Pure Appl. Geophys. 177 (2020), 1029–1041

https://doi.org/10.1007/s00024-019-02334-w

Pure and Applied Geophysics

Initializing Cross-Gradients Joint Inversion of Gravity and Magnetic Data with a Bayesian Surrogate Gravity Model

Emilia Fregoso,[1] Abel Palafox,[1] and Miguel Angel Moreles[1,2]

Abstract—Cross-gradients joint inversion of gravity and magnetic data is the focus of this work. Cross-gradients are introduced as a constraint in the minimization of a least square functional including the misfits of the available data. We propose to initialize the cross gradients iterations with a surrogate density model. The latter is constructed by means of Bayesian estimation in a low dimensional parameter space. To sample from the posterior, an affine invariant MCMC is also introduced. The proposed methodology is successfully tested on synthetic models consisting of isolated sources.

Key words: Join inversion, cross gradients, Bayesian estimation, surrogate model, gravity data, magnetic data.

1. Introduction

In this work we are concerned with joint inversion of gravity and magnetic data. Joint inversion is a strategy that combines information of different data sets to reduce the non uniqueness inherent in individual inverse problems.

In geophysics, there exist two preferred strategies to develop joint inversion. The first rests on proposing direct relationships between geophysical parameters for a specific site, when such dependency exists. Case studies that involve direct relationships between density and seismic velocity are described in Nielsen and Jacobsen (2000) and Moorkamp et al. (2013), or between conductivity and seismic velocity (Moorkamp et al. 2011).

The second strategy for joint inversion is more general and applies to any type of geophysical data. The basic idea is that physical properties tend to change at the same location and focuses on the search for structural similarities (Haber and Oldenburg 1997; Gallardo and Meju 2004). In the latter, the cross-gradient methodology is introduced. The success of this methodology in joint inversion is highly reported in the literature. A recent work is that of Joulidehsar et al. (2018). Therein a least-squares QR method is developed to speed up the calculation of cross-gradients joint inversion.

In practice, cross-gradients are introduced as a constraint in the minimization of a least square functional including the misfits of the available data. The minimization of the functional is by means of an iterative method. Hence the role of an initial guess is of great relevance.

The limited depth resolution of gravity and magnetic nonparametric underdetermined problems is a major issue that has been discussed by several authors (e.g., Fedi et al. 2005; Pilkington 2012; Paoletti et al. 2014).

Some works on depth resolution are Li and Oldenburg (1996, 1998). Therein, spatially dependent weighting functions are introduced and corresponding penalty coefficients, to ponder the importance of terms in the objective function. In particular, a depth weighting function whose parameters are adjusted for a good match between the function and the decay of the gravity kernel for a given mesh.

An alternative to inverse problems that has received great attention recently is the bayesian approach. An engaging description is presented in Stuart (2010). In this framework all variables are random. A probability density function, the prior, is

[1] Mathematics Department, Universidad de Guadalajara, CUCEI, Guadalajara, Jal, Mexico. E-mail: emilia.fbecerra@academicos.udg.mx; abel.palafox@academicos.udg.mx; moreles@cimat.mx

[2] *Present Address*: CIMAT, Jalisco S/N, Valenciana, Guanajuato, GTO, Mexico.

proposed for the parameters to be estimated. Then a noise model is introduced to determine the likelihood. The solutions of the inverse problem, that most likely explain the data, are characterized by the conditional probability density function, *the posterior*, obtained from Bayes' formula. Consequently, the Bayesian formulation of inverse problems requires numerical procedures for sampling from the posterior. The latter is carried out by a Markov Chain Monte Carlo (MCMC) method.

It is known that MCMC methods are computationally demanding, if the parameter space is high dimensional. Obtaining a sample from the posterior distribution by a MCMC method, makes the bayesian approach impractical.

Our contribution aims to alleviate this difficulty, as well as the need of an initial guess and the lack of depth resolution in the cross-gradient methodology when applied to gravity and magnetic data. Consequently, we propose to initialize the cross-gradients iterations with a surrogate density model. This is constructed by means of Bayesian estimation in a low dimensional parameter space. Also, an affine invariant MCMC is introduced to sample from the corresponding posterior distribution. The proposed methodology is successfully tested on synthetic models consisting of isolated sources. The models are the benchmark problems in Li and Oldenburg (1998). Namely, cubes buried and different depths, and a dipping dike buried at 50 m depth. The studied sources are embedded in a homogeneous half-space with null density and magnetization.

The geometry of the surrogate model is secondary to its description by a few parameters. In this work we choose a rectangular prism as surrogate model. It is expected that if the unknown model is also a prism, the recovery should be almost perfect. This is illustrated with two cubic models from Li and Oldenburg (1998).

We stress that our proposal is data driven. An extension of the surrogate model by introducing more physically motivated constraints, such as the penalty terms of Li and Oldenburg (1996, 1998), is to be developed elsewhere.

Monte Carlo inversion techniques on Geophysical inverse problems are not new, an earlier work is for instance Aarts and Korst (1988). Later, numerous techniques have been developed and applied by Earth scientists, see Sambridge and Mosegaard (2002) for a somewhat complete summary.

2. Problem Formulation and Methods

2.1. Structural Joint Inversion Coupled with Bayesian Estimation

Let us assume, as customary, that the subsurface model is an aggregate of rectangular prisms with constant geophysical properties built in a vector $\mathbf{m}$. This model leads to an anomaly observed at the surface and collected in the data vector $\mathbf{d}$. A forward map $\mathbf{G}$ from models to anomalies is constructed corresponding to the underlying geophysical property.

Consequently, the classical inverse problem is: given $\mathbf{d}$, determine $\mathbf{m}$ such that $\mathbf{G}(\mathbf{m}) = \mathbf{d}$. Or $\mathbf{Gm} = \mathbf{d}$ if $\mathbf{G}$ is linear.

It is common to have different sets of data at the same site. Here we assume that two sets of data are collected at the surface, $\mathbf{d}^{(1)}$ and $\mathbf{d}^{(2)}$, the gravity and magnetic anomalies.

We may pose the separated inversion problems: given $\mathbf{d}^{(i)}$, determine $\mathbf{m}^{(i)}$ such that $\mathbf{G}^{(i)}\mathbf{m}^{(i)} = \mathbf{d}^{(i)}$, $i = 1, 2$.

Separated inversion is used here only for comparison. Instead, the information of the different data sets is combined to reduce ill-posedness in a joint inversion framework. The classical strategy is to pose a smoothed and regularized least square functional, namely

$$\begin{aligned}\Phi(\mathbf{m}^{(1)}, \mathbf{m}^{(2)}) = &\left\| \begin{matrix} \mathbf{d}^{(1)} - \mathbf{G}^{(1)}\mathbf{m}^{(1)} \\ \\ \mathbf{d}^{(2)} - \mathbf{G}^{(2)}\mathbf{m}^{(2)} \end{matrix} \right\|^2_{\mathbf{C}_d^{-1}} \\ &+ \left\| \begin{matrix} \mathbf{m}^{(1)} - \mathbf{m}_R^{(1)} \\ \\ \mathbf{m}^{(2)} - \mathbf{m}_R^{(2)} \end{matrix} \right\|^2_{\mathbf{C}_R^{-1}} \\ &+ \left\| \begin{matrix} \alpha_1 \mathbf{L}\mathbf{m}^{(1)} \\ \\ \alpha_2 \mathbf{L}\mathbf{m}^{(2)} \end{matrix} \right\|^2 .\end{aligned} \tag{1}$$

For each geophysical data set (which are assumed to be uncorrelated), $\mathbf{C}_d$ and $\mathbf{C}_R$ correspond to block diagonal covariance matrices. Namely,

$$\mathbf{C}_d = \begin{bmatrix} \mathbf{C}_d^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_d^{(2)} \end{bmatrix}, \qquad \mathbf{C}_R = \begin{bmatrix} \mathbf{C}_R^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_R^{(2)} \end{bmatrix}.$$

The regularization terms involve a Laplacian-like smoothness matrix, $\mathbf{L}$, and regularization parameters α_1, α_2 for gravity and magnetics respectively. The reference models are $\mathbf{m}_R^{(1)}$ and $\mathbf{m}_R^{(2)}$.

The cross-gradient joint inversion methodology, Gallardo and Meju (2004), is based in the idea that changes of the physical properties $\mathbf{m}^{(1)}$, $\mathbf{m}^{(2)}$ are parallel. Consequently, the minimization of the functional (1) is constrained to

$$\mathbf{t}\left(\mathbf{m}^{(1)}, \mathbf{m}^{(2)}\right) \equiv \nabla \mathbf{m}^{(1)}(x, y, z) \times \nabla \mathbf{m}^{(2)}(x, y, z) = \mathbf{0}. \tag{2}$$

The minimization problem is solved by means of an iterative method, hence initial and reference models are required.

Our proposal is to construct a reference model using bayesian estimation for one of the given data sets. In the case of gravity and magnetic data, the former is chosen for bayesian estimation. It will become apparent that to estimate density on each prism is straightforward in this framework.

For later reference, we call the latter the *constrained (bayesian) joint inversion*, and the former simply the *unconstrained joint inversion*.

2.2. Bayesian Estimation of a Gravity Initial Model

Let us carry out the discussion in the context of the inverse gravity problem. In Bayesian estimation all the interesting quantities are random variables. Thus we assume the model

$$\mathbf{d} = \mathcal{G}(\mathbf{m}) + \eta.$$

Here $\mathbf{d}$ is the data, $\mathbf{m}$ is parameter to estimate, η is the random noise. In the bayesian jargon, the forward map $\mathbf{G}$ is called the observation operator, which in this section is denoted by $\mathcal{G}$.

We are interested in estimating $\mathbf{m}$, thus identifying the density and shape of source body that explains the gravity dataset.

This is the probability of $\mathbf{m}$ given $\mathbf{d}$. Hence, from Bayes formula we have

$$\pi(\mathbf{m}|\mathbf{d}) = \frac{\pi(\mathbf{d}|\mathbf{m})\pi_0(\mathbf{m})}{\pi(\mathbf{d})},$$

where $\pi(\mathbf{m}|\mathbf{d})$ is the posterior density of the parameters given the data, $\pi(\mathbf{d}|\mathbf{m})$ is the likelihood given by the noise model, $\pi_0(\mathbf{m})$ is a prior distribution, and $\pi(\mathbf{d})$ is a normalizing constant.

An important feature of bayesian estimation is to propose a prior probability density function, π_0, for the parameter $\mathbf{m}$. This prior encompasses all we know about $\mathbf{m}$ regardless or independent of the data.

In our case, we are assuming no knowledge about the horizontal position (x and y coordinates) of the source body. This means a uniform probability distribution on the horizontal components.

In respect to depth, we assume a prior distribution for the z component of the top of the source body, as the following triangular distribution

$$\pi_0(z) = \begin{cases} 2\frac{z-a}{(b-a)(c-a)} & z < c, \\ \frac{2}{(b-a)} & z = c, \\ 2\frac{b-z}{(b-a)(b-c)} & c < z < b, \\ 0 & \text{other case}, \end{cases} \tag{3}$$

where a, b are the lower and upper bounds of the z component, and c a fraction of $(b-a)$.

In our experiments, we consider $a = 0$, corresponding to zero height, $b = 500$ the largest depth and $c = 0.95(b-a)$ that places the guess value at around 95% of the depth range. The triangular distribution is shown in Fig. 1.

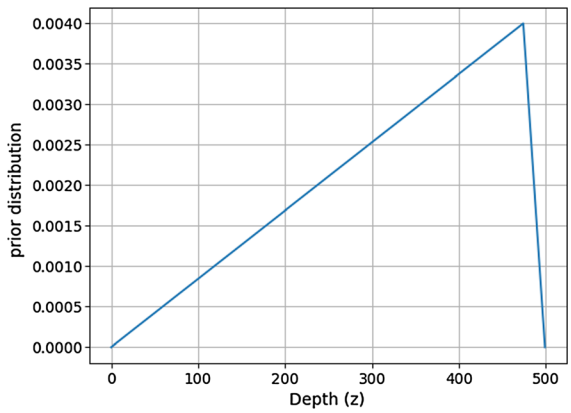

Figure 1
Prior model for source depth. The source object assumed as a prior to be located around 95% of the depth range from surface to bottom

In terms of the inversion process, this prior promotes the exploration of deeper sources instead of shallow ones in a non restrictive way. Its influence on the inversion results is not significant, a vague idea of the largest depth will suffice. Similar results are obtained even with a uniform distribution on the depth range.

Relation between data and parameter is established by the observation model and the noise model. In this sense, noise model should consider conditions that affects the data acquisition process, for example, measurement instruments precision, human errors, etc. In our problem, we assume additive noise η with

$$\mathbf{d} = \mathcal{G}(\mathbf{m}) + \eta,$$

and $\eta \sim N(0, \sigma^2)$ with σ known. It follows that

$$\begin{aligned} \eta &= \mathbf{d} - \mathcal{G}(\mathbf{m}), \\ \mathbf{d} - \mathcal{G}(\mathbf{m}) &\sim N(0, \sigma^2), \\ \pi(\mathbf{d}|\mathbf{m}) &\sim N(\mathcal{G}(\mathbf{m}), \sigma^2). \end{aligned}$$

Thus likelihood has the form

$$\pi(\mathbf{d}|\mathbf{m}) \propto e^{-\frac{||\mathbf{d}-\mathcal{G}(\mathbf{m})||^2}{2\sigma^2}},$$

and minus log-likelihood is proportional to the least squares functional

$$-\log \pi(\mathbf{d}|\mathbf{m}) \propto \frac{||\mathbf{d} - \mathcal{G}(\mathbf{m})||^2}{2\sigma^2}.$$

Estimating a high probability model $\mathbf{m}$ that explains the data $\mathbf{d}$, means maximizing the posterior distribution $\pi(\mathbf{m}|\mathbf{d})$. Equivalently, minimizing $-\log$ of posterior distribution.

$$\min_{\mathbf{m}}\{-\log \pi(\mathbf{m}|\mathbf{d})\} = \min_{\mathbf{m}}\{-\log \pi(\mathbf{d}|\mathbf{m}) - \log \pi_0(m)\}.$$

Let us note that the normalizing constant plays no role.

The energy function, used to explore minus logarithm of the posterior distribution is then given by

$$energy(\mathbf{m}) = \frac{||\mathbf{d} - \mathcal{G}(\mathbf{m})||^2}{2\sigma^2} - \log \pi_0(\mathbf{m}).$$

Obtaining a sample of this distribution can be accomplished by means of MCMC methods. The latter, construct a Markov chain whose stationary distribution is the posterior distribution. See Robert and Casella (2013) for further reading. We describe details of our MCMC design below.

2.3. MCMC Design for a Surrogate Gravity Model

It is known that MCMC methods are computationally demanding, since they require multiple forward map evaluations. Assuming a set of $n \times n$ observation points, and $n \times n \times n$ prisms, the evaluation of the forward map becomes a very expensive linear problem of order $O(n^5)$. This is a classical bayesian application in full dimension.

Since each prism becomes a variable (dimension) to be estimated by the MCMC, the inversion procedure would be designed to estimate n^3 variables. In a coarse grid this is of the order $O(1000)$, which is computationally prohibited. In any case the use of high performance computational methods is a must.

In order to reduce the dimension and the source ambiguity of gravity data inversion, we propose a surrogate model for source body estimation, which depends on a few parameters. Our representation consists in a single rectangular prism of constant density denoted below by R.

The prism R can be described using six parameters. Three for corner coordinates, x_R, y_R, z_R, and the three lengths of the sides of the rectangle along each axis, say $l_{x_R}, l_{y_R}, l_{z_R}$. The upper left frontal corner is set as reference. A last parameter is required for the density ρ_R. Consequently, we build a surrogate model determined by only seven parameters.

The next step is to propose a MCMC procedure with the property of affine-invariance; that is, the performance of the algorithm is invariant under affine transformations of the parameter space and independent of the aspect ratio. See Christen and Fox (2010) and Foreman-Mackey et al. (2013).

A simple path to affine invariance kernels is to define the transition kernel in terms of the estimating parameters. Our proposal is based on the following five movements:

– *Translate* The position of the corner of prism R at time t, denoted by $R^{(t)}$, is updated by selecting an axis and moving the entire prism as follows:

1. Randomly choose a coordinate $s^{(t)}$ from $\{x_{R^{(t)}}, y_{R^{(t)}}, z_{R^{(t)}}\}$. $s^{(t)}$ corresponds to the axis along the prism will be moved.
2. Set the scale of the translation as

$$d^{(t)} = \frac{l_{x_{R^{(t)}}} + l_{y_{R^{(t)}}} + l_{z_{R^{(t)}}}}{3}.$$

This way, the magnitude of movement corresponds to size of prism (small movement for small prism and viceversa).

3. Update $s^{(t+1)} = s^{(t)} + u$ with $u \sim U(-2d^{(t)}, 2d^{(t)})$. The selected coordinate changes its position by adding a uniform random variable u whose magnitude may be up to twice the mean length of the source object.

Note that, since point $(x_{R^{(t)}}, y_{R^{(t)}}, z_{R^{(t)}})$ is the reference corner, changing the position of any coordinate will change the position of the entire prism in consequence.

– *Resize* The lengths of the sides of prism $R^{(t)}$ are updated randomly as follows:

1. Randomly choose a side $l_{s^{(t)}}$ from $\{l_{x_{R^{(t)}}}, l_{y_{R^{(t)}}}, l_{z_{R^{(t)}}}\}$.
2. Set the scale of the resize as

$$l_{d^{(t)}} = \frac{l_{x_{R^{(t)}}} + l_{y_{R^{(t)}}} + l_{z_{R^{(t)}}}}{3}.$$

The resize magnitude is chosen in terms of the actual lengths. Then, large sized prism may have large resizing.

3. Update $l_{s^{(t+1)}} = l_u$ with $l_u \sim U(0, 2l_{d^{(t)}})$. The selected length is modified randomly as a uniform random variable. The length should be positive, negative values produce a reflected prism.

– *Move* The three coordinates of the reference corner are moved at the same time along a randomly chosen vector:

1. Randomly choose angles θ, ϕ from $U[0, 2\pi]$.
2. Set the magnitude of movement as

$$d^{(t)} = \frac{l_{x_{R^{(t)}}} + l_{y_{R^{(t)}}} + l_{z_{R^{(t)}}}}{3}.$$

3. Update

$$(x_{R^{(t+1)}}, y_{R^{(t+1)}}, z_{R^{(t+1)}}) = (x_{R^{(t)}}, y_{R^{(t)}}, z_{R^{(t)}}) + v,$$

where $v = d^{(t)} \cdot (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$

– *Shrink-Enlarge* The lengths of the prism are scaled by a randomly chosen factor:

1. Randomly chose the scaling factor k from $U[0.5, 2]$.
2. Update lenghts as

$$(l_{x_{R^{(t+1)}}}, l_{y_{R^{(t+1)}}}, l_{z_{R^{(t+1)}}}) = (kl_{x_{R^{(t)}}}, kl_{y_{R^{(t)}}}, kl_{z_{R^{(t)}}}).$$

Factor in range [0.5, 1] has the effect of shrinking, whereas factor in range [1, 2] produces enlargement of the prism.

– *Density* Proposals for $\rho_{R^{(t+1)}}$ are obtained from simulating directly from:

$$\rho_{R^{(t+1)}} \leftarrow \Gamma(2, 0.75),$$

where Γ refers to Gamma probability distribution. Since in practice density is positive and typical values of density are lower than 4, we establish parameters of shape 2 and scale 0.75.

The transitional kernel K of our MCMC is the mixture of these movements:

$$\begin{aligned} K(R^{(t)}, R^{(t+1)}) = {} & 0.2\, \textit{Translate}\,(R^{(t)}, R^{(t+1)}) \\ & + 0.1\, \textit{Resize}\,(R^{(t)}, R^{(t+1)}) \\ & + 0.2\, \textit{Move}\,(R^{(t)}, R^{(t+1)}) \\ & + 0.3\,(\textit{Shrink-Enlarge})\,(R^{(t)}, R^{(t+1)}) \\ & + 0.2\, \textit{Density}\,(R^{(t)}, R^{(t+1)}). \end{aligned}$$

This means that, on each time step, the transition from prism on time t, $R^{(t)}$, to prism on time $t+1$, $R^{(t+1)}$ is obtained by choosing a movement with probability according to its corresponding weight on the previous equation.

Let us note that the magnitude of Translate, Resize, Move and Shrink-Enlarge movements is obtained in terms of the magnitude of the prism. That is, the transition kernel does not depend on the parameters space but in the parameters themselves. Consequently, the affine invariant property holds as required.

Also, prism proposals R with large size, allow big changes of position and scale, whereas proposals of small size promote to have small changes along the

MCMC. In terms of our problem, affine invariant property means that the performance of the MCMC method remains the same regardless the initial guess or scale of target source.

The initial guess of the MCMC method is selected randomly on each run from:

$$\begin{aligned} x_R, y_R &\sim U(s_H, S_H), \\ z_R &\sim U(s_V, S_V), \\ l_{x_R}, l_{y_R}, l_{z_R} &\sim U(0, L), \\ \rho_R &\sim U(0, 1). \end{aligned}$$

It is expected that the user will provide informed guesses for the end values of each interval.

With all the components in place, a MCMC is run. The first iterations need not be in a region of high probability and are discarded, the so called burnin. The steps that follow are used to sample from the posterior.

Statistical information regarding this distribution may be obtained, such as conditional point estimators. For our purposes, the conditional posterior mean is the point estimator that we compute to yield an initial guess and a reference model.

Hereafter, we refer to the computed posterior mean obtained from MCMC as the *Bayesian estimation*.

2.4. Cross-Gradients Joint Inversion

Let $\mathbf{m}_R^{(1)}$ be the bayesian estimation model for density, obtained by the MCMC procedure, and propose an informed magnetic reference model $\mathbf{m}_R^{(2)}$. These are used for iterative schemes in separate and joint inversions.

For joint inversion we favor the strategy of cross-gradients. The gravity and magnetic anomalies are calculated following the scheme in Fregoso and Gallardo (2009), Fregoso et al. (2015). Gravity and magnetic responses are computed using the equations developed by Bhattacharyya (1966) and Bhaskara-Rao et al. (1990).

The iterative solution is obtained under the framework of the generalized least-squares formulation proposed by Tarantola and Valette (1982) by calculating

$$\hat{\mathbf{m}}_{k+1} = \hat{\mathbf{m}}_k + \mathbf{N}^{-1}\mathbf{n} - \mathbf{N}^{-1}(\mathbf{B}_k)^T\left(\mathbf{B}_k\mathbf{N}^{-1}(\mathbf{B}_k)^T\right)^{-1}(\mathbf{B}_k\mathbf{N}^{-1}\mathbf{n} + \mathbf{g}_m(\hat{\mathbf{m}}_k)) \tag{4}$$

where

$$\mathbf{N}^{-1} = \begin{bmatrix} \mathbf{G}^{(1)T}(\mathbf{C}_d^{(1)})^{-1}\mathbf{G}^{(1)} + \alpha_1^2\mathbf{L}^T\mathbf{L} + (\mathbf{C}_R^{(1)})^{-1} & 0 \\ 0 & \mathbf{G}^{(2)T}(\mathbf{C}_d^{(2)})^{-1}\mathbf{G}^{(2)} + \alpha_2^2\mathbf{L}^T\mathbf{L} + (\mathbf{C}_R^{(2)})^{-1} \end{bmatrix}^{-1} \tag{5}$$

and

$$\mathbf{n} = \begin{bmatrix} \mathbf{G}^{(1)T}(\mathbf{C}_d^{(1)})^{-1}[\mathbf{d}^{(1)} - \mathbf{G}^{(1)}\hat{\mathbf{m}}_k^{(1)}] - \alpha_1^2\mathbf{L}^T\mathbf{L}\hat{\mathbf{m}}_k^{(1)} + (\mathbf{C}_R^{(1)})^{-1}(\hat{\mathbf{m}}_k^{(1)} - \mathbf{m}_R^{(1)}) \\ \mathbf{G}^{(2)T}(\mathbf{C}_d^{(2)})^{-1})[\mathbf{d}^{(2)} - \mathbf{G}^{(2)}\hat{\mathbf{m}}_k^{(2)}] - \alpha_2^2\mathbf{L}^T\mathbf{L}\hat{\mathbf{m}}_k^{(2)} + (\mathbf{C}_R^{(2)})^{-1}(\hat{\mathbf{m}}_k^{(2)} - \mathbf{m}_R^{(2)}) \end{bmatrix}. \tag{6}$$

Here $\hat{\mathbf{m}}_k = [\mathbf{m}_k^{(1)}, \mathbf{m}_k^{(2)}]^T$, are the two sets of model parameters after k iterations. $\mathbf{G}^{(1)}$, $\mathbf{G}^{(2)}$ are the sensitivity matrices for each data type; $\mathbf{B}_k$ is the Jacobian matrix for the cross-gradient function and $\mathbf{g}_m(\hat{\mathbf{m}}_k)$ contains the three components of the cross-gradient vector.

The iteration (4) is run until a given tolerance, leading to estimated gravity and magnetic models.

3. Synthetic Models

Let us apply the methodology to realistic synthetic models consisting of isolated sources. In all cases the studied sources are embedded in a homogeneous half-space with null density and magnetization.

In our experiments, 10,000 iterations of MCMC are used, we reject the first 4000 iterations as burnin and the remaining 6000 iterations are subsampled with step 100.

The initial guess of the MCMC method is selected randomly on each run from the same distributions:

$$\begin{aligned} x_R, y_R &\sim U(200, 600), \\ z_R &\sim U(100, 250), \\ l_{x_R}, l_{y_R}, l_{z_R} &\sim U(0, 200), \\ \rho_R &\sim U(0, 1). \end{aligned}$$

As a result, we have a pseudo independent sample from the posterior distribution in a high probability region. The conditional mean is computed to obtain the density reference model $\mathbf{m}_R^{(1)}$.

3.1. Test Case 1

Here we consider two cubes from Li and Oldenburg (1998). Both cubes are 200 m on a side with $1.0\,\mathrm{g/cm^3}$ density. Also it is assumed that the cubes have a magnetization of $1.0\,\mathrm{A/m}$.

Gaussian noise with standard deviation of 0.02 mGal and 2 nT (approximately 1% of the maximum amplitudes of their respective anomalies), is added to the data.

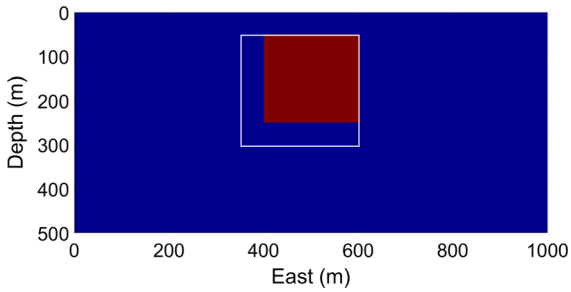

Figure 2
In white, the vertical cross section of density Bayesian estimation of the cube buried at 50 m depth. True cube in red

In Fig. 2 we show a vertical cross section for the cube buried at 50 m depth. The white box corresponds to the Bayesian estimation. It is observed that the latter encloses the true model.

Separate inversion is carried out. For density, the bayesian estimation is used as reference model. The corresponding covariance matrix is $\mathbf{C}_R^{(1)} = 0.1$. Whereas, the magnetization model is obtained using as reference model $\mathbf{m}_R^{(2)} = \mathbf{0}$, and a large enough $\mathbf{C}_R^{(2)}$ to have fully unconstrained experiments. Results in Fig. 3. It is observed that the upper boundary of the density model is lost. Magnetization yields spurious information at the deeper region.

As shown in Fig. 4, joint inversion recovers the model satisfactorily. The displayed results for joint inversion, are after 20 iterations. In this experiment, $\mathbf{C}_R^{(1)} = 0.1$ which implies that the density model is close to the reference model in all iterations. The value of the regularization parameter for both models is set as $\alpha_1^2 = \alpha_2^2 = 1 \times 10^9$ providing smooth models.

We analyze the suitability of the models in terms of the actual geophysical data, as in Fregoso and Gallardo (2009). That is, root mean squares are computed for data misfit. Namely,

$$rms_g = \frac{1}{\sqrt{n_g}} \left\| \mathbf{d}^{(1)} - \mathbf{G}^{(1)}\mathbf{m}^{(1)} \right\|_{(\mathbf{C}_d^{(1)})^{-1}},$$

the root mean squared associated to gravity. And

$$rms_m = \frac{1}{\sqrt{n_m}} \left\| \mathbf{d}^{(2)} - \mathbf{G}^{(2)}\mathbf{m}^{(2)} \right\|_{(\mathbf{C}_d^{(2)})^{-1}},$$

the corresponding for magnetization. Here n_g and n_m are the number of data for gravity and magnetization respectively.

In this case study we obtain $rms_g = 1.0$, and $rms_m = 1.09$ for separate inversion. For joint inversion, the values attained are $rms_g = 1.01$ and $rms_m = 1.18$. The values are reasonable in all cases, but the performance on source recovery is better when joint inversion with bayesian estimation is used.

The second cube in Li and Oldenburg (1998) we consider, is the one buried at 150 m depth. See Fig. 5. The true model is also enclosed by the white box corresponding to the Bayesian estimation. Now to the opposite side of the shallower cube, compare with Fig. 2.

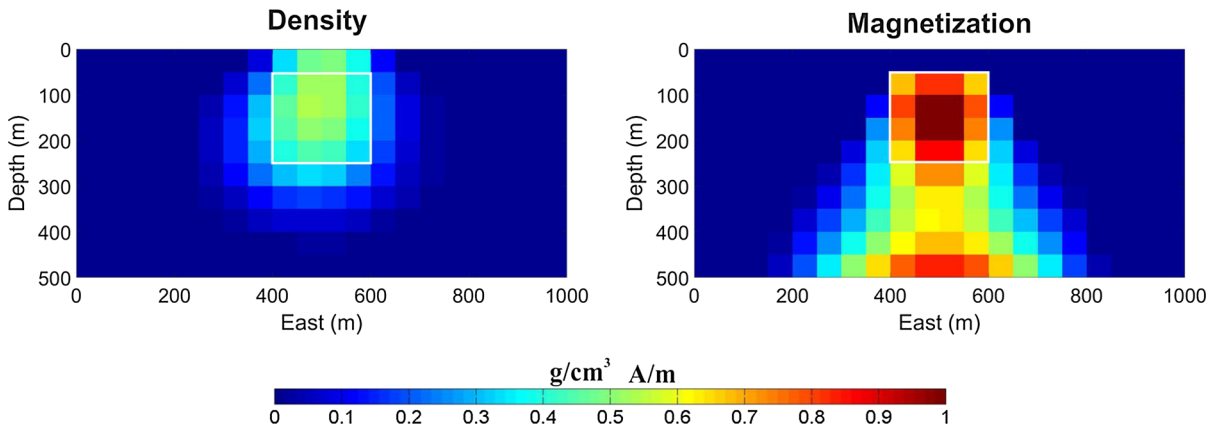

Figure 3
Vertical slices for separate inversion of the cube buried at 50 m depth. The true position of the cube is outlined by a white solid line

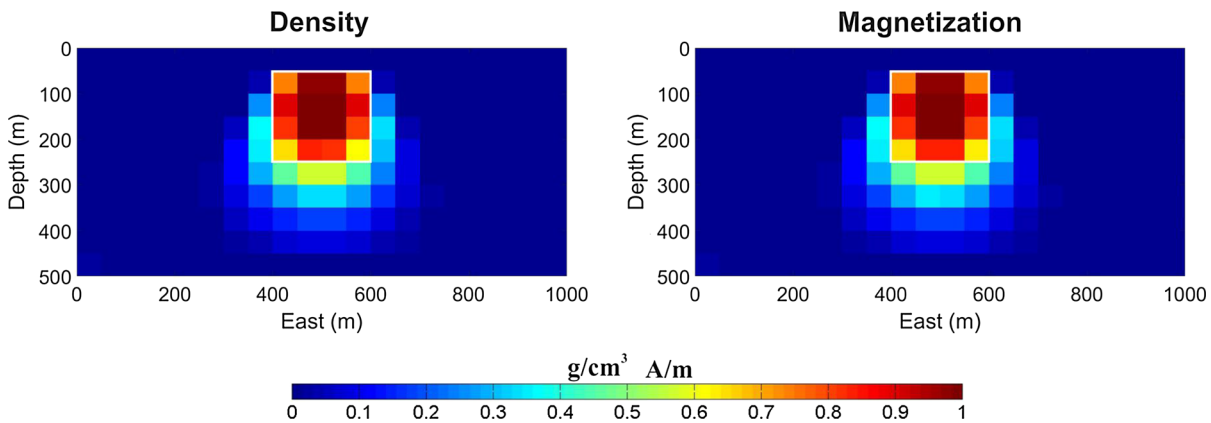

Figure 4
Vertical slices after joint inversion of the cube buried at 50 m depth. The true position of the cube is outlined by a white solid line

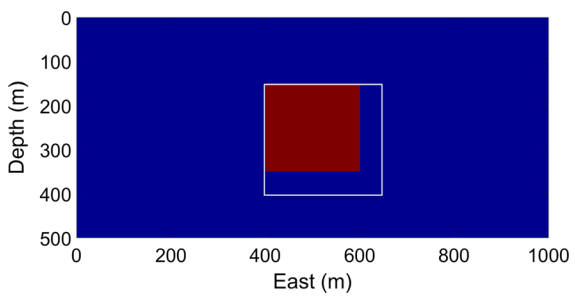

Figure 5
Vertical cross section of density Bayesian estimation in white of the cube buried at 150 m depth. True cube in red

As before, we compute the separate inversion and show results in Fig. 6. The reference model for magnetization is $\mathbf{m}_R^{(2)} = \mathbf{0}$, and covariance matrix $\mathbf{C}_R^{(2)}$ is chosen large enough to have fully unconstrained experiments. Results are unsatisfactory.

As expected, joint inversion initiated with bayesian informed referenced models, yields satisfactory results, see Fig. 7. The $rms_g = 0.97$ and $rms_m = 0.95$ for separate inversion, whereas $rms_g = 0.99$ and $rms_m = 0.95$ for joint inversion.

It is well known that joint inversions improve upon separate inversions. The difference here is the choice of reference models. For comparison let us show joint inversion without bayesian information. That is, the unconstrained joint inversion.

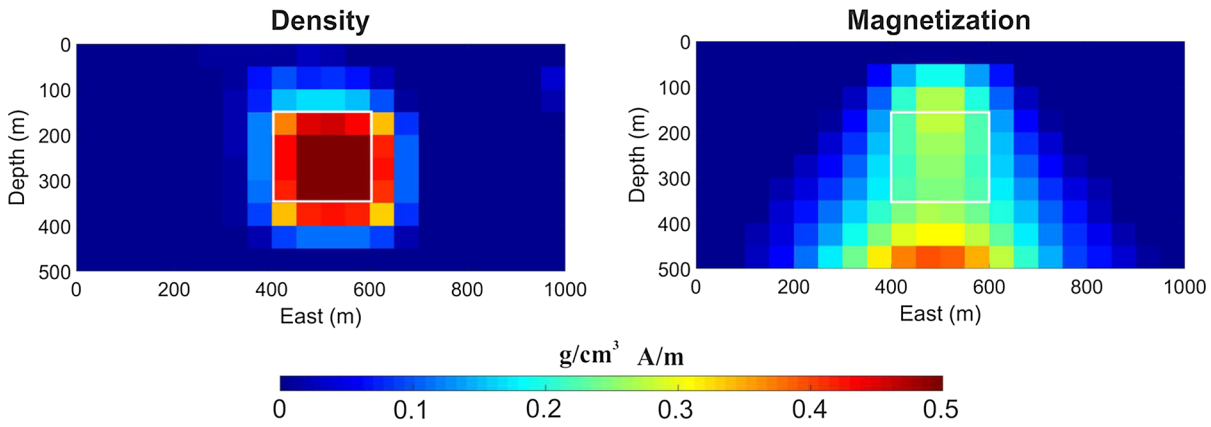

Figure 6
Vertical slices for separate inversion of the cube buried at 150 m depth. The true position of the cube is outlined by a white solid line

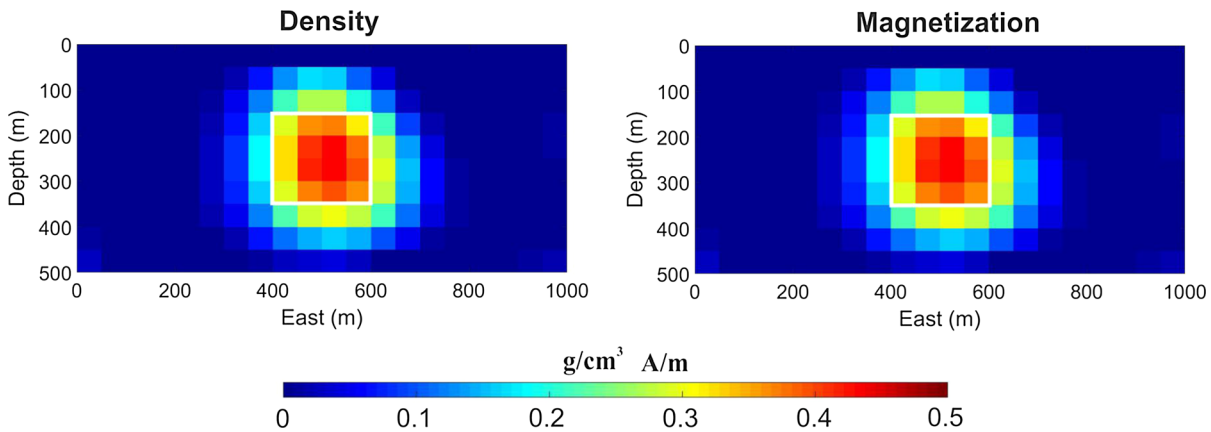

Figure 7
Vertical slices after joint inversion of the cube buried at 150 m depth. The true position of the cube is outlined by a white solid line

Figure 8 shows the results of the cross gradient methodology obtained after 20 iterations. The reference models are $\mathbf{m}_R^{(1)} = \mathbf{0}$, $\mathbf{m}_R^{(2)} = \mathbf{0}$. As before $\mathbf{C}_R^{(1)}$ and $\mathbf{C}_R^{(2)}$ are selected with large values. In this case $rms_g = 0.99$ and $rms_m = 0.95$. The initial models of the iterative process are the separate inverted models.

The accuracy of the estimated models is computed through *RMS* values between the estimated and test models for cross gradient joint inversions. Notice that we use capital letters to distinguish from the *rms* above. Naturally, the *RMS* is computed trough the whole region. For Bayesian joint inverted model, $RMS_g = RMS_m = 0.09$, comparatively to classical joint inversion where $RMS_g = RMS_m = 0.12$. Moreover, density and magnetization values attain larger magnitudes for the case of cross-gradient constrained by Bayesian density model, which is closest to real values comparatively to unconstrained cross-gradient methodology. This reflects that not only the spatial distribution is better recovered, but also the magnitude of the target.

3.2. Test Case 2

The second example is also from Li and Oldenburg (1998). It is a dipping dike model buried at 50 m depth, embedded in a homogeneous medium. The density and induced magnetization contrasts are $1\,\mathrm{g/cm^3}$ and $1\,\mathrm{A/m}$.

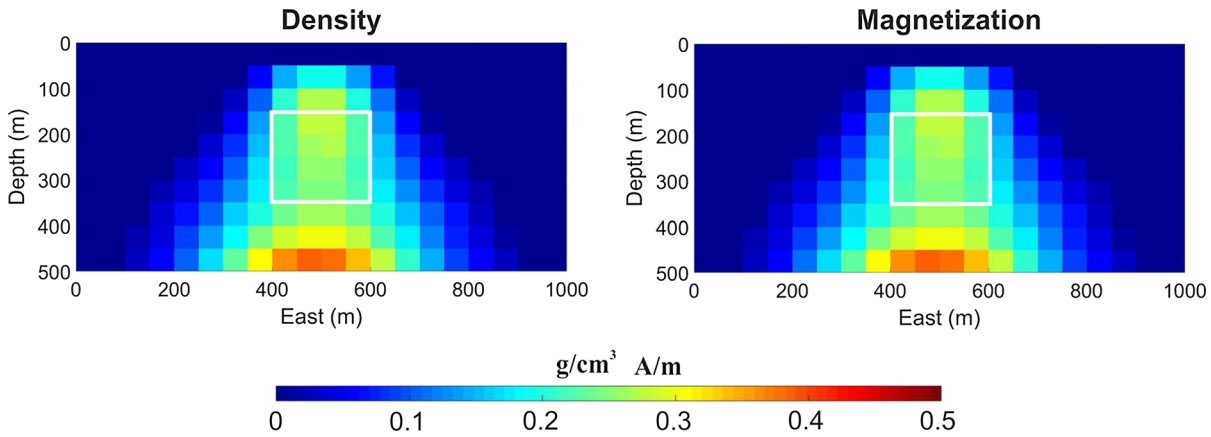

Figure 8
Vertical slices of classical cross gradient joint inversion of the cube buried at 150 m depth. The true position of the cube is outlined by a white solid line

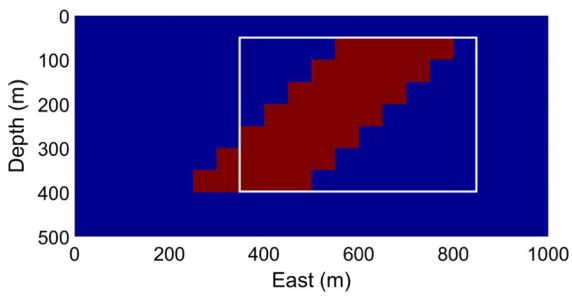

Figure 9
Vertical cross section for a dipping dike model (in red) buried at 50 m depth. The white box corresponds to the Bayesian estimation

The dimensions and the discretization of the inversion domain, in three directions, are equal to previous test case, as well as the simulated observations, which are located in the same spaced grid. Furthermore, Gaussian noise of 1% of the maximum amplitude of their respective anomalies, is added to the data. It corresponds to standard deviation of 0.03 mGal and 2 nT for the geomagnetic field with $I = 45^\circ$ and $D = 45^\circ$.

Running the MCMC procedure we are led to a density estimation as shown in Fig. 9.

Figures 10 and 11, show respectively, the separate, and joint inverted models after 20 iterations. The reference model corresponds to the Bayesian

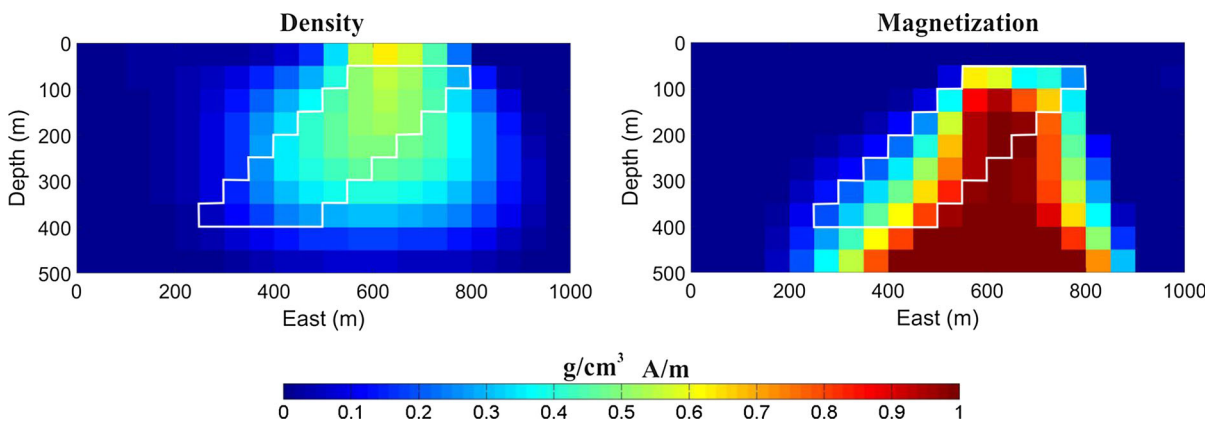

Figure 10
Vertical slices of density and magnetization separate inversion models. The true position of the dike is outlined by a white solid line

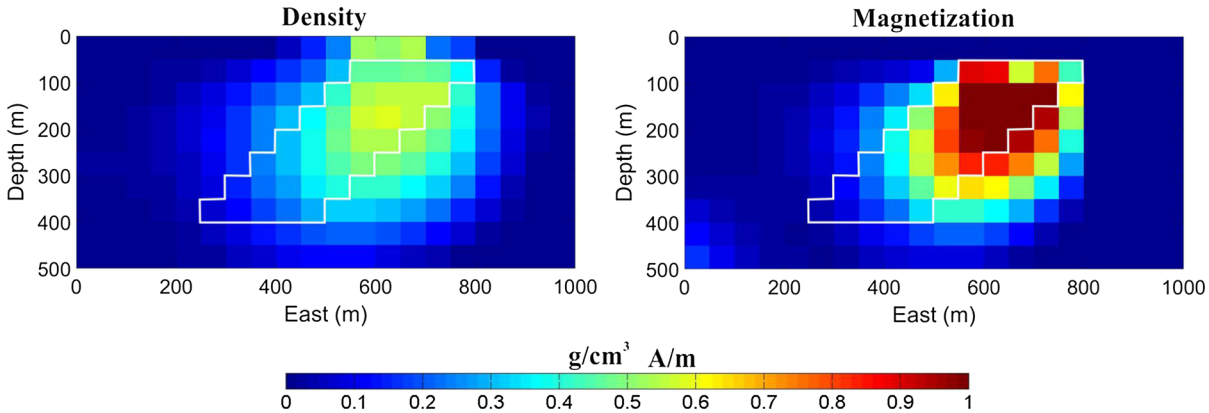

Figure 11
Vertical slices of density and magnetization joint inversion models with bayesian initialization. The true position of the dike is outlined by a white solid line

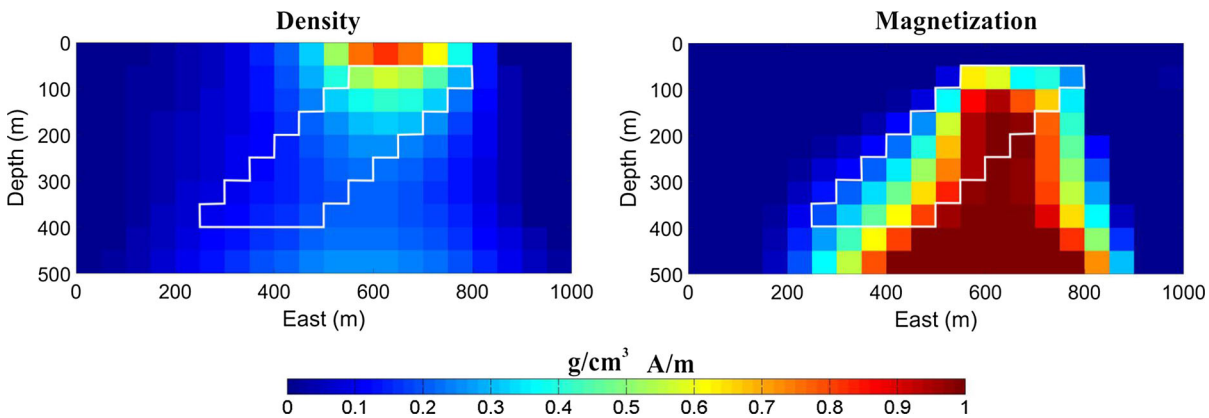

Figure 12
Vertical slices after unconstrained separate inversion of the dike buried at 50 m depth. The true position of the dike is outlined by a white solid line

estimated density model. The value of the regularization parameter are $\alpha_1^2 = 1 \times 10^{10}$ and $\alpha_2^2 = 1 \times 10^9$, thus providing smooth models.

The normalized data misfit attained for these models are $rms_g = 1.01$ and $rms_m = 1.09$ for separate inversion. For joint inversion, the values attained are $rms_g = 1.16$ and $rms_m = 1.28$.

For comparison we show in Fig. 12 the unconstrained separate inversion. Then in Fig. 13 the corresponding unconstrained cross-gradient results.

The improvement shown in Fig. 11 is highly satisfactory.

4. Conclusions and Future Work

We have introduced bayesian estimation to construct a surrogate density model to initialize the cross-gradients iteration for joint inversion of gravity and magnetic data. The initial model is a prism that recovers the density anomaly.

Reference models are commonly used in other joint inversion methodologies. An elliptic reference model is used in the first step of the algorithm in Li and Qian (2016). It is apparent that this, and other simple structures are easily implemented in our bayesian approach. In particular, one might start Li

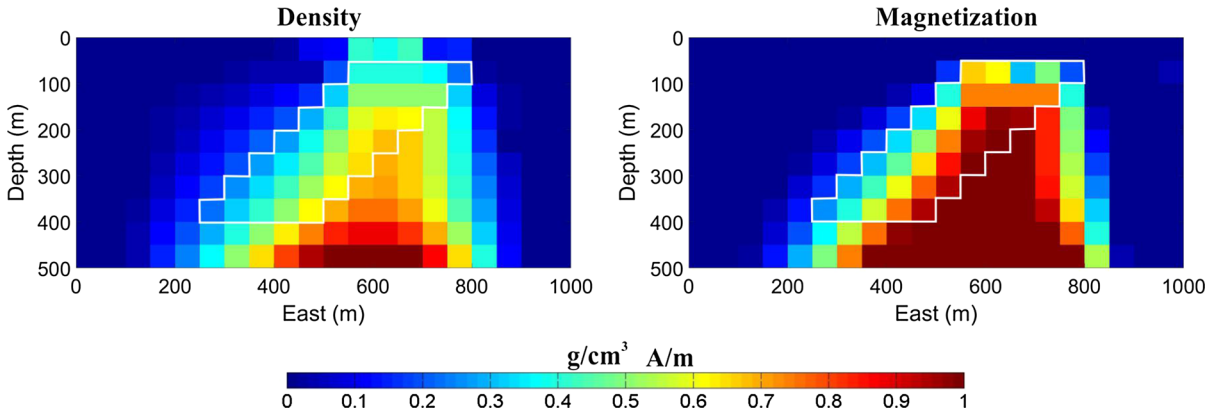

Figure 13
Vertical slices after unconstrained joint inversion of the dike buried at 50 m depth. The true position of the dike is outlined by a white solid line

and Qian algorithm with a bayesian elliptic density reference model.

Our surrogate model formulation is also efficient on memory usage. More precisely, for storing the entire MCMC chain, the amount of memory is about Tn^3p, with T number of MCMC iterations and p the size of a floating point number in bytes. Thus, for 25000 iterations of the MCMC in a classical setting, 8 bytes for floating point representation, and a coarse discretization of $n = 20$. A quick calculation yields that the memory usage raises to 1.5 Gb at least. In the case of the surrogate model, the memory usage is the order of 1.5 Mb.

An ongoing work, is the extension to the union of two or more simple structures as the surrogate model for initialization.

Of greater interest for future work is to consider more complex structures, not necessarily convex, for surrogate models.

In this study, we aimed for priors with no physical bias. The proposal (3), is straightforward and yields satisfactory depth information. The same principle is applied to the choice of the magnetization reference model. In our experiments $\mathbf{m}_R^{(2)} = \mathbf{0}$. This choice may be considered the worst case scenario. There are other natural choices to be tried, such as the same geometric structure of the density reference model. The basic principle is that more information on reference models, bias the inversion results towards these. In an actual application one ought to use all available information, but our intention is to show that the approach presented here is promising even with limited information on the second model.

Depth resolution is another important subject to pursue. Let us mention some directions.

Depth weighting functions are common practice in Geophysical inversion. Noteworthy is the proposal of Li and Oldenburg (1998). Therein the problem of inversion is posed as a minimization of a misfit functional corrected with several weighting functions. In particular, a depth weighting function of the form

$$w(z) = \frac{1}{(z + z_0)^{\beta/2}}.$$

It is suggested to set $\beta = 2$, and estimate z_0 in accord with the meshing of the model and the observation height of the data. This depth weighting function can be made into a prior, our preliminary exploration is promising, and shall be reported elsewhere.

In the context of cross-gradients, joint inversion of gravity data and seismic data adds to depth resolution. It is of interest to explore a comparative study of the strategies listed here.

Acknowledgements

This research was carried out while M. A. Moreles was a on sabbatical leave at the mathematics

Department of the Universidad de Guadalajara. Their hospitality is greatly appreciated. Also, M. A. Moreles would like to acknowledge the support of ECOS-NORD project number 000000000263116/M15M01. The authors wish to thank the anonymous referees. The manuscript was greatly improved by their very constructive comments.

References

Aarts, E., & Korst, J. (1988). *Simulated annealing and boltzmann machines*. New York, NY: Wiley.

Bhaskara-Rao, D., Prakash, M. J., & Ramesh-Babu, N. (1990). 3D and 2D modeling of gravity anomalies with variable density contrast. *Geophysical Prospecting*, *38*, 411–422.

Bhattacharyya, B. K. (1966). A method for computing the total magnetization vector and the dimensions of a rectangular block-shaped body from magnetic anomalies. *Geophysics*, *31*, 74–96.

Christen, A., & Fox, C. (2010). A general purpose sampling algorithm for continuous distributions (the t-walk). *Bayesian Analysis*, *5*(2), 263–281.

Fedi, M., Hansen, P. C., & Paoletti, V. (2005). Tutorial: Analysis of depth resolution in potential-field inversion. *Geophysics*, *70*(6), A1–A11. https://doi.org/10.1190/1.2122408.

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013). *Emcee: The MCMC Hammer*. Publications of the Astronomical Society of the Pacific 125(925). IOP Publishing.

Fregoso, E., & Gallardo, L. A. (2009). Cross-gradients joint 3D inversion with applications to gravity and magnetic data. *Geophysics*, *74*(4), L31–L42. https://doi.org/10.1190/1.3119263.

Fregoso, E., Gallardo, L. A., & García-Abdeslem, J. (2015). Structural joint inversion coupled with Euler deconvolution of isolated gravity and magnetic anomalies. *Geophysics*, *80*(2), G67–G79.

Gallardo L. A., & Meju., M. A. (2004). Joint two-dimensional DC resistivity and seismic travel time inversion with cross-gradients constraints. *Journal of Geophysical Research, 109*.

Haber, E., & Oldenburg, D. (1997). Joint inversion: A structural approach. *Inverse Problems*, *13*, 63–77.

Joulidehsar, F., Moradzadeh, A., & Ardejani, Doulati. (2018). An improved 3D joint inversion method of potential field data using cross-gradient constraint and LSQR method. *Pure and Applied Geophysics*,. https://doi.org/10.1007/s00024-018-1909-7.

Li, Y., & Oldenburg, D. W. (1996). 3-D inversion of magnetic data. *Geophysics*, *61*(2), 394–408.

Li, Y., & Oldenburg, D. W. (1998). 3-D inversion of gravity data. *Geophysics*, *63*(1), 109–119.

Li, W., & Qian, J. (2016). Joint inversion of gravity and traveltime data using a level-set-based structural parameterization. *Geophysics*, *81*(6), G107–G119.

Moorkamp, M. B., Heincke, M. B., Jegen, M., Roberts, A. W., & Hobbs, R. W. (2011). A framework for 3-D joint inversion of MT, gravity and seismic refraction data. *Geophysical Journal International*, *184*, 477–493. https://doi.org/10.1111/j.1365-246X.2010.04856.x.

Moorkamp, M., Roberts, A. W., Jegen, M., Heincke, B., & Hobbs, R. W. (2013). Verification of velocity-resistivity relationships derived from structural joint inversion with borehole data. *Geophysical Research Letters*, *40*, 3596–3601. https://doi.org/10.1002/grl.50696.

Nielsen, L., & Jacobsen, B. H. (2000). Integrated gravity and wide-angle seismic inversion for two-dimensional crustal modelling. *Geophysical Journal International*, *140*, 222–232. https://doi.org/10.1046/j.1365-246x.2000.00012.x.

Paoletti, V., Hansen, P. C., Hansen, M. F., & Fedi, M. (2014). A computationally efficient tool for assessing the depth resolution in large-scale potential-field inversion. *Geophysics*, *79*(4), A33–A38. https://doi.org/10.1190/GEO2014-0017.1.

Pilkington, M. (2012). Analysis of gravity gradiometer inverse problems using optimal design measures. *Geophysics*, *77*(2), G25–G31. https://doi.org/10.1190/geo2011-0317.1.

Robert, C., & Casella, G. (2013). *Monte Carlo statistical methods*. New York: Springer Science & Business Media.

Sambridge, M., & Mosegaard, K. (2002). Monte Carlo methods in geophyisical inverse problems. *Reviews of Geophysics*, *40*(3), 3-1.

Stuart, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numerica*, *19*, 451–559.

Tarantola, A., & Valette, B. (1982). Generalized non-linear inverse problems solved using the least-squares criterion. *Reviews of Geophysics and Space Physics*, *20*, 219–232.

(Received December 11, 2018, revised September 19, 2019, accepted September 23, 2019, Published online September 27, 2019)