A stochastic gradient descent approach with partitioned-truncated singular value decomposition for large-scale inverse problems of magnetic modulus data

Wenbin Li*, Kangzhi Wang and Tingting Fan

School of Science, Harbin Institute of Technology, Shenzhen, Shenzhen 518055, People’s Republic of China

E-mail: liwenbin@hit.edu.cn, 19S058004@stu.hit.edu.cn and 19S058010@stu.hit.edu.cn

Received 10 October 2021, revised 14 March 2022
Accepted for publication 25 April 2022
Published 20 May 2022

Abstract

We propose a stochastic gradient descent approach with partitioned-truncated singular value decomposition (SVD) for large-scale inverse problems of magnetic modulus data. Motivated by a uniqueness theorem in gravity inverse problem and realizing the similarity between gravity and magnetic inverse problems, we propose to solve the level-set function modeling the volume susceptibility distribution from the nonlinear magnetic modulus data. To deal with large-scale data, we employ a mini-batch stochastic gradient descent approach with random reshuffling when solving the optimization problem of the inverse problem. We propose a stepsize rule for the stochastic gradient descent according to the Courant–Friedrichs–Lewy condition of the evolution equation. In addition, we develop a partitioned-truncated SVD algorithm for the linear part of the inverse problem in the context of stochastic gradient descent. Numerical examples illustrate the efficacy of the proposed method, which turns out to have the capability of efficiently processing large-scale measurement data for the magnetic inverse problem. A possible generalization to the inverse problem of deep neural network is discussed at the end.

* Author to whom any correspondence should be addressed.

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Keywords: large-scale inverse problem, magnetic modulus data, stochastic gradient descent, partitioned-truncated SVD

(Some figures may appear in colour only in the online journal)

1. Introduction

Solution of magnetic inverse problem aims to reconstruct susceptibility distributions from magnetic measurements, which is an important task in detections and surveys employing magnetic approaches. Applications include exploration of geological resources [13], detection of unexploded ordnance [15], surveillance for naval warfare [22], and others. A main difficulty of solving the magnetic inverse problem comes from its ill-posedness, and there is a strong non-uniqueness of solution for a given set of magnetic measurement data. To deal with the non-uniqueness, level-set approaches have been proposed for the inverse problems of magnetic intensity data [9, 12] and magnetic gradient tensor data [10], where it assumes a given constant susceptibility supported in an unknown domain and the level-set function is used to represent topological shape of the unknown domain. In this work, we adopt the level-set formulation and develop efficient computational methods for massive datasets which were infeasible to process due to the restriction of computational resource. We will work on the large-scale inverse problem of nonlinear magnetic modulus data, and we propose a mini-batch stochastic gradient descent approach with partitioned-truncated singular value decomposition (SVD) for the solution of the level-set function modeling the volume susceptibility distribution.

When meeting with large-scale dataset in realistic magnetic inverse problem, we realize that the computational resource becomes inadequate as both the memory access and the computational time are demanding. A direct and trade-off strategy is to take a subsample of the measurement dataset, but the problem is how to choose an appropriate subsample preserving the data feature under the restriction of computational resource. As a result, we look for the employment of full dataset with limited computational resource, and the stochastic approach comes to our mind. The stochastic gradient methods have achieved great success in large-scale machine learning where massive datasets are involved in the training process; e.g., see [3] for a comprehensive review. The stochastic approaches use part of the dataset in every computation, which significantly saves memory usage and computational effort, whereas the full dataset is employed during each epoch if the without-replacement sampling strategy is adopted, or the full dataset is touched in expectation if the with-replacement sampling strategy is used [19]. Moreover, it is reported that with the help of noisy gradients, the stochastic approaches have the capability of escaping saddle points [5] and local minima [7] for non-convex optimization problems. Due to these advantages, we propose to introduce the stochastic strategy into the solution of the magnetic inverse problem. We use a mini-batch approach to reduce variances of stochastic gradients, and we formulate the iteration process by a partial differential equation and propose a stepsize rule according to its Courant–Friedrichs–Lewy (CFL) condition. In addition, a partitioned-truncated SVD algorithm is developed to efficiently compute the linear part of the inverse problem in the context of stochastic gradient approach. The details will be shown in the part of main algorithm.

We consider the nonlinear modulus data in the magnetic inverse problem for the following two reasons. Firstly, the modulus data are nonlinear stacking of multiple components of the magnetic vector field, which are considered to be less sensitive to systematic measurement noises than the directionally dependent vector data. Secondly, a well-posedness theory of gravity inverse problem suggests that the modulus of gravity field can uniquely determine gravity
source [6, 11], and since the magnetic inverse problem shares similarity to the gravity inverse problem, we believe that the measurement of magnetic modulus data is sufficient for the recovery of the susceptibility distribution. More details will be discussed in the formulation of the inverse problem of magnetic modulus data. In addition, we realize that the general form of the nonlinear magnetic inverse problem is similar to the architecture of neural network, and the proposed algorithm can be generalized to solve the inverse problem of deep neural network, which will be discussed in the section of conclusion and discussion.

2. Inverse problem of magnetic modulus data

Given an inducing magnetic field of strength $B^0$ and direction $\hat{B}^0$, the magnetic vector field resulting from a distribution of magnetic susceptibility $\kappa$ is modeled by the equation

$$B(r) = (B_1(r), B_2(r), B_3(r))^T = \frac{1}{4\pi} B^0 \int_{\Omega} K(r, \tilde{r}) \kappa(\tilde{r}) \, d\tilde{r},$$

with

$$K(r, \tilde{r}) = \nabla_r \left( \hat{B}^0 \cdot \nabla_r \left( \frac{1}{|r - \tilde{r}|} \right) \right),$$

where $r$ and $\tilde{r}$ denote 3D spatial coordinates: $r \in \Gamma$ (the data measurement boundary), $\tilde{r} \in \Omega$. The nonlinear magnetic modulus is given by,

$$d(r) = \left( \sum_{s=1}^{3} B_s(r)^2 \right)^{\frac{1}{2}};$$

since $d(r)$ is nonlinear stacking of multiple components of the magnetic vector field $B(r)$, it is considered to be less sensitive to systematic measurement noises than $B(r)$ which is direction-ally dependent. The nonlinear inverse problem reads as follows: given the modulus data $d(r)$ on the measurement boundary $\Gamma$, reconstruct the magnetic susceptibility parameter $\kappa(\tilde{r})$ in the domain $\Omega$.

It is well known that the solution of magnetic inverse problem is non-unique, and there exist infinitely many susceptibility distributions which can reproduce a given set of magnetic data. To deal with the non-uniqueness, we consider the volume susceptibility distribution in the following form,

$$\kappa(\tilde{r}) = \kappa_0 \chi_D(\tilde{r}),$$

where $\kappa_0$ is a constant that approximates the average value of the susceptibility, and $\chi_D$ is the characteristic function of a domain $D$: $\chi_D(\tilde{r}) = 1, \tilde{r} \in D; \chi_D(\tilde{r}) = 0, \tilde{r} \notin D$. We assume that the constant $\kappa_0$ is known as a priori information, and invert for the domain $D$ to recover the shape of the magnetic source. The setup is motivated by a well-posedness theory in gravity inverse problem [6, 11] which can be sated as follows.

**Theorem 3.1.** Let $\Omega_0$ be a convex domain in $\mathbb{R}^n$ with analytic (regular) boundary, and $\Gamma \subset \partial \Omega_0$ be a nonempty hyper-surface. $\Omega \subset \Omega_0$ is a bounded domain with connected $\mathbb{R}^n \setminus \overline{\Omega}$. $D \subset \Omega$ denotes the domain of mass distribution having piecewise smooth boundaries, and $U$ denotes the gravitational potential generated by the mass distribution $\mu = f \chi_D$. Given data as the modulus of gravity, $d = |\nabla U|$, on $\Gamma \subset \partial \Omega_0$ and given $f \geq 0$ in $\Omega$, the domain of mass distribution $D$ can be uniquely determined if one of the following constraints is satisfied:
(a) $D$ is star-shaped with respect to its center of gravity, and $f = C$, where $C$ is a constant;
(b) $D$ is convex in $x_1$, where $x_1$ denotes the first component of the spatial coordinate, and $f = C$, where $C$ is a constant;
(c) $D$ is convex in $x_1$, $f$ does not depend on $x_1$, $f \in C(\Omega)$, and $\Omega \subset \text{supp } f$;
(d) $D$ is convex, $f \in L_1(\Omega)$, and $0 < f$ on $\Omega$.

The theorem implies that the exterior gravity modulus $|\nabla U|$ is able to uniquely determine the domain $D$ if it satisfies some geometric constraints. Realizing that the magnetic inverse problem shares similarity to the gravity inverse problem, see, e.g., [10], we consider the volume susceptibility distribution as shown in equation (2.4), and expect that the domain $D$ can be well recovered from the magnetic modulus data $d$ as the constant value $\kappa_0$ is given.

Similar to our previous works [9, 10, 12], we adopt a level-set formulation for the volume susceptibility distribution,

$$\kappa(\tilde{r}) = \kappa_0 H(\phi(\tilde{r})), \quad (2.5)$$

where $\phi(\tilde{r})$ is the level-set function used to depict the topological shape of $D$, and $H(\cdot)$ is the Heaviside function: $H(x) = 1$, $x \geq 0$; $H(x) = 0$, $x < 0$. In level-set computation, $\phi(\tilde{r})$ is maintained to be a continuous signed-distance function to the boundary of $D$,

$$\phi(\tilde{r}) = \begin{cases} \text{dist}(\tilde{r}, \partial D), & \tilde{r} \in D \\ -\text{dist}(\tilde{r}, \partial D), & \tilde{r} \in \bar{D}^c \end{cases} \quad (2.6)$$

where $\text{dist}(\tilde{r}, \partial D)$ denotes the distance between $\tilde{r}$ and $\partial D$. As a result, equation (2.5) is a natural expression for the volume susceptibility distribution $\kappa$ as shown in equation (2.4), and the zero level-set $\{\tilde{r} | \phi(\tilde{r}) = 0\}$ indicates the location of the boundary $\partial D$. The inverse problem is then transformed into the following form: given the magnetic modulus data $d(\mathbf{r})$ on the measurement boundary $\Gamma$, recover the level-set function $\phi(\tilde{r})$ in $\Omega$ so that the magnetic susceptibility $\kappa(\tilde{r})$ is evaluated by equation (2.5).

3. Main algorithm
3.1. Data fitting and optimization

The inverse problem is solved by fitting the magnetic modulus data with an optimal solution of the level-set function. Let $d'(\mathbf{r})$ denote the measured modulus data at $\mathbf{r} \in \Gamma$, and $d(\phi, \mathbf{r})$ denote the predicted data with the value of $\phi(\tilde{r})$. The usually used $L^2$ fitting function is defined as follows,

$$E(\phi) = \frac{1}{|\Gamma|} \int_\Gamma \frac{(d(\phi, \mathbf{r}) - d'(\mathbf{r}))^2}{2} d\mathbf{r}, \quad (3.1)$$

where $|\Gamma|$ denotes the area of the measurement surface $\Gamma$. Equation (3.1) can be viewed as an expectation of the data misfit on $\Gamma$:

$$\mathbb{E}(f(\phi, \mathbf{r})) = \int_\Gamma f(\phi, \mathbf{r}) p(\mathbf{r}) d\mathbf{r} \quad (3.2)$$

with
\[ p(r) = \frac{1}{|\Gamma|} \quad \text{and} \quad f(\phi, r) = \frac{1}{2} (d(\phi, r) - d^*(r))^2. \] (3.3)

A regularization term is added to the data fitting function, so that minimizing the total objective function leads to the admissible solution of $\phi$. We propose to use the $H^1$ semi-norm of $\phi$ as the regularization term:
\[ E_r(\phi) = \frac{1}{2} \int_\Omega |\nabla \phi(\tilde{r})|^2 \, d\tilde{r}. \] (3.4)

As discussed in [11], minimizing $E_r(\phi)$ has the effect of shrinking the measure (area or length) of the interface characterized by the zero level-set $\{\tilde{r} | \phi(\tilde{r}) = 0\}$, so that it avoids forming sharp oscillations on the interface.

The level-set function $\phi(\tilde{r})$ is then recovered by solving the optimization problem:
\[ \arg \min_{\phi} E_r(\phi) := \mathbb{E}(f(\phi, r)) + \alpha E_r(\phi), \] (3.5)
where $\alpha$ is a weighting parameter that controls the amount of regularization applied.

### 3.2. Mini-batch stochastic gradient descent with random reshuffling

The gradient-descent based approach is favorable for solving the optimization problem of equation (3.5) with large scale. The derivative of $E_r(\phi)$ is given by the formula,
\[ \frac{\partial E_r(\phi)}{\partial \phi} = \frac{\partial \mathbb{E}(f(\phi, r))}{\partial \phi} + \alpha \frac{\partial E_r(\phi)}{\partial \phi} = \frac{\partial \mathbb{E}(f(\phi, r))}{\partial \phi} - \alpha \Delta \phi \] (3.6)

with
\[ \frac{\partial \mathbb{E}(f(\phi, r))}{\partial \phi} = \int_{\Gamma} \frac{\partial f(\phi, r)}{\partial \phi} p(r) \, d\Gamma; \] (3.7)

the level-set function $\phi$ is updated according to the negative gradient direction,
\[ \frac{\partial \phi}{\partial t} = - \frac{\partial E_r(\phi)}{\partial \phi} = - \frac{\partial \mathbb{E}(f(\phi, r))}{\partial \phi} + \alpha \Delta \phi, \] (3.8)

where $t$ denotes the pseudo time variable that enables the evolution of $\phi$. As $f(\phi, r)$ is defined in the form of equation (3.3), we have
\[ \frac{\partial f(\phi, r)}{\partial \phi} = (d(\phi, r) - d^*(r)) \frac{\partial d(\phi, r)}{\partial \phi}. \] (3.9)

Considering that $d(\phi, r)$ is related to $\phi$ through equations (2.1)–(2.3) and equation (2.5), formally we can compute $\frac{\partial d}{\partial \phi}$ by the chain rule,
\[ \frac{\partial d(\phi, r)}{\partial \phi} = \sum_{s=1}^3 \frac{\partial d}{\partial B_s} \frac{\partial B_s}{\partial \phi} \frac{\partial \phi}{\partial \phi} = \frac{B^0 \kappa_0}{4\pi} \delta(\tilde{r}) \sum_{s=1}^3 \frac{B_s(r)}{d(r)} K_s(r, \tilde{r}). \] (3.10)

Here, $\delta(\phi)$ denotes the Dirac delta function arising from the derivative of the Heaviside function $H(\phi)$, and $K_s(r, \tilde{r})$ denotes the $s$th component of the magnetic integral kernel $K(r, \tilde{r})$ so that
\[ (K_1(r, \tilde{r}), K_2(r, \tilde{r}), K_3(r, \tilde{r})) = K(r, \tilde{r}). \] (3.11)
Substituting equation (3.10) into equation (3.9), we have

$$\frac{\partial f(\phi, \mathbf{r})}{\partial \phi} = \frac{B^h_0 N_0}{4\pi} \delta(\phi(\mathbf{r})) \sum_{s=1}^{3} \frac{B_s(\mathbf{r})}{d(\mathbf{r})} \left( d(\phi, \mathbf{r}) - d^*(\mathbf{r}) \right) K_s(\mathbf{r}, \tilde{\mathbf{r}}).$$

(3.12)

The derivative of the expected data misfit $E(f(\phi, \mathbf{r}))$ is then obtained by putting equation (3.12) into equation (3.7).

In physical surveys, the measurement data are sampled at discrete coordinates. Denoting the set of measurement points by $\{\mathbf{r}_i, i = 1, 2, \ldots, M\}$, the expected data misfit $E(f(\phi, \mathbf{r}))$ reduces to the empirical data misfit:

$$f^M(\phi) = \frac{1}{M} \sum_{i=1}^{M} f(\phi, \mathbf{r}_i).$$

(3.13)

There will be difficulties in minimizing $f^M(\phi)$ directly when the size of data becomes large. To illustrate it, we derive $\frac{\partial f^M}{\partial \phi}$ from equations (3.13) and (3.12),

$$\frac{\partial f^M(\phi)}{\partial \phi} = \frac{B^h_0 N_0}{4\pi M} \delta(\phi(\tilde{\mathbf{r}})) \sum_{s=1}^{3} \frac{B_s(\mathbf{r})}{d(\mathbf{r})} \left( d(\phi, \mathbf{r}) - d^*(\mathbf{r}) \right) K_s(\mathbf{r}, \tilde{\mathbf{r}}).$$

(3.14)

Here, $\tilde{\mathbf{r}}$ denotes the grid points in the computational domain $\Omega$. In a regular-size computation, the number of grid points is in the order of $10^4$, e.g. 40 grids each direction in a 3D domain; if the number of data measurements is also in this order, i.e. $M \sim 10^4$, the dimension of the integral kernel $K_s(\mathbf{r}, \tilde{\mathbf{r}})$ will be greater than $10^8$. It becomes demanding in memory accessing and computational time as the derivative is computed repeatedly in the gradient-descent based iterative algorithm, and the computation will be infeasible if the size of data and the number of grid points grow larger.

We propose to utilize a mini-batch stochastic gradient descent approach [3, 8] to solve the optimization problem of equation (3.5) with large scale. Instead of using all the measurement data, we employ a part of them to realize the expected data misfit $E(f(\phi, \mathbf{r}))$ and its derivative:

$$f^{S_k}(\phi) = \frac{1}{|S_k|} \sum_{i \in S_k} f(\phi, \mathbf{r}_i),$$

(3.15)

$$\frac{\partial f^{S_k}(\phi)}{\partial \phi} = \frac{1}{|S_k|} \sum_{i \in S_k} \frac{\partial f(\phi, \mathbf{r}_i)}{\partial \phi},$$

(3.16)

where $S_k \subset \{1, 2, \ldots, M\}$ denotes a mini-batch of the index set, and $|S_k|$ denotes the batch size, i.e. the number of elements in $S_k$. Using $\frac{\partial f^{S_k}(\phi)}{\partial \phi}$ instead of $\frac{\partial f^M(\phi)}{\partial \phi}$ to evaluate $\frac{\partial f^{S_k}(\phi)}{\partial \phi}$ in equation (3.8), the level-set function $\phi$ is then evolved according to the following formula,

$$\frac{\partial \phi}{\partial t} = \frac{\partial f^{S_k}(\phi)}{\partial \phi} + \alpha \Delta \phi.$$

(3.17)

The sampling strategies of $S_k$ can be generally divided into two categories: with-replacement sampling and without-replacement sampling [1, 2, 19]. In the with-replacement sampling, $S_k$ is uniformly drawn from the data index set $\{1, 2, \ldots, M\}$ for each iteration, so that $E(f^{S_k}(\phi)) = E(f(\phi, \mathbf{r}))$ and $E\left(\frac{\partial f^{S_k}(\phi)}{\partial \phi}\right) = \frac{\partial f^M(\phi)}{\partial \phi}$. In this case, the evolution equation (3.17) can be seen as a noisy version of the exact updating as shown in equation (3.8); for instance, if one sets
Algorithm 1. Mini-batch stochastic gradient descent algorithm with random reshuffling for the inverse problem of magnetic modulus data.

1. Initialize the level-set function \( \phi \); set iteration number \( k = 0 \).
2. for epoch \( e = 1, 2, \ldots, N_e \) do
3. Sample a random permutation \( \{ \sigma(1), \sigma(2), \ldots, \sigma(M) \} \) of the index set \( \{1, 2, \ldots, M\} \).
4. With a given batch size \( b \), partition the shuffled index set into \( n = \frac{M}{b} \) subsets: \( S_k = \{ \sigma((k-1)b + 1), \ldots, \sigma(kb) \}, k = 1, 2, \ldots, n \).
5. for \( k = 1 \) to \( n \) do
6. Update \( \phi \) according to equation (3.17): \( \phi := \phi + \Delta t \left( \frac{\partial \phi}{\partial t} \right) \).
7. Reinitialize \( \phi \) to maintain the signed distance property.
8. Iteration number \( k := k + 1 \).
9. end for
10. end for

\(|S_k| = 1\), i.e. uniformly picking single sample from the data set for each iteration, the algorithm reduces to the classical stochastic gradient descent approach, which is also referred to as stochastic approximation in some context \([4, 16, 18]\). The with-replacement sampling strategy is preferred by theoretical analysis since it realizes the exact updating rule in expectation. However, it suffers from practical drawbacks such as requiring truly random data access and hence longer runtime \([19]\). In practice, it is common to use the without-replacement sampling strategy, where the measurement data are employed sequentially in some random or even deterministic order.

We propose to use the without-replacement sampling with random reshuffling in our mini-batch stochastic gradient descent algorithm. The details are summarized in algorithm 1. In every epoch, the data set is randomly permuted and then partitioned into mini-batches, and the iterative algorithm employs the mini-batches successively in order to evaluate the stochastic gradient descent. The shuffling strategy ensures that every sampling point is reached during each epoch. The mini-batch strategy has the capability of reducing variances when estimating the stochastic gradients \([3, 8]\). In step 7 of algorithm 1, the level-set function \( \phi \) is reinitialized to maintain the shape of the signed-distance function as shown in equation (2.6). This is a standard operation in level-set computation, which aims to make \( \phi \) sufficiently smooth without changing the location of the zero level set. For the details, we refer readers to our previous works \([9, 11]\) and textbooks on level-set methods \([17]\).

3.3. Stepsize rule

The choice of stepsize is an important aspect in the stochastic gradient descent algorithm \([16, 20]\). An inappropriate choice of stepsize can have a disastrous effect on the performance of convergence. In this work, we propose a stepsize rule according to the CFL condition of the evolution equation (3.17).

Similar to equation (3.14), we have

\[
\frac{\partial f^{S_k}(\phi)}{\partial \phi} = \frac{B^0 s_0}{4\pi |S_k|} \delta(\phi(\bar{r})) \sum_{i=1}^{3} \sum_{\bar{r} \in S_k} B_i(\bar{r}) \left( d(\phi(\bar{r}), r_i) - d^*(r_i) \right) K_i(\bar{r}, \tilde{r}).
\] (3.18)

We use a numerical version of the Dirac delta function \([21]\),

\[
\delta(\phi) \doteq \delta_\epsilon(\phi) = \chi_{T \epsilon} |\nabla \phi|.
\] (3.19)
where $\chi_{T^e}$ denotes the characteristic function of $T^e$, and $T^e = \{ \tilde{r} : |\phi(\tilde{r})| < \epsilon \}$ is a neighborhood of the zero level set. Substituting (3.19) into (3.18), the stochastic gradient can be evaluated in the following way,

$$\frac{\partial f^{S_k}(\phi)}{\partial \phi} = V_n |\nabla \phi|$$ (3.20)

with

$$V_n := \frac{B^0 \kappa_0}{4\pi |S_k|} \chi^\rho \sum_{i=1}^3 \sum_{r_i \in S_k} \frac{B_i(r_i)}{d(r_i)} \left( d(\phi, r_i) - d^*(r_i) \right) K_s(r_i, \tilde{r}).$$ (3.21)

The evolution equation (3.17) then reduces to

$$\frac{\partial \phi}{\partial t} = -V_n |\nabla \phi| + \alpha \Delta \phi,$$ (3.22)

which can be viewed as a Hamilton–Jacobi equation with an artificial viscosity term. Discretizing $\frac{\partial \phi}{\partial t}$ using a direct forward Euler scheme: $\frac{\phi_{n+1} - \phi_n}{\Delta t} \approx \frac{\partial \phi}{\partial t}$, the CFL condition for stability is

$$\Delta t \left( \frac{\max |V_n|}{\min \{\Delta x, \Delta y, \Delta z\}} + \frac{2\alpha}{\Delta x^2} + \frac{2\alpha}{\Delta y^2} + \frac{2\alpha}{\Delta z^2} \right) < 1.$$ (3.23)

The stepsize of the mini-batch stochastic gradient descent algorithm is taken as

$$\Delta t = C \left( \frac{\max |V_n|}{\min \{\Delta x, \Delta y, \Delta z\}} + \frac{2\alpha}{\Delta x^2} + \frac{2\alpha}{\Delta y^2} + \frac{2\alpha}{\Delta z^2} \right)^{-1},$$ (3.24)

where $C$ denotes a constant with $C \in (0, 1)$. In practice, as the regularization parameter $\alpha$ can be very small, $\min \{\Delta x, \Delta y, \Delta z\} \ll \max |V_n|$, one can drop the terms including $\alpha$ in equation (3.24), and the stepsize is taken as

$$\Delta t = C \left( \frac{\max |V_n|}{\min \{\Delta x, \Delta y, \Delta z\}} \right)^{-1}.$$ (3.25)

### 3.4. Partitioned-truncated SVD for matrix multiplications in the stochastic gradient descent

Large-scale matrix multiplications arise from the evaluation of gradient direction and the discretization of magnetic integral, as shown in equations (3.21) and (2.1) respectively. Considering formula (2.2) and formula (3.11), the magnetic integral kernel has the following form,

$$(K_1(\mathbf{r}, \tilde{\mathbf{r}}), K_2(\mathbf{r}, \tilde{\mathbf{r}}), K_3(\mathbf{r}, \tilde{\mathbf{r}})) = \frac{1}{|\mathbf{r} - \tilde{\mathbf{r}}|^3} \left[ \frac{3 (\tilde{\mathbf{B}}^0 \cdot (\mathbf{r} - \tilde{\mathbf{r}})) (\mathbf{r} - \tilde{\mathbf{r}})}{|\mathbf{r} - \tilde{\mathbf{r}}|^2} - \tilde{\mathbf{B}}^0 \right],$$ (3.26)

which decays rapidly as the distance $|\mathbf{r} - \tilde{\mathbf{r}}|$ is increasing. This decaying property is common to the integral kernel of potential-field data, i.e. magnetic and gravity data. We have proposed a partitioned-truncated SVD algorithm to develop low rank approximation to the decaying integral kernel in several different setups and for various types of potential-field data [9, 11, 14]. Here, we propose a variation of the partitioned-truncated SVD so that it is compatible with the mini-batch stochastic gradient descent approach.
3.4.1. Matrix multiplications involving the decaying kernel. To evaluate the stochastic gradient as shown in (3.21), we consider the following formulation,

$$G^S_i(\tilde{r}) := \chi_T \sum_{i \in S_k} \frac{B_i(\hat{r}_i)}{D(\hat{r}_i)} \left( d(\phi, \hat{r}_i) - d^*(\hat{r}_i) \right) K_i(\hat{r}_i, \tilde{r}), \quad S_k \subset \{1, 2, \ldots, M\}, \quad s \in \{1, 2, 3\}. \tag{3.27}$$

To evaluate $G^S_i$, we need to compute the magnetic components $B_i(\hat{r}_i)$: $i \in S_k$, $s \in \{1, 2, 3\}$, and we extract the following integral,

$$I^S_i(\hat{r}_i) := \int_{\Omega} K_i(\hat{r}_i, \tilde{r}) \kappa(\tilde{r}) \, d\tilde{r}, \quad i \in S_k \subset \{1, 2, \ldots, M\}, \quad s \in \{1, 2, 3\}. \tag{3.28}$$

Let the computational domain $\Omega \subset \mathbb{R}^3$ be a rectangular domain, $\Omega = [0, X] \times [0, Y] \times [-Z, 0]$, and let the measurement boundary $\Gamma$ satisfy $\Gamma \subset \{ r = (x, y, z) | z > 0 \}$. This setup corresponds to a common situation where the magnetic measurement is above the survey domain. Suppose that the domain $\Omega$ is uniformly discretized into $N = n_x n_y n_z$ grid points: $\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_N$, and recall that there are $M$ measurement points on $\Gamma$ and the mini-batch size $|S_k| = b$. After discretization, the evaluation of $I^S_i(\hat{r}_i)$ and $G^S_i(\tilde{r})$ reduces to the evaluation of matrix multiplications,

$$I^S_i = K^S_i \kappa^T,$$  \tag{3.29} 

$$G^S_i = d^S_i K^S_i D, \tag{3.30}$$

where

$$K^S_i = \begin{bmatrix} K_i(\hat{r}_i, \tilde{r}_j) \end{bmatrix}_{j \times N}, \quad i \in S_k, \quad j = 1, \ldots, N,$$

$$d^S_i = \begin{bmatrix} B_i(\hat{r}_i) \frac{D(\hat{r}_i)}{D(\hat{r}_i)} \left( d(\phi, \hat{r}_i) - d^*(\hat{r}_i) \right) \end{bmatrix}_{1 \times b}, \quad i \in S_k,$$

$$\kappa = [\kappa(\hat{r}_1), \ldots, \kappa(\hat{r}_N)],$$

$$D = \text{diag} \{\chi_T(\hat{r}_1), \ldots, \chi_T(\hat{r}_N)\}.$$

3.4.2. Preconditioning on the full kernel matrix. The multiplication matrix $K^S_i$ appearing in (3.29) and (3.30) is a sub-matrix of the full kernel matrix $K_i = [K_i(\hat{r}_i, \tilde{r}_j)]_{M \times N}$, $i = 1, \ldots, M$, $j = 1, \ldots, N$. Considering that the full kernel matrix $K_i$ is always constant once the measurement points $r_i$ and the grid points $\hat{r}_j$ are fixed, we perform preconditioning including partitioning and truncating on $K_i$.

The component $K_i(\hat{r}_i, \tilde{r}_j)$ decays rapidly as $|\hat{r}_i - \tilde{r}_j|$ increases. Since $r_i \in \Gamma \subset \{ r = (x, y, z) | z > 0 \}$, we order the grid points $\{\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_N\}$ according to their $z$-coordinates so that $0 \geq \tilde{z}_1 \geq \tilde{z}_2 \geq \cdots \geq \tilde{z}_N$, where we denote $\tilde{r}_j = (\tilde{x}_j, \tilde{y}_j, \tilde{z}_j)$. The set of grid points in the computational domain is then partitioned into $n_z$ subsets,

$$\{\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_N\} = \bigcup_{h=1}^{n_z} \{\tilde{r}_j\} | 1 \leq j \leq \tilde{M} := n_z n_y,$$  \tag{3.31} 

where $\tilde{r}_j = \tilde{r}_{j + (h-1)\Delta z}$ and so $\Delta z = -(h-1)\Delta z$. The full kernel matrix $K_i = [K_i(\hat{r}_i, \tilde{r}_j)]_{M \times N}$ is accordingly partitioned into $n_z$ sub-matrices along columns,
where Equations (3.32) and (3.38) imply the

\[
K_j = [K_j^1, K_j^2, \ldots, K_j^n], \quad \text{where } K_j^h = [K_j(r_i, \tilde{r}_j^h)]_{M \times M}, \quad 1 \leq h \leq n_z. \quad (3.32)
\]

In (3.32), every sub-matrix \(K_j^h\) corresponds to the grid points in the same depth. We perform the SVD for \(K_j^h\),

\[
K_j^h = U_j^h S_j^h (V_j^h)^T, \quad 1 \leq h \leq n_z, \quad s \in \{1, 2, 3\}, \quad (3.33)
\]

where \(U_j^h \in \mathbb{R}^{M \times M}, \ V_j^h \in \mathbb{R}^{\tilde{M} \times \tilde{M}}\) are unitary matrices, and

\[
S_j^h = \text{diag}\left\{ \sigma_1^{(s,h)}, \sigma_2^{(s,h)}, \ldots, \sigma_{\min\{M, \tilde{M}\}}^{(s,h)} \right\} \in \mathbb{R}^{M \times \tilde{M}}
\]

with the singular values \(\sigma_1^{(s,h)} \geq \sigma_2^{(s,h)} \geq \cdots \geq \sigma_{\min\{M, \tilde{M}\}}^{(s,h)} \geq 0\). Then we truncate \(S_j^h\) by choosing a thresholding parameter \(\epsilon_{\text{svd}}\) and setting

\[
\tilde{\sigma}_i^{(s,h)} = \begin{cases} \sigma_i^{(s,h)}, & \sigma_i^{(s,h)} \geq \epsilon_{\text{svd}}, \\ 0, & \sigma_i^{(s,h)} < \epsilon_{\text{svd}} \end{cases}, \quad 1 \leq i \leq \min\{M, \tilde{M}\}. \quad (3.35)
\]

Denoting \(\tau_{sh}\) the number of non-zero \(\tilde{\sigma}_i^{(s,h)}\), we have the truncated SVD for \(K_j^h\) in the following way,

\[
K_j^h \approx \tilde{U}_j^h \tilde{S}_j^h (\tilde{V}_j^h)^T, \quad 1 \leq h \leq n_z, \quad s \in \{1, 2, 3\}, \quad (3.36)
\]

where \(\tilde{S}_j^h = \text{diag}\left\{ \tilde{\sigma}_1^{(s,h)}, \tilde{\sigma}_2^{(s,h)}, \ldots, \tilde{\sigma}_{\tau_{sh}}^{(s,h)} \right\} \in \mathbb{R}^{\tau_{sh} \times \tau_{sh}}, \ \tilde{U}_j^h \in \mathbb{R}^{M \times \tau_{sh}}\) is composed of the first \(\tau_{sh}\) columns of \(U_j^h\), and \(\tilde{V}_j^h \in \mathbb{R}^{\tilde{M} \times \tau_{sh}}\) is composed of the first \(\tau_{sh}\) columns of \(V_j^h\). The truncating process reduces the dimension of the matrices.

Finally we compute

\[
\tilde{Q}_j^h = \tilde{S}_j^h (\tilde{V}_j^h)^T, \quad (3.37)
\]

so that

\[
K_j^h \approx \tilde{U}_j^h \tilde{Q}_j^h, \quad 1 \leq h \leq n_z, \quad s \in \{1, 2, 3\}. \quad (3.38)
\]

The preconditioning is completed and the matrices \(\tilde{U}_j^h, \tilde{Q}_j^h\) are stored for later use.

3.4.3. Evaluation of the stochastic gradient. Equations (3.32) and (3.38) imply the partitioned-truncated SVD for \(K_j^S\):

\[
K_j^S = [K_j^{S_1}, K_j^{S_2}, \ldots, K_j^{S_n}], \quad (3.39)
\]

\[
K_j^{h, S_i} \approx \tilde{U}_j^{S_i} \tilde{Q}_j^{S_i}, \quad 1 \leq h \leq n_z, \quad s \in \{1, 2, 3\}. \quad (3.40)
\]

where \(\tilde{U}_j^{S_i}\) is constructed by taking the rows of \(\tilde{U}_j^h\) with indices \(i \in S_i\).

To evaluate the matrix multiplications (3.29) and (3.30), we perform partitions on \(\kappa\) and \(D\) in the same way as the partition of \(K_j^S\),

\[
\kappa = [\kappa_1, \kappa_2, \ldots, \kappa_n], \quad (3.41)
\]

\[
D = \text{diag}\{D_1, D_2, \ldots, D_n\}. \quad (3.42)
\]
Finally, we propose the formulas of matrix multiplications for the stochastic gradient:

\[
I_s^h = \sum_{k=1}^{n_c} K_s^{h,k} \left( (\kappa_0)T \right) \approx \sum_{k=1}^{n_c} \tilde{U}_s^{h,k} \left( \tilde{Q}_s(k)T \right),
\]

\[
G_s^{h,k} = \left[ G_s^{x_1,k}, G_s^{x_2,k}, \ldots, G_s^{x_{n_c},k} \right], \quad \text{with} \quad G_s^{h,k} \approx \left( d_s^h \tilde{U}_s^{h,k} \right) \left( \tilde{Q}_s^h D_h \right).
\]

Realizing that \( D_h \) is a diagonal matrix with either 0 or 1 entries, \( \tilde{Q}_s^h D_h \) in (3.44) can be simply obtained by taking the columns of \( \tilde{Q}_s^h \) with the same indices of the non-zero diagonal entries in \( D_h \) and setting the other columns to be 0.

### 3.4.4. Estimate of error.

The approximations in equations (3.43) and (3.44) have the following error estimates.

**Proposition 4.1.** Let \( \| \cdot \|_2 \) denote the \( l^2 \)-norm of a column or row vector, then

\[
\left\| I_s^h - \sum_{k=1}^{n_c} \tilde{U}_s^{h,k} \left( \tilde{Q}_s(k)T \right) \right\|_2 < \epsilon_{\text{svd}}, \quad \left( 3.45 \right)
\]

\[
\left\| G_s^{h,k} - \left( d_s^h \tilde{U}_s^{h,k} \right) \left( \tilde{Q}_s^h D_h \right) \right\|_2 < \epsilon_{\text{svd}} \left\| d_s^k \right\|_2, \quad \left( 3.46 \right)
\]

**Proof.**

(a) Firstly, we show that

\[
\left\| K_s^{h,k} - \tilde{U}_s^{h,k} \tilde{Q}_s^h \right\|_2 < \epsilon_{\text{svd}}. \quad \text{By abuse of notation here,} \quad \| \cdot \|_2 \text{ denotes the induced } l^2 \text{-norm or spectral norm for a matrix.}
\]

Since \( K_s^{h,k} \) is constructed by taking the rows of \( K_s^h \) with indices \( i \in S_k \), equation (3.33) implies that

\[
K_s^{h,k} = U_s^{h,k} S_h^k V_h^k \text{T}, \quad \left( 3.47 \right)
\]

where \( U_s^{h,k} \) is constructed by taking the rows of \( U_s^h \) with indices \( i \in S_k \). Let

\[
\hat{S}_h^k = \text{diag} \left\{ \sigma_0, \sigma_2, \ldots, \sigma_{\min(M, \hat{M})} \right\} \in \mathbb{R}^{M \times \hat{M}} \quad \left( 3.48 \right)
\]

with \( \hat{S}_h^k \) defined by equation (3.35). \( \tilde{U}_s^{h,k} \tilde{Q}_s^h \) can be expressed in the following way,

\[
\tilde{U}_s^{h,k} \tilde{Q}_s^h = \tilde{U}_s^{h,k} \tilde{S}_h^k (V_h^k) \text{T} = U_s^{h,k} \hat{S}_h^k (V_h^k) \text{T}. \quad \left( 3.49 \right)
\]

Subtracting (3.49) from (3.47), we have

\[
\left\| K_s^{h,k} - \tilde{U}_s^{h,k} \tilde{Q}_s^h \right\|_2 = \left\| U_s^{h,k} \left( \hat{S}_h^k - S_h^k \right) (V_h^k) \text{T} \right\|_2 \leq \| U_s^{h,k} \|_2 \| \hat{S}_h^k - S_h^k \|_2 \| V_h^k \|_2. \quad \left( 3.50 \right)
\]

Since \( U_s^{h,k} \in \mathbb{R}^{h \times \hat{M}} \) is constructed by taking some rows from the unitary matrix \( U_s^h \in \mathbb{R}^{M \times M} \), it holds that
\[ \|\mathbf{U}_s^{h,\mathbf{S}_i}\|_2 = \sqrt{\rho\left(\mathbf{U}_s^{h,\mathbf{S}_i}(\mathbf{U}_s^{h,\mathbf{S}_i})^T\right)} = \sqrt{\rho(\mathbf{E})} = 1, \tag{3.51} \]

where \( \mathbf{E} \in \mathbb{R}^{h \times h} \) denotes the identity matrix, and \( \rho(\cdot) \) denotes the spectral radius of a matrix. Substituting (3.51) into (3.50), and realizing that \( \|\mathbf{S}_i^h - \hat{\mathbf{S}}_i^h\|_2 < \epsilon_{\text{svd}} \) and \( \|\mathbf{V}_h^h\|_2 = 1 \), we have

\[ \|\mathbf{K}_h^{h,\mathbf{S}_i} - \tilde{\mathbf{U}}_i^{h,\mathbf{S}_i} \hat{\mathbf{Q}}_i^h\|_2 < \epsilon_{\text{svd}}. \tag{3.52} \]

(b) Then the following estimates are immediate.

\[ \left\| \mathbf{F}_s^h - \sum_{h=1}^{n_r} \mathbf{U}_s^{h,\mathbf{S}_i}\left(\hat{\mathbf{Q}}_i^h(\kappa_0)^T\right) \right\|_2 \leq \sum_{h=1}^{n_r} \left\| \mathbf{K}_h^{h,\mathbf{S}_i} - \tilde{\mathbf{U}}_i^{h,\mathbf{S}_i} \hat{\mathbf{Q}}_i^h \right\|_2 \left\| \kappa_0 \right\|_2 < \epsilon_{\text{svd}} \sum_{h=1}^{n_r} \left\| \kappa_0 \right\|_2; \]

\[ \left\| \mathbf{d}_s^h - \mathbf{d}_s^{h,\mathbf{S}_i}\left(\hat{\mathbf{Q}}_i^h\mathbf{D}_h\right) \right\|_2 \leq \left\| \mathbf{d}_s^h \right\|_2 \left\| \mathbf{K}_h^{h,\mathbf{S}_i} - \tilde{\mathbf{U}}_i^{h,\mathbf{S}_i} \hat{\mathbf{Q}}_i^h \right\|_2 \left\| \mathbf{D}_h \right\|_2 < \epsilon_{\text{svd}} \left\| \mathbf{d}_s^h \right\|_2. \]

\[ \square \]

4. Results

We include some test examples to illustrate the efficacy of the proposed algorithm. The measurement magnetic modulus data are generated using equations (2.1)–(2.3). We discretize the magnetic integral (2.1) directly and we are not using the partitioned-truncated SVD when generating the measurement data. 5% Gaussian noises are added to each component \( B_s \),

\[ B_s^* = B_s (1 + \eta \cdot \mathcal{N}(0, 1)), \]

where \( \eta = 5\% \) and \( \mathcal{N}(0, 1) \) denotes the Gaussian noises with 0 mean and a standard deviation of 1. Then the synthetic nonlinear modulus data are generated according to equation (2.3),

\[ d^* (\mathbf{r}) = \left( \sum_{s=1}^{3} B_s^*(\mathbf{r})^2 \right)^{\frac{1}{2}}. \]

In the test examples, the inducing magnetic field has the strength of \( B_0 = 5.95 \times 10^4 \) nT, and the direction \( \hat{B}_0 \) is prescribed by inclination \( I_0 \) and declination \( D_0 \) so that

\[ \hat{B}_0 = (\cos I_0 \sin D_0, \cos I_0 \cos D_0, -\sin I_0), \]

where we designate the direction of \( x \)-axis as east, the direction of \( y \)-axis as north, and the direction of \( z \)-axis as negative depth. The average susceptibility value in the level-set formulation (2.5) is given as \( \kappa_0 = 0.05 \). In the inversion algorithm, we use the partitioned-truncated SVD when evaluating the predicted data and the stochastic gradient, where the thresholding parameter for truncating is taken as \( \epsilon_{\text{svd}} = 10^{-5} \).
Figure 1. Setup of examples in the shallow region. (a) Measurement points along $\Gamma = [0 \ 1] \times [0 \ 1] \times \{z = 0.1\}$ km; (b) initial guess.

Figure 2. Example 1. (a) True model; (b) magnetic modulus data with Gaussian noises.

4.1. Examples in a shallow region

Magnetic susceptibilities are reconstructed in the computational domain $\Omega = [0 \ 1] \times [0 \ 1] \times [-0.5 \ 0]$ km, and we discretize $\Omega$ uniformly into $41 \times 41 \times 21$ mesh grids. The measurement boundary is taken as $\Gamma = [0 \ 1] \times [0 \ 1] \times \{z = 0.1\}$ km, along which 10,000 measurement points are randomly distributed; figure 1(a) shows the picture of measurement points along $\Gamma$. In the inversion algorithm, we set the mini-batch size to be $b = 200$ for the stochastic gradient descent. The initial guess of the level-set function $\phi$ is set as follows,

$$
\phi_{\text{initial}} = 1 - \sqrt{\frac{(x - 0.5)^2}{0.35^2} + \frac{(y - 0.5)^2}{0.35^2} + \frac{(z - 0.25)^2}{0.15^2}};
$$

the zero level-set of the initial structure is an ellipsoid as shown in figure 1(b).

4.1.1. Example 1. As shown in figure 2(a), the true model of the volume susceptibility consists of three magnetic sources, where the cube is centered at $(0.75, 0.2, -0.3)$ with side length
Figure 3. Example 1. Inversion results. (a) Recovered solution; (b) data discrepancy \(d - d^*\); (c) and (d) cross-sections along \(x = 0.25\) and \(x = 0.75\), respectively, where the dashed line indicates the true model and the solid line indicates the recovered solution.

Figure 4. Example 1. Performance of the inversion algorithm. (a) Mini-batch misfit function \(f_{Sk}\); (b) \(L^2\)-difference between the recovered susceptibility \(\kappa\) and the true susceptibility \(\kappa^*\): \(\|\kappa - \kappa^*\|_2\). The misfit function \(f_{Sk}\) is oscillating partly due to the nonlinearity of level-set inversion and partly due to the noisy nature of stochastic gradient descent.
Figure 5. Example 2. (a) True model; (b) magnetic modulus data with Gaussian noises.

Figure 6. Example 2. Inversion results. (a) Recovered solution; (b) data discrepancy $d - d^*$; (c) and (d) cross-sections along $x = 0.25$ and $x = 0.75$, respectively, where the dashed line indicates the true model and the solid line indicates the recovered solution.

0.15, and the sphere is centered at $(0.8, 0.75, -0.2)$ with radius 0.1. We assume an inducing field with inclination and declination $(I_0, D_0) = (75^\circ, 25^\circ)$. Figure 2(b) plots the magnetic modulus data on the measurement surface. Figure 3 provides the inversion results. Figure 3(a) shows the recovered solution using the proposed stochastic gradient descent algorithm with partitioned-truncated SVD; figure 3(b) plots the data discrepancy $d - d^*$; figures 3(c) and (d)
plot the cross-sections along $x = 0.25$ and $x = 0.75$, respectively, where the dashed line indicates the true model and the solid line indicates the recovered solution. The solution adequately reproduces the magnetic modulus data, although we employ the partitioned-truncated SVD to approximate matrix multiplications and we never use all the 10,000 measurement data at once in the inversion algorithm. The recovered susceptibility structure matches well with the true model, where both the shape and location of the magnetic sources are successfully recovered. In figure 4, we provide more information to illustrate the performance of the inversion algorithm. Figure 4(a) presents the evolution of the mini-batch misfit function $f^S_k$ as defined in equation (3.15); figure 4(b) plots $L^2$-difference between the recovered susceptibility $\kappa$ and the true susceptibility $\kappa^*$. The misfit function $f^S_k$ is oscillating partly due to the nonlinearity of level-set inversion and partly due to the noisy nature of stochastic gradient descent. On the other hand, the performance of the $L^2$-difference $\|\kappa - \kappa^*\|_2$ implies a smooth convergence.
Figure 9. Example 3. Inversion results. (a) Recovered solution; (b) data discrepancy $d - d^*$; (c) and (d) cross-sections along $x = 0.25$ and $x = 0.75$, respectively, where the dashed line indicates the true model and the solid line indicates the recovered solution.

Figure 10. Example 3. Performance of the inversion algorithm. (a) Mini-batch misfit function $f_{Sk}$; (b) $\|\kappa - \kappa^*\|_2$.

4.1.2. Example 2. Figure 5(a) shows the true model of the volume susceptibility, where two dipping prisms are distributed in the computational domain. The inducing field has the inclination and declination $(I_0, D_0) = (75^\circ, 25^\circ)$. Figure 5(b) plots the magnetic modulus data on the measurement surface. Figure 6 provides the inversion results, where figure 6(a) shows the
Figure 11. Setup of examples in a cubic domain. (a) Measurement points along \( \Gamma = [-1, 2] \times [-1, 2] \times \{z = 0.1\} \) km; (b) initial guess.

Figure 12. Example 4. (a) True model; (b) magnetic modulus data with Gaussian noises.

recovered solution, figure 6(b) plots the data discrepancy \( d - d^* \), and figures 6(c) and (d) plot the cross-sections along \( x = 0.25 \) and \( x = 0.75 \), respectively. The magnetic modulus data are adequately reproduced. The recovered solution correctly captures the shape and location of the magnetic prisms, where the depth resolution looks perfect in the displayed cross-sections. In figure 7, we present the mini-batch misfit function \( f_{SB} \) and the \( L^2 \)-difference \( \| \kappa - \kappa^* \|_2 \). The performance of \( \| \kappa - \kappa^* \|_2 \) shows a smooth convergence although the mini-batch misfit function \( f_{SB} \) is oscillating.

4.1.3. Example 3. As shown in figure 8(a), we consider a susceptibility model including four distinct magnetic sources. The cube is centered at (0.75, 0.2, -0.2) with side length 0.15, the sphere is centered at (0.8, 0.75, -0.35) with radius 0.1, and the two dipping prisms with different depths are enclosed in the region \( \{(x, y, z) | x \in (0.1, 0.3)\} \). We assume an inducing field with inclination and declination \( (I^0, D^0) = (75^\circ, 25^\circ) \). Figure 8(b) shows the magnetic modulus...
Figure 13. Example 4. Inversion results. (a) Recovered solution; (b) data discrepancy $d - d^*$; (c) and (d) cross-sections along $x = 0.25$ and $x = 0.75$, respectively, where the dashed line indicates the true model and the solid line indicates the recovered solution.

data on the measurement surface. The inversion results are shown in figure 9, where figure 9(a) shows the recovered solution, figure 9(b) plots the data discrepancy, and figures 9(c) and (d) plot the cross-sections along $x = 0.25$ and $x = 0.75$, respectively. In addition, we present the mini-batch misfit function $f_{Sk}$ and the $l^2$-difference $\|\kappa - \kappa^*\|_2$ in figure 10. We conclude that the solution successfully recovers the susceptibility model and adequately reproduces the noisy magnetic modulus data. The performance of $\|\kappa - \kappa^*\|_2$ shows a smooth convergence although the mini-batch misfit function $f_{Sk}$ is oscillating.

4.2. Examples in a cubic domain

In the following examples, we implement the inversion algorithm in a cubic domain with larger depth. Depth resolution is generally a difficult task in magnetic inverse problem. Since the measurement surface is above the computational domain and the magnetic integral kernel decays rapidly as the distance between source and measurement increases, the responses of deep structures are likely to be suppressed by those of shallow structures and noise contaminations. As a result, depth resolution is an important criterion to evaluate the performance of a magnetic inversion algorithm. In the following examples, we include susceptibility distributions with
larger variances in depth, which aims to test the capability of depth resolution for the proposed inversion algorithm.

The computational domain is \( \Omega = [0, 1] \times [0, 1] \times [-10] \) km, which is uniformly discretized into \( 41 \times 41 \times 41 \) mesh grids. The measurement boundary is taken as \( \Gamma = [-1.2] \times [-1.2] \times \{z = 0.1\} \) km, along which 10 000 measurement points are randomly distributed; figure 11(a) plots the measurement points along \( \Gamma \). Again, we set the mini-batch size to be \( b = 200 \) for the stochastic gradient descent. The initial guess of the level-set function \( \phi \) is taken as:

\[
\phi_{\text{initial}} = 0.4 - \sqrt{(x - 0.5)^2 + (y - 0.5)^2 + (z - 0.5)^2},
\]

and so the zero level-set of the initial structure is a sphere as shown in figure 11(b).

4.2.1. Example 4. Figure 12(a) shows the susceptibility model, where two magnetic sources are located at different depths of the computational domain. The dipping prism is located in
Figure 16. Example 5. Inversion results. (a) Recovered solution; (b) data discrepancy \( d - d^* \); (c) and (d) cross-sections along \( x = 0.20 \) and \( x = 0.65 \), respectively, where the dashed line indicates the true model and the solid line indicates the recovered solution.

the shallow region, and the sphere is centered at \((0.75, 0.75, -0.7)\) with radius 0.1. We assume an inducing field with inclination and declination \((I_0, D_0) = (90^\circ, 0^\circ)\). Figure 12(b) shows the magnetic modulus data on the measurement surface. This is a typical example where the responses of deep structures are inapparent in the data profile. Figure 13 provides the inversion results using the mini-batch stochastic gradient descent algorithm with partitioned-truncated SVD. Figure 13(a) shows the recovered solution; figure 13(b) plots the data discrepancy \( d - d^* \); figure 13(c) plots the cross-section of the solution along \( x = 0.25 \); figure 13(d) plots the cross-section along \( x = 0.75 \). The performances of the mini-batch misfit function \( f^{Sk} \) and the \( \ell^2 \)-difference \( \| \kappa - \kappa^* \|_2 \) are presented in figures 14(a) and (b), respectively.

The solution successfully recovers the susceptibility model and adequately reproduces the magnetic modulus data. The depth resolution is amazing, given that the susceptibility model has large variance in depth and the spherical source in deep region has invisible response in the data profile. This test example illustrates that the inversion algorithm does not loose depth resolution in the solution, although the mini-batch stochastic gradient descent only employs part of data at each iteration and the partitioned-truncated SVD provides inaccurate approximation for matrix multiplications. In fact, since the stochastic gradient descent has the capability of
4.2.2. Example 5. In this example, we consider a susceptibility model including three magnetic sources with different depths. The true model is shown in figure 15(a). The measurement magnetic modulus data with Gaussian noises are shown in figure 15(b), where we assume an inducing field with inclination and declination \((I_0, D_0) = (75^\circ, 25^\circ)\). The inversion results are shown in figure 16, where figure 16(a) shows the recovered solution, figure 16(b) plots the data discrepancy, and figures 16(c) and (d) plot two cross-sections of the solution. The performances of the mini-batch misfit function \(f_{\beta_k}\) and the \(L^2\)-difference \(\|\kappa - \kappa^*\|_2\) are presented in figure 17. We conclude that the inversion algorithm successfully recovers the susceptibility model and adequately reproduces the measurement data. The depth resolution is reasonable if not perfect, given that the susceptibility model has distinct variance in depth, and the task of recovering depth is generally difficult. With the techniques of stochastic gradient descent and partitioned-truncated SVD, the proposed inversion algorithm is able to deal with large-scale measurement data efficiently in the magnetic inverse problem.

5. Conclusion and discussion

We have proposed a mini-batch stochastic gradient descent approach with partitioned-truncated SVD for large-scale inverse problems of magnetic modulus data. We employ a level-set formulation for susceptibility model and recover the volume susceptibility distribution from nonlinear magnetic modulus data by solving the level-set function. To deal with massive amount of measurement data, we propose a stochastic gradient descent approach to solve the level-set optimization problem with large scale. When evaluating the stochastic gradients, we consider a mini-batch strategy to reduce variances, and we use the without-replacement sampling with random reshuffling to explore full data set at every epoch. Realizing that the iteration process can be formulated as a Hamilton–Jacobi equation, we propose a step-size rule for the stochastic gradient descent according to the CFL condition of the evolution PDE. To further improve the computational efficiency, we take advantage of the decaying property of the magnetic integral kernel, and propose a partitioned-truncated SVD for matrix multiplications in the context of stochastic gradient descent.
Numerical examples are included to illustrate the efficacy of the proposed method. The solutions successfully recover the susceptibility models and adequately reproduce the magnetic modulus data. The inversion algorithm is able to reconstruct susceptibility distributions with large variances in depth. As an important criterion to evaluate the performance of the magnetic inversion algorithm, the depth resolution is very good in all the test examples. We believe that the stochastic gradient approach can help to improve depth resolution due to its capability of escaping saddle points and local minima. With the techniques of stochastic gradient descent and partitioned-truncated SVD, the proposed method enables us to efficiently process large-scale magnetic data which was infeasible due to the restriction of computational resource.

The inverse problem of magnetic modulus data can be formulated as the general form

\[ d = \sigma_2 \circ L \circ \sigma_1(\phi) : \text{reconstruc} \phi \text{ from } d, \]  

where \( \sigma_1 \) denotes the nonlinear level-set formulation as shown in (2.5), \( L \) denotes the linear integral as shown in (2.1), and \( \sigma_2 \) denotes the nonlinear stacking as shown in (2.3); after discretization, \( L \) is the linear operator of matrix multiplications. The forward operator in formula (5.1) is similar to the architecture of deep neural network:

\[ \Psi = \sigma_N \circ L_N \circ \ldots \circ \sigma_i \circ L_i \circ \ldots \circ \sigma_1 \circ L_1, \]

where \( \sigma_i \)'s denote nonlinear activating functions, and \( L_i \)'s denote linear operators including affine mappings, convolutions, etc. In this work, we have proposed an efficient stochastic gradient descent approach with partitioned-truncated SVD for the inverse problem of (5.1). The algorithm can be generalized to solve the inverse problem of the deep neural network as shown in (5.2), provided that the linear operators \( L_i \) have some sort of low-rank property. The inverse problem should be stated as follows: given a well-trained deep neural network \( \Psi = \sigma_N \circ L_N \circ \ldots \circ \sigma_1 \circ L_1 \) and its output response \( d = \Psi(\phi) \), reconstruct the input signal \( \phi \).

Further explorations will be performed on this subject in the future work.

Acknowledgments

Wenbin Li is supported by NSFC (Grant No. 41804096), Natural Science Foundation of Guangdong Province (Grant No. 2018A030313341), and Natural Science Foundation of Shenzhen (Grant No. JCYJ20190806144005645).

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Wenbin Li https://orcid.org/0000-0002-1125-2364
References

[1] Bottou L 2009 Curiously fast convergence of some stochastic gradient descent algorithms Proc. Symp. Learning and Data Science vol 8 (Paris) pp 2624–33
[2] Bottou L 2012 Stochastic gradient descent tricks Neural Networks: Tricks of the Trade (Berlin: Springer) pp 421–36
[3] Bottou L, Curtis F E and Nocedal J 2018 Optimization methods for large-scale machine learning SIAM Rev. 60 223–311
[4] Fu M C 2002 Feature article: optimization for simulation: theory vs practice INFORMS J. Comput. 14 192–215
[5] Ge R, Huang F, Jin C and Yang Y 2015 Escaping from saddle points—online stochastic gradient for tensor decomposition Conf. Learning Theory (PMLR) pp 797–842
[6] Isakov V 1990 Inverse Source Problems (Providence, RI: American Mathematical Society)
[7] Kleinberg R, Li Y and Yang Y 2018 An alternative view: when does SGD escape local minima? Int. Conf. Machine Learning (PMLR) pp 2698–707
[8] Mu L, Zhang T, Chen Y and Alexander Smola J 2014 Efficient mini-batch training for stochastic optimization Proc. 20th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining pp 661–70
[9] Li W, Lu W, Qian J and Li Y 2017 A multiple level-set method for 3D inversion of magnetic data Geophysics 82 J61–81
[10] Li W and Qian J 2020 A level-set algorithm for the inverse problem of full magnetic gradient tensor data Appl. Math. Lett. 107 106416
[11] Li W and Qian J 2021 Simultaneously recovering both domain and varying density in inverse gravimetry by efficient level-set methods Inverse Problems Imaging 15 387–413
[12] Li W, Qian J and Li Y 2020 Joint inversion of surface and borehole magnetic data: a level-set approach Geophysics 85 J15–32
[13] Li Y and Oldenburg D W 1996 3D inversion of magnetic data Geophysics 61 394–408
[14] Lu W, Leung S and Qian J 2015 An improved fast local level set method for three-dimensional inverse gravimetry Inverse Problems Imaging 9 479–509
[15] Munschy M, Boulanger D, Ulrich P and Boufflane M 2007 Magnetic mapping for the detection and characterization of UXO: use of multi-sensor fluxgate three-axis magnetometers and methods of interpretation J. Appl. Geophys. 61 168–83
[16] Nemirovski A, Juditsky A, Lan G and Shapiro A 2009 Robust stochastic approximation approach to stochastic programming SIAM J. Optim. 19 1574–609
[17] Stanley O and Fedkiw R 2006 Level Set Methods and Dynamic Implicit Surfaces vol 153 (Berlin: Springer)
[18] Robbins H and Monro S 1951 A stochastic approximation method Ann. Math. Stat. 22 400–7
[19] Safran I and Shamir O 2020 How good is SGD with random shuffling? Conf. Learning Theory (PMLR) pp 3250–84
[20] Tan C, Ma S, Dai Y-H and Qian Y 2016 Barzilai–Borwein step size for stochastic gradient descent Proc. 30th Int. Conf. Neural Information Processing Systems pp 685–93
[21] Zhao H-K, Chan T, Merriman B and Osher S 1996 A variational level set approach to multiphase motion J. Comput. Phys. 127 179–95
[22] Zhdanov M S, Cuma M, Wilson G A and Polomé L 2012 3D magnetization vector inversion for SQUID-based full tensor magnetic gradiometry SEG Technical Program Expanded Abstracts 2012 (Society of Exploration Geophysicists) pp 1–5