Efficiently Implementing and Balancing the Mixed L_p-Norm Joint Inversion of Gravity and Magnetic Data

Saeed Vatankhah, Xingguo Huang, Rosemary A. Renaut, Kevin Mickus, Hojjat Kabirzadeh, and Jun Lin

***Abstract*— The mixed L_p-norm, $0 \leq p \leq 2$, stabilization algorithm is flexible for constructing a suite of subsurface models with either distinct, or a combination of, smooth, sparse, or blocky structures. This general-purpose algorithm can be used for the inversion of data from regions with different subsurface characteristics. Model interpretation is improved by the simultaneous inversion of multiple datasets using a joint inversion approach. An effective and general algorithm is presented for the mixed L_p-norm joint inversion of gravity and magnetic datasets. The imposition of the structural cross-gradient enforces similarity between the reconstructed models. For efficiency, the implementation relies on three crucial realistic details: 1) the data are assumed to be on a uniform grid providing sensitivity matrices that decompose into the block Toeplitz Toeplitz block form for each depth layer of the model domain and yield efficiency in storage and computation via 2-D fast Fourier transforms; 2) matrix-free implementation for calculating derivatives of parameters reduces memory and computational overhead; and 3) an alternating updating algorithm is employed. Balancing of the data misfit terms is imposed to assure that the gravity and magnetic datasets fit with respect to their individual noise levels without overfitting of either model. Strategies to find all weighting parameters within the objective function are described. The algorithm is validated on two synthetic but complicated models. It is applied to invert gravity and magnetic data acquired over two kimberlite pipes in Botswana, producing models that are in good agreement with borehole information available in the survey area.**

***Index Terms*— Gravity, joint inversion, kimberlite, magnetic, mixed L_p-norm.**

I. Introduction

DUE to the availability of different datasets for a survey area and greater accessibility to computers with increased computational power, it is important to develop efficient algorithms for the joint inversion of multiple geophysical datasets. Equipped with the complementary information about the subsurface target(s) provided by different datasets, the standard issues of nonuniqueness for inversion of geophysical data are reduced, and simultaneous joint inversion techniques are stabilized [22]. Several approaches, dependent on coupling mechanisms for different model parameters, have been developed for the joint inversion of geophysical datasets. If there are explicitly known, or implicitly assumed empirical, relationships between different sets of model parameters, these relationships can be used to link different model parameters [1], [22], [28], [37]. In the absence of empirical relationships or the provision of a priori information on the geometries of model parameters, direct spatial coupling of the model parameters can be obtained by imposing structural constraints [13], [50]. Of these, the cross-gradient coupling enforces that changes in models, as represented by their spatial gradients, occur at similar locations. Algorithms using these constraints have been widely used in the joint inversion literature, e.g., [12], [14], [15], [27], [38], and [41], and yield reconstructed models that exhibit significant structural similarity [13], [16]. Gramian constraints can also be used within a joint inversion to provide correlated reconstructions for different model parameters, or their attributes, without a priori known relationship(s) [18], [25], [45], [50], [51]. Depending on the available information, these strategies can be used independently, or they can be combined [22]. Here, we apply the cross-gradient constraint that is an accepted approach for imposing structural similarity between density and magnetic susceptibility models for the joint inversion of gravity and magnetic datasets.

Many different stabilization techniques can be used for the inversion of potential field datasets. Choosing from these different approaches should not be arbitrary. Rather, consideration has to be paid to the goals of the inversion. For example, does one aim to recover a model that is simple and smooth with minimal structure, or does one anticipate that the model should be sparse, either in the parameter space, or with respect to perhaps the gradient of the model parameters? In the former case, the obtained model would only be able to represent the important and large-scale features of the subsurface target(s). With this aim, one would use a stabilizer that employs the L_2-norm of the gradient of the model parameters, usually called the maximum smoothness stabilizer [8], [23], [31].

Manuscript received 4 December 2022; revised 5 March 2023 and 11 May 2023; accepted 30 June 2023. Date of publication 6 July 2023; date of current version 19 July 2023. This work was supported by the Youth Talent Project of the China Association for Science and Technology under Grant 2020QNRC004. The work of Rosemary A. Renaut was supported by the NSF under Grant DMS-1913136 and Grant DMS-2152704. *(Corresponding author: Xingguo Huang.)*

Saeed Vatankhah, Xingguo Huang, and Jun Lin are with the College of Instrumentation and Electrical Engineering, Jilin University, Changchun 130012, China (e-mail: svatan@ut.ac.ir; xingguohuang@jlu.edu.cn; xingguo.huang19@gmail.com).

Rosemary A. Renaut is with the School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85287 USA.

Kevin Mickus is with the Department of Geography, Geology and Planning, Missouri State University, Springfield, MO 65897 USA.

Hojjat Kabirzadeh is with the Department of Geomatics Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada.

Digital Object Identifier 10.1109/TGRS.2023.3292889

In contrast, if the aim is to obtain a solution that is compact with well-defined edges to the structures, a stabilizer that uses the L_1- or L_0-norm of the model parameters or gradient of the model parameters would be appropriate [2], [3], [4], [9], [10], [19], [32], [42]. Such stabilizers are suitable if the goal is to reconstruct localized, or discontinuous, sources. In either case, the inversion is stabilized using an L_p-norm, $0 \leq p \leq 2$, which is also easily extended for incorporation into a joint inversion algorithm [41]. Multiple L_p stabilization terms can also be applied, so as to recover models with such different characteristics [11], [36], [48]. Consequently, it is possible to generate a suite of different models that all approximate the observed data with respect to a given noise level and can provide potential candidates for the unknown subsurface models. These approaches can be included within an algorithm for the joint inversion of multiple datasets. Here, we present a general algorithm for the joint inversion of gravity and magnetic data based on a mixed L_p-norm stabilization with coupling between the models using a cross-gradient constraint term to enforce structural similarity between models.

A mixed L_p-norm joint inversion algorithm presents multiple computational challenges that go beyond those for the independent inversion of individual datasets. Effective strategies to overcome the major challenges that build on the successful approaches for the single-dataset case are presented. The issues with memory and computational cost for the inversion of a single large-scale potential field dataset are compounded for a joint inversion algorithm. Moreover, an algorithm that correctly accounts for the physical properties of more than one set of parameters on a given volume presents additional complexity due to the larger number of terms and, consequently, the associated additional weighting parameters, which arise in the objective function that describes the appropriate formulation. For large-scale problems, even in the single-dataset case, it is important to optimize the costs of generating, storing, and using the dense sensitivity matrices for efficient matrix-vector operation regardless of the iterative or direct algorithm used to find the solution of the inverse problem [24]. Without careful attention to the implementation details, it is not feasible to obtain solutions on typical desktop or laptop computers. For joint inversion, these factors become even more crucial.

Several strategies have been developed to mitigate the difficulties associated with inverting large-scale datasets. For example, a compression technique to speed up the computations and decrease the memory requirements can be applied [33]. Using the fast wavelet transform on the sensitivity matrix, with thresholding applied to eliminate the small wavelet coefficients, so as to obtain a sparse minimal memory representation of the matrix has also been considered [24]. Not only does the wavelet approach reduce storage requirements because the matrix is compressed but also the computational cost is lowered through the use of fast matrix-vector multiplications in the wavelet domain. The overall costs for carrying out matrix-vector operations may be reduced by either reducing the number of such operations and/or by projecting the original problem to a smaller subspace. Algorithms that fall into this category are, for example, iterative solvers or randomization algorithms that use subspace projection. Given a suitable projected space, a subspace solution can be obtained cheaply under the assumption that the characteristics of the full-space problem are sufficiently captured [29], [30], [42], [43]. These solvers can also be used in the context of approaches that speed up the underlying matrix-vector operations. For example, taking advantage of the underlying structure of the sensitivity matrices is advantageous for the solution of large-scale inverse problems [34].

When the data are obtained on a uniform grid and an associated uniform volume discretization is defined for each depth layer of the volume, then the sensitivity matrix exhibits a block Toeplitz Toeplitz block (`BTTB`) structure for each set of unknowns by depth in the volume [7], [49]. These block matrices are symmetric for the gravity problem but unsymmetric for the magnetic problem. In either case, the resulting `BTTB` matrix can be embedded into a block circulant circulant block (`BCCB`) matrix that facilitates the use of the `2DFFT` for all matrix-vector operations [47], and there is no need to store either the `BTTB` or `BCCB` matrix. To perform the embedding, one needs only the first column, or for the unsymmetric case both the first row and the first column, in order to fully describe the underlying operations with the matrix as a `2DFFT`. Not only is there a consequent significant savings in memory but also the use of the `2DFFT` for all matrix-vector operations improves (reduces) the computational costs significantly within any solver, yielding an efficient strategy for the inversion of large-scale potential field data [34]. The strategy has also extended successfully for joint inversion of gravity and magnetic datasets [41]. Here, we apply the `BTTB` structure of sensitivity matrices in combination with a mixed L_p-norm joint inversion algorithm. The feasibility to use this approach on the large scale is further enhanced by adopting an inversion algorithm that alternates the updates for the density and magnetic susceptibility while keeping the coupling of the two models through the cross-gradient constraint. This further facilitates the inversion of large datasets in a reasonable time and with reduced storage requirements. The alternating algorithm applies regardless of the presence of the `BTTB` structure of the sensitivity matrices, and significant savings in computation time are observed.

For the joint inversion of gravity and magnetic datasets, it is standard to seek a solution for which the overall data misfit for both models satisfies a χ^2 criteria as a measurement of the convergence of the algorithm. When this data fit term is measured for the two models, it is possible that the iteration will terminate when one or other of the models has a much reduced individual data fit term. This will occur if the convergence for one model is faster than for the other and will correspond to over fitting of the model that converged faster. This overfitting can lead to greater error in the specific reconstructed models. To avoid the imbalance between the accuracies of each model, it is appropriate to balance the individual contributions of each data fit term to the overall data fit used for the convergence test. Here, we apply weighting parameters for each data misfit term and present an effective strategy to determine the relative weighting parameters. In the algorithm, all components of the objective function, including

the data misfits, stabilizers, and cross-gradient coupling, are weighted with suitable parameters that are updated through the iteration. This yields a joint inversion algorithm for which the fit for each data type is suited to its individual noise level.

A. Overview of Main Scientific Contributions

An efficient general algorithm for mixed L_p-norm joint inversion with a cross-gradient coupling is presented. The approach includes the imposition of an anisotropic gradient norm constraint, combined with a norm constraint on the model parameters. This extends the work in which the efficient joint alternating direction (AD) algorithm was introduced without balancing the data fits and without the mixed-formulation, namely, for either a single L_p-norm constraint on the model parameters or on a single isotropic gradient of the parameters (no experiments were reported using the isotropic gradient norm) [41], [44, eqs. (5)–(7)]. The implementation uses a matrix-free and storage-efficient calculation of all derivatives of the parameters, as well as the entries of the coupling matrices. Furthermore, a novel strategy to automatically balance the data fit terms to limit either over or underfitting of either dataset is presented. The weighting of the data misfit terms is designed to ensure that both datasets fit within their respective noise levels. This approach makes it completely feasible to address many options for mixed and practical inversion of real datasets. Different choices for the norm constraint terms, which are flexible to any choice of L_p-norm with $0 \leq p \leq 2$, can be used to provide sparse or smooth solutions and allow the use of isotropic or anisotropic gradient terms.

B. Organization

Section II addresses the main background for all the components of the joint inversion algorithm. The problem formulation is given in Section II-A, and a brief explanation of the mixed L_p-norm inversion methodology for the inversion of a single dataset is provided in Section II-B with all the major details that are standard for independent inversions. The coupled cross-gradient joint inversion formulation is provided in Section II-C, with algorithm considerations in Section II-D. Effective determination of the various important weighting parameters that define a specific inversion is discussed in Section II-E. The resulting mixed L_p cross-gradient coupled joint inversion algorithm is validated using synthetic datasets in Section III. The selection of options for the tunable weighting parameters is applied on two different models to showcase the properties of the algorithm that is given in the Appendix. Results on the application of the algorithm on real gravity and magnetic data obtained over two kimberlite pipes in Botswana are given in Section IV.

II. Joint Inversion Methodology

The joint inversion of the data depends on both the framework that is used for the independent inversion of two distinct datasets and the approach by which the two models are coupled. The fast inversion of a single potential field dataset is by now well-studied; for example, earlier works are provided in [10], [19], [23], and [31] with recent extensions that discuss efficiency and inclusion of more general regularization terms [7], [11], [17], [24], [33], [34], [35], [36], [42], [43], [44], [48], [49]. For the joint inversion, the use of a cross-gradient coupling constraint for potential field datasets has also been described in the literature [12], [13], [14], [15], [27], [38], [41]. Our main contribution is the extension to apply the cross-gradient coupling constraint with a mixed regularization term that includes an anisotropic gradient constraint and a stabilization constraint on the model. Balancing is applied for the data fit terms that arise. Combining these terms together introduces multiple weighting parameters that need to be tuned to obtain a suitable solution. To be complete, we present the basic information in Section II-A, a brief review of the approach for the single data case in Section II-B, a description of the major requirements for the joint algorithm in Section II-C, and an overview leading to the final objective function in Section II-D. The AD algorithm is detailed in Algorithm 1 in the Appendix along with a discussion of important computational considerations.

A. Problem Formulation

We assume the linear forward projection problem

$$\mathbf{d}^{\text{obs}} = \mathbf{G}\mathbf{m} \tag{1}$$

where $\mathbf{G} \in \mathcal{R}^{m \times n}$ is a linear forward modeling operator that projects the model parameters collected in vector $\mathbf{m} \in \mathcal{R}^n$ to observed measured data $\mathbf{d}^{\text{obs}} \in \mathcal{R}^m$ at, or above, the surface, with $m \ll n$. In obtaining this formulation, we make a number of assumptions: 1) the observed data are distributed on a uniform grid at a fixed height, denoted by z_{av}, relative to the surface of a volume for which the subsurface target(s) should be identified [5]; 2) each measurement station is at the center of the top face of a cuboid defined for the first layer of the volume; 3) the volume discretization is comprised of cuboids that are all of the same size; and 4) the unknown physical properties of the cuboids contained in vector $\mathbf{m}$ are assumed to be constant and are associated with the value at the geometric center of each cuboid. The assumption of a uniform grid can be achieved by interpolating the data from nonuniform locations to uniform locations. It has been shown that it is reasonable to use interpolation because the associated errors that are introduced have a lesser effect on the solution than the practical errors in the measurements [6]. The advantage of the use of these assumptions is that we obtain the BTTB structure of the matrices by depth layer, which facilitates both memory and cost reduction in the use of the sensitivity matrices, as noted in Section I and further described in [7] and [47].

The forward model [see (1)] applies for both gravity and magnetic problems, in which the unknowns represent the subsurface densities $\mathbf{m}_1$ and magnetic susceptibilities $\mathbf{m}_2$, respectively, and the matrix $\mathbf{G}$ is modified according to the model, specified consistently as $\mathbf{G}_1$ and $\mathbf{G}_2$.[1] For the magnetic case, we suppose that there is no remanent magnetization, and the self-demagnetization effects are negligible.

[1]Throughout this article, we use subscripts 1 and 2 to indicate terms that are related to the gravity and magnetic models, respectively.

B. Inversion of a Single Dataset Using a Mixed L_p Stabilization

An approach to find a single model $\mathbf{m}$ given $\mathbf{d}^{\text{obs}}$ uses a mixed L_p-norm regularization, as described further in the following [11], [36], [48]:

$$\mathbf{m}^* = \arg\min_{\mathbf{m}}\{\Phi(\mathbf{m})\} = \arg\min_{\mathbf{m}}\left\{\phi_{\text{datafit}}(\mathbf{m}) + \beta^2\phi_{\text{reg}}(\mathbf{m})\right\} \tag{2}$$

where $\phi_{\text{datafit}}(\mathbf{m})$ is a data fit term, $\phi_{\text{reg}}(\mathbf{m})$ comprises stabilization terms, and $\beta > 0$ is a regularization parameter that trades off the relative contributions of the two terms. The weighted data fit is given by

$$\phi_{\text{datafit}}(\mathbf{m}) = \|\mathbf{W}_{\mathbf{d}^{\text{obs}}}\left(\mathbf{d}^{\text{obs}} - \mathbf{G}\mathbf{m}\right)\|_2^2. \tag{3}$$

The diagonal matrix $\mathbf{W}_{\mathbf{d}^{\text{obs}}}$ is the inverse square root of the covariance for the noise in the measured data $\mathbf{d}^{\text{obs}}$. The noise is assumed independent but potentially colored so that the diagonal entries in $\mathbf{W}_{\mathbf{d}^{\text{obs}}}$ are the inverses of the standard deviation estimates for the noise per data point.

The mixed regularization term given by

$$\begin{aligned}\phi_{\text{reg}}(\mathbf{m}) &= \alpha_s\|\mathbf{m} - \mathbf{m}^{\text{apr}}\|_{p_s} + \sum_{j=\text{x,y,z}}\alpha_j\ \|\mathbf{D}_j\mathbf{m}\|_{p_j}\\ &= \phi_s(\mathbf{m}) + \phi_D(\mathbf{m})\end{aligned} \tag{4}$$

depends on the norm parameters $0 \le p_s, p_j \le 2$, typically assumed to be integers, which can be chosen differently for all components.[2] Subscript s indicates variables that are associated with the *smallness* component $\phi_s(\mathbf{m})$. This measures the size of a model with respect to an initial model, $\mathbf{m}^{\text{apr}}$ [11]. The starting model can be a known reference model for the candidate approximation of $\mathbf{m}$. It is also possible to set $\mathbf{m}^{\text{apr}} = \mathbf{0}$ if no prior information is available. In (4), $\phi_D(\mathbf{m})$ is used to denote the terms that measure the smoothness of the subsurface model along the three orthogonal directions [21], [23]. $\mathbf{D}_j\mathbf{m}$ yields an approximation to a given derivative, dependent on j and the intended order of the derivative. For $p_j = 1$, $\phi_D(\mathbf{m})$ is an anisotropic total variation stabilizer that provides a *blocky* model. It is also possible to replace the sum in $\phi_D(\mathbf{m})$ by a single term to yield an isotropic gradient constraint [41, eqs. (5)–(7)]. Also, it is possible to use only $\phi_{\text{S}}(\mathbf{m})$ to provide a sparse solution, but, as our results will demonstrate, the use of the mixed L_p-norm provides a result that is more robustly sparse, as discussed in [11] for the case of inverting a single dataset.

While the direct solution of (4 for general $p_s, p_j \neq 2$ poses challenges, a widely used and well-known strategy yields an iterative approximation for L_p-norm of a vector $\mathbf{x}$, $\|\mathbf{x}\|_p$, using an L_2-norm[3] [10], [11], [20], [42]

$$\|\mathbf{x}\|_p \approx \|\mathbf{W}\mathbf{x}\|_2^2, \quad \mathbf{W} = \text{diag}\left(\left(\mathbf{x}^2 + \epsilon^2\right)^{\frac{p-2}{4}}\right). \tag{5}$$

[2] $p = 0$ is not a norm in the mathematical sense but can still be used to generate solutions that are compact in that it gives the total number of nonzero elements for the given model.

[3] Note that, here and throughout as helpful, we give the form for $\mathbf{W}$ by using the inverse operation $1/a^{-c} = a^c$. Furthermore, we use the notation that the product of vectors refers to componentwise products.

Applying the method of iteratively reweighted least squares (IRLS), the solution of (2) is obtained by minimizing the objective function that is solved at each iteration k

$$\begin{aligned}\Phi^{(k)}(\mathbf{m}) = &\|\mathbf{W}_{\mathbf{d}^{\text{obs}}}\left(\mathbf{d}^{\text{obs}} - \mathbf{G}\mathbf{m}\right)\|_2^2\\ &+ \beta^2\alpha_s\|\mathbf{W}_{\text{dh}}\mathbf{W}_s^{(k)}\left(\mathbf{m} - \mathbf{m}^{(k-1)}\right)\|_2^2\\ &+ \beta^2\sum_{j=\text{x,y,z}}\alpha_j\ \|\mathbf{W}_{\text{dh}}\mathbf{W}_{\mathbf{D}_j}^{(k)}\mathbf{D}_j\mathbf{m}\|_2^2.\end{aligned} \tag{6}$$

This depends on weighting parameters $\alpha_s \ge 0$ and $\alpha_j \ge 0$.

The nonlinear diagonal weighting matrices that define the smallness and gradient terms are obtained from (5) and are given by

$$\mathbf{W}_s^{(k)} = \text{diag}\left(\left(\left(\mathbf{m}^{(k-1)} - \mathbf{m}^{(k-2)}\right)^2 + \epsilon_s^2\right)^{\frac{p_s-2}{4}}\right) \tag{7}$$

and

$$\mathbf{W}_{\mathbf{D}_j}^{(k)} = \text{diag}\left(\left(\left(\mathbf{D}_j\mathbf{m}^{(k-1)}\right)^2 + \epsilon_j^2\right)^{\frac{p_j-2}{4}}\right), \quad j = \text{x, y, z.} \tag{8}$$

The safety parameters $0 < \epsilon_s, \epsilon_j \ll 1$ are introduced to assure that the given matrices are nonsingular. The matrix $\mathbf{W}_{\text{dh}} = \mathbf{W}_{\text{hard}}\mathbf{W}_{\text{depth}}$ is a constant diagonal weighting matrix. Matrix $\mathbf{W}_{\text{hard}}$ is used to introduce hard constraints into the model. If some of the model parameters are already known, perhaps from drill-hole data or prior geological information, then the respective diagonal entries are set to large values for weighting of the known model parameters, and the relevant entries of $\mathbf{m}^{\text{apr}}$ are assigned to these known values. Otherwise, the diagonal entries in $\mathbf{W}_{\text{hard}}$ are set to 1. Thus, the known model parameters are kept fixed during the iterative minimization [5]. The diagonal depth weighting matrix $\mathbf{W}_{\text{depth}} = (z + z_{\text{av}})^{-(\nu/2)}$ is used to counteract the rapid decay of the kernels with depth [23]. Parameter z is the mean depth of the cuboid, and ν is an adjustable parameter that is determined in such a way that the resulting weighting matrix captures the decaying behavior of the model kernel.

For potential field inversion, it is also important to have knowledge of approximate upper and lower bounds for the physical properties that can be imposed iteratively. These bounds are used to maintain realistic solutions through the IRLS inversion algorithm and provide another way by which prior geophysical or geological knowledge is imposed. At each iteration, if an estimated model parameter falls outside the predefined bounds, it will be returned back to the nearest bound [5]. Imposing suitable and correct bounds is very relevant when sparsity regularization is used [19], [32].

C. Joint Inversion of Two Datasets

We rewrite the mixed regularization term that occurs in (6) as

$$\begin{aligned}\phi_{\text{reg}}^{(k)}(\mathbf{m}) \approx &\alpha_s\|\mathbf{W}_{\mathbf{m}}^{(k)}\left(\mathbf{m} - \mathbf{m}^{(k-1)}\right)\|_2^2\\ &+ \sum_{j=\text{x,y,z}}\alpha_j\ \|\mathbf{W}_{\mathbf{D}_j\mathbf{m}}^{(k)}\mathbf{m}\|_2^2\end{aligned} \tag{9}$$

where

$$\mathbf{W}_{\mathbf{m}}^{(k)} = \mathbf{W}_{\text{dh}}\mathbf{W}_s^{(k)} \quad \text{and} \quad \mathbf{W}_{\mathbf{D}_j\mathbf{m}}^{(k)} = \mathbf{W}_{\text{dh}}\mathbf{W}_{\mathbf{D}_j}^{(k)}\mathbf{D}_j. \tag{10}$$

In each case, the iteration matrices are calculated using (7) and (8) with the appropriate choice of $\mathbf{m}_1$, or $\mathbf{m}_2$, and the depth matrices and hard constraint matrices are chosen dependent on the data, and a choice of ν where generally $\nu_1 \neq \nu_2$. Safety parameters ϵ_s and ϵ_j are replaced by distinct ϵ_{s1}, ϵ_{s2} and ϵ_{j1}, ϵ_{j2}, respectively, as needed to tailor the tolerances to the model parameters. Moreover, all weights can be adjusted for each model. This yields a joint, but uncoupled formulation for the gravity and magnetic models

$$\begin{aligned}\Phi(\mathbf{m}) &= \left(\gamma_1^2\ \phi_{\text{datafit}}(\mathbf{m}_1) + \beta_1^2\ \phi_{\text{reg}}(\mathbf{m}_1)\right) \\ &\quad + \left(\gamma_2^2\ \phi_{\text{datafit}}(\mathbf{m}_2) + \beta_2^2\ \phi_{\text{reg}}(\mathbf{m}_2)\right) \\ &= \Phi_1(\mathbf{m}_1) + \Phi_2(\mathbf{m}_2). \end{aligned} \tag{11}$$

In our implementation, the coupled joint problem is obtained by imposing a correlation between the models. Equation (11) is augmented by the nonlinear cross-gradient term, $\|\mathbf{t}\|_2^2$, yielding

$$\Phi_{\text{Coupled}}(\mathbf{m}) = \Phi_1(\mathbf{m}_1) + \Phi_2(\mathbf{m}_2) + \lambda^2\|\mathbf{t}\|_2^2 \tag{12}$$

where $\lambda > 0$ is a Lagrange parameter

$$\mathbf{t} = \nabla\mathbf{m}_1(x, y, z) \times \nabla\mathbf{m}_2(x, y, z) \in \mathcal{R}^{3n} \tag{13}$$

and ∇ indicates the gradient operator. Structural similarity is achieved when $\mathbf{t} = 0$ [13]. To find $\mathbf{m}_1$ and $\mathbf{m}_2$ that achieve a minimum for Φ_{Coupled} in (12), we follow the approach of Gallardo and Meju [13], [14] in which a first-order Taylor expansion is applied to linearize the cross-gradient constraint. The linearized version of Φ_{Coupled} is

$$\Phi_{\text{Coupled}}(\mathbf{m}) = \Phi_1(\mathbf{m}_1) + \Phi_2(\mathbf{m}_2) + \lambda^2\|\mathbf{t} + \mathbf{B}\left(\mathbf{m} - \mathbf{m}^{\text{apr}}\right)\|_2^2 \tag{14}$$

where $\mathbf{B}$ is the Jacobian matrix of the discrete approximation for the cross-gradient function

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{1\text{x}} & \mathbf{B}_{2\text{x}} \\ \mathbf{B}_{1\text{y}} & \mathbf{B}_{2\text{y}} \\ \mathbf{B}_{1\text{z}} & \mathbf{B}_{2\text{z}} \end{pmatrix} = (\mathbf{B}_1, \mathbf{B}_2) \in \mathcal{R}^{3n\times 2n}. \tag{15}$$

Block matrices $\mathbf{B}_{ij}$, $i = 1, 2$ and $j = \text{x}, \text{y}, \text{z}$, are the components of $\mathbf{B}$ for gravity and magnetic problems, respectively, in the three orthogonal directions [12], [13], [14].

Taking $\nabla_{\mathbf{m}}\Phi_{\text{Coupled}}(\mathbf{m}) = \mathbf{0}$ defines an iterative scheme to update the model parameters as the solution of a linear symmetric positive definite (SPD) system

$$\mathbf{E}^{(k)}\mathbf{m}^{(k)} = \mathbf{f}^{(k)}, \quad k = 0, 1, \ldots \tag{16}$$

where the terms are updated for each iteration k. The system is defined by

$$\begin{aligned}\mathbf{E}^{(k)} &= \gamma^2(\mathbf{W}_{\mathbf{d}^{\text{obs}}}\mathbf{G})^\top\mathbf{W}_{\mathbf{d}^{\text{obs}}}\mathbf{G} + \lambda^2\left(\mathbf{B}^{(k)}\right)^\top\mathbf{B}^{(k)} \\ &\quad + \beta^2\left(\alpha_s\left(\mathbf{W}_{\mathbf{m}}^{(k)}\right)^\top\mathbf{W}_{\mathbf{m}}^{(k)} + \sum_{j=\text{x},\text{y},\text{z}}\alpha_j\left(\mathbf{W}_{\mathbf{D}_j\mathbf{m}}^{(k)}\right)^\top\mathbf{W}_{\mathbf{D}_j\mathbf{m}}^{(k)}\right)\end{aligned} \tag{17}$$

and

$$\begin{aligned}\mathbf{f}^{(k)} &= \gamma^2(\mathbf{W}_{\mathbf{d}^{\text{obs}}}\mathbf{G})^\top\mathbf{W}_{\mathbf{d}^{\text{obs}}}\mathbf{d}^{\text{obs}} + \lambda^2\left(\mathbf{B}^{(k)}\right)^\top\mathbf{B}^{(k)}\mathbf{m}^{(k-1)} \\ &\quad - \lambda^2\left(\mathbf{B}^{(k)}\right)^\top\mathbf{t}^{(k)} + \beta^2\left(\alpha_s\left(\mathbf{W}_{\mathbf{m}}^{(k)}\right)^\top\mathbf{W}_{\mathbf{m}}^{(k)}\right)\mathbf{m}^{(k-1)}.\end{aligned} \tag{18}$$

The coupling of the model parameters arises because

$$\mathbf{B}^\top\mathbf{B} = \begin{pmatrix} \mathbf{B}_1^\top\mathbf{B}_1 & \mathbf{B}_1^\top\mathbf{B}_2 \\ \mathbf{B}_2^\top\mathbf{B}_1 & \mathbf{B}_2^\top\mathbf{B}_2 \end{pmatrix} \tag{19}$$

does not have a block diagonal form. On the other hand, if we ignore the off-diagonal entries, $\mathbf{B}_1^\top\mathbf{B}_2$ and $\mathbf{B}_2^\top\mathbf{B}_1$, we obtain the two independent systems of equations equivalent to (16) for each of $\mathbf{m}_1$ and $\mathbf{m}_2$. The system is described by

$$\begin{aligned}\mathbf{E}_i^{(k)} &= \gamma_i^2\mathbf{G}_i^\top\mathbf{W}_{\mathbf{d}_i^{\text{obs}}}^\top\mathbf{W}_{\mathbf{d}_i^{\text{obs}}}\mathbf{G}_i + \lambda_i^2\left(\mathbf{B}_i^{(k)}\right)^\top\mathbf{B}_i^{(k)} \\ &\quad + \beta_i^2\alpha_s\left(\mathbf{W}_{\mathbf{m}_i}^{(k)}\right)^\top\left(\mathbf{W}_{\mathbf{m}_i}^{(k)}\right) \\ &\quad + \beta_i^2\sum_{j=\text{x},\text{y},\text{z}}\alpha_j\left(\mathbf{W}_{\mathbf{D}_j\mathbf{m}_i}^{(k)}\right)^\top\mathbf{W}_{\mathbf{D}_j\mathbf{m}_i}^{(k)}\end{aligned} \tag{20}$$

and

$$\begin{aligned}\mathbf{f}_i^{(k)} &= \gamma_i^2\mathbf{G}_i^\top\mathbf{W}_{\mathbf{d}_i^{\text{obs}}}^\top\mathbf{W}_{\mathbf{d}_i^{\text{obs}}}\mathbf{d}_i^{\text{obs}} + \lambda_i^2\left(\mathbf{B}_i^{(k)}\right)^\top\mathbf{B}_i^{(k)}\mathbf{m}_i^{(k-1)} \\ &\quad - \lambda_i^2\left(\mathbf{B}_i^{(k)}\right)^\top\mathbf{t}^{(k)} + \beta_i^2\left(\alpha_s\ \left(\mathbf{W}_{\mathbf{m}_i}^{(k)}\right)^\top\mathbf{W}_{\mathbf{m}_i}^{(k)}\right)\mathbf{m}_i^{(k-1)}.\end{aligned} \tag{21}$$

The Jacobian matrix $\mathbf{B}_i$ and vector $\mathbf{t}$ are updated between the solution of the two systems. Equations (20) and (21) are of the same format as the full system that is of size $2n \times 2n$ but adapted for system i of size $n \times n$. It has been shown that significant computational savings are achieved for the solution of (17) and (18) without the anisotropic constraint [41]. While ignoring off-diagonal entries in (19) can introduce a small error into the final solution, computational savings in memory and time are, indeed, observed by adopting AD Algorithm 1 to solve (20) and (21).

D. Algorithmic Considerations for the Joint Inversion Objective Function With the Cross-Gradient Constraint

The AD algorithm, as given in Algorithm 1, is terminated when the χ^2 test is satisfied for each data fit term. This requires that

$$(\chi^2)^{(k)} = \|\mathbf{W}_{\mathbf{d}^{\text{obs}}}\left(\mathbf{d}^{\text{obs}} - \mathbf{G}\mathbf{m}^{(k)}\right)\|_2^2 \leq m + \sqrt{2m} \tag{22}$$

or equivalently

$$\omega_i^{(k)} = \frac{\left(\chi_i^2\right)^{(k)}}{m + \sqrt{2m}} \leq 1 \tag{23}$$

for both $\mathbf{m}_1$ and $\mathbf{m}_2$. As a safety check, the algorithm will also terminate if $k = \text{K}_{\max}$ is achieved before the χ^2 test is satisfied.

In (14) defined for each iteration k, the data misfit terms are weighted with parameters $\gamma_i > 0$, $i = 1, 2$. The motivation for weighting each misfit term is that it is typically difficult to balance the relative values of the two misfit terms [22]. The convergence of one data fit term is generally faster than the other. This means that, at convergence with a combined measure of the χ^2 fit given by (22), it is feasible that one dataset

will be oversmoothed due to overfitting, while the other may be noisier due to underfitting. Applying weighting parameters γ_i to balance the two data misfits can mitigate this problem and can impose that both misfits are met to their individual noise levels. This requires a strategy to determine the γ_i parameters dynamically, along with all the other regularization parameters α_s, α_i, β_i, and λ_i. This presents a greater challenge than the approach in [22] in which only the data misfit weightings are changed. We note that the parameters ν_i, ϵ_j, and ϵ_s, defined in Section II-B, also have important effects on the solution.

E. Estimating the Algorithm Parameters

In Algorithm 1, the parameters γ_i, α_s, α_j, β_i, and λ_i, $i = 1, 2$, determine the relative weights for the data misfits, stabilization terms, and cross-gradient constraints within the minimization of (14). Consequently, the choices for all of these parameters have a significant impact on the quality of the obtained results.

1) Relative Weighting Parameters on Regularization Terms $\beta_i^{(k)}$: We use the `cooling approach` adopted in [39]. For large initial $\beta_i^{(0)}$, β_i and $i = 1, 2$ are adjusted slowly according to

$$\beta_i^{(k+1)} = \begin{cases} \beta_i^{(k)} q_i, & \omega_i^{(k)} > 1 \\ \beta_i^{(k)}, & \texttt{otherwise}. \end{cases} \tag{24}$$

The decay rate parameters $0 < q_i \leq 1$ impact the rate at which the $\beta_i^{(k)}$ decays with iteration k, and the use of different starting values and decay rates allows an algorithm that can be adapted to the properties of the two model sensitivity matrices. Prior experience with these two models tells us that, while the magnetic model is better conditioned than the gravity model [39, Figs. 8 and 9, eq. (8), and tables], the largest singular value for $\mathbf{G}_2$ is much larger. Therefore, it is important that $\beta_2^{(0)} > \beta_1^{(0)}$. If we do not adjust these parameters appropriately to the largest singular values, the impact of filtering the components will not be equivalent across the two models. In view of this, it is also important that β_2 is not allowed to decrease as rapidly as β_1. In contrast, it is also possible to use variable parameters q_i that are small when the initial iterations indicate predicted data that are far from the observed data. Then, the data misfits are large, and q_i should be small to force faster convergence for the data misfit term.

2) Relative Weighting Parameters on Data Fit Terms $\gamma_i^{(k)}$: Initially, we take $\gamma_i^{(0)} = 1$. Suppose that, during the `cooling approach`, $\omega_i^{(k)} < 1$ for one dataset but not the other. At this point, parameter γ_i is adjusted as

$$\gamma_i^{(k+1)} = \frac{1}{1 + \left(1 - \left(\omega_i^{(k)}\right)^2\right)}, \quad \text{where} \quad \omega_i^{(k)} \leq 1. \tag{25}$$

This provides a weighting $\gamma_i \leq 1$, for the next iteration, which is proportional to the distance from the target misfit at the current iteration. The relative values of the two data misfit terms are adjusted to guarantee the approximate fitting of each dataset within its noise level.

3) Regularization Parameters α_s and α_j: When implemented in a true mathematical formulation, each term of the anisotropic regularizer in ϕ_{reg} depends on an approximation to the derivative of the model. The smallness term, in contrast, provides differences in the model from one iteration to the next and is not scaled by the grid size. Assuming that the derivative is order 1 and applied mathematically with the grid scaling included, then, with $p_s = p_j$, this means that we will have $\alpha_s \approx \alpha_j h_j^{-p_s}$, for $j = \text{x, y}$, where h_j is the dimension of the cell in the jth dimension. In the z dimension, however, the depth weighting directly impacts this approximation, and the relation cannot be uniformly applied at all depths. Taking the three terms together, and assuming a uniform grid size in each dimension, we can see that using $\alpha_s \approx \alpha_j h_j^{-p_s}$ would overweight the L_p contribution. Instead, it is reasonable to roughly average the contributions from the x, y directions and to use a much smaller contribution to the z direction so that solutions that are inappropriately concentrated at depth are not obtained. Using $\alpha_s = 1$, this yields choice $\alpha_j \approx 0.4 h_j^{p_j}$ for the east and north dimensions with the reduced contribution $\alpha_z \approx 0.01\alpha_x$.

When $p_s \neq p_j$, the relation between the various terms also depends on the anticipated size of the model parameter. For example, when $p_j = 2$ but $p_s = 1$, we find $\alpha_j \boldsymbol{\delta}\mathbf{m} \approx \alpha_s h_j^2$ for each cell. However, based on the bound constraints applied for our experiments, we expect $\mathbf{m}_1 \approx 10$, and we would expect the change $\boldsymbol{\delta}\mathbf{m}$ to scale similarly, which yields $\alpha_j^2 \approx 10\alpha_j^1$, where $\alpha_j^1 \propto (h_j^2)$. On the other hand, when $p_j = 0$ but $p_s = 1$, we obtain $\alpha_j \approx \alpha_s \boldsymbol{\delta}\mathbf{m}$, which is independent of h_j with $\alpha_j < \alpha_s$, and $\alpha_j^1 \approx 10\alpha_j^2$. Notice that, at convergence for $\mathbf{m}$, we expect that $\boldsymbol{\delta}\mathbf{m}$ to be small. These estimates guide the initial choices that are used in the experiments, which are then refined to provide iterations that converge.

We note that, for nonuniform cuboids, different grid sizes h_j, $j = \text{x, y, and z}$, for the volume discretization, the arguments need to be modified appropriately. Allowing the use of separate parameters α_s and α_j for all of the contributions to the mixed regularization terms permits differential weighting for the anisotropic derivatives and the smallness component.

4) Safety Parameters ϵ_1, ϵ_2, ϵ_{s1}, and ϵ_{s2}: It has been already demonstrated in the literature that these are small positive parameters that are chosen to assure that the weight matrices are not singular. When the L_0-norm is imposed, it is standard to use a slightly larger parameter [35], [44].

5) Lagrange Parameters λ_i: It is clear that the limiting case $\lambda_i = 0$ provides solutions without a structural similarity constraint, and λ_i should be large in order to impose structural similarity. However, in considering the size of λ_i, it is important to consider that it weights the term $\|\mathbf{t}+\mathbf{B}_i(\mathbf{m}_i - \mathbf{m}_i^{(k-1)})\|_2^2$. Again, assuming that the derivatives are calculated directly, and then by the definitions of $\mathbf{t}$ and $\mathbf{B}$, we can see that each of these has entries that are inversely proportional to the second power of the grid size in each dimension, h_j^{-2}, equivalent to h_j^{-4} in the two-norm. Hence, if h_j is relatively large, the impact of the cross-gradient term will not be significant unless λ_i is taken large enough to balance against the regularization and data fit terms. However, if λ_i is too large, it is also possible that the data fit χ^2 measure is not satisfied, and the algorithm

TABLE I
DATASET PARAMETERS FOR SIMULATED AND REAL DATA

Test	(m, n)	$\mathrm{SNR}_1, \mathrm{SNR}_2$	$\mathbf{d}_1^{\mathrm{obs}}$: τ_1, τ_2	$\mathbf{d}_2^{\mathrm{obs}}$: τ_1, τ_2
1-6	$(800, 8000)$	$26.72, 26.28$	$0.02, 0.01$	$0.01, 0.01$
7	$(5000, 40000)$	$25.83, 25.56$	$0.02, 0.01$	$0.01, 0.01$
8	$(884, 25600)$	NA	$0.01, 0.01$	$0.01, 0.03$

TABLE II
ALGORITHM PARAMETER CHOICES COMMON TO ALL REPORTED RESULTS. ϵ_{s1} AND ϵ_{s2} ARE THE SAFETY PARAMETERS IN THE SMALLNESS REGULARIZATION WEIGHTING MATRIX [SEE (7)]

(β_1, β_2)	(q_1, q_2)	(ν_1, ν_2)	$(\epsilon_{s1}, \epsilon_{s2})$	$\mathrm{K}_{\max}$
(2e4, 4e4)	(.90, .95)	(1.6, 2.8)	(1e-9, 1e-10)	200

does not converge within a reasonable number of iterations. Lack of convergence can be seen in increasing or oscillating values of $\omega_i^{(k)}$.

It can take more than one run of the algorithm to find the suitable compromise value for λ_i that is neither too small nor too large. We follow the strategy in [39] in which we initially use a relatively small value for λ_i. If the algorithm converges quickly but without achieving a good similarity, λ_i is increased. The smallest values of λ_i that provides similarity between the reconstructed models are selected. When the algorithm is implemented for the full system without the AD, the same approach is used for $\lambda = \lambda_i$.

6) Impact of Relative Weightings: Large values of β_i give more weight to the importance of minimizing the regularization terms and less weight to the data fit. This can lead to solutions that are oversmoothed, while the data misfit criterion is not satisfied. Small values for β_i will likely yield models that satisfy the data misfit but are noise-contaminated and not stabilized. In contrast, small values of γ_i give less weight to the data misfits. In our algorithm, γ_i and β_i are iteration dependent, but they do not change simultaneously. On the other hand, λ_i is taken to be constant, and the impact of the cross-gradient term gradually increases as $\beta_i^{(k)}$ decreases. An alternative balancing of the terms is also possible [11].

III. EXPERIMENTAL VERIFICATION OF THE MIXED AND BALANCED JOINT INVERSION WITH CROSS-GRADIENT COUPLING

The presented formulation is validated for synthetic examples discussed in Sections III-A–III-C and a practical dataset described in Section IV. For the simulations, we assume: 1) the density contrast and magnetic susceptibility of the subsurface structures are $\rho = 0.6$ g cm^{-3} and $\kappa = 0.06$ (SI unit), respectively; 2) the bodies are embedded in a homogeneous nonsusceptible background medium; and 3) for all implementations, $\mathbf{m}_i^{\mathrm{apr}} = 0$. The experiments are performed using diagonal matrix $\mathbf{W}_{\mathbf{d}^{\mathrm{obs}}}$ for the noise added to the true data using ℓ^{th} entry $\tau_1 \, |(\mathbf{d}_i^{\mathrm{obs}})_\ell| + \tau_2 \, \max|\mathbf{d}_i^{\mathrm{obs}}|$, $i = 1, 2$ and $\ell = 1, \ldots, m$. The pairs (τ_1, τ_2) for the experiments and resulting SNR in each case are given in Table I. For the experiments, we evaluate the performance of the algorithm in terms of the weighted χ^2 measures (23), the required number of iterations, and, when available, the relative errors of the reconstructed models given by

$$\mathrm{RE_i} = \frac{||\mathbf{m}_\mathrm{i}^{\mathrm{exact}} - \mathbf{m}_\mathrm{i}^{(\mathrm{K})}||_2}{||\mathbf{m}_\mathrm{i}^{\mathrm{exact}}||_2}, \quad \mathrm{i} = 1, \, 2 \tag{26}$$

where $\mathbf{m}^{\mathrm{exact}}$ is the exact model and $\mathbf{m}^{(K)}$ is the reconstructed model at the final iteration $K \leq \mathrm{K}_{\max}$.

In all cases, there are a number of algorithm parameters that need to be chosen. The parameters that are fixed for all experiments are given in Table II. Parameters used to test the algorithm with respect to the impact on the resulting characteristics of the models and the convergence properties of the algorithm are given in Table III. Upper and lower bounds, $\rho_{\min} \leq \mathbf{m}_1 \leq \rho_{\max}$ (g cm^{-3}) and $\kappa_{\min} \leq \mathbf{m}_2 \leq \kappa_{\max}$ (SI unit), are imposed at all iterations of the inversion algorithm. Different choices for p_s and p_j are assessed for the joint inversion algorithm, extending the approach for independent mixed L_p-norm inversions given in [11] and [48]. Moreover, we aim to demonstrate that the AD strategy, which has already been seen to be successful for the single L_p-norm inversion, is also valid within the mixed formulation for providing a robust and computationally efficient inversion of two datasets.

For all implementations, we use an iMac 2017 computer with a 4.2-GHz Quad-Core Intel Core i7 processor and 32-GB RAM. The results for all experiments are given in Table IV, with timings obtained for the same problems also given in parentheses for implementation on a MacBook Pro with M1 Max processor and 32-GB RAM. We note that the timings and number of iterations may not be the same across each environment. It is the comparison of robustness across environments and the relative increase in cost for the larger problems that are most important in assessing that the algorithm is robust for multiple environments and problems.

A. Simulations

The first example is a small but complicated model consisting of two dipping dikes for which it is feasible in terms of computational time to evaluate the algorithm for multiple and different mixed L_p-norm components. It is also possible to validate the AD algorithm versus the full system implementation. These tests for the dipping dyke are referred to as Tests 1–6 in Table IV. It is also feasible in this case to compare with the standard independent inversion for the inversion of each dataset. Results using the mixed L_p-norm stabilization without balancing, as described in (2), $\gamma_1 = \gamma_2 = 1$ and without the cross-coupling constraint $\lambda_1 = \lambda_2 = 0$, are also presented. We note also that tests with a different configuration for a dipping dike using the cross-gradient constraint, without balancing or mixed terms, were given in [40]. Results were chosen to also demonstrate the impact of the cross-gradient when one structure is not magnetic; in this case, a vertical dike is not magnetic and demonstrated that the cross-gradient coupling does not force the existence of the vertical structure for the magnetic domain [40, Figs. 16 and 18].

The second example, Test 7, is a larger model consisting of multiple bodies that are designed to validate the algorithm

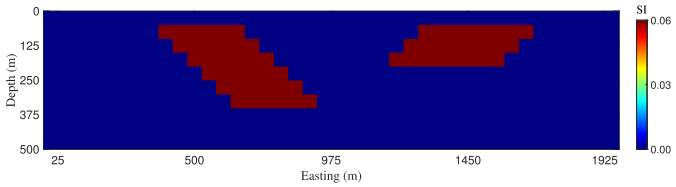

Fig. 1. Cross section of the synthetic model consisting of two dipping dikes that are used in Tests 1–6. The dikes have a value of 0.6g/cm^3 for the density distribution and 0.06 in SI units for the magnetic susceptibility distribution. We illustrate the structure of the magnetic susceptibility distribution.

for larger problems. The application of the hard constraint matrix is demonstrated in this example. In all cases, the chosen parameters are given in Tables II and III and the results in Table IV.

The experiments presented in Section III-B are chosen to demonstrate that the joint inversion algorithm can be used to generate many different models all of which satisfy the observed data to the given noise level. Thus, for practical applications, it is feasible to use different stabilizing terms, dependent on available a priori information as to what may be desired in the model. The practicality of the solution to a large problem is demonstrated in Section III-C.

B. Two-Dipping-Dike Model

Fig. 1 illustrates the cross section of the model with two dipping dikes, each with a top depth 50 m, but with extensions in depth for the left and right dikes of 350 and 200 m, respectively. The gravity and magnetic data for the model are generated on the surface on a 40 × 20 uniform grid with a grid spacing of 50 m. The noisy datasets are illustrated in Fig. 2. The subsurface volume is discretized into 40 × 20 × 10 = 8000 cuboids of size 50 m, for a model that extends to depth 500 m.

*1) Test 1 (*AD *Algorithm for Mixed-L_p Inversion, $p_x = p_y = p_z = 2$ and $p_s = 1$):* It has been suggested that these choices are ideal, especially for the reconstruction of a dipping structure [48]. Convergence is achieved in about 5–6 min. Structurally, the reconstructed density and magnetic susceptibility models illustrated in Fig. 3 are similar. The upper depths and dip of the structures are approximately reconstructed, but the extensions in depth are not consistent with the original synthetic models. Parameters ω_i at the final iteration indicate that the gravity solution is close to the target value, but the noise in the magnetic solution is further reduced. The algorithm was also implemented for the same parameters as for Test 1 (indicated as Test 1a) but without balancing, i.e., $\gamma_i = 1$, for all iterations. The algorithm terminates at the same number of iterations, but ω_2 is further reduced to 0.71, which suggests overfitting of the magnetic solutions and validates the use of balancing in the algorithm. Furthermore, the relative errors using the joint inversion are smaller than those achieved by the independent inversions, which supports the use of the joint inversion. Indeed, the independent inversion incorrectly predicts increased susceptibility at greater depth.

*2) Test 2 (*FULL *Algorithm for Mixed-L_p Inversion, $p_x = p_y = p_z = 2$ and $p_s = 1$):* The number of iterations to convergence and the relative errors are comparable to Test 1, but the cost increases fourfold to 1785 s. The reconstructed models illustrated in Fig. 4 cannot be determined to be any improvement on those in Fig. 3. Given the commensurate increase in computational cost without apparent improved solutions, hereafter, we will use the AD strategy in the presented simulations.

3) Test 3 (L_1-Norm for All Terms, $p_s = p_x = p_y = p_z = 1$): This implementation should yield a *blocky* structure. The algorithm termination is comparable to Test 1 with reduced relative errors that are less than those achieved using independent inversion. In each case, we see that $\omega_2 < \omega_1$. From the cross sections in Fig. 5, we see that the magnetic susceptibility model is a slightly better representation of dipping dikes compared to the density model. This is reflected also in the relative errors given in Table IV. The results demonstrate that the arguments given in Section II-E for selecting the regularization parameters α_s and α_j are appropriate and are equally valid for the independent inversions. Again, in comparison to the joint inversion, the independent inversion incorrectly predicts variation in the susceptibility at greater depth.

4) Test 4 (Mixed L_p, $p_s = 1$, and $p_x = p_y = p_z = 0$): The L_0-norm applied to the gradient terms should provide a *blocky* solution. We have observed that, when an L_0-norm is applied, the solution is more sensitive to both the weighting parameters and the choice of ϵ_s and ϵ_j used within the iterative reweighting matrices. This occurs due to the use of the second root in (7) and (8) rather than the fourth root, which is required for L_1-norm components. Generally, a good set of parameters allows the algorithm to converge within a reasonable number of iterations. Hence, this sensitivity becomes apparent if the algorithm is not converging. In such cases, it is sufficient to run the algorithm under several configurations of the parameter set and to, thus, find suitable parameters that provide a good convergence behavior. From the density and magnetic susceptibility models shown in Fig. 6, we see that the results are quite *blocky*, particularly for the independent inversion, which overestimates the depth of the magnetic structure. From this result, we conclude that, although it is more challenging to find converged reconstructions using the L_0-norm, it is possible to reconstruct the dip of structure if suitable weighting parameters are selected. This contrasts with the suggestion in [48].

5) Test 5 (L_1-Norm for All Terms, $p_s = p_x = p_y = p_z = 1$, but Increased Upper Bounds $\rho_{\max} = 1.5$ and $\kappa_{\max} = 0.15$): This is the same configuration as Test 3 except for the upper bounds. The timing and required number of iterations are comparable to the Test 1 and 3 results, but the relative errors are higher for both joint and independent inversions. This is reflected in the cross sections illustrated in Fig. 7. In all cases, it is clear that the estimated parameters are too high at the centers of the target structures and are not consistent with the original model. The obtained reconstructions are not satisfactory. On the other hand, again, the joint algorithm does yield better results in terms of relative error estimates than the solutions obtained with independent inversion. Note that, if we apply the L_1- or L_0-norm of the *smallness* term without the anisotropic components (see [32]), the obtained models become much more focused due to the incorrect bounds.

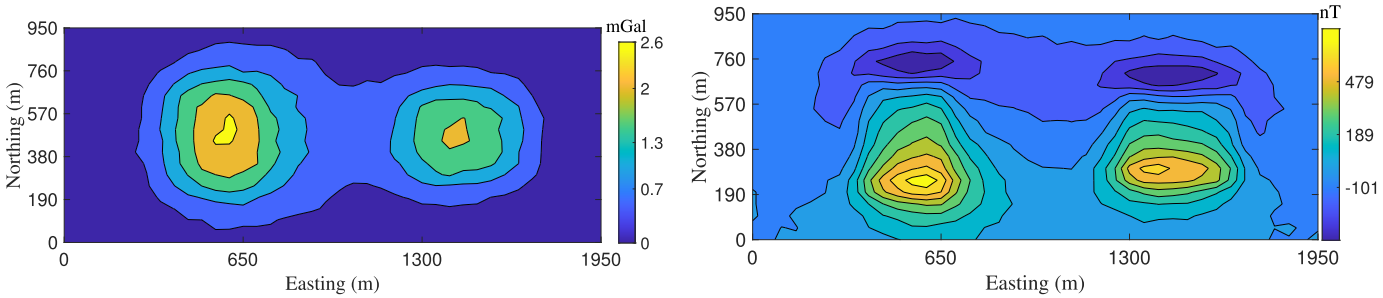

Fig. 2. Anomaly produced by the two dipping dikes shown in Fig. 1 and contaminated with noise. (a) Vertical components of the gravity field. (b) Total magnetic field.

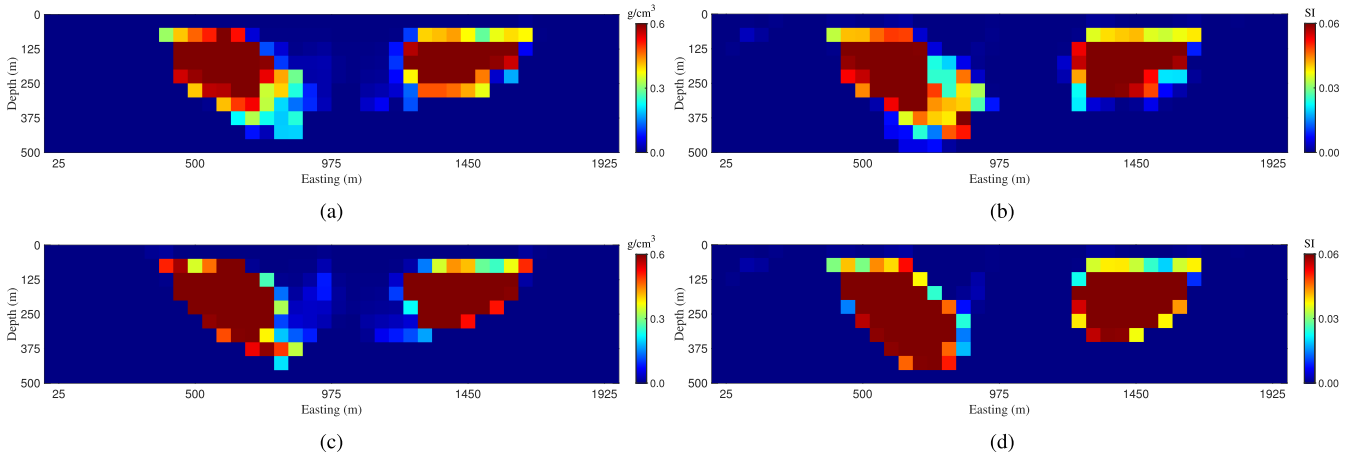

Fig. 3. Cross sections of the reconstructed models using the AD algorithm for Test 1 for the datasets shown in Fig. 2. The results show the joint inversion to give the density distribution and magnetic susceptibility distribution in (a) and (b) and the equivalent results for the independent inversions in (c) and (d), respectively.

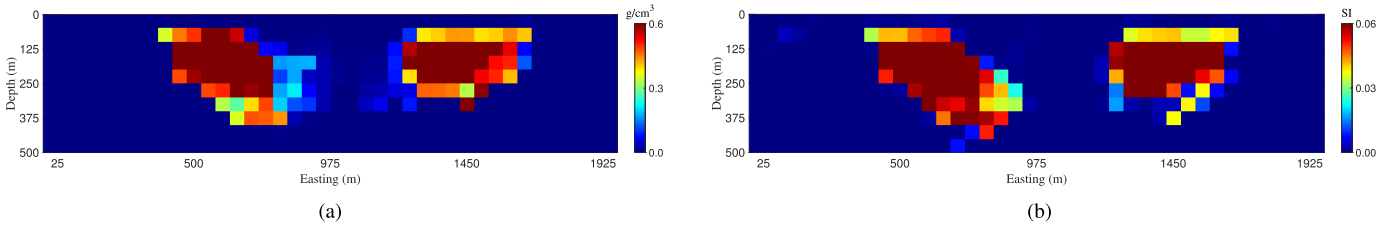

Fig. 4. Cross sections of the reconstructed models for the datasets shown in Fig. 2. (a) Density distribution. (b) Magnetic susceptibility distribution. The Full system is used with the parameters, as given in Tables II and III, for Test 2.

6) Test 6 (L_2-Norm for All Terms, $p_s = p_x = p_y = p_z = 2$): Here, the use of L_2-norms contrasts with the L_1-norms used for Test 5 and should provide a smooth model of the subsurface. The smoothness is immediately evident in Fig. 8. The algorithm requires more iterations to converge, but the algorithm, whether joint or independent, is very fast, indicating that the solution solves at each iteration are fast. In particular, for each outer iteration, two different linear systems are solved using the PCG algorithm to a specified tolerance and not with a set number of iterations for the PCG. We notice that the solutions, in this case, are more robust to the imposition of the incorrect bounds than the results using the L_1-norms. The solutions by independent inversion are also smooth but are less robust to the incorrect bounds.

7) Summary of Tests 1–6: Overall, we determine that: 1) the joint inversion outperforms the independent inversion, particularly in terms of depth resolution; 2) the AD algorithm is an effective approach that avoids the higher cost of the FULL algorithm; and 3) the use of the mixed formulation with L_1 in all cases does indeed provide solutions that are better at predicting the structures, particularly at depth.

C. Model of Multiple Bodies

Fig. 9 illustrates a model consisting of six subsurface targets with different shapes, dimensions, and depths, for which three depth sections are illustrated in Fig. 10. Gravity and magnetic data are generated on a grid with 100 × 50 points and a grid spacing of 100 m, and the volume is extended to the depth of 800 m, yielding 40 000 cuboids of size 100 m. The noise-contaminated data are illustrated in Fig. 11. For this model, the results of the inversion using the L_1-norm for

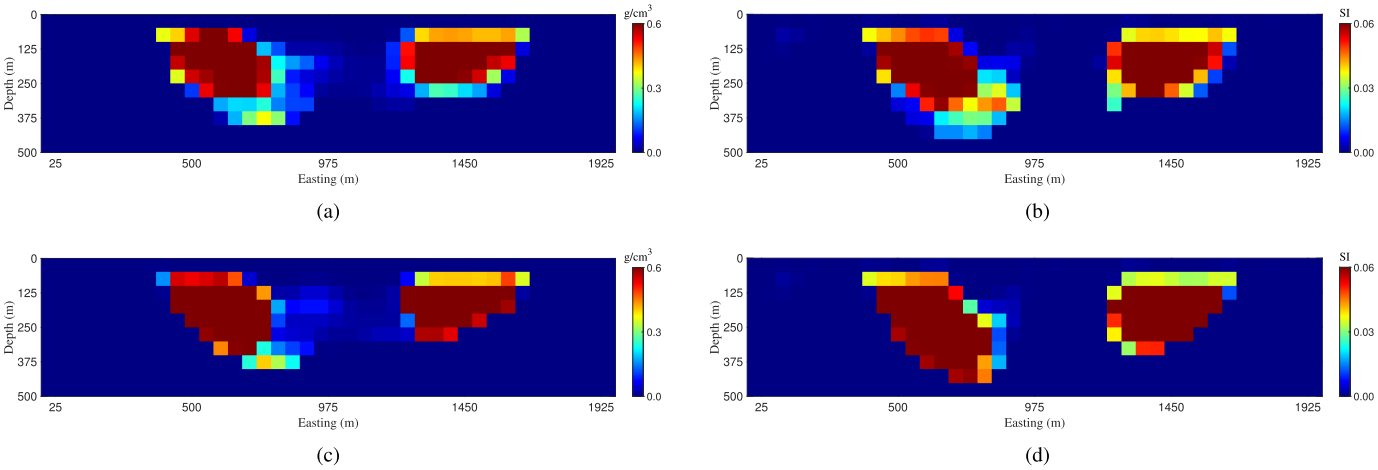

Fig. 5. Cross sections of the reconstructed models using the AD algorithm for Test 3 for the datasets shown in Fig. 2. The results show the joint inversion to give the density distribution and magnetic susceptibility distribution in (a) and (b) and the equivalent results for the independent inversions in (c) and (d), respectively.

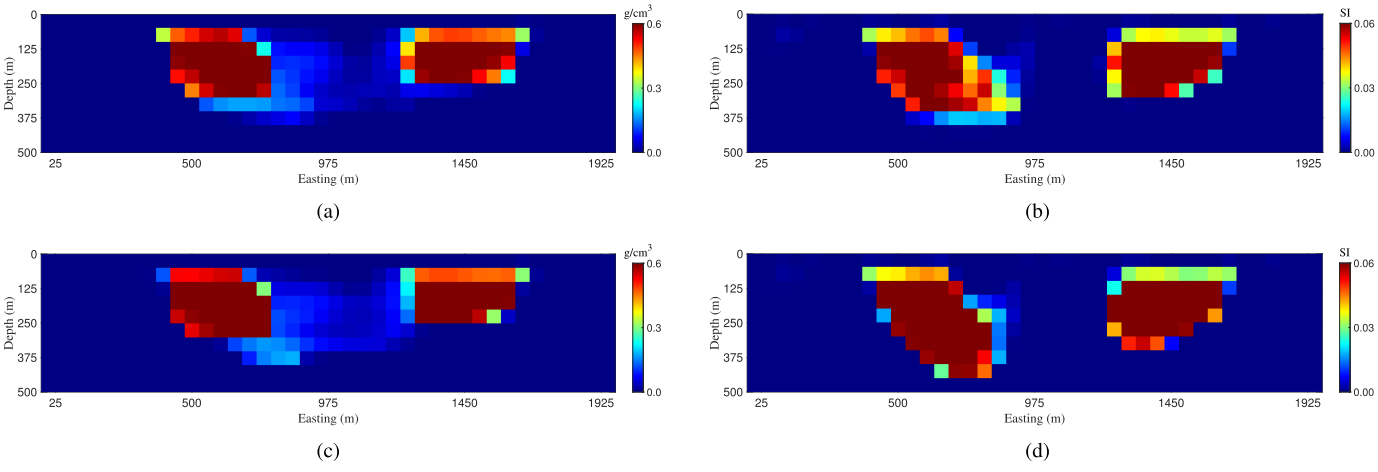

Fig. 6. Cross sections of the reconstructed models using the AD algorithm for Test 4 for the datasets shown in Fig. 2. The results show the joint inversion to give the density distribution and magnetic susceptibility distribution in (a) and (b) and the equivalent results for the independent inversions in (c) and (d), respectively.

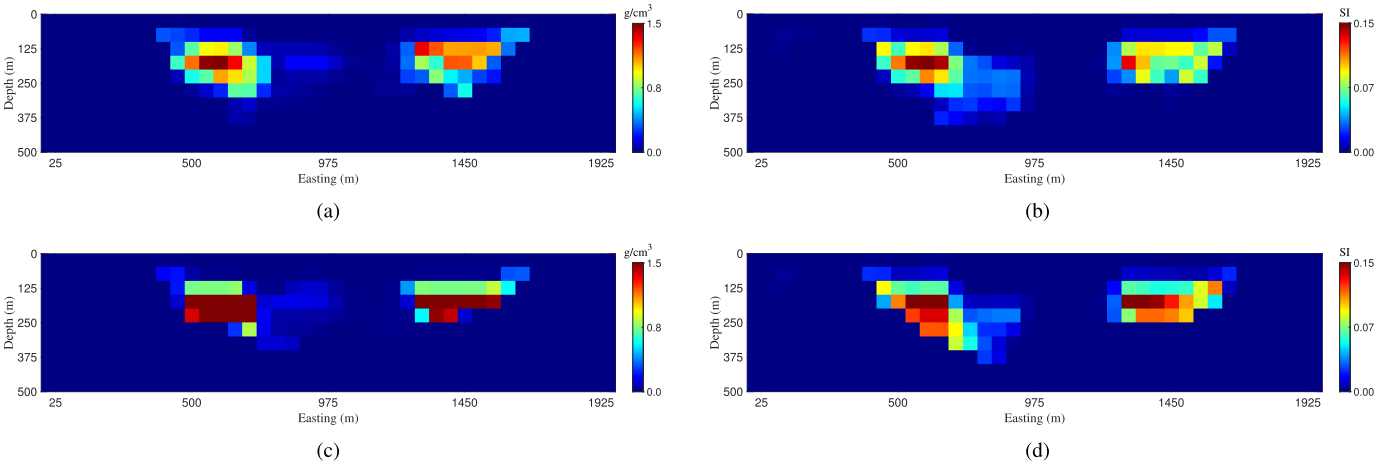

Fig. 7. Cross sections of the reconstructed models using the AD algorithm for Test 5 for the datasets shown in Fig. 2. The results show the joint inversion to give the density distribution and magnetic susceptibility distribution in (a) and (b) and the equivalent results for the independent inversions in (c) and (d), respectively.

the *smallness* term, without the anisotropic components, are presented in [41].

1) Test 7 (Large Model With Hard Constraint Matrices): Matrices $(\mathbf{W}_{\text{hard}})_1$ and $(\mathbf{W}_{\text{hard}})_2$ are tested by assuming that there is a priori information on the physical properties of four cuboids that are located in the southwest corner of the small cube in the second layer. This layer extends from 100 to 200 m. We also suppose that the cuboids above and below

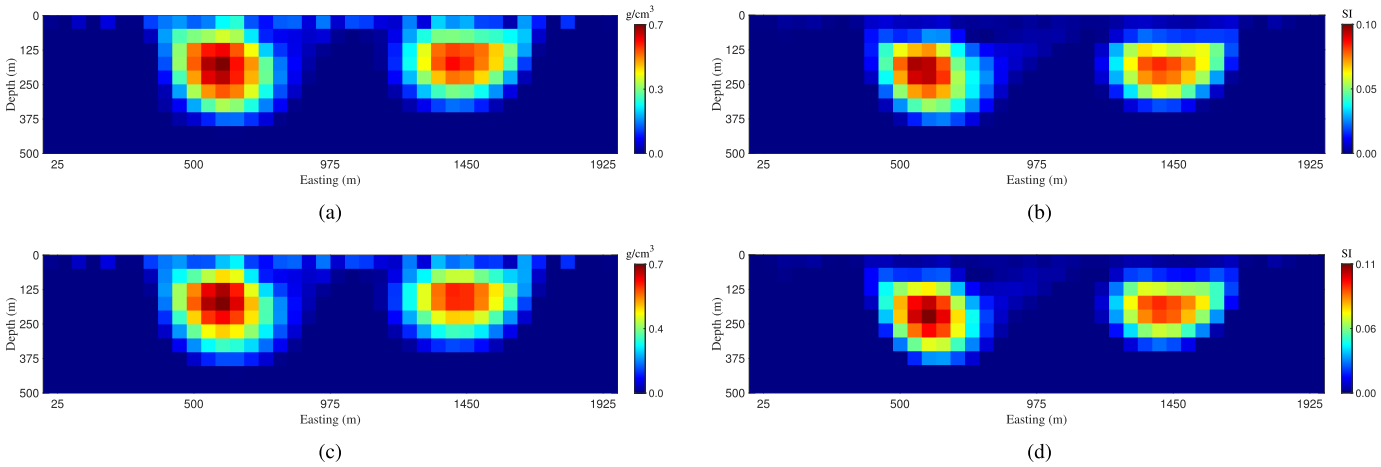

Fig. 8. Cross sections of the reconstructed models using the AD algorithm for Test 6 for the datasets shown in Fig. 2. The results show the joint inversion to give the density distribution and magnetic susceptibility distribution in (a) and (b) and the equivalent results for the independent inversions in (c) and (d), respectively.

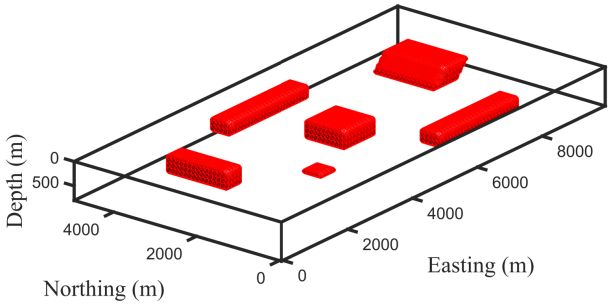

Fig. 9. Synthetic model consisting of six targets with different geometries and depths for Test 7.

these known cuboids, in the first and third layers, do not have any contrast with the background. These known density and magnetic susceptibility values are imposed in the initial model, which is otherwise 0, and by setting the relevant entries in $(\mathbf{W}_{\text{hard}})_1$ and $(\mathbf{W}_{\text{hard}})_2$ to 100.

From the results given in Table IV, we see that the magnetic problem converges in fewer iterations than the gravity problem. The time increases commensurate with the increased size of the problem but is still reasonable and less than an hour. From Figs. 12 and 13, it is evident that sharp and focused reconstructions for the subsurface targets are obtained. The models are in good agreement with the original model and have nearly similar structures. Although an incorrect extension in the depth direction is seen in other parts of the small cube, the cuboids with known values do not show any depth extension. This example indicates that it is possible to incorporate information into the joint inversion algorithm in order to increase the reliability of the reconstructed models. Furthermore, the results presented in Figs 12 and 13 are better than the results for the same model and a comparable SNR, as noted with only the L_1 constraint on the model parameters, which were presented in [41]. There, the results were given with and without the joint inversion with the cross-gradient coupling, as shown in Figs. 5 and 6 and 8 and 9, respectively. Compared with Figs 12 and 13, we see that we now achieve better resolution at depth 400 m using the cross-gradient coupling with balancing, particularly for the susceptibility, despite the incorrect extensions of the smaller structures. This validates using the hard constraint and the mixed and balanced stabilization for joint inversion.

IV. Real Data

The AD joint inversion algorithm is applied on ground gravity and magnetic data acquired over two kimberlite pipes in the Orapa kimberlite field (OKF) in north-central Botswana. The OKF is a well-known diamond-producing region and lies within the Kalahari basin, which trends northeast across central Botswana. The Kalahari basin is comprised of the Karoo Supergroup, which consists of mostly clastic sediments interbedded with basaltic layers. Underlying the Karoo Supergroup are Archean-aged lithologies that include biotite gneisses, amphibolites, and metamorphosed basalts (greenstones) [26]. Of importance to our study are basalt flows of the Stromberg Group that overlies most of the Karoo sedimentary sequences. Overlying all of the sedimentary and basaltic layers is a thin layer of calcrete and silcrete. More information on the geology of the region can be found in [26]. Recently, several small kimberlite pipes have been discovered in the region using geophysical methods. Two of these pipes, hereafter called BK54 and BK55, were found using ground gravity and magnetic surveys. Subsequent drilling confirmed the presence of the kimberlitic material [26]. Residual gravity and magnetic anomalies over BK54 and BK55 are illustrated in Fig. 14. The residual magnetic anomaly, clearly, has positive anomalies over the BK54 and BK55 pipes. On the other hand, the BK55 pipe is only characterized by a small amplitude gravity maximum, while the BK54 pipe does not produce a gravity maximum but has a small amplitude gravity minimum. This may be due to the crater facies containing less dense material or may also be due to a lack of density contrast between the BK54 pipe and the surrounding sandstones [26]. Based on available boreholes, the kimberlitic material was found at a depth between approximately 80–150 m in BK54 and 125–220 m for BK55 [26]. One borehole, drilled until

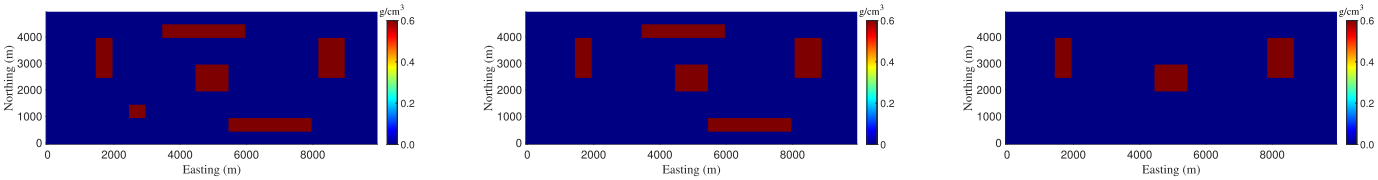

Fig. 10. Depth sections of the synthetic model shown in Fig. 9. Here, the density distribution of the model is presented at depths of (a) 100, (b) 200, and (c) 400 m.

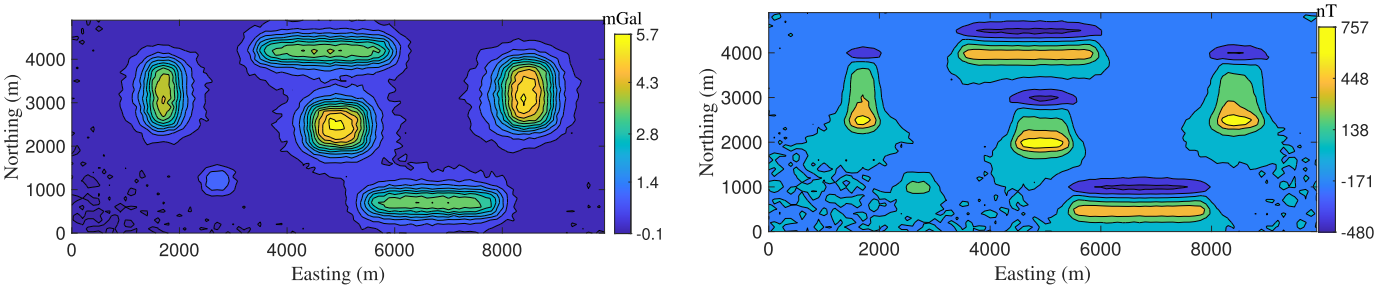

Fig. 11. Anomalies produced by the model shown in Fig. 9 and contaminated with noise. (a) Vertical components of the gravity field. (b) Total magnetic field.

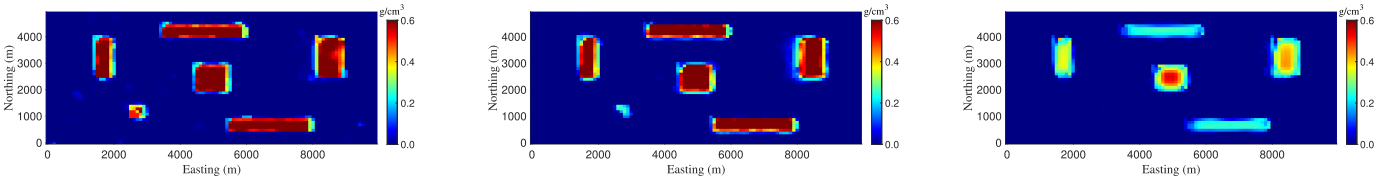

Fig. 12. Depth sections of the reconstructed density model using the joint inversion algorithm at depths of (a) 100, (b) 200, and (c) 400 m for the data shown in Fig. 11.

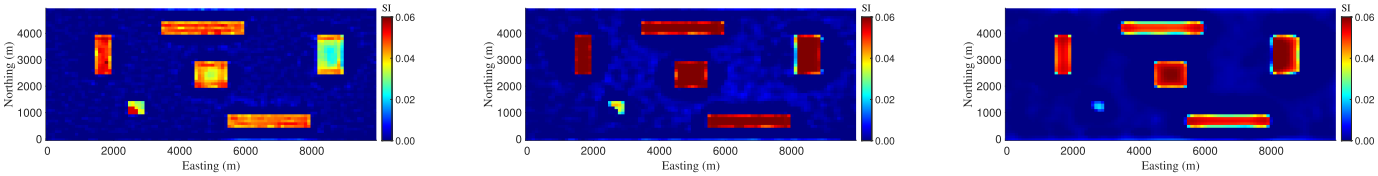

Fig. 13. Depth sections of the reconstructed magnetic susceptibility model using the joint inversion algorithm applied on the data shown in Fig. 11. The depth sections are presented at depths of (a) 100, (b) 200, and (c) 400 m.

depth 221 m, did not encounter the bottom of the BK55 pipe, which indicates that more extension is possible.

For the joint inversion of this data, we used the same setup for the inversion, as presented in [46]. Specifically, the gravity and magnetic anomalies were gridded on a uniform grid of size $26 \times 34 = 884$ data points with grid spacing 15 m. The subsurface volume is discretized with uniform cells of sizes 15 m in each dimension to a depth of 300 m corresponding to 20 depth layers. To avoid possible distortion in the reconstruction along the boundaries, the domain was padded to a width of 45 m in the east, west, south, and north directions in the grid domain. This yields a model with $32 \times 40 \times 20 = 25\,600$ cuboids. The lower and upper bounds for this case, as indicated in Table III, are obtained from [26]. Furthermore, the choice $p_s = p_x = p_y = p_z = 1$ was made with the goal to obtain a compact and focused reconstruction for the identification of kimberlite pipes [11]. The use of Algorithm 1 for the inversion of the data contrasts with the inversion using the cross-gradient coupling and L_1-norm constraint on each of the model parameters, as presented in [46].

Convergence was achieved at 78 iterations, requiring a clock time of 2286 s, considerably faster than the 6461 s for convergence at 86 iterations, for the inversion presented in [46]. The improved (reduced) cost is in part due to the use of a more powerful laptop for the current set of results, but note also that the problem is larger due to the use of the anisotropic norm constraint and the L_1-norm on the model parameters. Compared to the synthetic studies, the timing is consistent; the synthetic multiple-body

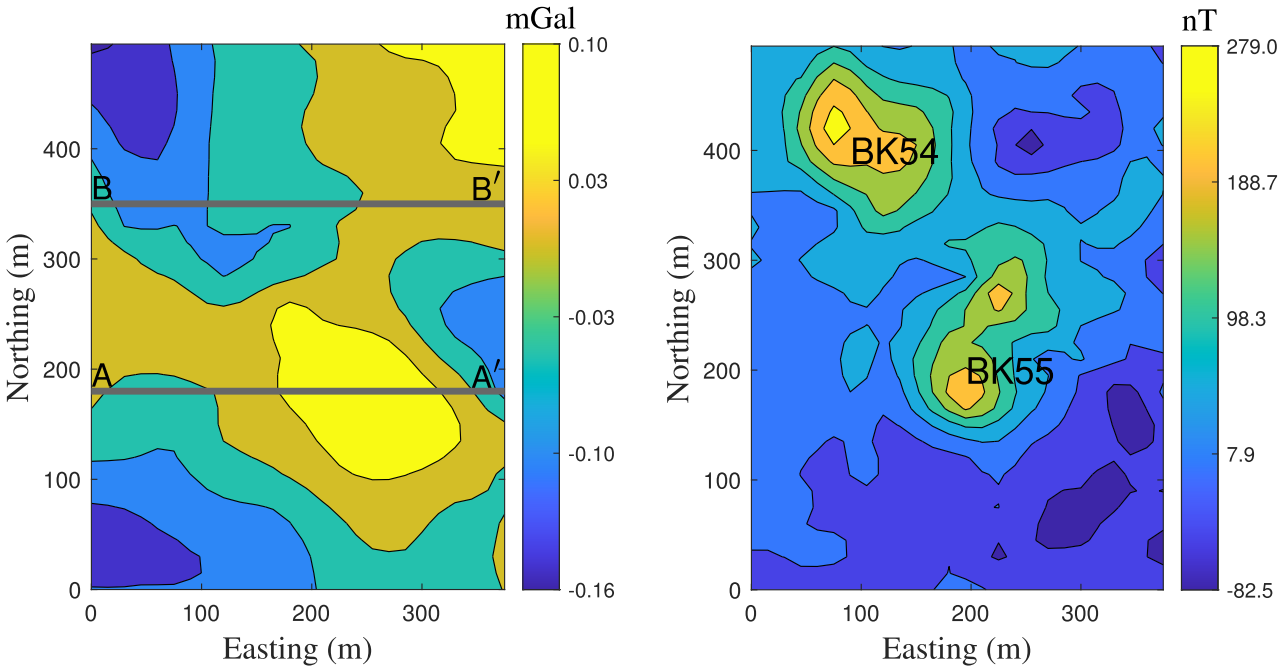

Fig. 14. Gravity and magnetic anomalies over BK54 and BK55. (a) Residual gravity anomaly. (b) Residual magnetic anomaly. Lines AA′ and BB′ are the locations for which the cross sections of the reconstructed models are illustrated.

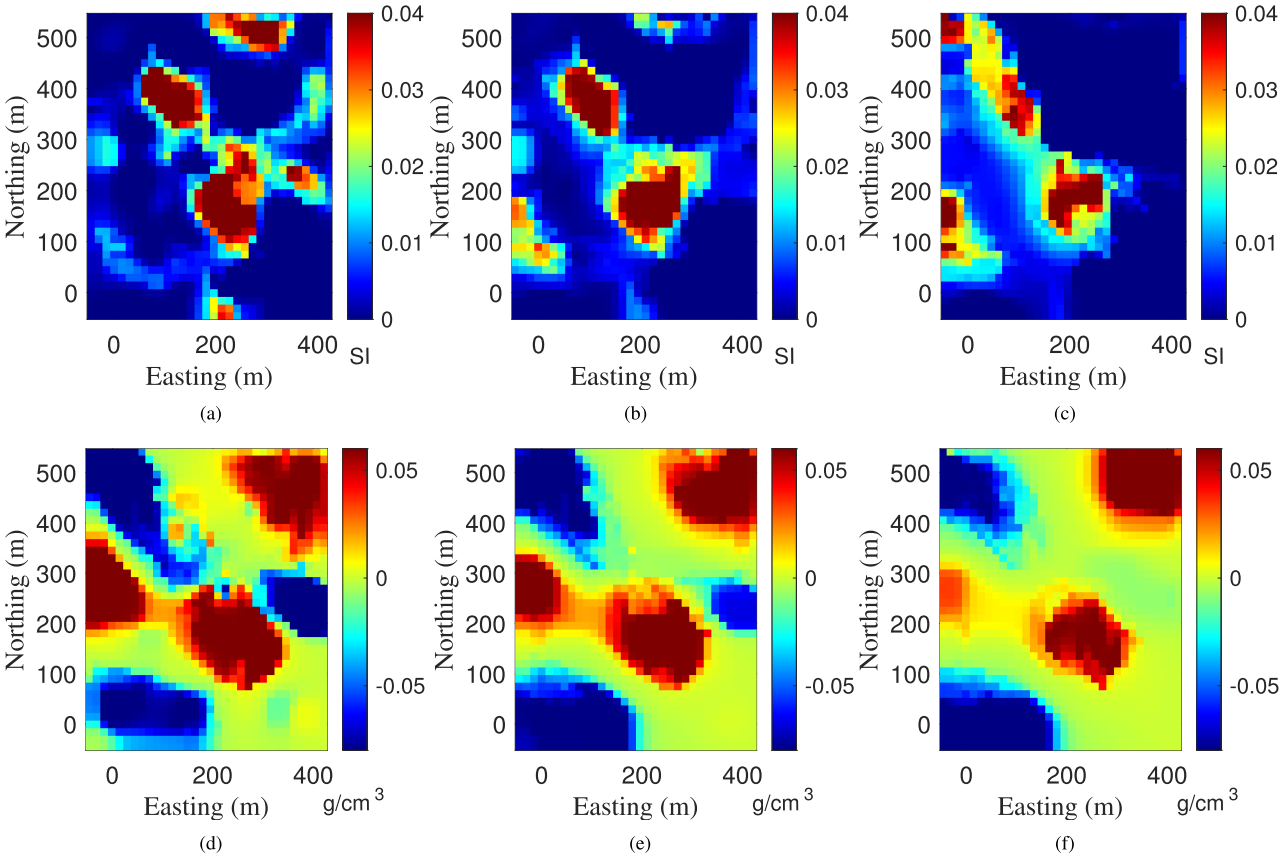

Fig. 15. Depth-sections of the reconstructed models using the joint AD algorithm for the real datasets shown in Fig. 14. The parameters used in the inversion are given in Tables II and III (Test R). (a)–(c) Magnetic susceptibility model and (d)–(f) density model at depths of 70, 110, and 150 m, respectively.

model is a little larger and requires a larger clock time to converge. At convergence, ω_1 and ω_2 are .76 and .91, respectively, when balancing is applied, but, without balancing, the algorithm still terminates at iteration 78 but with $\omega_1 = 0.34$ and $\omega_2 = 0.92$. This indicates overfitting occurs without balancing.

These results clearly indicate that our strategy for weighting the data misfit terms is appropriate and can provide an

TABLE III

SUMMARY OF THE EXPERIMENTS FOR THE JOINT INVERSION ALGORITHM APPLIED ON DATA PRODUCED BY THE SYNTHETIC MODELS 1 AND 2 AND FOR THE REAL DATASETS OF THE OKF IN BOTSWANA, HERE INDICATED USING R. D, I, AND F ARE THE DECLINATION, INCLINATION, AND INTENSITY OF THE GEOMAGNETIC FIELD, RESPECTIVELY. AD AND F INDICATE THE APPLICATION OF THE AD ALGORITHM AND THE USE OF THE FULL SYSTEM, RESPECTIVELY. ADP INDICATES THAT PADDING IS USED WITH THE AD ALGORITHM FOR A UNIFORM PADDING OF THREE CELLS IN EAST AND NORTH DIRECTIONS. ϵ_j IS THE SAFETY PARAMETER USED IN THE TV TERMS [SEE (8)]. $\rho_{\min}$, $\rho_{\max}$, $\kappa_{\min}$, AND $\kappa_{\max}$, RESPECTIVELY, ARE THE BOUNDS ON THE VARIABLES. λ_1 AND λ_2 ARE THE LAGRANGE PARAMETERS WEIGHTING THE CROSS-GRADIENT COUPLING. REGULARIZATION IS IMPOSED WITH PARAMETERS α_s, $\alpha_x^i = \alpha_y^i$, AND $\alpha_z^i = 0.01\alpha_x^i$ FOR p_s AND p_j REGULARIZATION NORMS

Test	D, I, F	Alg	(ϵ_1, ϵ_2)	$(\rho_{\min}, \rho_{\max})$	$(\kappa_{\min}, \kappa_{\max})$	(λ_1, λ_2)	(p_s, p_j)	$(\alpha_s, \alpha_x^1, \alpha_x^2)$
1	2°, 50°, 47000 nT	AD	(1e-9, 1e-9)	(0, 0.6)	(0, 0.06)	(5e5, 5e5)	(1, 2)	(1, 1e3, 1e4)
2	2°, 50°, 47000 nT	F	(1e-9, 1e-9)	(0, 0.6)	(0, 0.06)	(5e5, 5e5)	(1, 2)	(1, 1e3, 1e4)
3	2°, 50°, 47000 nT	AD	(1e-9, 1e-9)	(0, 0.6)	(0, 0.06)	(5e5, 5e5)	(1, 1)	(1, 20, 20)
4	2°, 50°, 47000 nT	AD	(1e-6, 1e-7)	(0, 0.6)	(0, 0.06)	(5e5, 5e5)	(1, 0)	(1, 2.4e-1, 2.4e-2)
5	2°, 50°, 47000 nT	AD	(1e-9, 1e-9)	(0, 1.5)	(0, 0.15)	(5e5, 5e5)	(1, 1)	(1, 20, 20)
6	2°, 50°, 47000 nT	AD	(1e-9, 1e-9)	(0, 1.5)	(0, 0.15)	(1e4, 1e4)	(2, 2)	(.1, 1e2, 1e2)
7	2°, 50°, 47000 nT	AD	(1e-9, 1e-9)	(0, 0.6)	(0, 0.06)	(3e6, 3e6)	(1, 1)	(1, 40, 40)
R	−13.5°, −62.1°, 28625 nT	ADP	(1e-9, 1e-9)	(−0.08, 0.06)	(0, 0.04)	(5e5, 5e5)	(1, 1)	(1, 6, 6)

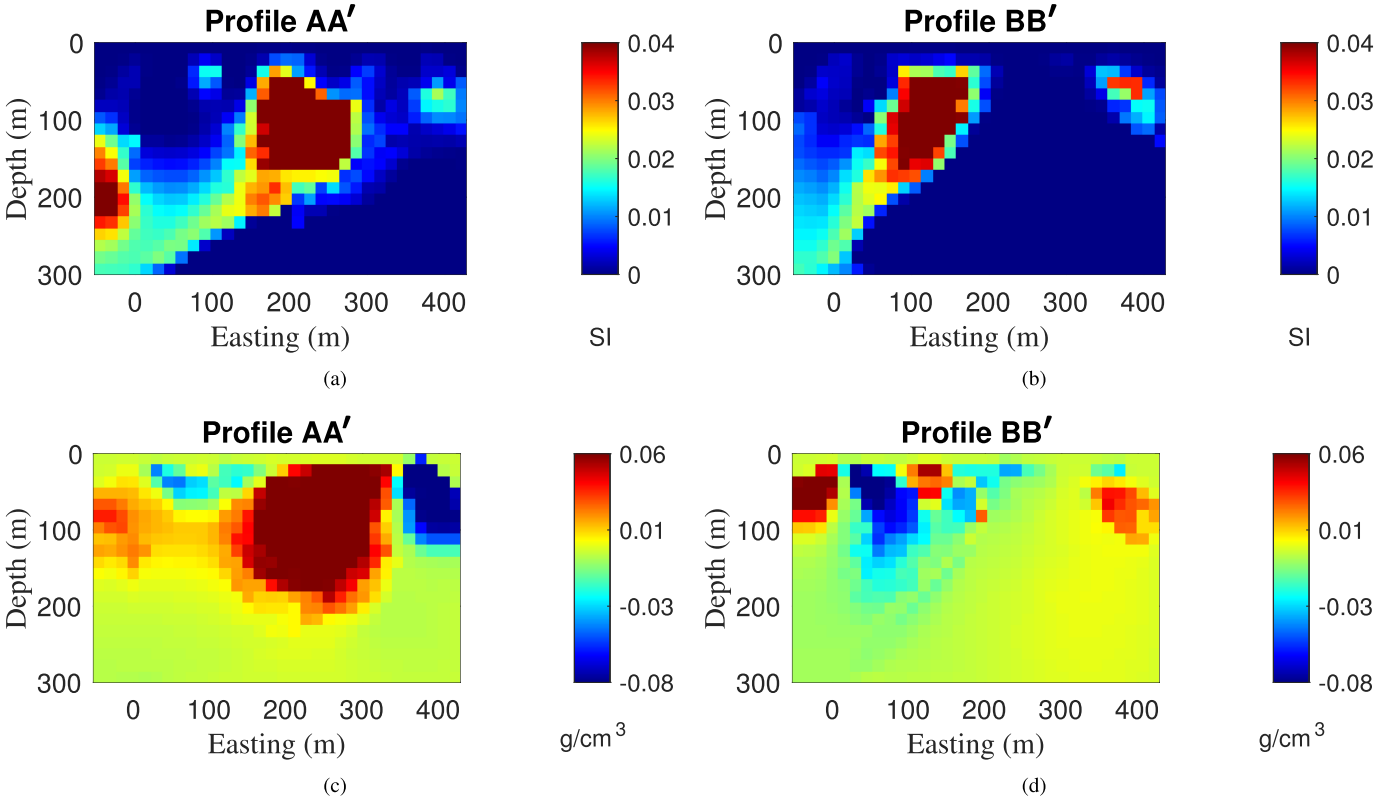

Fig. 16. Cross sections of the reconstructed models using the joint AD algorithm for the real datasets shown in Fig. 14. The parameters used in the inversion are given in Tables II and III (Test R). (a) and (b) Magnetic susceptibility model and (c) and (d) density model with cross sections at Northing = 180 m and Northing = 350 m, respectively.

approximately reasonable balance between two data misfit terms. The depth and cross sections of the reconstructed density and magnetic susceptibility models are illustrated in Figs. 15 and 16. Although the upper depths of the reconstructed models, for BK54 and BK55, are not consistent with the information from boreholes, the extensions of the targets approximately match the estimates indicated by the boreholes. From the reconstructed density model, we see that there are lower density materials associated with BK54. The lack of a large dense body for the BK54 kimberlite but having a high magnetic susceptibility could be indicative of the diatreme portion for most of the kimberlite. The diatreme portion consists of highly brecciated material, which would reduce the density of the material but still contains a large amount of magnetic minerals. The amount of brecciation is within the middle of the kimberlite, and the higher density region near the surface above the known position of the kimberlite may be due to the overlying basalt. The higher density and magnetic susceptibility regions at the west of the models [see Fig. 16(a) and (c)] may be due to basalt layers within the Karoo Supergroup sediments or Archean lithologies, which are known to exist just to the north of the study area [26]. The results presented in Figs. 15 and 16 are for the same depth and cross sections as [46, Figs. 10 and 11]. We see overall that

TABLE IV

RESULTS OF THE JOINT INVERSION ALGORITHM FOR THE TESTS DETAILED IN TABLES II AND III IN ADDITION TO THE RESULTS WITH NO BALANCING INDICATED AS 1a AND Ra, RESPECTIVELY. HERE, THE RESULTS SHOW THE ITERATIONS AT WHICH THE χ^2 TEST WAS SATISFIED FOR EACH DATASET, INDICATED BY K_1 AND K_2, RESPECTIVELY, AS WELL AS THE ERRORS AND χ^2 MEASURES AT CONVERGENCE AT THE FINAL ITERATION. THE TIME FOR THE RUN IS ALSO REPORTED IN SECONDS FOR THE IMAC AND IN PARENTHESES FOR THE MAC LAPTOP. IN THE SECOND BLOCK ARE THE SAME EXPERIMENTS ON THE IMAC BUT WITHOUT THE CROSS-GRADIENT COUPLING AND WITH THE RESULTS GIVEN AT THE INDIVIDUAL CONVERGENCE ITERATIONS K_1 AND K_2

Test	(K_1, K_2)	RE_1	RE_2	ω_1	ω_2	Time (s)	(K_1, K_2)	RE_1	RE_2	ω_1	ω_2	Time (s)	Figure
1	(58, 42)	0.55	0.61	0.99	0.82	433(406)	(60, 42)	0.58	0.70	0.97	0.98	372	3
1a	(58, 42)	0.55	0.59	1.00	0.71	605(382)				NA			
2	(59, 42)	0.57	0.63	0.92	0.83	1554(1646)	(60, 42)	0.58	0.66	0.97	0.98	1802	4
3	(59, 44)	0.55	0.55	0.96	0.86	367(365)	(59, 43)	0.57	0.67	0.95	0.98	287	5
4	(60, 42)	0.55	0.58	0.96	0.88	296(265)	(59, 42)	0.59	0.68	0.98	0.98	232	6
5	(60, 46)	0.74	0.83	0.95	0.85	415(403)	(60, 42)	1.00	0.98	0.96	0.96	319	7
6	(75, 68)	0.65	0.70	0.90	0.94	177(161)	(74, 69)	0.65	0.72	0.84	0.91	120	8
7	(54, 27)	0.41	.63	0.99	0.79	2669(2428)				NA			
R	(58, 78)	NA	NA	0.76	0.91	2286(2077)				NA			
Ra	(58, 78)	NA	NA	0.34	0.91	2328(2177)				NA			

the use of balancing and the mixed formulation yields results that are overall less smoothed.

V. CONCLUSION

An efficient strategy for mixed L_p-norm, $0 \leq p \leq 2$, joint inversion of gravity and magnetic datasets, combined with a cross-gradient structural coupling constraint between density and magnetic susceptibility models, has been presented and validated. The algorithm uses weighting parameters for all components. The data misfits are weighted with parameters that are proportional to the distance from the target misfit. This provides an effective strategy to balance the two terms so that each model is fit to its own noise level, which is important in a joint inversion algorithm. The strategy used here provides a good balance between the two terms. Furthermore, a general and robust approach to determine all weighting parameters is discussed. It is also demonstrated that the use of alternating updates for the density and magnetic susceptibility models yields a very effective approach for the joint inversion of datasets at significantly reduced computational cost compared to an approach using the full systems at each iteration while providing solutions that are comparable in quality to the solves without the AD. The application of the BTTB structure of the sensitivity matrices, and the matrix-free calculation of derivatives, is also crucial in providing an algorithm that is memory efficient for the solution of large problems. Together, the memory and computationally efficient implementation make it feasible to make robust estimates of subsurface targets.

Two synthetic datasets were used to assess, and highlight, the performance of the algorithm with respect to the selection of different stabilizers and regularization parameters. We showed the robust performance of the algorithm by presenting the results of applying the algorithm for different combinations of the L_p-norm components of the stabilizers. If suitable weighting parameters are chosen, the algorithm can provide smooth, sparse, or blocky models, dependent on the L_p-norm components applied within the stabilization term. Furthermore, a combination of both sparse and blocky stabilizers yields good reconstructions of the subsurface. Moreover, by also presenting results for mixed L_p stabilization used with independent inversions, we verify that the joint inversion generally outperforms the independent algorithm in terms of reduced relative error at a comparable cost.

The algorithm was applied to real data obtained over two kimberlite pipes, BK54 and BK55, in Botswana. The magnetic data clearly indicate two pipes, but the BK54 pipe does not appear on the residual gravity anomaly. The algorithm provides an acceptable reconstruction for both pipes. Although the depths to the top of the pipes are not consistent with the drill hole information, the extensions in depth are close to those estimated by boreholes.

APPENDIX

Practically, we note that all diagonal weighting matrices only require storage of their n diagonal entries, and the multiplications $\mathbf{W}_{\text{dh}}\mathbf{W}_s^{(k)}$ and $\mathbf{W}_{\text{dh}}\mathbf{W}_{\mathbf{D}_j}^{(k)}$ are accomplished in $\mathcal{O}(n)$ operations for each iteration. Moreover, the calculation of updating matrices $\mathbf{W}_s^{(k)}$ and $\mathbf{W}_{\mathbf{D}_j}^{(k)}$ is accomplished in $\mathcal{O}(n)$ operations per iteration, given $\mathbf{m}^{(k-1)}$ and $\mathbf{D}_j\mathbf{m}^{(k-1)}$. For our implementation, $\mathbf{D}_j\mathbf{m}^{(k-1)}$ is obtained using a matrix-free implementation in which, for the first-order derivative operator, the calculation of the x or y derivative for $\mathbf{m}$ on a given depth slice can be obtained using a suitable differencing operation on that slice. This is equivalent to forming the first-order difference approximation for the first-order derivative, which is obtained using the bidiagonal matrix with -1 on the diagonal and 1 above the diagonal. The derivative in depth is obtained in the same way but now applied for a constant slice in either x- or y-direction. Overall, derivatives in all directions are accomplished using the differencing operations along with suitable permutations of the 3-D array for storing the parameter vector $\mathbf{m}$. Furthermore, the efficient computation of the entries in the sparse matrix $\mathbf{B}_i$ is described in [41].

In the implementation, it is not necessary to scale the derivative calculations by the grid size in order to form what we would call true mathematical derivatives. However, in this case, given that the weighting parameter estimates provided in Section II-E assume that the grid sizes are employed, it is then necessary to move the scaling to the calculations of these parameters. Equivalently, λ_i is scaled by h_j^{-2}, α_j by $h_j^{-p_j}$, and also the safety parameters scale up by h_j^2. Indeed, doing the scaling to the parameters rather than the derivative calculations

Algorithm 1 AD Mixed L_p-Norm Joint Inversion of Gravity and Magnetic Datasets With the Cross-Gradient Constraint

Input: Data vectors $\mathbf{d}_i^{\text{obs}}$ and $\mathbf{W}_{\mathbf{d}_i^{\text{obs}}}$.
Prior information: vectors defining $\mathbf{m}_i^{\text{apr}}$, and $(\mathbf{W}_{\text{hard}})_i$
Bound constraints: $\rho_{\min}$, $\rho_{\max}$, $\kappa_{\min}$, $\kappa_{\max}$
Functions to calculate operations with matrices $\mathbf{G}_i$, $\mathbf{D}_j$.
Variables to define the functions.
Parameters: ϵ_s, ϵ_j, p_s, p_j, ν_i, α_s, α_j, β_i and λ_i for $i = 1$, 2 and $j = \text{x, y, z}$
Initial regularization parameter, $\beta_i^{(1)}$ and decay rate q_i for $i = 1, 2$
Maximum number of iterations: $\text{K}_{\max}$

1: Calculate $(\mathbf{W}_{\text{depth}})_i$ and then form $(\mathbf{W}_{\text{dh}})_i = (\mathbf{W}_{\text{depth}})_i(\mathbf{W}_{\text{hard}})_i$, for $i = 1, 2$.
2: Set $\gamma_i = 1$, $i = 1, 2$, and $k = 1$.
3: Initialize $\mathbf{m}_i^{(k)} = \mathbf{m}_i^{\text{apr}}$, $\mathbf{W}_{si}^{(k)} = \mathbf{I}_n$, $\mathbf{W}_{\mathbf{D}_{ji}}^{(k)} = \mathbf{I}_n$, for $i = 1, 2$, and $j = \text{x, y, z}$.
4: **while** Not converged, $(\chi_i^2)^{(k)}$ for both data sets not satisfied, and $k < \text{K}_{\max}$ **do**
5: $k = k + 1$.
6: Compute $\mathbf{t}^{(k)}$ and $\mathbf{B}_1^{(k)}$.
7: Solve $\mathbf{E}_1^{(k)}\mathbf{m}_1^{(k)} = \mathbf{f}_1^{(k)}$, defined by Eqs. (20) and (21).
8: Impose constraint $\rho_{\min} \leq \mathbf{m}_1^{(k)} \leq \rho_{\max}$.
9: Update $\mathbf{t}^{(k)}$ and $\mathbf{B}_2^{(k)}$ using estimated $\mathbf{m}_1^{(k)}$.
10: Solve $\mathbf{E}_2^{(k)}\mathbf{m}_2^{(k)} = \mathbf{f}_2^{(k)}$, defined by Eqs. (20) and (21).
11: Impose constraint $\kappa_{\min} \leq \mathbf{m}_2^{(k)} \leq \kappa_{\max}$.
12: Test convergence criteria for both χ_1^2 and χ_2^2. Exit loop if both satisfied.
13: If both convergence criteria are not satisfied, set $\beta_i^{(k)} = \beta_i^{(k-1)} q_i$, for $i = 1, 2$.
14: If $\omega_i \leq 1$ for either $i = 1$ or $i = 2$, for that i keep β_i fixed and decrease γ_i using Eq. (25).
15: Update $\mathbf{W}_{si}^{(k)}$, and $\mathbf{W}_{\mathbf{D}_{ji}}^{(k)}$, for $i = 1, 2$ and $j = \text{x, y, z}$, Eqs. (7) and (8).
16: **end while**

Output: $\rho = \mathbf{m}_1^{(k)}$ and $\kappa = \mathbf{m}_2^{(k)}$.

reduces the computational cost marginally since scalars rather than vectors are scaled.

The use of the BTTB structure of the matrices $\mathbf{G}_1$ and $\mathbf{G}_2$ to improve the efficiency of operations with sensitivity matrices has been described in multiple papers [7], [17], [49] and adopted for the cross-gradient joint inversion in [41]. Here, we use an implementation for the BTTB structure, which further improves the efficiency. Specifically, noting that the underlying operations for $\mathbf{Gm}$ are implemented as a sum over the depth layers, where, for layer j, we have a BTTB-operation for $\mathbf{G}_j\mathbf{m}_j$, $\mathbf{Gm} = \sum_j \mathbf{G}_j\mathbf{m}_j$, we can use linearity and reduce the number of inverse transforms required. Previously, we obtained $\mathbf{G}_j\mathbf{m}_j$ by using the inverse 2DFFT of the convolution $\mathbf{G}_j\mathbf{m}_j$ for each j. However, by summing in the Fourier space before taking the inverse, the total number of inverses is reduced. The transpose is also implemented more efficiently. Thus, we note that all operations with matrices $\mathbf{G}_1$ and $\mathbf{G}_2$ are matrix-free and are implemented using fast Fourier transforms, as described in [34], with the use of the linearity to reduce the total number of transforms overall.

ACKNOWLEDGMENT

The authors would like to thank three anonymous reviewers. Their helpful comments guided improvements in the clarity of the manuscript.

REFERENCES

[1] A. Afnimar, K. Koketsu, and K. Nakagawa, "Joint inversion of refraction and gravity data for the three-dimensional topography of a sediment–basement interface," *Geophys. J. Int.*, vol. 151, no. 1, pp. 243–254, Oct. 2002, doi: 10.1046/j.1365-246X.2002.01772.x.
[2] J. B. Ajo-Franklin, B. J. Minsley, and T. M. Daley, "Applying compactness constraints to differential traveltime tomography," *Geophysics*, vol. 72, no. 4, pp. R67–R75, Jul. 2007. [Online]. Available: http://geophysics.geoscienceworld.org/content/72/4/R67.abstract
[3] V. C. F. Barbosa and J. B. C. Silva, "Generalized compact gravity inversion," *Geophysics*, vol. 59, no. 1, pp. 57–68, Jan. 1994.
[4] H. Bertete-Aguirre, E. Cherkaev, and M. Oristaglio, "Non-smooth gravity problem with total variation penalization functional," *Geophys. J. Int.*, vol. 149, no. 2, pp. 499–507, May 2002, doi: 10.1046/j.1365-246X.2002.01664.x.
[5] O. Boulanger and M. Chouteau, "Constraints in 3D gravity inversion," *Geophys. Prospecting*, vol. 49, no. 2, pp. 265–280, Mar. 2001.
[6] C. E. Bruun and T. B. Nielsen, "Algorithms and software for large scale geophysical reconstructions," M.S. thesis, Tech. Univ. Denmark, Lyngby, Denmark, 2007. [Online]. Available: https://www2.imm.dtu.dk/pubdb/edoc/imm5257.pdf
[7] L. Chen and L. Liu, "Fast and accurate forward modelling of gravity field using prismatic grids," *Geophys. J. Int.*, vol. 216, no. 2, pp. 1062–1071, Feb. 2019, doi: 10.1093/gji/ggy480.
[8] S. C. Constable, R. L. Parker, and C. G. Constable, "Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data," *Geophysics*, vol. 52, no. 3, pp. 289–300, Mar. 1987.
[9] C. G. Farquharson, "Constructing piecewise-constant models in multi-dimensional minimum-structure inversions," *Geophysics*, vol. 73, no. 1, pp. K1–K9, Jan. 2008, doi: 10.1190/1.2816650.
[10] C. G. Farquharson and D. W. Oldenburg, "Non-linear inversion using general measures of data misfit and model structure," *Geophys. J. Int.*, vol. 134, no. 1, pp. 213–227, Jul. 1998.
[11] D. Fournier and D. W. Oldenburg, "Inversion using spatially variable mixed ℓ_p norms," *Geophys. J. Int.*, vol. 218, no. 1, pp. 268–282, Jul. 2019.
[12] E. Fregoso and L. A. Gallardo, "Cross-gradients joint 3D inversion with applications to gravity and magnetic data," *Geophysics*, vol. 74, no. 4, pp. L31–L42, Jul. 2009.
[13] L. A. Gallardo and M. A. Meju, "Characterization of heterogeneous near-surface materials by joint 2D inversion of DC resistivity and seismic data," *Geophys. Res. Lett.*, vol. 30, no. 13, p. 1658, Jul. 2003.
[14] L. A. Gallardo, "Joint two-dimensional DC resistivity and seismic travel time inversion with cross-gradients constraints," *J. Geophys. Res.*, vol. 109, no. B3, 2004, Art. no. B03311.
[15] E. Haber and M. H. Gazit, "Model fusion and joint inversion," *Surveys Geophys.*, vol. 34, no. 5, pp. 675–695, Sep. 2013.
[16] E. Haber and D. Oldenburg, "Joint inversion: A structural approach," *Inverse Problems*, vol. 13, no. 1, pp. 63–77, Feb. 1997.
[17] J. D. Hogue, R. A. Renaut, and S. Vatankhah, "A tutorial and open source software for the efficient evaluation of gravity and magnetic kernels," *Comput. Geosci.*, vol. 144, Nov. 2020, Art. no. 104575. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0098300420305628
[18] M. Jorgensen and M. S. Zhdanov, "Imaging Yellowstone magmatic system by the joint Gramian inversion of gravity and magnetotelluric data," *Phys. Earth Planet. Interiors*, vol. 292, pp. 12–20, Jul. 2019.
[19] B. J. Last and K. Kubik, "Compact gravity inversion," *Geophysics*, vol. 48, no. 6, pp. 713–721, Jun. 1983.
[20] C. L. Lawson, *Contributions to the Theory of Linear Least Maximum Approximation*. Los Angeles, CA, USA: Univ. of California, Los Angeles, 1961. [Online]. Available: https://books.google.ca/books?id=b_CtbwAACAAJ

[21] P. G. Lelièvre and D. W. Oldenburg, "A comprehensive study of including structural orientation information in geophysical inversions," *Geophys. J. Int.*, vol. 178, no. 2, pp. 623–637, Aug. 2009, doi: 10.1111/j.1365-246X.2009.04188.x.

[22] P. G. Lelièvre, C. G. Farquharson, and C. A. Hurich, "Joint inversion of seismic traveltimes and gravity data on unstructured grids with application to mineral exploration," *Geophysics*, vol. 77, no. 1, pp. K1–K15, Jan. 2012, doi: 10.1190/geo2011-0154.1.

[23] Y. Li and D. W. Oldenburg, "3-D inversion of magnetic data," *Geophysics*, vol. 61, no. 2, pp. 394–408, Mar. 1996, doi: 10.1190/1.1443968.

[24] Y. Li and D. W. Oldenburg, "Fast inversion of large-scale magnetic data using wavelet transforms and a logarithmic barrier method," *Geophys. J. Int.*, vol. 152, no. 2, pp. 251–265, Feb. 2003, doi: 10.1046/j.1365-246X.2003.01766.x.

[25] W. Lin and M. S. Zhdanov, "Joint multinary inversion of gravity and magnetic data using gramian constraints," *Geophys. J. Int.*, vol. 215, no. 3, pp. 1540–1557, Aug. 2018.

[26] K. Matende and K. Mickus, "Magnetic and gravity investigation of kimberlites in north-central Botswana," *Geophysics*, vol. 86, no. 2, pp. B67–B78, Mar. 2021, doi: 10.1190/geo2020-0450.1.

[27] M. Moorkamp, B. Heincke, M. Jegen, A. W. Roberts, and R. W. Hobbs, "A framework for 3-D joint inversion of MT, gravity and seismic refraction data," *Geophys. J. Int.*, vol. 184, no. 1, pp. 477–493, Jan. 2011. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-246X.2010.04856.x

[28] L. Nielsen and B. H. Jacobsen, "Integrated gravity and wide-angle seismic inversion for two-dimensional crustal modelling," *Geophys. J. Int.*, vol. 140, no. 1, pp. 222–232, Jan. 2000.

[29] D. W. Oldenburg and Y. Li, "Subspace linear inverse method," *Inverse Problems*, vol. 10, no. 4, pp. 915–935, Aug. 1994, doi: 10.1088/0266-5611/10/4/011.

[30] D. W. Oldenburg, P. R. Mcgillivray, and R. G. Ellis, "Generalized subspace methods for large-scale inverse problems," *Geophys. J. Int.*, vol. 114, no. 1, pp. 12–20, Jul. 1993, doi: 10.1111/j.1365-246X.1993.tb01462.x.

[31] M. Pilkington, "3-D magnetic imaging using conjugate gradients," *Geophysics*, vol. 62, no. 4, pp. 1132–1142, Jul. 1997.

[32] O. Portniaguine and M. S. Zhdanov, "Focusing geophysical inversion images," *Geophysics*, vol. 64, no. 3, pp. 874–887, May 1999.

[33] O. Portniaguine and M. S. Zhdanov, "3-D magnetic inversion with data compression and image focusing," *Geophysics*, vol. 67, no. 5, pp. 1532–1541, Sep. 2002, doi: 10.1190/1.1512749.

[34] R. A. Renaut, J. D. Hogue, S. Vatankhah, and S. Liu, "A fast methodology for large-scale focusing inversion of gravity and magnetic data using the structured model matrix and the 2-D fast Fourier transform," *Geophys. J. Int.*, vol. 223, no. 1, pp. 1378–1397, May 2020, doi: 10.1093/gji/ggaa372.

[35] R. A. Renaut, S. Vatankhah, and V. E. Ardestani, "Hybrid and iteratively reweighted regularization by unbiased predictive risk and weighted GCV for projected systems," *SIAM J. Sci. Comput.*, vol. 39, no. 2, pp. B221–B243, Jan. 2017.

[36] J. Sun and Y. Li, "Adaptive ℓ_p inversion for simultaneous recovery of both blocky and smooth features in a geophysical model," *Geophys. J. Int.*, vol. 197, no. 2, pp. 882–899, May 2014.

[37] J. Sun and Y. Li, "Joint inversion of multiple geophysical and petrophysical data using generalized fuzzy clustering algorithms," *Geophys. J. Int.*, vol. 208, no. 2, pp. 1201–1216, Feb. 2017, doi: 10.1093/gji/ggw442.

[38] A. Tryggvason and N. Linde, "Local earthquake (LE) tomography with joint inversion for P- and S-wave velocities using structural constraints," *Geophys. Res. Lett.*, vol. 33, no. 7, pp. 1–5, 2006.

[39] S. Vatankhah, S. Liu, R. A. Renaut, X. Hu, and J. Baniamerian, "Improving the use of the randomized singular value decomposition for the inversion of gravity and magnetic data," *Geophysics*, vol. 85, no. 5, pp. G93–G107, Sep. 2020.

[40] S. Vatankhah, S. Liu, R. A. Renaut, X. Hu, J. D. Hogue, and M. Gharloghi, "Generalized L_p-norm joint inversion of gravity and magnetic data using cross-gradient constraint," 2020, *arXiv:2001.03579*.

[41] S. Vatankhah, S. Liu, R. A. Renaut, X. Hu, J. D. Hogue, and M. Gharloghi, "An efficient alternating algorithm for the ℓ_p-norm cross-gradient joint inversion of gravity and magnetic data using the 2-D fast Fourier transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4500416.

[42] S. Vatankhah, R. A. Renaut, and V. E. Ardestani, "3-D projected ℓ_1 inversion of gravity data using truncated unbiased predictive risk estimator for regularization parameter estimation," *Geophys. J. Int.*, vol. 210, no. 3, pp. 1872–1887, Sep. 2017.

[43] S. Vatankhah, R. A. Renaut, and V. E. Ardestani, "A fast algorithm for regularized focused 3D inversion of gravity data using randomized singular-value decomposition," *Geophysics*, vol. 83, no. 4, pp. G25–G34, 2018.

[44] S. Vatankhah, R. A. Renaut, and V. E. Ardestani, "Total variation regularization of the 3-D gravity inverse problem using a randomized generalized singular value decomposition," *Geophys. J. Int.*, vol. 213, no. 1, pp. 695–705, Apr. 2018, doi: 10.1093/gji/ggy014.

[45] S. Vatankhah, R. A. Renaut, X. Huang, K. Mickus, and M. Gharloghi, "Large-scale focusing joint inversion of gravity and magnetic data with gramian constraint," *Geophys. J. Int.*, vol. 230, no. 3, pp. 1585–1611, May 2022, doi: 10.1093/gji/ggac138.

[46] S. Vatankhah, R. A. Renaut, K. Mickus, S. Liu, and K. Matende, "A comparison of the joint and independent inversions for magnetic and gravity data over kimberlites in Botswana," *Geophys. Prospecting*, vol. 70, no. 9, pp. 1602–1616, Nov. 2022, doi: 10.1111/1365-2478.13265.

[47] C. Vogel, *Computational Methods for Inverse Problems*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2002. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/1.9780898717570

[48] X. Wei and J. Sun, "Uncertainty analysis of 3D potential-field deterministic inversion using mixed L_p norms," *Geophysics*, vol. 86, no. 6, pp. G133–G158, Nov. 2021, doi: 10.1190/geo2020-0672.1.

[49] Y. Zhang and Y. S. Wong, "BTTB-based numerical schemes for three-dimensional gravity field inversion," *Geophys. J. Int.*, vol. 203, no. 1, pp. 243–256, Aug. 2015, doi: 10.1093/gji/ggv301.

[50] M. S. Zhdanov, *Geophysical Inverse Theory and Regularization Problems*, vol. 36. Amsterdam, The Netherlands: Elsevier, 2002.

[51] M. S. Zhdanov, A. Gribenko, and G. Wilson, "Generalized joint inversion of multimodal geophysical data using Gramian constraints," *Geophys. Res. Lett.*, vol. 39, no. 9, p. 1233, May 2012.