtt{H}_k$ is the Hessian operator $\mathrm{Hess}_\beta(\mathbf{m}^{(k)})$, both of which are evaluated for the current model $\mathbf{m}^{(k)}$, and C_1 is a normalizing constant. For such a proposal density the acceptance rate should be very high (moreover, if the penalty is a quadratic function of the model vector then the acceptance rate is 100 %), and thus M–H should work much faster. This situation is illustrated in Fig. 5. But the question arises how to draw a new model $\mathbf{m}' = \mathbf{m}^{(k)} + \mathbf{s}$ from the proposal density defined by Eq. (26)? What follows is an explanation of a procedure which makes such a drawing feasible.

Let us first rewrite an argument of the exponent on the right-hand side (RHS) of Eq. (26) via perfect squares

$$\mathbf{g}_k^T\mathbf{s} + \tfrac{1}{2}\mathbf{s}^T\mathtt{H}_k\mathbf{s} = \tfrac{1}{2}\left|\mathbf{F}_k\mathbf{s} + \mathbf{F}_k^{-T}\mathbf{g}_k\right|^2 - \tfrac{1}{2}\left|\mathbf{F}_k^{-T}\mathbf{g}_k\right|^2, \tag{28}$$

where $\mathbf{F}_k$ is a matrix that satisfies the equality

$$\mathbf{F}_k^T\mathbf{F}_k = \mathtt{H}_k. \tag{29}$$

Then we can write the introduced proposal probability distribution

$$\mathrm{d}\tilde{\Pi}_k(\mathbf{s}) = \tilde{\pi}(\mathbf{m}^{(k)}, \mathbf{s})\; d^{N_\mathcal{M}}\mathbf{s}, \tag{30}$$

in the following form

$$\mathrm{d}\tilde{\Pi}_k(\mathbf{s}) = C_2 \exp\left(-\tfrac{1}{2}\left|\mathbf{F}_k\mathbf{s} + \mathbf{F}_k^{-T}\mathbf{g}_k\right|^2\right) d^{N_\mathcal{M}}\mathbf{s}. \tag{31}$$

Note that the second term in RHS of Eq. (28) reduces to constant C_2 since it does not depend on $\mathbf{s}$. Next, defining

$$\mathbf{g}_k' = -\mathbf{F}_k^{-T}\mathbf{g}_k, \tag{32}$$

and

$$\mathbf{y} = \mathbf{F}_k\mathbf{s}, \tag{33}$$

and using standard Gaussian distribution normalization coefficient, we obtain

$$\mathrm{d}\tilde{\Pi}_k(\mathbf{y}) = \frac{1}{(2\pi)^{N_\mathcal{M}/2}} \exp\left(-\tfrac{1}{2}\left|\mathbf{y} - \mathbf{g}_k'\right|^2\right) d^{N_\mathcal{M}}\mathbf{y}, \tag{34}$$

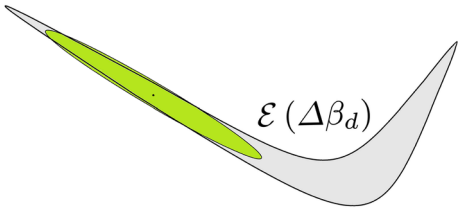

Fig. 5 Equivalence domain and a proposal density based on local quadratic approximation of penalty function (depicted as a *narrow ellipsoid*)

where $\mathbf{y}$ is drawn as

$$\mathbf{y} = \mathbf{g}'_k + \mathcal{N}_{N_\mathcal{M}}(\mathbf{0}, 1). \tag{35}$$

Here $\mathcal{N}_{N_\mathcal{M}}(\mathbf{0}, 1)$ is an $N_\mathcal{M}$-dimensional normal (Gaussian) distribution generator with zero mean and unit variance. Finally, using Eqs. (30), (31), (34) and the Jacobian, $\det \mathbf{F}_k$, of transform (33), we find constant C_2 in (31) and write

$$\tilde{\pi}(\mathbf{m}^{(k)},\ \mathbf{s}) = \frac{|\det \mathbf{F}_k|}{(2\pi)^{N_\mathcal{M}/2}} \exp\left(-\tfrac{1}{2}\left|\mathbf{F}_k\mathbf{s} + \mathbf{F}_k^{-T}\mathbf{g}_k\right|^2\right), \tag{36}$$

where $\mathbf{s}$ is drawn as the Newtonian step $\left(-\mathtt{H}_k^{-1}\mathbf{g}_k\right)$ plus a random *spherically distributed* vector $\mathcal{N}_{N_\mathcal{M}}(\mathbf{0}, 1)$ transformed to a random *ellipsoidally distributed* vector by linear operator $\mathbf{F}_k^{-1}$ as follows

$$\mathbf{s} = -\mathtt{H}_k^{-1}\mathbf{g}_k + \mathbf{F}_k^{-1}\mathcal{N}_{N_\mathcal{M}}(\mathbf{0}, 1) = \mathbf{F}_k^{-1} \cdot \left(-\mathbf{F}_k^{-T}\mathbf{g}_k + \mathcal{N}_{N_\mathcal{M}}(\mathbf{0}, 1)\right). \tag{37}$$

The RHS of the latter equation follows from Eq. (29). Here the term *spherically distributed* means that the distribution of the vector is invariant with respect to orthogonal transformations. The term *ellipsoidally distributed* stands for the random vector that is a linear transformation of some spherically distributed random vector. Note that we choose the normalization of the misfit that led to unit coefficient in front of the second (stochastic) term in the RHS of the latter equation (the more general case is described in Appendix 1). Equation (37) shows a way how to draw $\mathbf{s}$ but this way looks frightening since it involves the inverse of the Hessian and the inverse of "square root" of the Hessian. In the next section we explain how one can avoid calculations of these inverses by calculating instead only a limited number of the Hessian-vector products.

5 Low-rank Approximation of the Hessian-Based Proposal Density

Let us introduce the following definitions

$$\mathtt{H}_{d,k} := \mathrm{Hess}_{\beta_d}(\mathbf{m}^{(k)}), \tag{38}$$

$$\mathtt{H}_{\mathtt{reg}} := \mathrm{Hess}_{\beta_{\mathrm{reg}}}(\mathbf{m}^{(k)}). \tag{39}$$

$$\mathtt{H}_k := \mathrm{Hess}_{\beta}(\mathbf{m}^{(k)}) := \mathtt{H}_{d,k} + \lambda\mathtt{H}_{\mathtt{reg}} \tag{40}$$

Note that the symmetric positive matrix $\mathtt{H}_{\mathtt{reg}}$ does not depend on k [due to Eq. (18)]. Further, let us decompose the matrix $\mathtt{H}_{\mathtt{reg}}$ as follows

$$\mathtt{H}_{\mathtt{reg}} = \mathbf{L}^{-T}\mathbf{L}^{-1}. \tag{41}$$

The latter equation in particular means that $\mathbf{L}$ does not depend on k. Suppose we have a decomposition

$$\mathbf{L}^T\mathtt{H}_{d,k}\mathbf{L} \approx \mathbf{V}_k\mathbf{D}_k\mathbf{V}_k^T. \tag{42}$$

The latter can be approximately obtained from a number of steps (say, $n \le N_\mathcal{M}$ steps) of the Lanczos algorithm to matrix $\mathbf{L}^T\mathtt{H}_{d,k}\mathbf{L}$, thus here $\mathbf{D}_k$ is a diagonal $n \times n$-matrix

$$\mathbf{D}_k = \operatorname{diag}\left(d_1^{(k)}, \ldots, d_n^{(k)}\right) \equiv \begin{pmatrix} d_1^{(k)} & & \mathbf{0} \\ & \cdots & \\ \mathbf{0} & & d_n^{(k)} \end{pmatrix}, \tag{43}$$

and $\mathbf{V}_k$ is a semi-orthogonal $N_{\mathcal{M}} \times n$-matrix; that is, a matrix such that

$$\mathbf{V}_k^T \mathbf{V}_k = \mathbf{1}_n. \tag{44}$$

Here $\mathbf{1}_n$ is an identity $n \times n$-matrix. An equivalent definition of a semi-orthogonal $N_{\mathcal{M}} \times n$-matrix is that it can be complemented to an orthogonal $N_{\mathcal{M}} \times N_{\mathcal{M}}$-matrix.

For the sake of numerical efficiency, we propose to terminate the Lanczos process after $n \ll N_{\mathcal{M}}$ iterations thus obtaining a low-rank approximation (42) for the Hessian matrix of the data misfit.

Thus, defining an approximate to $\mathtt{H}_{d,k}$ as

$$\tilde{\mathtt{H}}_{d,k} := \mathbf{L}^{-T} \mathbf{V}_k \mathbf{D}_k \mathbf{V}_k^T \mathbf{L}^{-1}, \tag{45}$$

and an approximate penalty function Hessian as

$$\tilde{\mathtt{H}}_k = \tilde{\mathtt{H}}_{d,k} + \lambda \mathtt{H}_{\text{reg}} \approx \mathtt{H}(\mathbf{m}^{(k)}), \tag{46}$$

we obtain the following representation for the approximate penalty function Hessian

$$\tilde{\mathtt{H}}_k = \mathbf{L}^{-T}\left(\mathbf{V}_k \mathbf{D}_k \mathbf{V}_k^T + \lambda \mathbf{1}_{N_{\mathcal{M}}}\right)\mathbf{L}^{-1}, \tag{47}$$

where $\mathbf{1}_{N_{\mathcal{M}}}$ is an identity $N_{\mathcal{M}} \times N_{\mathcal{M}}$-matrix.

Following Eqs. (36) and (37) we in fact need not $\tilde{\mathtt{H}}_k$ but $\tilde{\mathbf{F}}_k^{-1}, \tilde{\mathbf{F}}_k^{-T}$ and $\det \tilde{\mathbf{F}}_k$, which satisfy

$$\tilde{\mathbf{F}}_k^T \tilde{\mathbf{F}}_k = \tilde{\mathtt{H}}_k. \tag{48}$$

In order to find them we introduce matrix $\mathbf{J}_k$ as

$$\mathbf{J}_k = \mathbf{V}_k \mathbf{D}_k \mathbf{V}_k^T + \lambda \mathbf{1}_{N_{\mathcal{M}}}. \tag{49}$$

From this representation we can construct the power $\mathbf{J}_k^x$ for any $x \in \mathbb{R}$ as follows

$$\begin{aligned} \mathbf{J}_k^x &= \mathbf{V}_k((\mathbf{D}_k + \lambda \mathbf{1}_n)^x - \lambda^x \mathbf{1}_n)\mathbf{V}_k^T + \lambda^x \mathbf{1}_{N_{\mathcal{M}}} \\ &= \mathbf{V}_k \operatorname{diag}\left((\lambda + d_j^{(k)})^x - \lambda^x\right)\mathbf{V}_k^T + \lambda^x \mathbf{1}_{N_{\mathcal{M}}}. \end{aligned} \tag{50}$$

It can be also shown that its determinant has a form

$$\det \mathbf{J}_k = \lambda^{N_{\mathcal{M}}-n} \det(\mathbf{D}_k + \lambda \mathbf{1}_n) = \lambda^{N_{\mathcal{M}}-n} \prod_{j=1}^{n} (\lambda + d_j^{(k)}). \tag{51}$$

The latter two results are achieved using $\mathbf{V}_k^T \mathbf{V}_k = \mathbf{1}_n$ and complementing $\mathbf{V}_k$ to an orthogonal matrix $N_{\mathcal{M}} \times N_{\mathcal{M}}$. From these two results, and defining $\tilde{\mathbf{F}}_k$ as

$$\tilde{\mathbf{F}}_k := \mathbf{J}_k^{\frac{1}{2}} \mathbf{L}^{-1}, \tag{52}$$

we obtain

$$\tilde{\mathbf{F}}_k^{-T} = \mathbf{J}_k^{-\frac{1}{2}}\mathbf{L}^T, \tag{53}$$

$$\tilde{\mathbf{F}}_k^{-1} = \mathbf{L}\mathbf{J}_k^{-\frac{1}{2}}, \tag{54}$$

$$\det\tilde{\mathbf{F}}_k = (\det\mathbf{J}_k)^{\frac{1}{2}}(\det\mathbf{L})^{-1}. \tag{55}$$

Introduced in this way $\tilde{\mathbf{F}}_k^T$ and $\tilde{\mathbf{F}}_k$ give Eq. (48), indeed

$$\tilde{\mathbf{F}}_k^T\tilde{\mathbf{F}}_k = \mathbf{L}^{-T}\mathbf{J}_k\mathbf{L}^{-1} = \mathbf{L}^{-T}\left(\mathbf{V}_k\mathbf{D}_k\mathbf{V}_k^T + \lambda\mathbf{1}_{N_\mathcal{M}}\right)\mathbf{L}^{-1} = \tilde{\mathtt{H}}_{d,k} + \lambda\mathtt{H}_{\mathtt{reg}} = \tilde{\mathtt{H}}_k. \tag{56}$$

Note that matrices $\mathbf{D}_k = \mathbf{D}(\mathbf{m}^{(k)})$ and $\mathbf{V}_k = \mathbf{V}(\mathbf{m}^{(k)})$ (see Eqs. 43, 44) are primary components of the algorithm, and they ensure fast performance of all needed matrix-vector operations (32), (58) as well as the determinant operation $\det\tilde{\mathbf{F}}_k$ from Eq. (36) (see Eqs. 50, 51, 53, 54). Matrices $\mathbf{J}_k = \mathbf{J}(\mathbf{m}^{(k)})$, $\mathbf{J}_k^x$ and $\tilde{\mathbf{F}}_k = \tilde{\mathbf{F}}(\mathbf{m}^{(k)})$ (see Eqs. 49, 50, 52) are not to be stored, and their operations are to be calculated on the fly.

Thus, to draw a sample $\mathbf{s}$ from low-rank approximation of $\tilde{\pi}(\mathbf{m}^{(k)}, \mathbf{s})$ the following procedure is applied:

- Instead of using Eq. (32) we calculate vector $\mathbf{g}'_k$ as

$$\mathbf{g}'_k = -\tilde{\mathbf{F}}_k^{-T}\mathbf{g}_k, \tag{57}$$

  with $\tilde{\mathbf{F}}_k^{-T}$ defined by Eq. (53).
- Next we draw a sample $\mathbf{y}$ using Eq. (35).
- Finally we calculate vector $\mathbf{s}$ from equation

$$\mathbf{s} = \tilde{\mathbf{F}}_k^{-1}\mathbf{y} \tag{58}$$

  with $\tilde{\mathbf{F}}_k^{-1}$ defined by Eq. (54).

Note that for reasons explained in Appendix 1 Eq. (35) can be modified as

$$\mathbf{y} = \mathbf{g}'_k + \sqrt{\kappa}\mathcal{N}_{N_\mathcal{M}}(\mathbf{0}, 1), \tag{59}$$

where κ is a positive real-valued constant.

6 Metropolis–Hastings Sampling Algorithm with the Hessian-Based Proposal Density

An overall Hessian-based M–H algorithm utilizes proposal density $\tilde{\pi}(\mathbf{m}^{(k)}, \mathbf{s})$ (see Eq. 36) built with the use of the second-order approximation of the penalty. In this algorithm we invoke matrix-vector Hessian operations only, rather than computing the full Hessian matrix. The details of the matrix-vector Hessian operations can be found in Appendix 3. The algorithm is presented as follows.

Algorithm 2 The M-H sampling algorithm with proposal density $\tilde{\pi}(\mathbf{m}^{(k)}, \mathbf{s})$ built with the use of the second-order approximation of the penalty function

- Choose$^{(A)}$ initial model $\mathbf{m}^{(0)} = \left(m_1^{(0)}, ..., m_{N_\mathcal{M}}^{(0)}\right)$
- Calculate$^{(B)}$ regularization term $\lambda\beta_{\text{reg}}(\mathbf{m}^{(0)})$
- Decompose$^{(C)}$ the regularization matrix as $\mathtt{H}_{\text{reg}} = \mathbf{L}^{-T}\mathbf{L}^{-1}$ (see eq. 41)
- Perform forward modelling to calculate $\beta_d(\mathbf{m}^{(0)})$
- Calculate posterior probability $\pi(\mathbf{m}^{(0)})/C_p = \exp\left(-\frac{1}{2}\beta(\mathbf{m}^{(0)})\right)$
- Perform forward modelling to calculate gradient $\mathbf{g}_0 = \nabla\,\beta(\mathbf{m}^{(0)})$ and low-rank approximation of the data misfit Hessian (56)
  - Define$^{(D)}$ $n < N_\mathcal{M}$
  - In n steps of the Lanczos algorithm (using only the Hessian-vector products) calculate a semi-orthogonal matrix $\mathbf{V}_0 \in \mathbb{R}^{N_\mathcal{M}\times n}$ and a low-rank Hessian representation $\mathbf{D}_0 = \text{diag}\left(d_j^{(0)}\right) \in \mathbb{R}^{n\times n}$, such that $\mathbf{L}^T\mathtt{H}_d(\mathbf{m}^{(0)})\mathbf{L} \approx \mathbf{V}_0\mathbf{D}_0\mathbf{V}_0^T$. Note that n columns of $\mathbf{V}_0$ are orthonormal vectors of the Lanczos basis.
- **for** k **in** 0 **to** $N_{it} - 1$ **do**
  - Draw$^{(E)}$ a new sample model $\mathbf{m}'$ from the proposal density $q(\mathbf{m}^{(k)}, \cdot)$:
    - Calculate $\mathbf{g}'_k = -\tilde{\mathbf{F}}_k^{-T}\mathbf{g}_k$ using eqs (32) and (53)
    - Draw vector $\mathbf{y} \sim \mathcal{N}_{N_\mathcal{M}}(\mathbf{0}, 1)$ from a standard Gaussian $N_\mathcal{M}$-dimensional sampler
    - $\mathbf{s} = \tilde{\mathbf{F}}_k^{-1} \cdot (\mathbf{y} + \mathbf{g}'_k)$ using eq. (54)
    - Calculate vector $\mathbf{m}' = \mathbf{m}^{(k)} + \mathbf{s}$
  - Calculate criterion $\alpha(\mathbf{m}^{(k)}, \mathbf{m}') = \min\left(1, \frac{\pi(\mathbf{m}')q(\mathbf{m}', \mathbf{m}^{(k)})}{\pi(\mathbf{m}^{(k)})q(\mathbf{m}^{(k)}, \mathbf{m}')}\right)$:
    - Perform forward modelling to calculate $\pi(\mathbf{m}')/C_0 = \exp\left(-\frac{1}{2}\beta(\mathbf{m}')\right)$
    - Calculate $\frac{\pi(\mathbf{m}')}{\pi(\mathbf{m}^{(k)})}$
    - Calculate$^{(F)}$ $\frac{q(\mathbf{m}', \mathbf{m}^{(k)})}{q(\mathbf{m}^{(k)}, \mathbf{m}')}$:
      1. Calculate $\det\tilde{\mathbf{F}}_k$ using eq. (55)
      2. Calculate $q(\mathbf{m}^{(k)}, \mathbf{m}') = \tilde{\pi}(\mathbf{m}^{(k)}, \mathbf{m}' - \mathbf{m}^{(k)})$ using eq. (36)
      3. Calculate the gradient $\mathbf{g}(\mathbf{m}') = \nabla\,\beta_d(\mathbf{m}') \in \mathbb{R}^{N_\mathcal{M}}$ for model $\mathbf{m}'$
      4. Calculate$^{(D)}$ $n' < N_\mathcal{M}$ and primary components $\mathbf{D}(\mathbf{m}') \in \mathbb{R}^{n'\times n'}$ and $\mathbf{V}(\mathbf{m}') \in \mathbb{R}^{N_\mathcal{M}\times n'}$ of the approximate Hessian matrix $\tilde{\mathtt{H}}(\mathbf{m}')$ (56), which ensures fast matrix-vector operations (40)-(56) for matrices $\mathbf{D}(\mathbf{m}')$, $\mathbf{V}(\mathbf{m}')$, $\tilde{\mathtt{H}}(\mathbf{m}')$, $\mathbf{J}(\mathbf{m}')$, $\mathbf{J}^x(\mathbf{m}')$, $\tilde{\mathbf{F}}(\mathbf{m}')$, and determinant $\det\tilde{\mathbf{F}}(\mathbf{m}')$ calculation
      5. Calculate $q(\mathbf{m}', \mathbf{m}^{(k)}) = \tilde{\pi}(\mathbf{m}', \mathbf{m}^{(k)} - \mathbf{m}')$
  - $\mathbf{m}^{(k+1)} = \begin{cases} \mathbf{m}' \text{ with probability } \alpha(\mathbf{m}^{(k)}, \mathbf{m}')^{(G)}, \\ \mathbf{m}^{(k)} \text{ with probability } 1 - \alpha(\mathbf{m}^{(k)}, \mathbf{m}'). \end{cases}$
- **end for**

Below are the explanatory notes on the code.

(A) It is very reasonable to take a (regularized) solution of the deterministic inverse problem as an initial model $\mathbf{m}^{(0)}$. Making this choice there is a hope of being inside, or in the vicinity of, the equivalence domain.

(B) A value of λ can be specified during deterministic inversion; however, for realistic study of the uncertainty it is necessary to define a value of '*the uncertainty* λ' significantly less than the value of λ used for the deterministic inversion.

(C) To decompose the regularization matrix as $\mathtt{H}_{\mathrm{reg}} = \mathbf{L}^{-T}\mathbf{L}^{-1}$ one can use either

- a Cholesky decomposition $\mathtt{H}_{\mathrm{reg}}^{-1} = \mathbf{L}\mathbf{L}^{T}$,
- or the spectral decomposition $\mathtt{H}_{\mathrm{reg}} = U\,\mathrm{diag}\left(h_j\right)U^{T}$ and a symmetric square root $\mathbf{L} = \mathtt{H}_{\mathrm{reg}}^{-\frac{1}{2}} = U\,\mathrm{diag}\left(h_j^{-\frac{1}{2}}\right)U^{T}$ (where $U^{T} = U^{-1}$).

(D) A simple choice for $n, n' < N_{\mathcal{M}}$ could be a fixed value for all iterations. A more complicated scenario could be dependency of n on the decay rate of the coefficients d_j, d'_j, where $\mathbf{D}' = \mathrm{diag}\left(d'_j\right)$.

(E) To reduce the Hessian calculation, we suggest to introduce a tunable branching parameter b_1 and draw b_1 new models $\mathbf{m}'_1, \ldots, \mathbf{m}'_{b_1}$ at this step from the proposal density $q(\mathbf{m}^{(k)}, \cdot)$. As a result we have a *tree*-like family of models, and the latest *branches* of the tree are not needed in the Hessian calculation to draw new models (as they are the latest). This saving is bigger, the larger the branching parameter b_1 is.

(F) To reduce the Hessian calculation, we suggest to introduce a tunable frequency $f_1 \in [0, 1]$ and assign the proposal density to be symmetric with probability f_1, thus having in this case $\frac{q(\mathbf{m}', \mathbf{m}^{(k)})}{q(\mathbf{m}^{(k)}, \mathbf{m}')} = 1$. For $f_1 = 1$ the value of the fraction is always 1; this means a maximum saving in the Hessian calculation but the M–H criterion is approximate in this case. For $f_1 = 0$ the value of the fraction is always to be calculated, which requires calculation of $\tilde{\mathbf{F}}(\mathbf{m}'), \mathbf{V}(\mathbf{m}')$, but the M–H criterion $\alpha(\mathbf{m}^{(k)}, \mathbf{m}')$ is precise in this case.
In addition, we suggest not calculating the fraction (but to assign it a value of 1) for the latest point in the chain [or for the latest *branches* in the *tree*-like family of models, see note (E)].

(G) The last step in the loop of the pseudo-code stands for the following action

- Draw a random number u uniformly distributed in the interval (0, 1)

if $u < \alpha(\mathbf{m}^{(k)}, \mathbf{m}')$ **then**
  "Accept": set $\mathbf{m}^{(k+1)} = \mathbf{m}'$
else
  "Reject": set $\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)}$
end if

7 Uncertainty Analysis Based on M–H Results

In this section we briefly outline possible methods to analyze the uncertainty of the inverse solution based on M–H results.

7.1 Marginal Distributions

First we suggest to study *local uncertainties* that can be calculated and plotted as follows. From the resulting stochastic chain, we calculate the marginal densities, $p_i(m_i)$, for each individual model parameter m_i (see illustration on the definition of these densities in Fig. 4). Note that model parameters stand for conductivity or log-conductivity of the i-th spatial cell of 3-D inverse domain. Having the density p_i we calculate mean value of the i-th parameter,

$$e_i = \int_{\mathbb{R}} m_i p_i(m_i) \mathrm{d}m_i, \tag{60}$$

its variance,

$$\mathcal{D}_i = \int_{\mathbb{R}} (m_i - e_i)^2 p_i(m_i) \mathrm{d}m_i, \tag{61}$$

as well as the other moments. All these moments can be portrayed as spatial maps, delivering an impression of how uncertain the resulting solution is.

However, marginal moments may not be sufficient to study the equivalence domain (ED), as the inverse problem might have much more complex uncertainties than the local ones mentioned above. As an example, let us recall the well-known magnetotelluric (MT) effect (S-equivalence) where a stack of n relatively thin layers $h_1, \cdots, h_n$ can be well resolved from the MT data vector only as a value of conductance S in the form

$$f(\sigma_1, \ldots, \sigma_n) := \sigma_1 h_1 + \cdots + \sigma_n h_n, \tag{62}$$

and any combination of values of $\sigma_1 \geq 0, \ldots, \sigma_n \geq 0$ satisfies the MT data vector provided that the values satisfy the following equation

$$f(\sigma_1, \ldots, \sigma_n) = S. \tag{63}$$

Let us refer to such layers (or cells) as *linked elements of the model*. In the next section we discuss a possible way to detect such linked elements.

7.2 Linked Elements of the Model

Let us first introduce the following terminology. We call *the weak forms* those eigenvectors $\mathbf{Ve}_j$ of the Hessian that have the smallest eigenvalues d_j. A set of the weak forms we denote in this section as $\mathbf{w}_1, \ldots, \mathbf{w}_n$. Note that weak forms determine *poorly resolved* directions in the model space.

Let us denote the linear span of the weak forms as

$$W = \text{ span } \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}, \tag{64}$$

and from the basis vectors $\mathbf{e}_i$ of model parameters m_i we build the following coordinate spaces

$$\mathcal{E}_i^1 = \text{ span } \{\mathbf{e}_i\}, \tag{65}$$

$$\mathcal{E}_{i,j}^2 = \text{ span } \{\mathbf{e}_i, \mathbf{e}_j\}, \tag{66}$$

and similarly for $\mathcal{E}^3_{i,j,k}$ etc. Next we introduce the following *linking angles*

$$\alpha^1_i = \angle\left(W, \mathcal{E}^1_i\right) = \arccos \max_{\mathbf{u}\in\mathcal{E}^1_i,\ \mathbf{w}\in W} \frac{\mathbf{u}\cdot\mathbf{w}}{|\mathbf{u}|\,|\mathbf{w}|}, \tag{67}$$

$$\alpha^2_{i,j} = \angle\left(W, \mathcal{E}^2_{i,j}\right) = \arccos \max_{\mathbf{u}\in\mathcal{E}^2_{i,j},\ \mathbf{w}\in W} \frac{\mathbf{u}\cdot\mathbf{w}}{|\mathbf{u}|\,|\mathbf{w}|}, \tag{68}$$

and similarly for $\alpha^3_{i,j,k}$ etc. (note that in the fractions above we take only nonzero vectors).

Then the condition $\alpha^1_i \to 0$ means that the i-th model parameter is *weak* by itself: there is no link with other parameters in terms of resolution. Such a situation, for example, arises in MT when a resistive layer underlies a conductive layer. In this case the resistive layer is poorly resolved. Let us imagine that $\alpha^1_i \not\to 0$ and $\alpha^1_j \not\to 0$. Then the condition $\alpha^2_{i,j} \to 0$ means that a combination of the i-th and the j-th parameters is poorly resolved. In this case we consider the i-th and the j-th parameters to be linked.

7.3 Study of Dimensionality of ED Using Metrics

This study might be performed using a Riemannian metrics and having the chain from the M–H sampling algorithm with the Hessian-based proposal density. Riemannian metrics gives a possibility to measure length, angle and volume in the model space. By fixing a radius R we construct a union $\mathcal{F}$ of all convex hulls of the part of the M–H chain that belongs to a ball of radius R. Then the dependence of the volume of $\mathcal{F}$ as $R \to 0$ allows us to estimate the dimension of the ED which can be considered as a degree of uncertainty of the solution. For example, if in MT three conducting layers are seen as one conductive layer of conductance S, then the dimension of ED (which is an analog of *null space*) is equal to 2.

7.4 Topological Study

Topological study can answer the following questions:

1. Whether the ED is a connected subset of the model space?
2. How many connected components exist in ED?

Note that the number of connected components is related to the number of local minima of the penalty function. Constructively, having the M–H chain from Sect. 6 we can build the set $\mathcal{F}$ (from Sect. 7.3) and answer these two questions with respect to $\mathcal{F}$ but relating the result to ED.

8 Uncertainty Analysis Based on the Results of the Deterministic Inversion

So far we discussed a stochastic approach to explore the ED. However the ED to some extent can be studied in the frame of a deterministic approach.

Let us assume that $\mathbf{m}^{(k)}$ is a solution of a regularized deterministic inversion with regularization parameter λ_{reg}. Let us then substantially decrease λ to make the inverse solution unstable. Further we run a Lanczos process (see Eqs. 40–56) for the Hessian

matrix $\tilde{\mathrm{H}}_k(\lambda)$ to get the largest eigenvalues d_j and their eigenvectors and for the shifted Hessian matrix $\tilde{\mathrm{H}}_k - \mu \mathbf{1}_{N_\mathcal{M}}$ (where μ is the maximum of d_j) to get the smallest eigenvalues d_j and their eigenvectors. As a result we will have an approximate local pattern of the uncertainty in the vicinity of the model $\mathbf{m}^{(k)}$, namely:

- the largest eigenvalues d_j and their eigenvectors $\mathbf{Ve}_j$ represent the so-called *strong* forms, i.e. combinations of the model parameters that are well determined from the data vector $\mathbf{\Phi}$. Note that strong forms determine *well-resolved* directions in the model space;
- the smallest eigenvalues d_j and their eigenvectors $\mathbf{Ve}_j$ represent the weak forms (already discussed in Sect. 7.2), i.e. combinations of the model parameters that are poorly resolved from the data vector $\mathbf{\Phi}$;
- intermediate eigenvalues and their eigenvectors which can be interpreted to relate to the one of the aforementioned forms depending on the problem set up, namely:
  - if importance is given to the structures that do exist in the model, then the intermediate forms should be considered as weak ones;
  - if importance is given to the structures that do not exist in the model, then the intermediate forms should be considered as strong ones;

With this local uncertainty analysis it is possible to represent a general solution as a sum of strong (fixed) part and an arbitrary linear combination of the weak forms.

It is evident that the implementation of both stochastic and deterministic approaches requires calculation of the data misfit gradients and the Hessian-vector products. The efficient way to perform these calculations is discussed in Sects. 9 and 10.

9 Calculation of Data Misfit Gradient

The efficient calculation of a data misfit gradient using adjoint formulation is a well-established approach in computational science and is described in numerous publications. As applied to EM studies it can be shown that $N_\Omega N_\mathcal{P}$ forward modellings are needed to calculate a data misfit gradient, where N_Ω and $N_\mathcal{P}$ are the numbers of frequencies and polarizations, respectively. We refer the reader to Appendix 2 in which we provide the key formulae in accordance with nomenclature of Pankratov and Kuvshinov (2010) and Pankratov and Kuvshinov (2015).

10 Calculation of the Hessian-Vector Products

The computation of the Hessian-vector products using adjoint formulation is also a rather well-established approach, especially in seismic inverse modelling (Santosa and Symes 1988; Epanomeritakis et al. 2008; Fichtner and Trampert 2011; Metivier et al. 2013, among others). Surprisingly, for EM modelling the literature discussing efficient calculation of the Hessian-vector products is very limited. Some work has been done by Newman and Hoversten (2000) who, however, confined the discussion to the case when the data are controlled-source electric or magnetic fields. Very recently Pankratov and Kuvshinov (2015) present a general formalism for the efficient calculation of the second derivatives of EM frequency-domain responses and the second derivatives of the misfit (Hessian matrix) with respect to variations of 3-D conductivity, using adjoint sources approach. They also

show how this technique can be implemented to calculate multiple Hessian-vector products very efficiently. Their finding is that a single Hessian-vector product can be calculated for a price of $O(N_\Omega(N_\mathcal{P} + N_\mathcal{S}))$ forward problem runs, where $N_\mathcal{S}$ is the number of observation sites. However, if such a product is calculated multiple times, as in the case of the Lanczos process, the price drops down to $2N_\mathcal{P}N_\Omega$ per product. Appendix 3 discusses this result in more detail and provides necessary formulae.

The formalism introduced in Pankratov and Kuvshinov (2015) allows one to work with any responses that arise in EM problem set-ups either with natural- or controlled-source excitations. Using this methodology one can readily obtain appropriate formulae for the specific sounding methods. To illustrate the concept authors provide such formulae for two EM techniques: magnetotellurics and controlled-source sounding with vertical magnetic dipole as a source.

11 Modern Trends in Developments of Efficient 3-D EM Forward Modelling Schemes

Implementation of a M–H concept, as shown in previous sections, also requires a huge amount of 3-D forward modellings. This prompts development of highly efficient and scalable modelling codes. There are three basic numerical simulation techniques for computing 3-D EM fields/responses, namely finite-difference (FD) methods, finite element (FE) methods and volume integral equation (IE) methods. For decades FD schemes (Mackie et al. 1994; Haber and Ascher 2001; Newman and Alumbaugh 2002, among others) dominated in EM; however, in recent years FE and IE methods have grown in popularity due to a number of advantages which we discuss below.

11.1 FE Method: Short Summary

It is well known that FE methods are a very convenient way to treat conductivity models that involve complicated shapes and curved interfaces, for instance, free surface topography or bathymetry, by utilizing unstructured meshes. It is also important that FE method for Maxwell's equations provides a well-elaborated theory (Hiptmair 2002; Monk 2003, among others) which comprises error estimation analysis (Beck et al. 2000; Bürg 2000, among others) and efficient adaptive schemes (e.g., Bürg 2013; Ren et al. 2013). The popularity of FE is also driven by the availability of open-source libraries (Kirk et al. 2006; Logg and Wells 2010, among others) and mesh generating software which simplify development of the codes for specific applications. There is a large number of papers published recently which use a variety of finite element formulations for 3-D EM modelling [see review paper by Börner (2010) and more recent works by Schwarzbach et al. (2011), Farquharson and Miensopust (2011), Puzyrev et al. (2013), Ren et al. (2013) and Um et al. (2013)].

Usually, FE discretizations of Maxwell's equations result in large sparse systems of linear equations. The solution of these systems is the most time-consuming part of the modelling codes. In addition the standard formulation of frequency-domain Maxwell's equations involves the operator $\nabla\times$ which inherits a large cokernel. In other words, numerical estimations of $\nabla \times \mathbf{E}$ and $\nabla \times \mathbf{H}$ do not satisfy the zero divergence equation. The standard recipe to overcome this problem is an application of a divergence correction scheme (Smith 1996; Farquharson and Miensopust 2011, among others). Although this

correction supresses to some extent the fictive effects in the solution, the original problem remains ill-conditioned and generic iterative methods applied to the resulting linear system converge slowly (Ernst and Gander 2011; Um et al. 2013). Furthermore, locally refined grids are often required to properly discretize large modelling domains. This also often results in performance losses (Mulder 2006; Farquharson and Miensopust 2011) due to increase in the condition number of the linear system matrix. Thus there is a demand for stable and high-performance FE algorithms to the aforementioned problems. Another claim is the scalability of new FE schemes, bearing in mind growing availability of massive parallel processor arrays. An important step forward has been made recently by Grayver and Burg (2014) and Grayver and Kolev (2015) who presented highly robust and scalable FE solvers for 3-D EM problems with large conductivity contrasts, wide range of frequencies, stretched grids and locally refined meshes.

11.2 Integral Equation (IE) Method: Detailed Summary

The application of IE to 3-D EM modelling dates to the works of Raiche (1974), Hohmann (1975) and Weidelt (1975). Since then the IE concept has progressed dramatically but there is still a general preconception in the EM community that IE-based solvers are the solvers of choice only for 3-D models with simple and compact geometries. In the following we will try to persuade the reader that recent developments of the IE concept, both theoretical and numerical, make modern IE realizations competitive with the most advanced FD- and FE-based codes, irrespective of the complexity and extension of 3-D conductivity models under consideration.

This is especially true for the particular class of IE with a contracting kernel [hereinafter denoted as contracting integral equation (CIE)] which possesses a remarkable property: its system matrix is well-conditioned irrespective of discretization, frequency and contrasts of conductivity. The next section summarizes the evolution of the CIE concept, and Sect. 11.2.2 discusses general advantages of the IE approach and modern developments. Note that the explanation of the CIE approach in a nutshell, and some theoretical aspects of CIE are presented in Appendices 4 and 5, respectively.

11.2.1 The Evolution of CIE Concept

The CIE concept was first introduced by Fainberg and Zinger (1980), who obtained an integral equation (IE) of a specific form and proved that this equation can be solved using simple iterates that always converge to the equation solution. The authors called this technique the iterative dissipative method (IDM) and showed that the optimal convergence rate of the IDM iterates is inversely proportional to the *lateral contrast* of the conductivity distribution.

Singer (1995) derived a new CIE and showed that the optimal convergence rate of simple iterates as applied to new IE is inversely proportional to the *square root of the lateral contrast* of the conductivity distribution. He called this technique modified iterative dissipative method (MIDM). Two remarks are relevant at this point: (a) originally in both methods displacement currents are ignored and only isotropic conductivities are considered; (b) in both methods optimal convergence rate is achieved by specific choice of the *reference* 1-D media which does not have to always coincide with the background 1-D section.

Pankratov et al. (1995, 1997) [and independently Singer and Fainberg (1995, 1997)] generalized the technique to media with complex-valued conductivities (to account for

displacement currents and polarization effects), and to media with tensor-valued conductivities (to account for conductivity anisotropy). This became possible due to a renewed way of looking on CIE as a spectrum-shifted equation, in which the shift is not a constant but a spatial distribution. An optimal choice of the shift is achieved by extracting perfect squares from the energy inequality, which provides faster convergence of the CIE iterates. In addition, Pankratov et al. (1995) proposed a general scheme (based on Lobachevsky–Bolyai geometry) to construct reference 1-D media which delivers an optimal convergence rate, irrespective of whether the conductivity is a real- or complex-valued function. Also Pankratov et al. (1995) noted that the simple iterates applied to solve the CIE equation can be considered as a partial sum of convergent Neumann series for this equation. In their nomenclature the technique has been named a modified Neumann series (MNS).

Further development of the CIE approach was made by Avdeev et al. (2000), who observed that the integral operator of CIE is well-conditioned irrespective of the physically feasible conductivity contrasts, and suggested that Krylov subspace iterates may be used to replace the Neumann series summation in the solution of the CIE. Later it was widely reported (Avdeev et al. 2002; Kuvshinov et al. 2005; Singer 2008, among others) that the implementation of Krylov subspace iterates further reduces the number of iterations needed to obtain the CIE solution. Although it was not proved theoretically, the common observation is that the rate of convergence—if Krylov subspace iterates are implemented—becomes inversely proportional to the *natural logarithm of the lateral contrast* of the conductivity distribution (see, e.g., Singer 2008). Now all codes based on CIE—either working in spherical (Koyama et al. 2006; Kuvshinov 2008; Sun and Egbert 2012) or Cartesian (Avdeev et al. 2002; Hursan and Zhdanov 2002; Singer 2008; Koyama et al. 2008; Avdeev and Knizhnik 2009; Kamm and Pedersen 2014) geometries—exploit Krylov subspace iterates.

It is interesting to note that Pankratov et al. (1995) stated in the conclusion of their paper that CIE approach has, in a certain sense, a universal character and can effectively be used, for example, in seismology and acoustics. Almost two decades later the CIE concept indeed attracts the attention of the seismic community. As far as we know Abubakar and Habashy (2013) were the first who presented a CIE formalism to solve 3-D seismic (acoustic approximation) wave propagation problems.

11.2.2 General Advantages of IE Approach and Modern Developments

The general computational advantages of IE-based solvers along with recent developments are summarized below.

- The two key components of all IE-based codes are the computation of the Green's functions and the numerical solution of IE, respectively. In many scenarios of practical interest CPU loads for the computation of Green's functions are comparable to, or outperform, the loads needed to obtain a numerical solution of IE [see some estimates for spherical geometry set up in Kelbert et al. (2014)]. This fact puts the IE-based codes at a computational advantage when repeated forward modellings are required, such as for multiple source calculations, and/or conductivity inverse problems. Indeed, as long as the background model remains unchanged, Green's tensors only need to be computed once and may be reused for multiple sources, as well as for varying 3-D conductivity distributions during inversion iterates.
- Certain properties of typical Green's functions allow efficient storage and fast application (convolution with some vector) of Green's operators, leading to fast

implementation of Krylov subspace iterates (Greenbaum 1997) to the solution of IE. One such property is the shift invariances of the Green's functions defined on lateral regular grids under invariant background geometries, e.g., shift invariances in lateral dimensions for horizontally layered structures under Cartesian geometry, and rotational invariance under spherically layered geometry. Such invariances allow performing linear (for Cartesian geometry) or circular (for spherical geometry) convolutions using fast Fourier transform (FFT). Hereinafter we will discuss mostly the results relevant to Cartesian geometry set up. In this case elements of Green's tensors can be written as

$$G(x, y, z, x', y', z') = G(x - x', y - y', z, z'), \tag{69}$$

and thus the loads to store and calculate elements of Green's tensors are proportional to $N_x \times N_y \times N_z^2$, and the loads to solve IE are proportional to $2N_x \times \ln(2N_x) \times 2N_y \ln(2N_y) \times N_z^2$, where N_x, N_y and N_z are dimensions in lateral and vertical directions, respectively.

Representation of the fields and conductivity on a uniform (in lateral directions) grid is a necessary prerequisite for the use of FFTs. However, in many practical cases, one needs large model dimensions (for example, to account for the regional effects, such as ocean-continent contrast), and relatively short wavelengths of the field to accurately compute EM responses in the region of observations/interest. In these cases, the use of a regular grid leads to an excessively large number of cells and thereby to a groundless increase in the required computational loads. This obstacle is tackled in many IE publications (Phillips and White 1996; Nie et al. 2013; Avdeev et al. 2002; Kamm and Pedersen 2014, among others), and probably the most natural solution to the problem is to use a concept of nested uniform grids (cf. Avdeev et al. 2002; Kamm and Pedersen 2014), which preserves the computation efficiency of the Green's tensor and its convolutions. The following idea underlies the concept. The initial CIE in the entire model volume is reduced to a system of integral equations specified in successively nested volumes. The resulting system is then solved by an iterative procedure gradually improving the accuracy of the solution at each iteration.

Quite recently Koyama et al. (2008) and Avdeev and Knizhnik (2009) showed that separability of the spectral Green's tensors allows for achieving linear dependence of IE computational loads on the vertical dimension N_z. Avdeev and Knizhnik (2009) demonstrated that in the course of IE solution the calculation and storage of $N_x \times N_y \times 5N_z$ elements of Green's tensors are required, rather than calculation and storage of the whole set of $N_x \times N_y \times N_z^2$ elements. Note that so far we did not assume that discretization in the vertical direction is uniform.

In some special cases one can apply FFT to all three space dimensions, for example if the background section is represented by a uniform space. In this case one can write elements of Green's tensors as

$$G(x, y, z, x', y', z') = G(x - x', y - y', z - z'). \tag{70}$$

One can also apply FFT in all directions if the background is a homogeneous half-space. In this case the elements of Green's tensors can be decomposed into direct waves and waves reflected from the interface (cf. Millard and Liu 2003; Kamm and Pedersen 2014)

$$G(x, y, z, x', y', z') = G_1(x - x', y - y', z - z') + G_2(x - x', y - y', z + z'). \tag{71}$$

Consequently, the FFT can be applied to both terms in the latter equation through the convolution and correlation theorems, which also gives linear dependence of IE computational loads on the vertical dimension N_z. However we notice here that using FFT in the vertical direction leads to uniform discretization in this direction.

- In contrast to FD and FE formulations that invoke large system matrices, the IE-based codes work with compact system matrices. The reason for compactness is that boundary conditions are exactly accounted for via Green's functions, and thus the modelling region is confined only to 3-D anomalies. By contrast, in the FD and FE codes one has to discretize a much larger volume both in lateral and vertical directions in order to enable the decay (or stabilization) of the fields at the boundaries of the domain. One can argue that this advantage is counterbalanced by the fact that IE matrices are dense but FD and FE matrices are sparse.
  However it is possible to surmount this IE shortcoming by exploiting the diagonal dominance of the Green's functions. The diagonal dominance is the result of a general phenomenon that EM field responses of a collection of sources reduce with distance. Acceleration methods based on such consideration are rather numerous, see Chew et al. (2014) for a review. Two such methods widely used in high-frequency EM forward modelling (e.g., optical and microwave scattering) are fast multipole algorithm and varieties (Rokhlin 1990; Lu and Chew 1994, among others), and matrix decomposition algorithm and varieties (Michielssen and Boag 1996; Rius et al. 2008, among others). Recently Sun and Kuvshinov (2015) apply a singular value decomposition (SVD)-based matrix compression method, similar to that developed in Rius et al. (2008), to an IE forward solver for global geomagnetic induction. After employing rotational invariance of the Green's function of a spherical Earth by applying FFT in the longitudinal direction, the reduced system is further compressed using a divide-and-conquer approach based on SVD low-rank approximations with controlled errors in terms of Frobenius norm, leading to a recursively compressed form of the system matrix that is computationally efficient.
- The IE approach in general is less demanding than FD and FE schemes, in terms of cell size, since it does not involve numerical differentiation.
- IE schemes allow for rather straightforward parallelization, for example, with respect to integration along layers containing inhomogeneities, with respect to convolution of different elements of Green's tensors with corresponding components of the input vector field, and/or with respect to performing distributed FFT.
- So far we have discussed the properties which both standard IE and CIE share. Regarding specifically the CIE its system matrices has a remarkable property—it is well conditioned by construction irrespective of discretization, frequency and contrasts of conductivity, and thus does not require preconditioning, whereas FD and FE do, especially for the models with large conductivity contrasts (see an estimate of condition number for CIE system in Appendix 5).

Note that the success of IE solvers is strongly dependent on having a robust and fast computation of Green's tensors in the background medium. Such computation is not a trivial task and requires a lot of analytical work. Avdeev et al. (1997) and Geraskin et al. (2015) provide a number of numerical recipes to compute Green's tensors (in Cartesian geometry) very efficiently, including implementation of fast Hankel transform (Christensen 1990) and fast and accurate integration of Green's functions.

Summarizing this section we claim that the CIE-based solvers, if properly coded, can constitute a competitive alternative to the most advanced FE- and FD-based solvers in

Table 1 CPU times (see columns 2–4; in seconds) of novel CIE-based code (Geraskin et al. 2015) with respect to an increasing number of CPUs (see column 1)

MPI processes	Green's tensors	CIE solution	Overall (Green + 2 × CIE)
16	234	276	786
32	99	133	365
64	67	76	219
128	34	37	108
256	21	21	63
512	14	10	34
1024	11	6	22

The computations were performed on Cray XC30 supercomputer "Piz Daint" of Swiss National Supercomputer Centre. The same numerical grid, $N_x \times N_y \times N_z = 512 \times 512 \times 40$, was exploited for all runs. The grid is equidistant in $x-$ and $y-$directions and covers only the domain where conductivity is non 1-D. Note that the number of degrees of freedom (DoF) in this CIE solution is $3 \times 2 \times N_x \times N_y \times N_z$ where 3 indicates three components of the (electric) field, and 2 – their real and imaginary parts. The CPU time in the last (fourth) column is for two plane-wave excitations (note that for MT scenario the Green's tensor for a given period is computed only once). The resulting system of equations $Ax = b$ was solved by Krylov subspace iterations, or, more explicitly, by FGMRES (Frayss et al. 2003). The CPU time for $Ax = b$ solution is summarized in the third column. It was assumed that the approximation to the solution, $x^{(n)}$, was obtained once the following inequality holded: $||Ax^{(n)} - b||/||b|| < \epsilon$, with $\epsilon = 10^{-8}$. In this particular case the total (for two plane-wave excitations) number of matrix-vector multiplications, Ax, was 139. *MPI* stands for Message Passing Interface

terms of accuracy, performance and scalability, irrespective of complexity, expansion and spatial detail of 3-D conductivity models under consideration. To support this statement Table 1 provides CPU time (in seconds) of novel CIE-based code by Geraskin et al. (2015) with respect to an increasing number of CPUs. Parallelization in this code involves a fragmentation of the model along one lateral direction and distributed convolution using distributed FFT. The model under consideration is COMMEMI 3D-II model (shown in Fig. 6), and computed quantities are magnetotelluric (MT) responses at period of 100 s. It is seen from the table that the run time of the code decreases proportionally to the number of processes, which means that the code is strongly scalable. Note that this model was excessively discretized by $N_x \times N_y \times N_z = 512 \times 512 \times 40$ cells in order to demonstrate the scalability of the code; one can obtain the results of comparable accuracy with much

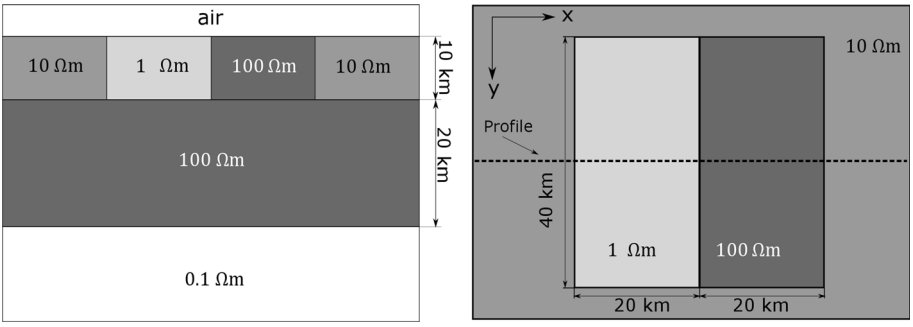

Fig. 6 COMMEMI 3D-II conductivity model. *Left* and *right plots* are side and plane views of the model, respectively

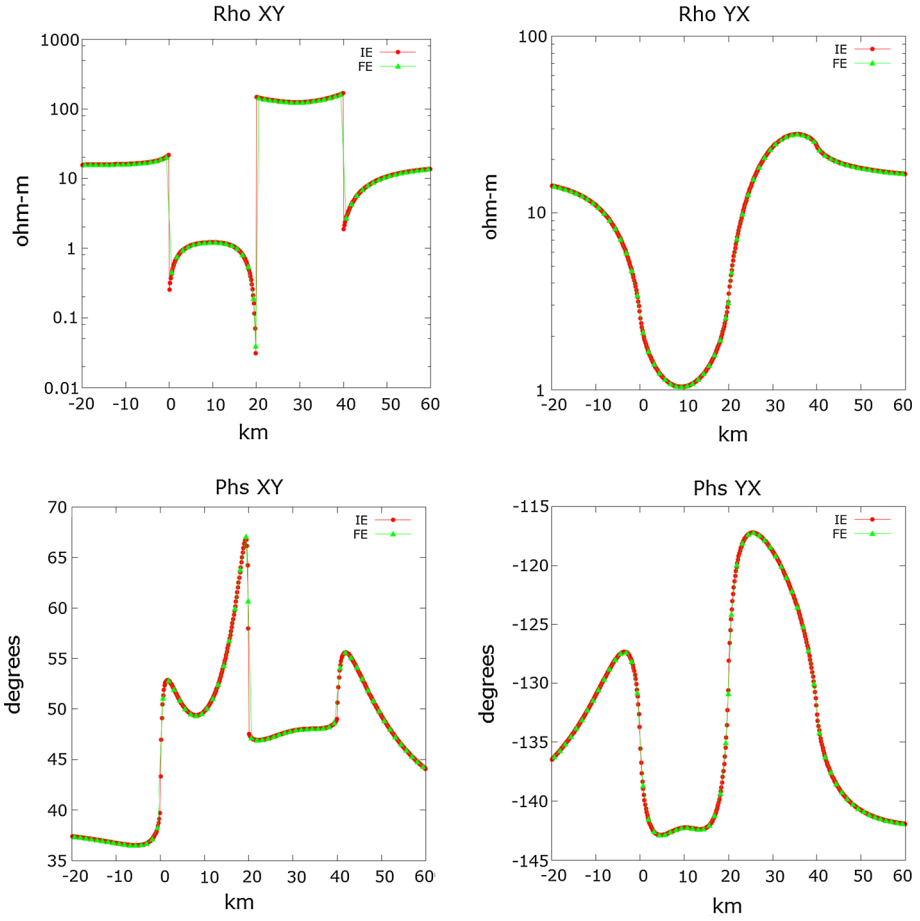

Fig. 7 MT responses at period of 100 s along profile depicted by *dashed line* in Fig. 6. *Left* and *right upper plots* show apparent resistivities, ρ_{xy} and ρ_{yx}, respectively. *Lower plots* demonstrate corresponding phases of impedances. The comparison is between IE (Geraskin et al. 2015) and FE (Grayver and Kolev 2015) results. See details in the text

smaller gridding of the model. An accuracy of the code is illustrated in Fig. 7 where IE results are compared with the results obtained by one of the most advanced 3-D FE code (Grayver and Kolev 2015). One can see that the agreement between IE and FE results is excellent.

12 Concluding Remarks

In this review we address the problem of uncertainty quantification in 3-D EM studies of the Earth. It is rather evident that a Bayesian (stochastic) approach is the most appropriate tool to quantify systematically such uncertainties. However, there is a common opinion that it is still prohibitive to perform uncertainty analysis in the framework of 3-D models due to tremendous computational loads. We argue that recent strides in 3-D forward

problem modelling and sampling algorithms (that are cornerstones of any stochastic inversion), and implementation of adjoint source formalism likely make this analysis tractable.

In the discussion of sampling algorithms, we concentrate on the most popular technique—the Metropolis–Hastings scheme and its recent (potentially less expensive) modification. We also discuss how the results of stochastic (and even deterministic) inversions can be used to quantify uncertainties of the recovered electrical conductivity models. Applying of the developed formalisms to practical scenarios is intentionally beyond the scope of the paper but will be the subject of a subsequent study.

As for 3-D forward modelling we made a decision to mostly address recent progress in integral equation solutions. Surprisingly, this method is overlooked in all recent reviews which are dedicated to a progress in 3-D forward problem solutions. We debate that the modern IE-based codes, if properly coded, could be a competitive alternative to the most advanced FE- and FD-based codes, irrespective of complexity, contrasts, expansion and spatial detail of 3-D conductivity models under consideration.

Most of this review (along with the appendices) is rather mathematically saturated but is written in a way to allow researchers to apply the discussed ideas and algorithms in a rather straightforward manner.

Acknowledgments The authors would like to thank Alexander Grayver for many seminal discussions over the course of this work, and Alexey Geraskin and Alexander Grayver for their input to Sect. 11. We wish to thank William Lowrie who helped us to improve the English presentation of this paper. We extend our gratitude to Chester Weiss and an anonymous reviewer for constructive comments on the manuscript. This work has been supported by the European Space Agency through ESTEC contract No. 4000102140/10/NL/JA and in part by the Russian Foundation for Basic Research under Grant No. 13-05-12111. Oleg Pankratov acknowledges the support of ETH during his stay in Zurich as a visiting professor.

Appendix 1: The Misfit Renormalization

Quadratic approximation $[-\frac{1}{2}(2\mathbf{g}_k^T\mathbf{s}+\mathbf{s}^T\mathrm{H}_k\mathbf{s})$, cf. Eq. (26)] of the misfit $\beta_d(\mathbf{m})$ might not work well in the desired vicinity of the current conductivity model $\mathbf{m}^{(k)}$. In this case we suggest to use the following renormalization of the misfit

$$\beta_d \mapsto \frac{\beta_d}{\kappa} \tag{72}$$

which tells us that the larger local uncertainty ellipsoid (explained later in this section) has more chances to catch the true minimum point as illustrated in Fig. 8.

Renormalization (72) leads to a modification in Eqs. (28), (32), (33), (58) and (37) as follows

$$\frac{\mathbf{g}_k^T}{\kappa}\mathbf{s}+\tfrac{1}{2}\mathbf{s}^T\frac{\mathrm{H}_k}{\kappa}\mathbf{s}=\tfrac{1}{2}\left|\mathbf{F}_k\frac{\mathbf{s}}{\sqrt{\kappa}}+\mathbf{F}_k^{-T}\frac{\mathbf{g}_k}{\sqrt{\kappa}}\right|^2-\tfrac{1}{2}\left|\mathbf{F}_k^{-T}\frac{\mathbf{g}_k}{\sqrt{\kappa}}\right|^2, \tag{73}$$

$$\mathbf{g}_k'=-\mathbf{F}_k^{-T}\frac{\mathbf{g}_k}{\sqrt{\kappa}}, \tag{74}$$

$$\mathbf{y}=\mathbf{F}_k\frac{\mathbf{s}}{\sqrt{\kappa}}, \tag{75}$$

$$\mathbf{s}=\sqrt{\kappa}\,\mathbf{F}_k^{-1}\mathbf{y} \tag{76}$$

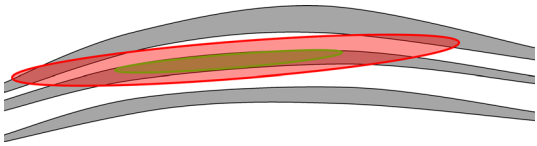

Fig. 8 Illustration of necessity to introduce renormalization factor κ, see details in the text

$$\mathbf{s} = -\mathtt{H}_k^{-1}\mathbf{g}_k + \sqrt{\kappa}\,\mathbf{F}_k^{-1}\mathcal{N}_{N_\mathcal{M}}(\mathbf{0}, 1), \tag{77}$$

and to corresponding equations for $\tilde{\mathbf{F}}$ (recall that $\tilde{\mathbf{F}}$ is a low-rank approximation for $\mathbf{F}$, see Sect. 5).

We see that the renormalization leads to the similarity transformation of the uncertainty ellipsoid, but it does not change the Newtonian step $\left(-\mathtt{H}_k^{-1}\mathbf{g}_k\right)$. We call here as *uncertainty ellipsoid* the following set

$$\mathcal{U} = -\mathtt{H}_k^{-1}\mathbf{g}_k + \sqrt{\kappa}\mathbf{F}_k^{-1}B(\mathbf{0}, 1), \tag{78}$$

where $B(\mathbf{0}, 1)$ is a unit sphere in $\mathbb{R}^{N_\mathcal{M}}$ centered at the origin.

Finally, the renormalization factor κ might be one of the parameters that are needed to keep the acceptance rate (percentage of the accepted models in the sample) to be in the desired interval; e.g., in the M–H method it is generally accepted that the rate should be between 20 and 70 %.

Appendix 2: Summary of Formulae to Calculate Data Misfit Gradient

This and follow-up appendices summarize the results presented in Pankratov and Kuvshinov (2015). But before we proceed with final formulae we introduce definitions to be used.

Green's Operators

Let us define an operator $\mathbf{G}^{\cdot\cdot}$ in 3-D space $\mathbb{R}^3$

$$\begin{pmatrix}\mathbf{E}\\ \mathbf{H}\end{pmatrix} = \mathbf{G}^{\cdot\cdot}\begin{pmatrix}\mathbf{j}^{\text{imp}}\\ \mathbf{h}^{\text{imp}}\end{pmatrix} \Leftrightarrow \begin{cases}\nabla\times\mathbf{H} = \sigma\mathbf{E} + \mathbf{j}^{\text{imp}},\\ \nabla\times\mathbf{E} = i\omega\mu\mathbf{H} + \mathbf{h}^{\text{imp}},\\ \mathbf{E}(\mathbf{r}), \mathbf{H}(\mathbf{r}) \longrightarrow 0 \quad \text{as} \quad |\mathbf{r}| \longrightarrow \infty,\end{cases} \tag{79}$$

where $\mathbf{E}$ and $\mathbf{H}$ are electric and magnetic fields, $\mathbf{j}^{\text{imp}}$ and $\mathbf{h}^{\text{imp}}$ are impressed (extraneous) electric and magnetic sources, respectively, $\mathbf{r} \in \mathbb{R}^3$ is a position vector, $i = \sqrt{-1}, \omega = 2\pi/\text{Period}$ is an angular frequency, $\sigma(\mathbf{r})$ and $\mu(\mathbf{r})$ are electric conductivity and magnetic permeability distributions in an Earth's model, respectively. In this appendix we assume that $\sigma(\mathbf{r})$ is a real-valued function. One can readily generalize the concept for complex-valued conductivity. The corresponding formulae are provided in the last appendix of Pankratov and Kuvshinov (2015). All fields, $\mathbf{E}, \mathbf{H}, \mathbf{j}^{\text{imp}}, \mathbf{h}^{\text{imp}}$, are complex-valued functions of ω and $\mathbf{r}$. In addition the fields $\mathbf{E}$ and $\mathbf{H}$ depend on σ and μ. We study the derivatives with respect to σ only. Green's operator $\mathbf{G}^{\cdot\cdot}$ depends on functional arguments $\mathbf{j}^{\text{imp}}$ and $\mathbf{h}^{\text{imp}}$.

Hereinafter the dependence of Green's operator on $\sigma, \mathbf{r}$, and ω is omitted but implied. Time dependence of fields is accounted for by $e^{-i\omega t}$, which reads, for example, for the electric field as $\breve{\mathbf{E}}(\mathbf{r}, t) = \int \mathbf{E}(\mathbf{r}, \omega) e^{-i\omega t} \mathrm{d}\omega$. At this stage we do not specify the coordinate system in $\mathbb{R}^3$; this means that $\mathbf{r}$ can be, for example, a triplet of Cartesian coordinates, (x, y, z), or a triplet of spherical coordinates, (r, θ, ϕ). As far as the column in the left-hand side (LHS) of Eq. (79) contains two fields, $\mathbf{E}$ and $\mathbf{H}$, operator $\mathbf{G}^{\cdot\cdot}$ can be represented via operators $\mathbf{G}^{e\cdot}, \mathbf{G}^{h\cdot}\, \mathbf{G}^{ee}, \mathbf{G}^{eh}, \mathbf{G}^{he}, \mathbf{G}^{hh}$ as follows

$$\mathbf{G}^{\cdot\cdot} = \begin{pmatrix} \mathbf{G}^{e\cdot} \\ \mathbf{G}^{h\cdot} \end{pmatrix} = \begin{pmatrix} \mathbf{G}^{ee} & \mathbf{G}^{eh} \\ \mathbf{G}^{he} & \mathbf{G}^{hh} \end{pmatrix}, \quad \mathbf{G}^{e\cdot} = \left(\mathbf{G}^{ee}, \mathbf{G}^{eh}\right), \quad \mathbf{G}^{h\cdot} = \left(\mathbf{G}^{he}, \mathbf{G}^{hh}\right), \tag{80}$$

where operators $\mathbf{G}^{e\cdot}$ and $\mathbf{G}^{h\cdot}$ are electric and magnetic components of $\mathbf{G}^{\cdot\cdot}$, operator $\mathbf{G}^{ee}$ is a restriction of $\mathbf{G}^{e\cdot}$ to electric sources etc.

Let us introduce an electromagnetic field, $\mathbf{u}$, as

$$\mathbf{u}(\mathbf{r}, \omega) = \begin{pmatrix} \mathbf{E}(\mathbf{r}, \omega) \\ \mathbf{H}(\mathbf{r}, \omega) \end{pmatrix}, \tag{81}$$

which is a complex-valued six-dimensional (6-D) vector. Let us denote the space of such vectors as $\mathcal{U} \cong \mathbb{C}^6$. Note that once we have chosen coordinates in 3-D space $\mathbb{R}^3$ with the following basis

$$\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \tag{82}$$

then we naturally and unambiguously have a coordinate system and basis $\mathbf{e}'_1, \cdots, \mathbf{e}'_6$ in 6-D complex space $\mathcal{U}$

$$\mathbf{u}(\mathbf{r}, \omega) = \sum_{\alpha=1}^{6} u_\alpha \mathbf{e}'_\alpha, \tag{83}$$

saying that $\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3$ are $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ for electric fields, whereas $\mathbf{e}'_4, \mathbf{e}'_5, \mathbf{e}'_6$ are $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ for magnetic fields, respectively.

Polarizations/Sources

Let

$$\left\{\mathbf{f}_p^{\mathrm{imp}}\right\}_{p \in \mathcal{P}}, \quad \mathcal{P} = \{1, 2, \ldots, N_{\mathcal{P}}\}, \tag{84}$$

be a set of linearly independent distributions (in space and frequency) of the impressed sources, $\mathbf{f}_p^{\mathrm{imp}}$. For example, in magnetotelluric (MT) studies, $N_{\mathcal{P}} = 2$, and $\mathbf{f}_1^{\mathrm{imp}}$ and $\mathbf{f}_2^{\mathrm{imp}}$ correspond to the plane waves of different orientations. Each $\mathbf{f}_p^{\mathrm{imp}}$ produces electric, $\mathbf{E}_p$, and magnetic, $\mathbf{H}_p$, fields that constitute an EM field $\mathbf{u}_p$ that can be written via $\mathbf{G}^{\cdot\cdot}$ operator (104) as

$$\mathbf{u}_p = \mathbf{G}^{\cdot\cdot}\left(\mathbf{f}_p^{\mathrm{imp}}\right). \tag{85}$$

Inversion Domain and Parameterization

As far as the inversion is usually done numerically, let the inversion domain, V^{inv}, be represented as

$$V^{\text{inv}} = \bigcup_{l=1}^{N_{\mathcal{M}}} V_l, \tag{86}$$

where $\{V_l\}_{l\in\mathcal{M}}$, $\mathcal{M} = \{1, \ldots, N_{\mathcal{M}}\}$, represent a set of elementary volumes V_l, and within each volume V_l let the conductivity be a constant $\sigma(\mathbf{r}) = \sigma_l$. We assemble this conductivity distribution in the following vector

$$\boldsymbol{\sigma} = \left(\sigma_1, \ldots, \sigma_{N_{\mathcal{M}}}\right)^T, \tag{87}$$

and introduce model parameterization as

$$\mathbf{m} = \left(m_1, \ldots, m_{N_{\mathcal{M}}}\right)^T, \quad m_l = v^{-1}(\sigma_l), \quad l \in \mathcal{M}, \tag{88}$$

where function $\mathbf{m} = v^{-1}(\boldsymbol{\sigma})$ can be implemented, for example, to preserve conductivity to be positive. Note that a popular choice is $\mathbf{m} = \ln \boldsymbol{\sigma}$. We also remark that some volumes V_l might be cells (or combinations of cells) of the 3-D part of the model.

Observation Sites, Frequencies, Response Functions and Misfit

Let

$$\Phi_g, \Delta\Phi_g, \qquad g \in \mathcal{G} = \{1, 2, \ldots, N_{\mathcal{G}}\}, \tag{89}$$

be the experimental responses and their uncertainties, respectively, and let $N_{\mathcal{G}}$ be the number of all responses. Let $\mathbf{r}_g$, and ω_g be the spatial location and the frequency, respectively, at which the response Φ_g has been obtained.

Let $\mathcal{S}$ be a set of observation sites

$$\mathcal{S} = \left\{\mathbf{r}_g \,|\, g \in \mathcal{G}\right\} = \{\mathbf{s}_1, \ldots, \mathbf{s}_{N_S}\}, \tag{90}$$

where $\mathbf{s}_1, \ldots, \mathbf{s}_{N_S}$ are different observation sites, and $N_{\mathcal{S}}$ is the number of sites.

Let Ω be a set of observation frequencies

$$\Omega = \left\{\omega_g \,|\, g \in \mathcal{G}\right\} = \{f_1, \ldots, f_{N_\Omega}\}, \tag{91}$$

where $f_1, \ldots, f_{N_\Omega}$ are different observation frequencies, and N_Ω is the number of frequencies. The definitions (89)–(91) are introduced in this specific way intentionally in order to stress the fact that in practice an actual set of experimental responses to be used for inversion varies with frequency and site.

For each $g \in \mathcal{G}$, the predicted response, θ_g, can be written in the following form

$$\theta_g(\mathbf{m}) = \Psi_g\left(\mathbf{u}_1(\mathbf{m}, \mathbf{r}_g, \omega_g), \mathbf{u}_2(\mathbf{m}, \mathbf{r}_g, \omega_g), \ldots, \mathbf{u}_{N_\mathcal{P}}(\mathbf{m}, \mathbf{r}_g, \omega_g)\right). \tag{92}$$

Finally the misfit is introduced as

$$\beta_d(\mathbf{m}) = \sum_{g\in\mathcal{G}} \left| \frac{\theta_g(\mathbf{m}) - \Phi_g}{\Delta\Phi_g} \right|^2. \tag{93}$$

Following Pankratov and Kuvshinov (2015) we write the elements of data misfit gradient as

$$\frac{\partial \beta_d}{\partial m_l} = 2\nu_l' \,\mathrm{Re} \sum_{\omega\in\Omega}\sum_{p\in\mathcal{P}} \int\limits_{V_l} \mathbf{E}_p(\omega)\cdot \mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(\omega)\right)\,\mathrm{d}v, \quad l=1,2,\ldots,N_{\mathcal{M}}, \tag{94}$$

where an *adjoint source* $\mathbf{J}_p^M$ is given by

$$\mathbf{J}_p^M(\omega) = \sum_{g:\omega_g=\omega} \frac{(\theta_g-\Phi_g)^*}{|\Delta\Phi_g|^2}\frac{\partial \Psi_g}{\partial \mathbf{u}_p}\delta_{\mathbf{r}_g}\Big|_{\omega}, \tag{95}$$

where * stands for complex conjugation. Table 2 summarizes the steps needed to calculate the misfit gradient. From the Eq. (94) it is seen that we need $2N_{\mathcal{P}}N_{\Omega}$ forward modellings in total to calculate the data misfit gradient.

Appendix 3: Summary of Formulae to Calculate the Hessian-Vector Products

We are interested to calculate $\mathrm{Hess}_{\beta_d}\,\mathbf{a}_k, k=1,\ldots,K$, where $\mathbf{a}_k$ we represent as

$$\mathbf{a}_k = \sum_{m=1}^{N_{\mathcal{M}}} a_{km}\mathbf{1}_{V_m}(\mathbf{r}), \tag{96}$$

where $\mathbf{1}_{V_l}(\mathbf{r})$ is an indicator function given by

$$\mathbf{1}_{V_m}(\mathbf{r}) = \begin{cases} 1, & \mathbf{r}\in V_m, \\ 0, & \mathbf{r}\notin V_m. \end{cases} \tag{97}$$

Following Pankratov and Kuvshinov (2015), the l-th element of the Hessian-vector product $\mathrm{Hess}_{\beta_d}\,\mathbf{a}_k$ (which is a vector) has a form

$$\left[\mathrm{Hess}_{\beta_d}\,\mathbf{a}_k\right]_l = \mathrm{Re}(\mathcal{B}_{kl}^A + \mathcal{B}_{kl}^L), \quad l=1,2,\ldots N_{\mathcal{M}}, \tag{98}$$

where $\mathcal{B}_{kl}^A$ and $\mathcal{B}_{kl}^L$ are

$$\mathcal{B}_{kl}^A = 2\sum_{\omega\in\Omega}\sum_{p\in\mathcal{P}}\int\limits_{V_l} \nu_l'\mathbf{E}_p\cdot\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^B(\nu'\mathbf{a}_k)\right)\mathrm{d}v\Big|_{\omega}, \tag{99}$$

Table 2 The steps needed to calculate the gradient of the data misfit β_d

The term	Indices range	# of forward runs
$\mathbf{u}_p(\omega)=\mathbf{G}^{\cdot e}\left(\mathbf{f}_p^{\mathrm{imp}}\right)$	$p\in\mathcal{P}, \omega\in\Omega$	$N_{\mathcal{P}}N_{\Omega}$
$\mathbf{E}_p(\omega), \theta_g(\mathbf{m}), \frac{\partial\Psi_g}{\partial\mathbf{u}_p}, \mathbf{J}_p^M(\omega)$		0
$\mathbf{G}^{e\cdot}\left(\mathbf{J}_p^M(\omega)\right)$	$p\in\mathcal{P}, \omega\in\Omega$	$N_{\mathcal{P}}N_{\Omega}$
The total number of forward modellings		$2N_{\mathcal{P}}N_{\Omega}$

$$\begin{aligned}\mathcal{B}^L_{kl} = 2\sum_{\omega\in\Omega}\sum_{p\in\mathcal{P}}\int\limits_{V_l} v'_l\mathbf{E}_p\cdot\mathbf{G}^{e\cdot}\left(\mathbf{J}^{\Psi}_p(v'\mathbf{a}_k)\right)\mathrm{d}v\Big|_{\omega}\\ + 2\sum_{\omega\in\Omega}\sum_{p\in\mathcal{P}}\int\limits_{V_l}\left\{\left(v'_l\mathbf{G}^{ee}\left(v'\mathbf{a}_k\,\mathbf{E}_p\right)+v''\mathbf{a}_k\,\mathbf{E}_p\right)\cdot\mathbf{G}^{e\cdot}\left(\mathbf{J}^{M}_p\right)\right.\\ \left. + v'_l\mathbf{E}_p\mathbf{G}^{ee}\left(v'\mathbf{a}_k\,\mathbf{G}^{e\cdot}\left(\mathbf{J}^{M}_p\right)\right)\right\}\mathrm{d}v\Big|_{\omega},\end{aligned} \tag{100}$$

where $\mathbf{J}^{M}_p(\omega)$ is defined in Eq. (95) and $\mathbf{G}^{e\cdot}\left(\mathbf{J}^{B}_p(\mathbf{a}_k,\omega)\right)$ and $\mathbf{G}^{e\cdot}\left(\mathbf{J}^{\Psi}_p(\mathbf{a}_k,\omega)\right)$ are as follows

$$\begin{aligned}&\mathbf{G}^{e\cdot}\left(\mathbf{J}^{B}_p(\mathbf{a}_k,\omega)\right)\\ &= \sum_{g:\,\omega_g=\omega}\sum_{q\in\mathcal{P}}\frac{1}{|\Delta\Phi_g|^2}\left\{\sum_{m=1}^{N_{\mathcal{M}}}\int\limits_{V_m} a_{km}\,\mathbf{E}_q\cdot\mathbf{G}^{e\cdot}\left(\frac{\partial\Psi_g}{\partial\mathbf{u}_q}\,\delta_{\mathbf{r}_g}\right)\right\}^{*}\mathbf{G}^{e\cdot}\left(\frac{\partial\Psi_g}{\partial\mathbf{u}_p}\,\delta_{\mathbf{r}_g}\right)\Bigg|_{\omega},\end{aligned} \tag{101}$$

$$\mathbf{G}^{e\cdot}\left(\mathbf{J}^{\Psi}_p(\mathbf{a}_k,\omega)\right) = \sum_{g:\omega_g=\omega}\sum_{q\in\mathcal{P}}\frac{(\theta_g-\Phi_g)^{*}}{|\Delta\Phi_g|^2}\mathbf{G}^{e\cdot}\left(\frac{\partial^2\Psi_g}{\partial\mathbf{u}_p\partial\mathbf{u}_q}\,\mathbf{G}^{\cdot e}\left(\mathbf{a}_k\mathbf{E}_q\right)\delta_{\mathbf{r}_g}\right)\Bigg|_{\omega}. \tag{102}$$

We make here three notes.

- Term $\int\limits_{V_l}\mathbf{E}_p\cdot\mathbf{G}^{e\cdot}\left(\mathbf{J}^{\Psi}_p(\mathbf{1}_{V_k})\right)\mathrm{d}v$ vanishes if the response Ψ is a linear function of EM field $\mathbf{u}$ (e.g., for most of the CSEM methods).
- Term $\int\limits_{V_l}\delta_{lk}v''_l\mathbf{E}_p(\omega)\cdot\mathbf{G}^{e\cdot}\left(\mathbf{J}^{M}_p(\omega)\right)\mathrm{d}v$ vanishes if $\boldsymbol{\sigma}=\mathbf{m}$.
- One can readily generalize the concept for complex-valued conductivity σ. The corresponding formulae are provided in the last appendix of Pankratov and Kuvshinov (2015).

Table 3 provides a number of forward modellings needed to calculate the Hessian-vector product K times. As seen from the table, a single Hessian-vector product can be calculated

Table 3 The steps needed to calculate the Hessian-vector products $\mathrm{Hess}_{\beta_d}\,\mathbf{a}_k, k=1,\ldots,K$

The term	Indices range	# of forward runs
$\mathbf{u}_p(\omega)=\mathbf{G}^{\cdot e}\left(\mathbf{f}^{\mathrm{imp}}_p(\omega)\right)$	$p\in\mathcal{P},\omega\in\Omega$	$N_{\mathcal{P}}N_{\Omega}$
$\mathbf{E}_p(\omega),\theta_g(\mathbf{m}),\frac{\partial\Psi_g}{\partial\mathbf{u}_p}$		0
$\mathbf{G}^{\cdot\cdot}\left(\mathbf{e}'_{\alpha}\delta_{\mathbf{s}_a}\right)\big\vert_{\omega}$	$\alpha=1,\ldots,6,\,\mathbf{s}_a\in\mathcal{S},\omega\in\Omega$	$\tilde{N}\,(\le 6N_{\mathcal{S}}N_{\Omega})$
$\mathbf{G}^{e\cdot}\left(\mathbf{J}^{B}_p(\mathbf{a}_k,\omega)\right)$		0
$\mathbf{G}^{e\cdot}\left(\mathbf{J}^{M}_p(\omega)\right)$		0
$\mathbf{G}^{\cdot e}\left(\mathbf{a}_k\,\mathbf{E}_p\right),\mathbf{G}^{ee}\left(\mathbf{a}_k\,\mathbf{G}^{e\cdot}\left(\mathbf{J}^{M}_p\right)\right)$	$k=1,\ldots,K,p\in\mathcal{P},\omega\in\Omega$	$2KN_{\mathcal{P}}N_{\Omega}$
$\mathbf{G}^{e\cdot}\left(\mathbf{J}^{\Psi}_p(\mathbf{a}_k,\omega)\right)$		0
The total # of forward runs		$\tilde{N}+(2K+1)N_{\mathcal{P}}N_{\Omega}$

for a price of $O(N_\Omega N_S)$ forward problem runs. Moreover, if such a product is calculated multiple times the price drops down to $2N_P N_\Omega$ per product. Note also that computation of the Hessian itself is merely calculation of the Hessian-vector product for $K = N_\mathcal{M}$ times with respective vectors $\mathbf{a}_k = \mathbf{1}_{v_k}, k = 1, \ldots, N_\mathcal{M}$.

Appendix 4: Contracting Integral Equation in a Nutshell

Let $\sigma(\mathbf{r})$ be a desired 3-D model of complex-valued conductivity, including the real part term $\gamma(\mathbf{r}) = \mathrm{Re}\,\sigma(\mathbf{r}) > 0$ as well as the imaginary part term $\eta(\mathbf{r}) = \mathrm{Im}\,\sigma(\mathbf{r})$ that describes displacement currents and/or induced polarization effects. Let the model be excited by electric source $\mathbf{j}^{\mathrm{imp}}$. Let us search for the electric field excited by $\mathbf{j}^{\mathrm{imp}}$ in the model σ. This electric field is the *electric field solution* of Maxwell's equations

$$\mathbf{E} = \mathbf{G}^{ee}(\mathbf{j}^{\mathrm{imp}}) \Leftrightarrow \begin{cases} \nabla \times \mathbf{H} = \sigma \mathbf{E} + \mathbf{j}^{\mathrm{imp}}, \\ \nabla \times \mathbf{E} = i\omega\mu \mathbf{H}, \\ \mathbf{E}(\mathbf{r}), \mathbf{H}(\mathbf{r}) \longrightarrow 0 \quad \text{as} \quad |\mathbf{r}| \longrightarrow \infty. \end{cases} \tag{103}$$

Let $\sigma_b(\mathbf{r})$ be any model where we can evaluate Green's operator $\mathbf{G}^{ee}_{\sigma_b}$ that is electric field solution of the following Maxwell's equations

$$\mathbf{E}_b = \mathbf{G}^{ee}_{\sigma_b}(\mathbf{j}^{\mathrm{imp}}) \Leftrightarrow \begin{cases} \nabla \times \mathbf{H}_b = \sigma_b \mathbf{E}_b + \mathbf{j}^{\mathrm{imp}}, \\ \nabla \times \mathbf{E}_b = i\omega\mu \mathbf{H}_b, \\ \mathbf{E}_b(\mathbf{r}), \mathbf{H}_b(\mathbf{r}) \longrightarrow 0 \quad \text{as} \quad |\mathbf{r}| \longrightarrow \infty. \end{cases} \tag{104}$$

We refer to σ_b as a *reference model*, e.g., it could be a *background (host) model*. In further discussion we assume that σ_b describes one-dimensional (1-D) conductivity section, i.e. $\sigma_b \equiv \sigma_b(z)$. With such introduced reference model, an action of operator $\mathbf{G}^{ee}_{\sigma_b}$ on the field $\mathbf{j}^{\mathrm{imp}}$ is represented by the following convolution integral

$$\mathbf{E}_b = \mathbf{G}^{ee}_{\sigma_b}(\mathbf{j}^{\mathrm{imp}}) = \int\limits_{\mathbb{V}} \mathbf{G}^{ee}_{\sigma_b}(x - x', y - y', z, z')\, \mathbf{j}^{\mathrm{imp}}(x', y', z')\mathrm{d}v', \tag{105}$$

where $\mathbb{V}$ is the volume, occupied by $\mathbf{j}^{\mathrm{imp}}$. Note, that analogously we can obtain "reference" magnetic field, $\mathbf{H}_b$, via corresponding Green's operator $\mathbf{G}^{he}_{\sigma_b}$, namely, $\mathbf{H}_b = \mathbf{G}^{he}_{\sigma_b}(\mathbf{j}^{\mathrm{imp}})$.

Let σ_a be an anomalous conductivity distribution

$$\sigma_a(\mathbf{r}) = \sigma(\mathbf{r}) - \sigma_b(\mathbf{r}). \tag{106}$$

By trivial manipulations with Eqs. (103) and (104) one can arrive at the scattering (integral) equation with respect to $\mathbf{E}$

$$\mathbf{E} = A\mathbf{E} + \mathbf{E}_b, \tag{107}$$

where

$$A = \mathbf{G}^{ee}_{\sigma_b} \circ \sigma_a. \tag{108}$$

Note that the composition operator $A = \mathbf{G}^{ee}_{\sigma_b} \circ \sigma_a$ acts on $\mathbf{E}$ as $A\mathbf{E} = \mathbf{G}^{ee}_{\sigma_b}(\sigma_a \mathbf{E})$. The solution of Eq. (107) can be written in the following form

$$\mathbf{E} = (\mathbf{1} - A)^{-1}\mathbf{E}_b, \tag{109}$$

where $\mathbf{1}$ is a unit operator. If operator A is contracting (which means that $\|A\| < 1$) then $(\mathbf{1} - A)^{-1}$ can be represented as

$$(\mathbf{1} - A)^{-1} = \mathbf{1} + A + A^2 + \cdots, \tag{110}$$

and thus the solution of Eq. (109) reads as the following Neumann series

$$\mathbf{E} = (\mathbf{1} - A)^{-1}\mathbf{E}_b = \mathbf{E}_b + A\mathbf{E}_b + A(A(\mathbf{E}_b)) + \cdots. \tag{111}$$

Generally, contracting properties of operator $\mathbf{G}^{ee}_{\sigma_b} \circ \sigma_a$ are not known to us.

In Pankratov et al. (1995) it is shown that the energy inequality for Maxwell's equations can be expressed as

$$\|\mathbf{K}_{\sigma_b}\| \leq 1, \tag{112}$$

or in alternative form

$$\|\mathbf{K}_{\sigma_b}(\chi)\| \leq \|\chi\|, \quad \text{for any vector field } \chi, \tag{113}$$

where

$$\mathbf{K}_{\sigma_b} = \mathbf{1} + 2\sqrt{\gamma_b} \circ \mathbf{G}^{ee}_{\sigma_b} \circ \sqrt{\gamma_b}. \tag{114}$$

Here $\gamma_b = \operatorname{Re}\sigma_b$ is the real part of the reference conductivity. Energy inequality (113) is sharp in the following sense: it turns into equality for a reference model with real-valued conductivity ($\operatorname{Im}\sigma_b = 0$). In the presence of the imaginary part of conductivity, some part of the energy can radiate into the space, which makes (113) an inequality.

Using inequality (112) it is possible to obtain a new scattering equation with contracting operator. Let us first rewrite scattering equation (107) in the form

$$x = Ax + b, \tag{115}$$

where $x = \mathbf{E}$ and $b = \mathbf{E}_b$. Let us renormalize it as

$$x = P\chi + Q \tag{116}$$

and

$$x + \lambda x = Ax + \lambda x + b, \tag{117}$$

with some unknown multipliers $P(\mathbf{r}), Q(\mathbf{r}), \lambda(\mathbf{r})$. Such renormalization modifies scattering equation (115) to a new scattering equation

$$\chi = B\chi + \beta \tag{118}$$

with new linear operator

$$B = \frac{1}{P}\frac{1}{1+\lambda}(\lambda \mathbf{1} + A) \circ P \tag{119}$$

and new source term

$$\beta = \frac{1}{P}\frac{1}{1+\lambda}(AQ - Q + b). \tag{120}$$

Let us then require that new scattering operator B be expressed as a composition of $\mathbf{K}_{\sigma_b}$ with some multiplication operator R

$$B = \mathbf{K}_{\sigma_b} \circ R, \tag{121}$$

and hence we obtain

$$P(\mathbf{r}) = \frac{2g\sqrt{\gamma_b}}{\sigma + \sigma_b^*}, \quad Q(\mathbf{r}) = g\frac{\mathbf{j}^{\text{imp}}}{\sigma + \sigma_b^*}, \quad \lambda(\mathbf{r}) = \frac{\sigma - \sigma_b}{2\gamma_b}, \quad R(\mathbf{r}) = \frac{\sigma - \sigma_b}{\sigma + \sigma_b^*}. \tag{122}$$

Here $g \neq 0$ is an undefined constant, thus we can assign

$$g = 1, \tag{123}$$

as it is done in all works on CIE.

Deducing that $\frac{1}{P}\frac{1}{1+\lambda} = \sqrt{\gamma_b}$ and substituting expressions (122)–(123) into Eqs. (116) and (120), we get the expressions for $\mathbf{E}$ and β as follows

$$\mathbf{E} = \frac{2\sqrt{\gamma_b}\chi + \mathbf{j}^{\text{imp}}}{\sigma + \sigma_b^*}, \tag{124}$$

$$\beta = \sqrt{\gamma_b}\left(\mathbf{E}_b - \frac{\mathbf{j}^{\text{imp}}}{\sigma + \sigma_b^*} + \mathbf{G}_{\sigma_b}^{ee}\left(\frac{\sigma - \sigma_b}{\sigma + \sigma_b^*}\mathbf{j}^{\text{imp}}\right)\right). \tag{125}$$

It is proven in Pankratov et al. (1995) that

$$\|R\| \equiv \max|R(\mathbf{r})| = q < 1, \tag{126}$$

for feasible conductivity distributions ($\operatorname{Re}\sigma > 0, \operatorname{Re}\sigma_b > 0$), thus together with Eq. (113) it implies that new scattering operator (121) is contracting

$$\|B\| \leq q < 1, \tag{127}$$

and the Neumann series for Eq. (118)

$$\chi = (\mathbf{1} - B)^{-1}\beta = \beta + B\beta + B^2\beta + \cdots, \tag{128}$$

is always convergent. We call new scattering equation (118) the *contracting integral equation* (CIE). An optimum choice of the reference conductivity, as well as an estimate of condition number of CIE system operator, $\mathbf{1} - B$, is discussed in the next appendix.

Appendix 5: Choice of Optimal Reference Model and Estimate of Condition Number for CIE System Operator

In the previous appendix we considered always contracting series (128) to solve Maxwell's equations (103). Now we are interested in a choice of "optimum" reference conductivity, i.e. the conductivity that delivers the fastest convergence of the series (128). Let us imagine that we want to obtain the solution with desired accuracy ε, i.e.

$$\varepsilon = \frac{\|\chi - \chi_n\|}{\|\chi\|}, \tag{129}$$

where

$$\chi_n = \beta + B\beta + \cdots + B^{n-1}\beta. \tag{130}$$

Using the relation

$$\|\chi - \chi_n\| = \|B^n\chi\| \sim q^n\|\chi\|, \tag{131}$$

we obtain that the number of iterations is governed by the following approximate equality

$$n \sim \ln\varepsilon / \ln q. \tag{132}$$

From this relation and Eq. (126) it follows that the minimum number of iterations is achieved for the reference conductivity $\sigma_b(z) = \sigma_{\text{opt}}(z)$ such that

$$\max_{\sigma\in M(z)} |R(\sigma, \sigma_{\text{opt}}(z))| = \min_{\sigma_b\in\Pi} \max_{\sigma\in M(z)} |R(\sigma, \sigma_b)|, \quad \text{for each } z, \tag{133}$$

where

$$R(\sigma, \sigma_b) := \frac{\sigma - \sigma_b}{\sigma + \sigma_b^*}. \tag{134}$$

Here Π stands for the right half plane of complex variable σ

$$\Pi = \{\sigma \in \mathbb{C} \,|\, \mathrm{Re}\,\sigma > 0\}, \tag{135}$$

and $M(z) \subset \Pi$ is the range of values of $\sigma(x, y, z)$ at horizontal plane $z = Const$

$$M(z) = \{\sigma(x, y, z) \in \mathbb{C} \,|\, x, y \in \mathbb{R}\}. \tag{136}$$

Under these assumptions, it turns out that

$$q_{\text{opt}} = \max_z \min_{\sigma_b\in\Pi} \max_{\sigma\in M(z)} |R(\sigma, \sigma_b)|. \tag{137}$$

Now we notice that

$$|R(\sigma, \sigma_b)| = \left|\frac{\sigma - \sigma_b}{\sigma + \sigma_b^*}\right| = \tanh\frac{1}{2}\mathcal{S}(\sigma, \sigma_b), \tag{138}$$

where

$$\mathcal{S}(\sigma, \sigma_b) = 2\,\mathrm{artanh}\left|\frac{\sigma - \sigma_b}{\sigma + \sigma_b^*}\right| = 2\,\mathrm{arsinh}\frac{|\sigma - \sigma_b|}{2\sqrt{\gamma}\sqrt{\gamma_b}}, \tag{139}$$

is the Lobachevsky–Bolyai (Prasolov 2004) distance in the right half plane Π. Using Lobachevsky–Bolyai geometry formalism, with some efforts, it can be shown that

$$q_{\text{opt}}(z) = \tanh\left(\frac{1}{4}\operatorname{diam}\mathcal{O}(M(z))\right), \tag{140}$$

where $\mathcal{O}(M(z))$ is the minimum-size Lobachevsky–Bolyai circle that contains the set $M(z)$. Moreover it can be shown that

$$\operatorname{diam}(M) \le \operatorname{diam}(\mathcal{O}(M)) \le 2\,\operatorname{arsinh}\left(\frac{\sinh\left(\frac{1}{2}\operatorname{diam}(M)\right)}{\sqrt{3}/2}\right). \tag{141}$$

Here $\operatorname{diam}(M)$ is the diameter of set M in the Lobachevsky–Bolyai half plane Π

$$\operatorname{diam} M = \max_{\sigma,\sigma_b\in M} \mathcal{S}(\sigma,\sigma_b). \tag{142}$$

In addition it can be proven that $2\,\operatorname{arsinh}\left(\frac{\sinh x}{\sqrt{3}/2}\right)$ is monotonically increasing and convex upward (for $x > 0$) function, from which the following inequality is valid

$$2\,\operatorname{arsinh}\left(\frac{\sinh\left(\frac{1}{2}\operatorname{diam}(M)\right)}{\sqrt{3}/2}\right) \le \min\left(\frac{\operatorname{diam}(M)}{\sqrt{3}/2},\ \operatorname{diam}(M) - 2\ln\frac{\sqrt{3}}{2}\right). \tag{143}$$

From Eqs. (140) to (143) it follows that the optimum number of iterations for the desired accuracy ε is specified by the following inequality

$$E \le n_{\text{opt}} \le 2E/\sqrt{3}, \quad E = \max_z \tilde{E}(z)\cdot\ln\frac{1}{\varepsilon}, \tag{144}$$

where

$$\tilde{E}(z) \sim \frac{1}{2}\sqrt{\left(\sqrt{\frac{\gamma_1}{\gamma_2}} - \sqrt{\frac{\gamma_2}{\gamma_1}}\right)^2 + \frac{(\xi_1-\xi_2)^2}{\gamma_1\gamma_2}} = \frac{|\sigma_1-\sigma_2|}{2\sqrt{\gamma_1}\sqrt{\gamma_2}} = \sinh\frac{1}{2}\mathcal{S}(\sigma_1,\sigma_2). \tag{145}$$

Here conductivity pair $\sigma_1 = \gamma_1 + i\xi_1, \sigma_2 = \gamma_2 + i\xi_2$ from $M(z)$ is the most mutually distant (in terms of Lobachevsky–Bolyai geometry) pair, i.e. a pair that delivers a maximum to expression in (145).

Using the developed formalism, it is also possible to estimate condition number κ of CIE system operator $\mathbf{1} - B$ which obeys the following inequality

$$\kappa \le \frac{4}{\sqrt{3}}\max_z \tilde{E}(z). \tag{146}$$

Let us illustrate an application of formulae (144)–(145) for the Cole-Cole induced polarization (IP) conductivity model

$$\sigma(x,y,z,\omega) = \sigma_\infty(x,y,z)\left(1 - \frac{\eta(x,y,z)}{1+(i\omega\tau(x,y,z))^c}\right), \tag{147}$$

with the following typically adopted parameters

$$0 \le \eta(x,y,z) \le \eta_{\max} \approx 0.5, \qquad c = 1/2. \tag{148}$$

Let us decompose the conductivity into real and imaginary parts as follows

$$\sigma = \gamma + i\xi, \tag{149}$$

and let us vary location (x, y, z) in a thin layer at depth z, thus denoting in this layer

$$\gamma_{\min}(z,\omega) = \min_{x,y} \gamma(x,y,z,\omega), \quad \gamma_{\max}(z,\omega) = \max_{x,y} \gamma(x,y,z,\omega), \tag{150}$$

and in the same manner for the imaginary part, thus getting that for all locations in the layer z, complex-valued conductivity values belong to the following (Euclidean) rectangle

$$\begin{cases} \gamma_{\min}(z,\omega) \le \gamma(x,y,z,\omega) \le \gamma_{\max}(z,\omega), \\ \xi_{\min}(z,\omega) \le \xi(x,y,z,\omega) \le \xi_{\max}(z,\omega). \end{cases} \tag{151}$$

Let us evaluate $\tilde{E}(z)$ for the Euclidean rectangle (151) using formulae (144)–(145). Value of $\tilde{E}(z)$ is then a maximum value of (145) for any pair of complex numbers $\gamma_1 + i\xi_1, \gamma_2 + i\xi_2$ that belong to Euclidean rectangle (151) as follows

$$\tilde{E}(z) = \max\bigl(\tilde{E}_1(z), \tilde{E}_2(z)\bigr), \tag{152}$$

where

$$\tilde{E}_1(z) \sim \frac{\xi_{\max} - \xi_{\min}}{\gamma_{\min}}, \qquad \tilde{E}_2(z) \sim \frac{1}{2}\sqrt{\frac{\gamma_{\max}}{\gamma_{\min}} + \frac{(\xi_{\max} - \xi_{\min})^2}{\gamma_{\max}\gamma_{\min}}}. \tag{153}$$

Next, taking into account Cole–Cole Eqs. (147)–(148) we get $|\xi_{\max}| \sim 0.1|\gamma_{\max}|$ and thus

$$\tilde{E}(z) \sim \frac{1}{2}\max\left(0.1\,\frac{\gamma_{\max}}{\gamma_{\min}}, \sqrt{1.01}\sqrt{\frac{\gamma_{\max}}{\gamma_{\min}}}\right). \tag{154}$$

We see that for a high-contrasting inductive polarized z-layer the value of $\tilde{E}(z)$ grows as the contrast of the real part of conductivity, $K := \frac{\gamma_{\max}}{\gamma_{\min}}$. The other consequence is that the IP contrast in Cole–Cole model plays significant role in finding the optimum model, if the contrast of the real part of conductivity is greater than 100. The latter value is a threshold: for $K \ge 100$ we get $\tilde{E}(z) \sim \frac{1}{20}K$, whereas for $K \le 100$ we have $\tilde{E}(z) \sim \frac{1}{2}\sqrt{K}$.

Final remark of this section is that if conductivity is a real-valued function (which is the most common case in EM studies), then for $\tilde{E}(z)$ the following equality holds

$$\tilde{E}(z) = \frac{1}{2}\sqrt{\frac{\gamma_{\max}(z)}{\gamma_{\min}(z)}}, \tag{155}$$

where

$$\gamma_{\min}(z) = \min_{x,y} \gamma(x,y,z), \quad \gamma_{\max}(z) = \max_{x,y} \gamma(x,y,z). \tag{156}$$

Equation (155) means that the number of iterations and condition number for the optimal model are proportional to the square root of the maximum *lateral* contrast in the model. As for the optimum conductivity, $\sigma_{\text{opt}}(z)$, it is equal in this case to the conductivity of the host section outside the depths occupied by the inhomogeneities, but at depths with laterally inhomogeneous distribution of conductivity it has the form

$$\sigma_{\text{opt}}(z) \equiv \gamma_{\text{opt}}(z) = \sqrt{\gamma_{\min}(z)\gamma_{\max}(z)}. \tag{157}$$

References

Abubakar A, Habashy T (2013) Three-dimensional visco-acoustic modeling using a renormalized integral equation iterative solver. J Comput Phys 249:1–12

Avdeev D, Knizhnik S (2009) 3D integral equation modeling with a linear dependence on dimensions. Geophysics 74:89–94

Avdeev D, Kuvshinov A, Epova K (2002) Three-dimensional modelling of electromagnetic modelling of electromagnetic logs from inclined-horizontal wells. Phys Solid Earth 38:975–980

Avdeev D, Kuvshinov A, Pankratov O, Newman G (1997) High-performance three-dimensional electromagnetic modeling using modified Neumann series. Wide-band numerical solution and examples. J Geomagn Geoelectr 49:1519–1539

Avdeev D, Kuvshinov A, Pankratov O, Newman G (2000) 3-D EM modelling using fast integral equation approach with Krylov subspaces accelerator. In: 2nd EAGE conference and technical exhibition, vol 2, Scotland, Scotland

Avdeev D, Kuvshinov A, Pankratov O, Newman G (2002) Three-dimensional induction logging problems, part I: an integral equation solution and model comparisons. Geophysics 67(2):413–426

Bayes T (1763) An essay towards solving a problem in the doctrine of chances. Philos Trans R Soc Lond 53:370–418

Beck R, Hiptmair R, Hoppe RH, Wohlmuth B (2000) Residual based a posteriori error estimators for eddy current computation. Math Model Numer Anal 34:159–182

Bodin T, Sambridge M, Rawlinson N, Arroucau P (2012) Transdimensional tomography with unknown data noise. Geophys J Int 189:1536–1556

Börner RU (2010) Numerical modelling in geo-electromagnetics: advances and challenges. Surv Geophys 31:225–245

Brown V, Hoversten M, Key K, Chen J (2012) Resolution of reservoir scale electrical anisotropy from marine CSEM data. Water Resour Res 77(2):E147–E158

Bürg M (2000) A residual-based a posteriori error estimator for the hp-finite element method for Maxwells equations. Appl Numer Math 62:922–940

Bürg M (2013) Convergence of an automatic hp-adaptive finite element strategy for Maxwells equations. Appl Numer Math 72:188–206

Chen J, Hoversten GM, Key K, Nordquist G, Cumming W (2011) Stochastic inversion of magnetotelluric data using a sharp boundary parameterization and application to a geothermal site. Geophysics 77(4):E265–E279

Chen J, Hoversten GM, Vasco D, Rubin Y, Hou Z (2007) A Bayesian model for gas saturation estimation using marine seismic AVA and CSEM data. Geophysics 72(2):WA85–WA95

Chew W, Jin J, Lu C, Michielssen E, Song J (2014) Fast solution methods in electromagnetics. IEEE Trans Antenna Propag 45:533–543

Christensen N (1990) Optimized fast Hankel transform filters. Geophys Prospect 38:545–568

Epanomeritakis I, Akcelik V, Ghattas O, Bielak J (2008) A Newton-CG method for large-scale three-dimensional elastic full waveform seismic inversion. Inverse Probl 24:1–26

Ernst OG, Gander MJ (2011) Why it is difficult to solve Helmholtz problems with classical iterative methods. Numer Anal Multiscale Probl 83:325–361

Fainberg E, Zinger B (1980) Electromagnetic induction in a nonuniform spherical model of the Earth. Ann Geophys 36:127–134

Farquharson CG, Miensopust MP (2011) Three-dimensional finite-element modelling of magnetotelluric data with a divergence correction. J Appl Geophys 75:699–710

Fichtner A, Trampert J (2011) Hessian kernels of seismic data functionals based upon adjoint techniques. Geophys J Int 185:775–798

Frayss V, Giraud L, Gratton S, Langou J (2003) A set of GMRES routines for real and complex arithmetics on high performance computers. CERFACS technical report TR/PA/03/3

Geraskin A, Kruglyakov M, Kuvshinov A (2015) Novel robust and scalable 3-D forward solver based on contracting integral equation method and modern programming technologies. Comput Geosci (submitted)

Gilks WR, Richardson S, Spiegelhalter D (eds) (1996) Markov chain Monte Carlo in practice. Chapman and Hall, London, pp 1–485

Grandis H, Menvielle M, Roussignol M (1999) Bayesian inversion with Markov chains-I. The magnetotelluric one-dimensional case. GJI 138:757–768

Grandis H, Sumintaredja P, Irawan D (2012) A template for 1-D inversion of geo-electromagnetic data using MCMC method. EMSEV 2012, Gotemba Kogen Resort, Gotemba, Japan, October 14, 2012, pp 1–4

Grayver A, Burg M (2014) Robust and scalable 3-D geo-electromagnetic modelling approach using the finite element method. Geophys J Int. doi:10.1093/gji/ggu119
Grayver A, Kolev T (2015) Large-scale 3D geo-electromagnetic modeling using parallel adaptive high-order finite element method. Geophysics 80(6):277–291
Greenbaum A (1997) Iterative methods for solving linear systems. SIAM, Philadelphia, USA
Haber E, Ascher UM (2001) Fast finite volume simulation of 3D electromagnetic problems with highly discontinuous coefficients. SIAM J Sci Comput 22(6):1943–1961
Hastings WK (1970) Monte carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109
Hiptmair R (2002) Finite elements in computational electromagnetism. Acta Numer 11:237–339
Hohmann G (1975) Three-dimensional induced polarization and electromagnetic modeling. J Geophys 40:309–324
Hursan G, Zhdanov M (2002) Contraction integral equation method in three-dimensional electromagnetic modeling. Radio Sci 37:2001R. doi:10.1029/S002513
Kaipio J, Somersalo E (2005) Statistical and computational inverse problems. Springer, New York
Kamm J, Pedersen L (2014) Inversion of airborne tensor VLF data using integral equations. Geophys J Int. doi:10.1093/gji/ggu161
Kelbert A, Kuvshinov A, Velimsky J, Koyama T, Ribaudo J, Sun J, Martinec Z, Weiss C (2014) Global 3-D electromagnetic forward modelling: a benchmark study. Geophys J Int 197:785–814
Kirk BS, Peterson JW, Stogner RH, Carey GF (2006) libmesh: a C++ library for parallel adaptive mesh refinement/coarsening simulations. Eng Comput 22:237–254
Kolmogorov AN (1956) Foundations of the theory of probability, 2nd edn. Chelsea, New York
Koyama T, Shimizu H, Utada H, Ichiki M, Ohtani E, Hae R (2006) Water content in the mantle transition zone beneath the North Pacific derived from the electrical conductivity anomaly. AGU Geophys Monogr Ser 168:171–179
Koyama T, Utada H, Avdeev D (2008) Fast and memory-saved 3-D forward modeling code for MT by using integral equation method. In: Abstract book. 19th workshop on electromagnetic induction in the Earth, China
Kuvshinov A (2008) 3-D global induction in the oceans and solid Earth: recent progress in modeling magnetic and electric fields from sources of magnetospheric, ionospheric and oceanic origin. Surv Geophys. doi:10.1007/s10712-008-9045-z
Kuvshinov AV, Utada H, Avdeev D, Koyama T (2005) 3-D modelling and analysis of Dst C-responses in the North Pacific Ocean region, revisited. Geophys J Int 160:505–526
Logg A, Wells GN (2010) Dolfin: automated finite element computing. ACM Trans Math Softw 37:20:1–20:28
Lu CC, Chew W (1994) A multilevel algorithm for solving boundary-value scattering. Microwave Opt Technol Lett 7:466–470
Mackie R, Smith J, Madden T (1994) 3-Dimensional electromagnetic modeling using finite-difference equation—the magnetotelluric example. Radio Sci 29(4):923–935
Martin J, Wilcox LC, Burstedde C, Ghattas O (2012) A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. SIAM J Sci Comput 34(3):A1460–A1487
Metivier L, Brossier R, Virieux J, Operto S (2013) Full waveform inversion and the truncated Newton method. SIAM J Sci Comput 35:401–437
Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21:1087–1092
Michielssen E, Boag A (1996) A multilevel matrix decomposition algorithm for analyzing scattering from large structures. IEEE Trans Antennas Propag 44:1086–1093
Millard X, Liu QH (2003) A fast volume integral equation solver for electromagnetic scattering from large inhomogeneous objects in planarly layered media. IEEE Trans Antennas Propag 51:2393–2401
Minsley BJ (2011) A trans-dimensional Bayesian Markov chain Monte Carlo algorithm for model assessment using frequency-domain electromagnetic data. GJI 187:252–272
Monk P (2003) Finite element methods for Maxwells equations. Oxford University Press, Oxford
Mulder W (2006) A multigrid solver for 3D electromagnetic diffusion. Geophys Prospect 54:633–649
Newman G, Alumbaugh D (2002) Three-dimensional induction logging problems, part 2: a finite-difference solution. Geophysics 61:484–491
Newman GA, Hoversten GM (2000) Solution strategies for two- and three-dimensional electromagnetic inverse problems. Inverse Probl 16:1357–1375
Nie X, Li LW, Yuan N, Yeo TS (2013) A fast integral equation solver for 3D induction well logging in formations with large conductivity contrasts. J Comput Phys 61:645–657

Pankratov O, Avdeev D, Kuvshinov A (1995) Electromagnetic field scattering in a homogeneous Earth: a solution to the forward problem. Phys Solid Earth 31:201–209
Pankratov O, Kuvshinov A (2010) General formalism for the efficient calculation of derivatives of EM frequency domain responses and derivatives of the misfit. Geophys J Int 181:229–249
Pankratov O, Kuvshinov A (2015) General formalism for the efficient calculation of the Hessian matrix of EM data misfit and Hessian-vector products based upon adjoint sources approach. Geophys J Int 200:1449–1465
Pankratov O, Kuvshinov A, Avdeev D (1997) High-performance three-dimensional electromagnetic modeling using modified Neumann series. Anisotropic case. J Geomag Geoelectr 49:1541–1547
Phillips JR, White J (1996) A precorrected-FFT method for electrostatic analysis of complicated 3-D structures. IEEE Trans Comput Aided Des Integr Circuits Syst 16:1059–1071
Prasolov VV (2004) Lobachevsky geometry. MCCME Publishing House, Moscow **(in Russian)**
Puzyrev V, Koldan J, de la Puente J, Houzeaux G, Vazquez M, Cela JM (2013) Efficient pre-conditioned iterative solution strategies for the electromagnetic diffusion in the earth: finite-element frequency-domain approach. Geophys J Int 193:678–693
Raiche A (1974) An integral equation approach to three-dimensional modelling. Geophys J R Astr Soc 36:363–376
Ren Z, Kalscheuer T, Greenhalgh S, Maurer H (2013) A goal-oriented adaptive finite-element approach for plane wave 3-D electromagnetic modelling. Geophys J Int 194:700–718
Rius JM, Parron J, Heldring A, Tamayo J, Ubeda E (2008) Fast iterative solution of integral equations with method of moments and matrix decomposition algorithm - singular value decomposition. IEEE Trans Anteenas Propag 56:2314–2324
Rokhlin V (1990) Rapid solution of integral equations of scattering theory in two dimensions. J. Comp. Phys. 36:414–439
Rosas-Carbajal M, Linde N, Kalscheuer T, Vrugt JA (2013) Two-dimensional probabilistic inversion of plane-wave electromagnetic data: methodology, model constraints and joint inversion with electrical resistivity data. GJI 196:1–17
Santosa F, Symes WW (1988) Computation of the Hessian for least-squares solutions of inverse problems of reflection seismology. Inverse Probl 4:211–213
Schwarzbach C, Börner RU, Spitzer K (2011) Three-dimensional adaptive higher order finite element simulation for geo-electromagnetics—a marine CSEM example. Geophys J Int 187:63–74
Singer B (1995) Method for solution of Maxwell's equations in non-uniform media. Geophys J Int 120:590–598
Singer B (2008) Electromagnetic integral equation approach based on contraction operator and solution optimization in Krylov subspace. Geophys J Int 175:857–884
Singer B, Fainberg E (1995) Generalization of the iterative dissipative method for modeling electromagnetic fields in nonuniform media with displacement currents. J Appl Geophys 34:41–46
Singer B, Fainberg E (1997) Fast and stable method for 3D modeling of electromagnetic field. Explor Geophys 34:130–135
Smith TJ (1996) Conservative modeling of 3-D electromagnetic fields, part II: biconjugate gradient solution as an accelerator. Geophysics 61:1319–1324
Sun J, Egbert G (2012) A thin-sheet model for global electromagnetic induction. Geophys J Int 189:343–356
Sun J, Kuvshinov A (2015) Accelerating EM integral equation forward solver for global geomagnetic induction using SVD based matrix compression method. Geophys J Int 1200:1003–1009
Tarantola A (2005) Inverse problem theory and methods for model parameter estimation. Society for Industrial and Applied Mathematics, Philadelphia
Um ES, Commer M, Newman GA (2013) Efficient pre-conditioned iterative solution strategies for the electromagnetic diffusion in the earth: finite-element frequency-domain approach. Geophys J Int 193:1460–1473
Weidelt P (1975) Electromagnetic induction in three-dimensional structures. J Geophys 41:85–109