Variational data assimilation for the initial-value dynamo problem

Kuan Li and Andrew Jackson
Institute of Geophysics, ETH Zurich, CH-8092 Zurich, Switzerland

Philip W. Livermore
School of Earth and Environment, Leeds University, Leeds, GB-LS2 9JT, United Kingdom
(Received 24 April 2011; revised manuscript received 16 August 2011; published 23 November 2011)

The secular variation of the geomagnetic field as observed at the Earth's surface results from the complex magnetohydrodynamics taking place in the fluid core of the Earth. One way to analyze this system is to use the data in concert with an underlying dynamical model of the system through the technique of variational data assimilation, in much the same way as is employed in meteorology and oceanography. The aim is to discover an optimal initial condition that leads to a trajectory of the system in agreement with observations. Taking the Earth's core to be an electrically conducting fluid sphere in which convection takes place, we develop the continuous adjoint forms of the magnetohydrodynamic equations that govern the dynamical system together with the corresponding numerical algorithms appropriate for a fully spectral method. These adjoint equations enable a computationally fast iterative improvement of the initial condition that determines the system evolution. The initial condition depends on the three dimensional form of quantities such as the magnetic field in the entire sphere. For the magnetic field, conservation of the divergence-free condition for the adjoint magnetic field requires the introduction of an adjoint pressure term satisfying a zero boundary condition. We thus find that solving the forward and adjoint dynamo system requires different numerical algorithms. In this paper, an efficient algorithm for numerically solving this problem is developed and tested for two illustrative problems in a whole sphere: one is a kinematic problem with prescribed velocity field, and the second is associated with the Hall-effect dynamo, exhibiting considerable nonlinearity. The algorithm exhibits reliable numerical accuracy and stability. Using both the analytical and the numerical techniques of this paper, the adjoint dynamo system can be solved directly with the same order of computational complexity as that required to solve the forward problem. These numerical techniques form a foundation for ultimate application to observations of the geomagnetic field over the time scale of centuries.

DOI: 10.1103/PhysRevE.84.056321 PACS number(s): 47.65.−d, 05.45.−a, 41.20.−q

I. INTRODUCTION

In 1919, the British scientist Larmor first proposed that the magnetic fields of the Sun and Earth are generated and sustained by a complex and nonlinear magnetohydrodynamical (MHD) process [1], known as dynamo action. In the Earth's dynamo system, the magnetic field is generated by, for example, a buoyancy-driven convecting flow of molten iron, confined to the spherical shell between the Earth's inner core and overlying solid mantle. In order for the geomagnetic field to be sustained, the amplifying influence of induction due to the electrically conducting flow must overcome the natural tendency of the magnetic field to decay (e.g., [2]). In general, the magnetic field reacts back on the convecting iron through the Lorentz force resulting in a highly complex nonlinear system.

In order to understand more aspects of the Earth's dynamo system, such as the spatial distributions of the velocity, magnetic, and temperature fields [3], the convection pattern [4], possible MHD wave propagation in the core [5], and magnetic secular variation, more studies from different perspectives are required, for instance, by tackling the inverse problem. The Earth's dynamo system continuously creates a magnetic field, which is recorded in rocks, lavas, and sediments and also observed by observatories and satellites. These observations are represented by two models: *gufm1* (for the past 400 years) [6] and *cals7k* (for past 7000 years) [7]. For Earth, the electrical conductivity of the mantle is several orders of magnitude weaker than that of the core; hence Earth's mantle is assumed as an insulator and thus, the surface observed data represented by *gufm1* and *cals7k* can be downward projected onto the core mantle boundary (CMB). Continuous data in time of the radial component of the magnetic field at the CMB are thus given as a function of spherical harmonic degree l and order m, where $l \leqslant 13$ [8]. These data provide information on the dynamo system and help to infer the inner working of Earth's core. Notice that the magnetic field is divergence-free and Earth's mantle is assumed as an insulator, thus a scalar observation in the radial direction is equivalent to observation of the magnetic field in all three directions at the CMB.

When simplifying assumptions about the material properties of the core are made, the evolution of the magnetic field becomes an initial-value problem, whose trajectory at all times is governed by the initial conditions. Such a system lends itself well to the technique of data assimilation, in either its variational or sequential forms. Data assimilation is a powerful tool for understanding the evolution of a time-varying physical system. For example, it can be used to determine properties of control parameters, e.g., in seismology [9,10], in meteorology [11], and recently in fluid control [12] for retrieving unknown control parameters or the initial conditions of the dynamic system. In meteorology, it has become a standard approach to use observations in conjunction with a dynamical model to attempt to find the correct initial condition that leads to an evolution of the atmosphere in best agreement with the observations. The final state in this evolution, derived

using observations necessarily earlier than the present, is then evolved into the future, generating predictions of the future state of the atmosphere, the weather forecast.

Despite the wide application of variational data assimilation in other fields of Earth science, its application to Earth's dynamo system only recently began [13]. In this landmark paper [13], the authors carried out a one-dimensional (1D) toy model study as a proof of the possibility of applying a discrete variational method to the MHD system. This work was further advanced by Canet *et al.* [14], who developed a variational formalism for a specific dynamical model of the core, the quasigeostrophic model. This work focuses on short time scale dynamics, and thus approximates the core as a perfect electrical conductor; is has been the basis for the recent remarkable discovery of torsional oscillations in the core with time scales on the order of six years [15]. The common thread to these works is the setting of the problem in a discrete form and the derivation of the adjoint system under this discrete form. Because of the large storage requirement and heavy computational burden in solving the discrete adjoint nonlinear problem [13], it is our purpose to develop the adjoint dynamo system in terms of continuous functions and also design the corresponding algorithms, which carry out the computation of the adjoint model more efficiently.

An allied, equally influential strand of data assimilation has been developed in a geomagnetic context in parallel with the variational approach, based on sequential data assimilation (see, e.g., [16]). This effort began with the work of Sun *et al.* [17] and Liu *et al.* [18] and has been recently reviewed in Fournier *et al.* [19] and Kuang and Tangborn [20]. This approach is based on the use of classical sequential data assimilation methods being applied to a geodynamo model [21] to create a framework for guiding the trajectory via observational constraints [22,23]. The results are so encouraging that the approach has culminated in a sequential data assimilation model [24] contributing to the predicted secular variation of the most recent International Geomagnetic Reference Field [25].

A criterion must be introduced measuring the agreement between physical observations and predictions of the model. There is considerable flexibility in the way this agreement is measured [e.g., [10,26] but a very convenient measure is based on a weighted squared deviation, often termed χ^2. Regardless of the definition of goodness of fit between the observations and the model prediction, the process of estimating the parameters governing a dynamical system by taking into account observations is usually termed *data assimilation* (4Dvar).

The functional χ^2 is called the misfit and the optimal prediction occurs at the global minimum of χ^2; finding this minimum is, in general, a nonlinear process. Generally speaking, the adjoint assimilation relies on deriving the adjoint system of the forward model, which is, mathematically speaking, the total derivative of χ^2 with respect to the unknowns constrained by the forward model. The adjoint model is driven by the combination of observations and predictions backward in time, and its value at the initial time is the downhill direction of the misfit with respect to the unknown control parameters which define the initial condition. The current values of the unknown control parameters are updated recursively in this downhill direction. Hence, a most important aspect of the analysis involves deriving and computing the adjoint model.

The format of the paper is as follows. In Sec. II, we discuss the fundamental mathematics, including the general mathematical framework of the adjoint method for the initial-value problem, the choice of the boundary condition of the adjoint field, and the matrix representation of the linear operator and its adjoint. Section III describes the applications to a kinematic dynamo and to an illustrative Hall-effect nonlinear problem, from which we understand the adjoint form of the divergence-free equation of magnetic induction; we also develop the proper numerical algorithm for computing the adjoint system. Some numerical results are reported in Sec. IV. Based on these results, we derive the adjoint system of the dynamo problem in Sec. V. Finally, in Sec. VI, we draw conclusions and discuss the outlook for further work.

II. MEASUREMENTS AND AN UNDERLYING DYNAMICAL MODEL

A. Geomagnetic data

Variations of the magnetic field have been recorded either by direct human observation for the last few hundred years [6] or through the magnetization of rocks and artifacts over the last few thousand years [7]. These observations have sensitivity to the magnetic field originating in the core of the Earth. We adopt the standard approximation of assuming the mantle to be an electrical insulator. Then the vector observations of the magnetic field at the Earth's surface $\mathbf{B}$ are related to the radial component of the magnetic field at the core surface through a Green's function appropriate for the solution of Laplace's equation in a spherical geometry, subject to Neumann boundary conditions [27,28]. The ill-posedness of this problem is clearly understood, and its remedies through the technique of regularization [e.g., [29] are routinely implemented. Geomagnetic data are inherently imperfect measurements and are subject to the presence of noise. Thus one normally chooses to fit the data under a least-squares criterion where each datum is weighted inversely by its estimated error.

For the purposes of our study we do not deal with the intricacies of dealing with real data; instead we assume this problem has been solved and that what is available is the spherical harmonic expansion of the radial magnetic field B_r at the core-mantle boundary, or equivalently the values of the spectral components of the poloidal scalar on the core surface [see Eq. (14)]. This shortcut of sidestepping the difficulties attendant to real data is in accord with the previous initial work of others [e.g., [22]].

B. The variational data assimilation method for initial-value problems

Consider a time-evolving physical system given by the following governing equation:

$$\frac{\partial}{\partial t}\mathbf{P} = X(\mathbf{P}), \tag{1}$$

where $\mathbf{P}(\mathbf{r},t)$ is a vector from which observables can be extracted, satisfying a given boundary condition, and X is

a spatial operator. Depending on the physical system, $X(\bullet)$ could be a linear or a nonlinear operator acting on a target field $\bullet$. Thus in the examples detailed in this paper, for a diffusion process, X is a linear operator, namely, $\nabla^2\bullet$; for a kinematic dynamo, X is a linear operator given by $\nabla\times(\mathbf{u}\times\bullet)+\nabla^2\bullet$, where $\mathbf{u}$ is the velocity of the flow; and for a Hall-effect problem, X is nonlinear given by $\nabla\times[(\mathbf{u}+\nabla\times\bullet)\times\bullet]+\nabla^2\bullet$.

The dynamical system is defined in the volume V and time interval τ. The vector $\mathbf{P}$ is a function of the location $\mathbf{r}$ and time t and is uniquely determined by the initial and boundary conditions. Without losing generality, we assume the scalar square integrable function $y(\mathbf{r},t)$ is a measured property of the dynamical system.

Our aim is to find the predictions from $\mathbf{P}(\mathbf{r},t)$ that agree best with the observed data y. First, we define the inner product in $\mathbf{L}^2$ (Lebesgue square integrable) as the integral over the volume. For $f,g\in\mathbf{L}^2$, the inner product is

$$\langle f,g\rangle=\int_V wfg\,dV, \tag{2}$$

where w is the integration weight, and hence we define a misfit χ^2, measuring the disagreement between observations and predictions within the observation time window $t\in[0,\tau]$, as

$$\begin{aligned}\chi^2&=\frac{1}{2}\int_V\int_{t=0}^{t=\tau}w\,[O(\mathbf{P})-y]^2\,dV\,dt\\&=\frac{1}{2}\int_{t=0}^{\tau}\langle O(\mathbf{P})-y,O(\mathbf{P})-y\rangle dt,\end{aligned} \tag{3}$$

where O is the observation operator, which generates the prediction $O(\mathbf{P})$ at the same positions in space and time as the measurements $y(\mathbf{r},t)$.

In Eq. (3), the misfit χ^2 depends purely on the initial condition, and an optimal solution (i.e., a local or global minimum) for the initial condition $\mathbf{P}_0=\mathbf{P}(\mathbf{r},t=0)$ is given by $\nabla_{\mathbf{P}_0}\chi^2=0$, where $\nabla_{\mathbf{P}_0}\chi^2$ is known as the gradient of χ^2 with respect to the initial condition $\mathbf{P}_0$.

However, in general, it is hard to compute $\nabla_{\mathbf{P}_0}\chi^2$ directly, since $\mathbf{P}$ is governed by the dynamic system (1) and the associated boundary conditions. One standard approach is to introduce a Lagrange multiplier $\mathbf{P}^\dagger$ [16,30] and define a new constrained functional χ^2 by augmenting Eq. (3) with the dynamical constraints to give

$$\begin{aligned}\chi^2=&\frac{1}{2}\int_{t=0}^{\tau}\langle O(\mathbf{P})-y,O(\mathbf{P})-y\rangle dt\\&+\int_{t=0}^{\tau}\left\langle\mathbf{P}^\dagger,\frac{\partial}{\partial t}\mathbf{P}-X(\mathbf{P})\right\rangle dt.\end{aligned} \tag{4}$$

Mathematically speaking, $\mathbf{P}^\dagger$ is also known as the adjoint field of $\mathbf{P}$.

1. Continuous approach

The differential of a functional in a Banach space is called Gâteaux differentiation [31] and is consequently so defined in a Hilbert space too, e.g. $\mathbf{L}^2$. For example, the Gâteaux differential of $O(\mathbf{P})$ in the direction of $\mathbf{q}$ is defined as

$$D\,(O(\mathbf{P}))=\lim_{\epsilon\to0}\frac{O(\mathbf{P}+\epsilon\mathbf{q})-O(\mathbf{P})}{\epsilon}=O(\mathbf{q}). \tag{5}$$

where D stands for Gâteaux differentiation and in order to keep our discussion simple, we assume that the observation operator O is linear in our paper. Notice that $\mathbf{q}$ is an arbitrary direction and can be equivalently represented considering O as the identity operator as $\mathbf{q}=\lim_{\epsilon\to0}\frac{\mathbf{P}+\epsilon\mathbf{q}-\mathbf{P}}{\epsilon}=D\mathbf{P}$, according to the definition of Gâteaux differential in (5). Therefore, we can write $D\,(O(\mathbf{P}))=O(\mathbf{q})=O(D\mathbf{P})$. Differentiating the constrained misfit χ^2 in the direction $D\mathbf{P}$, we have

$$\begin{aligned}D\chi^2=&\int_{t=0}^{\tau}\langle O(D\mathbf{P}),O(\mathbf{P})-y\rangle dt\\&+\int_{t=0}^{\tau}\left\langle\mathbf{P}^\dagger,\frac{\partial}{\partial t}D\mathbf{P}-X'(D\mathbf{P})\right\rangle dt,\end{aligned} \tag{6}$$

where X' is the linearized differential of X, known as the tangent linear operator [13].

Integrating Eq. (6) by parts, we look for adjoint operators $\frac{\partial}{\partial t}$ and $[X']^\dagger$ and find

$$\begin{aligned}D\chi^2=&[\langle D\mathbf{P},\mathbf{P}^\dagger\rangle]_{t=0}^{\tau}\\&+\int_{t=0}^{\tau}\left\langle D\mathbf{P},-\frac{\partial}{\partial t}\mathbf{P}^\dagger-[X']^\dagger(\mathbf{P}^\dagger)+O^\dagger\,[O(\mathbf{P})-y]\right\rangle dt,\end{aligned}$$

where $O^\dagger$ is the adjoint observation operator satisfying

$$\langle O(D\mathbf{P}),O(\mathbf{P})-y\rangle=\langle D\mathbf{P},O^\dagger\,[O(\mathbf{P})-y]\rangle$$

and $[X']^\dagger$ is the adjoint operator of X' satisfying

$$\langle\mathbf{P}^\dagger,X'(D\mathbf{P})\rangle=\langle[X']^\dagger(\mathbf{P}^\dagger),D\mathbf{P}\rangle.$$

The boundary condition on $\mathbf{P}^\dagger$ must be chosen; see Sec. II D. Further imposing the terminal condition that $\mathbf{P}^\dagger=0$ at $t=\tau$, we have

$$\begin{aligned}D\chi^2=&-\langle\mathbf{P}_0^\dagger,D\mathbf{P}_0\rangle\\&+\int_{t=0}^{\tau}\left\langle D\mathbf{P},-\frac{\partial}{\partial t}\mathbf{P}^\dagger-[X']^\dagger(\mathbf{P}^\dagger)+O^\dagger[O(\mathbf{P})-y]\right\rangle dt.\end{aligned}$$

In a Hilbert space, Gâteaux differentiation can be further written in the form of an inner product [31]. For example, the Gâteaux differentiation of the misfit χ^2 can be written as $D\chi^2=\langle\nabla_{\mathbf{P}_0}\chi^2,D\mathbf{P}_0\rangle$. Hence the downhill direction with respect to the current prediction of the initial condition is

$$\nabla_{\mathbf{P}_0}\chi^2=-\mathbf{P}_0^\dagger, \tag{7}$$

where $\mathbf{P}^\dagger$ satisfies the following adjoint system:

$$-\frac{\partial}{\partial t}\mathbf{P}^\dagger=[X']^\dagger(\mathbf{P}^\dagger)-O^\dagger\,[O(\mathbf{P})-y]\,. \tag{8}$$

Notice that in mathematics, the term adjoint, denoted as †, is applied in several situations. The Lagrange multiplier $\mathbf{P}^\dagger$ is also known as the adjoint field of $\mathbf{P}$; the operator $[X']^\dagger$ is known as the adjoint operator of X'. Furthermore, in matrix language, the adjoint of a matrix $\mathcal{H}$ is also named as the transpose conjugate of the matrix $\mathcal{H}$, denoted as $\mathcal{H}^\dagger$.

2. *Discrete approach*

Let $\{\mathbf{q}_i\}$ be a complete set of vectors $i \to \infty$ that can be used to represent $\mathbf{P}$ in Eq. (1). The basis $\mathbf{q}_i$ is orthogonal with an inner product $\langle\ \rangle_2$ given by $\langle \mathbf{q}_i, \mathbf{q}_j \rangle_2 = \int_V w_2 \mathbf{q}_i \cdot \mathbf{q}_j \, dV = \delta_{i,j}$, where the integration is over the volume V and w_2 is a weight function.

The differential equation (1) can be approximated by a finite dimensional system and written in matrix form. And hence the forward problem (1) can be written in the spatially discretized form as

$$\frac{\partial}{\partial t}\mathbf{f} = \mathcal{M}_{\mathbf{f}}\mathbf{f}, \tag{9}$$

where $\mathbf{f} = [\langle \mathbf{q}_1, \mathbf{P} \rangle_2, \langle \mathbf{q}_2, \mathbf{P} \rangle_2, \ldots]^T$ is the discrete form of $\mathbf{P}$ and the matrix $\mathcal{M}_{\mathbf{f}}$ ($[\mathcal{M}_{\mathbf{f}}]_{i,j} = \langle \mathbf{q}_i, X\mathbf{q}_j \rangle_2$) is the matrix representation of X. Notice that if the operator X is nonlinear, the matrix representation of the operator X depends on the current state of $\mathbf{f}$.

Guaranteed by the Riesz representation theorem, the adjoint of bounded operators, such as matrices, always uniquely exists [32]. Hence, one can approximate the adjoint tangent linear operator $[X']^\dagger$ in Eq. (8) by their matrix representations. Let us denote the matrix representation of $[X']^\dagger$ by Λ, where $\Lambda_{i,j} = \langle \mathbf{q}_i, [X']^\dagger \mathbf{q}_j \rangle_2$. The adjoint system in Eqs. (7) and (8) can be written as

$$\begin{aligned} \nabla_{\mathbf{f}_0}\chi^2 &= -\mathbf{f}_0^\dagger, \\ -\frac{\partial}{\partial t}\mathbf{f}^\dagger &= \Lambda\mathbf{f}^\dagger - \mathcal{O}^\dagger(\mathcal{O}\mathbf{f} - \mathbf{y}), \end{aligned} \tag{10}$$

where $\mathbf{f}^\dagger = [\langle \mathbf{q}_1, \mathbf{P}^\dagger \rangle_2, \langle \mathbf{q}_2, \mathbf{P}^\dagger \rangle_2, \ldots]^T$ is the discrete form of $\mathbf{P}^\dagger$, $\mathcal{O}$ is the discrete form of the observation operator O, and $\mathbf{y}$ is the set of observations.

Notice that we derive the continuous adjoint system using the inner product $\langle\ \rangle$ and we discretize the continuous system using $\langle\ \rangle_2$. If these two inner products are identical, i.e., $\langle\ \rangle = \langle\ \rangle_2$ (requiring $w = w_2$), one can easily show

$$\mathcal{Z}_{i,j} = \langle \mathbf{q}_i, X'\mathbf{q}_j \rangle = \langle [X']^\dagger \mathbf{q}_i, \mathbf{q}_j \rangle = \Lambda_{j,i} \tag{11}$$

by the definition of the adjoint operator, where $\mathcal{Z}$ is the matrix representation of the tangent linear X'. Hence $\Lambda = \mathcal{Z}^\dagger$ and furthermore, the continuous adjoint and discrete adjoint are entirely numerically equivalent.

C. Numerical approaches and their computational complexity

We have shown the equivalence of the continuous and discrete approaches in solving the adjoint system in Sec. II B 2, if X' is discretized and $[X']^\dagger$ is derived and discretized in the same Hilbert space. It means that the adjoint operator can be treated as a black box. Instead of deriving the analytical form of the adjoint operator $[X']^\dagger$, one could solve the adjoint system by simply taking the transpose conjugate of the matrix representation of X' [13].

Generally speaking, ordinary differential equations (ODEs) or partial differential equations (PDEs) can be solved by the following two different approaches using spectral methods [33]: (1) a matrix-free algorithm, which requires computing the forward and inverse spectral transforms and (2) a matrix-based algorithm, which requires computing the matrix representation of the operators in ODEs or PDEs.

Suppose we would like to evolve a three-dimensional (3D) geodynamo and its adjoint system with the spatial truncation $N_{\max} = L_{\max} = m_{\max} = k$, where $N_{\max}$, $L_{\max}$, and $m_{\max}$ are the maximal degree and order in the radial, colatitude, and longitude directions in spherical coordinates. Using the first method, at each time step, the spatial transform requires εk^4 operations [34,35], where ε is a prefactor. In contrast, the matrix-based algorithm requires computing the matrix representation of the spatial operators of dimension $2k^2(k+2)$ [35]. Since the system is nonlinear, at each time step the matrix representations of the nonlinear operators have to be recomputed, and computing each column requires $\varepsilon_2 k^4$ operations. Therefore, at each time step the computational complexity in computing the matrix representations is bounded by $\varepsilon_3 k^7$, which is k^3 slower than method (1), where ε_3 is a prefactor. Typical spatial resolutions for geodynamo calculations are several hundred basis functions in each spatial direction [36]. If $k = 100$, solving the geodynamo and its adjoint system using method (1) is 10^6 times faster than method (2). Furthermore, the matrix dimension for $k = 100$ resolution is about 2×10^6 and occupies more than 20 terabytes in storage, which is extremely difficult for current computer clusters to handle. Therefore, method (2) is not directly applicable to large complex systems like the geodynamo problem. The conventional way to solve the adjoint system is to mechanically adjoint the forward code [37,38]. However, the geodynamo is a very complex system and the computer code is programed and parallelized in a sophisticated way. It would be very challenging to adjoint the dynamo codes by hand [19]. Therefore, we are interested in developing the continuous adjoint dynamo system and the corresponding numerical algorithm.

D. Updating the estimated initial condition

Having the derivative of χ^2 in hand (Sec. II B), we can now optimize the initial condition, so that the fit to the data is improved. This can be carried out recursively by updating the current initial condition estimate $\mathbf{P}_0^{(n)}$ using a simple descent algorithm,

$$\mathbf{P}_0^{(n+1)} = \mathbf{P}_0^{(n)} - \mu\nabla_{\mathbf{P}_0^{(n)}}\chi^2 = \mathbf{P}_0^{(n)} + \mu\mathbf{P}_0^\dagger, \tag{12}$$

where μ is the searching length and $\mathbf{P}_0^\dagger$ is the solution of the adjoint system (8) at the nth iteration of $\mathbf{P}_0$.

1. *Constraints on the adjoint field* $\mathbf{P}^\dagger$ *in 4Dvar*

It must be noted that the boundary condition of $\mathbf{P}^\dagger$ is not arbitrary. In Eq. (12), the uphill direction of the misfit is given as $\nabla_{\mathbf{P}_0}\chi^2 = -\mathbf{P}_0^\dagger$, which leads to an update of the boundary condition of $\mathbf{P}_0^{(n)}$. Thus when deriving the adjoint system or directly computing the matrix representation of the adjoint operator, there are two principles that need to be satisfied:

(1) If the field $\mathbf{P}$ is a divergence-free or curl-free field, i.e., $\nabla \cdot \mathbf{P} = 0$ or $\nabla \times \mathbf{P} = 0$, the adjoint field $\mathbf{P}^\dagger$ has to be divergence-free or curl-free.

(2) The term adjoint field is not allowed to change the boundary condition of the estimated initial condition of the

forward problem. Let us use $K(\mathbf{P}) = c$ to represent the boundary condition of $\mathbf{P}$. For both linear homogeneous and inhomogeneous boundary conditions of $\mathbf{P}$, one can have $K(\mathbf{P} + \mu\mathbf{P}^\dagger) = K(\mathbf{P}) + \mu K(\mathbf{P}^\dagger) = c$ for any μ. Since $K(\mathbf{P}) = c$, the adjoint field $\mathbf{P}^\dagger$ has to vanish as $K(\mathbf{P}^\dagger) = 0$. For example, if the field $\frac{\partial}{\partial r}\mathbf{P} + \mathbf{P} = \mathbf{H}$ at the boundary $r = 1$, the adjoint field $\mathbf{P}^\dagger$ has to satisfy $\frac{\partial}{\partial r}\mathbf{P}^\dagger + \mathbf{P}^\dagger = 0$ at $r = 1$.

We use a Galerkin method in this paper (see Sec. III A and Appendix A) in which every one of the radial basis functions satisfies the boundary conditions. This gives us the flexibility to use a more general updating scheme than Eq. (12), whereby μ in Eq. (12) is replaced by a preconditioning matrix [39].

III. DERIVATION OF ADJOINT OPERATORS FOR INDUCTIVE SYSTEMS

In this section we introduce two problems of increasing nonlinearity to illustrate how the adjoint systems can be derived. We evolve magnetic fields subject to specified physical laws, and show how their initial conditions can be recovered. The first example has a linear spatial operator X, and the second has a nonlinear operator. The motivation for these examples is the following. The example with a linear operator X (the induction equation with a prescribed flow $\mathbf{u}$) is the simplest possible physically relevant example. Despite this, it of course exhibits considerable complexity in its evolution. We devise our second problem to have nonlinear evolution, since the spatial operator in the given Hall-effect problem now depends quadratically on the magnetic field. These examples serve to illustrate the critical problems inherent in the full dynamo equations, which are ultimately tackled in Sec. V.

A. Example 1: The kinematic dynamo and its adjoint

We nondimensionalize the magnetic induction equation (e.g., [2]) by choosing U and L for the characteristic velocity and length scales. Defining a magnetic Reynolds number, $R_m = UL/\eta$, the induction equation including a possible α-effect term (where α is a mean-field α effect parametrizing the interactions of the small scales) is written as [2]

$$\frac{\partial \mathbf{B}}{\partial t} = R_m \nabla \times [\mathbf{u} \times \mathbf{B} + \alpha(\mathbf{r})\mathbf{B}] + \nabla^2 \mathbf{B}, \tag{13}$$

where $\mathbf{B}$ is the magnetic field, $\mathbf{u}$ is the velocity field, and the general tensor α has been assumed to be isotropic. In our study, the kinematic dynamo is defined in a unit sphere and we assume the flow is incompressible, thus both $\mathbf{u}$ and $\mathbf{B}$ are divergence-free and can be uniquely represented using a poloidal and toroidal decomposition. For example, in a spherical geometry with coordinates (r,θ,ϕ), the magnetic field can be written as

$$\mathbf{B} = \sum_{(n,l,m)} \left(a_{(n,l,m)}\, {}_n\mathbf{S}_l^m + b_{(n,l,m)}\, {}_n\mathbf{T}_l^m \right),$$

where $\{a_{(n,l,m)}\}$ and $\{b_{(n,l,m)}\}$ are spectral coefficients, the poloidal and toroidal vectors are given by [28]

$$\begin{aligned} {}_n\mathbf{S}_l^m &= \nabla \times \nabla \times \left[\Phi_n^l(r) Y_l^m(\theta,\phi)\hat{\mathbf{r}}\right], \\ {}_n\mathbf{T}_l^m &= \nabla \times \left[\Psi_n^l(r) Y_l^m(\theta,\phi)\hat{\mathbf{r}}\right], \end{aligned} \tag{14}$$

and where $\hat{\mathbf{r}}$ is the unit vector in the r direction; Φ_n^l and Ψ_n^l are radial basis functions with degree n less than or equal to $N_{\max}$, the radial truncation level. In colatitude θ and longitude ϕ, the spherical harmonics $Y_l^m(\theta,\phi)$ [28] that we use are real and fully normalized in solid angle and have degree l and order m satisfying the truncation $0 \leqslant m \leqslant l \leqslant L_{\max}$, for some specified truncation $L_{\max}$. We may further indicate either a cosine or sine azimuthal dependence of a nonaxisymmetric harmonic by the addition of either "s" or "c" as a superscript: e.g., ${}_n\mathbf{S}_l^{mc}$.

When the sphere is surrounded by an electrical insulator, the radial basis functions must obey the following boundary conditions [40]:

$$\begin{aligned} d\Phi_n^l/dr + l\Phi_n^l &= 0 \quad \text{at} \quad r = 1, \\ \Psi_n^l &= 0 \quad \text{at} \quad r = 1, \end{aligned} \tag{15}$$

where the definitions of Φ_n^l and Ψ_n^l can be found in Appendix A.

For the sake of simplicity, we initially treat the case where $\alpha = 0$ and use nonslip boundary conditions for $\mathbf{u}$, thus $\mathbf{u} = 0$ at $r = 1$. Let us denote $L_1(\mathbf{B}) = R_m \nabla \times (\mathbf{u} \times \mathbf{B})$ and $L_2(\mathbf{B}) = \nabla^2 \mathbf{B}$. Namikawa and Matsushita [41] found the adjoint kinematic dynamo system, where the inner product is defined by the energy norm, and in mathematical language the adjoint operators can be written as

$$L_1^\dagger(\mathbf{B}^\dagger) = R_m(\nabla \times \mathbf{B}^\dagger) \times \mathbf{u} \quad \text{and} \quad L_2^\dagger(\mathbf{B}^\dagger) = \nabla^2 \mathbf{B}^\dagger, \tag{16}$$

where the diffusion operator L_2 is self-adjoint and the boundary condition for $\mathbf{B}^\dagger$ stays the same as that for $\mathbf{B}$, namely, the electrical insulating boundary condition (15). However, of most interest is the adjoint operator $L_1^\dagger$. Using the vector identity $\nabla \cdot (\mathbf{a} \times \mathbf{b}) = -(\nabla \times \mathbf{b}) \cdot \mathbf{a} + (\nabla \times \mathbf{a}) \cdot \mathbf{b}$ and integrating by parts, we have

$$\begin{aligned} \langle \mathbf{B}^\dagger, \nabla \times (\mathbf{u} \times \mathbf{B}) \rangle &= \int_V \mathbf{B}^\dagger \cdot [\nabla \times (\mathbf{u} \times \mathbf{B})]\, dV \\ &= \int_\Sigma [(\mathbf{u} \times \mathbf{B}) \times \mathbf{B}^\dagger] \cdot d\boldsymbol{\Sigma} \\ &\quad + \int_V \mathbf{B} \cdot [(\nabla \times \mathbf{B}^\dagger) \times \mathbf{u}] dV, \end{aligned} \tag{17}$$

where the inner product $\langle\ \rangle$ is defined in Eq. (2) with unit integration weight $w = 1$ and $d\boldsymbol{\Sigma}$ is the surface element. Since $\mathbf{u} = 0$ at $r = 1$, the surface integral vanishes, thus we have

$$\langle \mathbf{B}^\dagger, L_1(\mathbf{B}) \rangle = R_m \int_V \mathbf{B} \cdot [(\nabla \times \mathbf{B}^\dagger) \times \mathbf{u}] dV = \langle L_1^\dagger(\mathbf{B}^\dagger), \mathbf{B} \rangle. \tag{18}$$

Notice that $\nabla \cdot \mathbf{B} = 0$ and $\nabla \cdot L_1(\mathbf{B}) = 0$; hence L_1 is the linear mapping from divergence-free field back to itself. Following our remarks in Sec. II D 1, the adjoint operator $L_1^\dagger$ has to be divergence-free for the initial-value problem. However, the adjoint term $(\nabla \times \mathbf{B}^\dagger) \times \mathbf{u}$ does not satisfy such a condition. Also notice that $\mathbf{B}$ is uniquely represented by the poloidal and toroidal fields (14), hence the linear operator L_1 is densely defined in $\mathbf{L}^2$ and its adjoint, $L_1^\dagger = R_m(\nabla \times \bullet) \times \mathbf{u}$, is unique [42]. Therefore, we need to introduce another

constraint, i.e., $\nabla \cdot \mathbf{B} = 0$ to the misfit functional χ^2 in order to annihilate the non-divergence-free part. We write

$$\chi^2 = \frac{1}{2}\int_{t=0}^{\tau} \langle O(\mathbf{B}) - y, O(\mathbf{B}) - y\rangle dt + \int_{t=0}^{\tau} \left\langle \mathbf{B}^\dagger, \left(\frac{\partial}{\partial t} - L_1 - L_2\right)(\mathbf{B})\right\rangle dt + \int_{t=0}^{\tau} \langle p^\dagger, \nabla \cdot \mathbf{B}\rangle dt,$$

where $p^\dagger$ is a new Lagrange multiplier.

Taking the Gâteaux differentiation, integrating by parts and introducing the terminal boundary condition of $\mathbf{B}^\dagger = 0$ at $t = \tau$, we have

$$\nabla_{\mathbf{B}_0}\chi^2 = -\mathbf{B}_0^\dagger,$$

$$0 = \int_{t=0}^{\tau}\left\langle D\mathbf{B}, \frac{\partial \mathbf{B}^\dagger}{\partial t}\right\rangle dt + \int_{t=0}^{\tau}\langle D\mathbf{B}, O^\dagger[O(\mathbf{B}) - y]\rangle dt + \int_{t=0}^{\tau}\langle D\mathbf{B}, R_m(\nabla \times \mathbf{B}^\dagger) \times \mathbf{u} + \nabla^2\mathbf{B}^\dagger\rangle dt + \int_{t=0}^{\tau}\langle D\mathbf{B}, -\nabla p^\dagger + \hat{\mathbf{r}}\, p^\dagger \delta(r-1)\rangle dt, \quad (19)$$

where the derivation of the last term on the right-hand side of Eq. (19) is

$$\langle \nabla \cdot D\mathbf{B}, p^\dagger\rangle = \int_V (\nabla \cdot D\mathbf{B}) p^\dagger\, dV = -\int_V D\mathbf{B} \cdot \nabla p^\dagger dV + \int_V \nabla \cdot (D\mathbf{B} p^\dagger) dV = -\int_V D\mathbf{B} \cdot \nabla p^\dagger dV + \int_V D\mathbf{B} \cdot \hat{\mathbf{r}}\, p^\dagger \delta(r-1) dV = \langle D\mathbf{B}, -\nabla p^\dagger + \hat{\mathbf{r}}\, p^\dagger \delta(r-1)\rangle. \quad (20)$$

Notice that Eq. (20) is true for any arbitrary boundary condition of $p^\dagger$. However, the zero boundary condition, i.e., $p^\dagger|_{r=1} = 0$ is the simplest one, since the boundary term $\int_\Sigma p^\dagger D\mathbf{B} \cdot \hat{\mathbf{r}} d\Sigma = \langle D\mathbf{B} \cdot \hat{\mathbf{r}}\, p^\dagger \delta(r-1)\rangle$ on the right-hand side of Eq. (19) vanishes, where Σ is the core surface. Therefore the complete form of the adjoint induction operator is

$$L_1^\dagger(\mathbf{B}^\dagger) = R_m(\nabla \times \mathbf{B}^\dagger) \times \mathbf{u} - \nabla p^\dagger, \quad (21)$$

where the Lagrange multiplier $p^\dagger$ plays the role of the adjoint pressure and

$$\nabla^2 p^\dagger = R_m \nabla \cdot [(\nabla \times \mathbf{B}^\dagger) \times \mathbf{u}]. \quad (22)$$

Therefore, the adjoint dynamo model can be written as

$$\nabla_{\mathbf{B}_0}\chi^2 = -\mathbf{B}_0^\dagger,$$

$$-\frac{\partial \mathbf{B}^\dagger}{\partial t} = R_m(\nabla \times \mathbf{B}^\dagger) \times \mathbf{u} - \nabla p^\dagger + \nabla^2\mathbf{B}^\dagger - O^\dagger[O(\mathbf{B}) - y],$$

where the adjoint pressure term satisfies Eq. (22) with the zero boundary condition $p^\dagger|_{r=1} = 0$. For a general 3D vector field, there is a subset $\mathbf{s}$, where each element in $\mathbf{s}$ is both divergence-free and curl-free, i.e., $\nabla \times \nabla \times [r^{l+1}Y_l^m(\theta,\phi)\hat{\mathbf{r}}] \in \mathbf{s}$ in spherical coordinates. Hence, the adjoint pressure together with its associated boundary condition determines how much of the field in $\mathbf{s}$ belongs to the adjoint induction term $L_1^\dagger(\mathbf{B}^\dagger)$ and how much of that needs to be removed together with the purely curl-free field from $L_1^\dagger(\mathbf{B}^\dagger)$ in order to maintain the divergence-free condition (21). It is pertinent to note that adjoint dynamo equations were obtained by Roberts [43], Gibson and Roberts [44], and Kono and Roberts [45], but for an adjoint magnetic field in a finite conductor that obeys different boundary conditions to the original magnetic field. Thus these interesting adjoint equations cannot be applied to the problem at hand.

In some situations, the flow is assumed to be stress-free at the boundary or completely inviscid, thus the velocity field does not vanish at the boundary, and only a nonpenetration condition applies, i.e., $u_r = 0$. Thus the adjoint operator $L_1^\dagger$ has to contain a boundary term. Converting the surface term into a flux injection term, we can rewrite the surface integral in Eq. (17) as

$$\int_\Sigma [(\mathbf{u} \times \mathbf{B}) \times \mathbf{B}^\dagger] \cdot \hat{\mathbf{r}}\, d\Sigma = \int_V \mathbf{B} \cdot [(\mathbf{u} \cdot \mathbf{B}^\dagger)\hat{\mathbf{r}} - \mathbf{B}^\dagger u_r]\delta(r-1) dV = \int_V \mathbf{B} \cdot \hat{\mathbf{r}}(\mathbf{u} \cdot \mathbf{B}^\dagger)\delta(r-1) dV,$$

where $\hat{\mathbf{r}}(\mathbf{u} \cdot \mathbf{B}^\dagger)$ is the flux injection which drives the adjoint system.

We now consider the modifications to the inductive term that are required when the α-effect term is present. The adjoint system with an α effect can be written as

$$\int_V \mathbf{B}^\dagger \cdot \nabla \times (\alpha\mathbf{B}) dV = \int_\Sigma (\alpha\mathbf{B} \times \mathbf{B}^\dagger) \cdot d\boldsymbol{\Sigma} + \int_V \mathbf{B} \cdot \alpha(\nabla \times \mathbf{B}^\dagger) dV, \quad (23)$$

where the boundary term can be further written as a flux injection term as

$$\int_\Sigma (\alpha\mathbf{B} \times \mathbf{B}^\dagger) \cdot d\boldsymbol{\Sigma} = \int_V \mathbf{B} \cdot \alpha(\mathbf{B}^\dagger \times \hat{\mathbf{r}})\delta(r-1) dV.$$

Hence, the adjoint operator of the α-effect term reads

$$L_\alpha^\dagger(\mathbf{B}^\dagger) = R_m[\alpha(\mathbf{B}^\dagger \times \hat{\mathbf{r}})\delta(r-1) + \alpha(\nabla \times \mathbf{B}^\dagger)].$$

Equation (23) is considerably simplified when α is a constant: the surface term vanishes, and indeed the operator L_α is self-adjoint as we now demonstrate. Using the poloidal and toroidal decomposition, we find that the two integrals $\alpha\int_V \mathbf{B}^\dagger \cdot (\nabla \times \mathbf{B}) dV$ and $\alpha\int_V \mathbf{B} \cdot \left(\nabla \times \mathbf{B}^\dagger\right) dV$ are nonzero, only if the following two conditions are satisfied [40]: (1) $l = l'$ and $m = m'$, where l and m are the spherical harmonic degree and order of $\mathbf{B}^\dagger$, and l' and m' are the spherical harmonic degree and order of $\mathbf{B}$ and (2) if $\mathbf{B}^\dagger$ is poloidal, $\mathbf{B}$ has to be toroidal, and vice versa.

Since between $\mathbf{B}$ and $\mathbf{B}^\dagger$, one field is toroidal and thus has to vanish at $r = 1$, the surface integral vanishes at $r = 1$. Thus the operator L_α is self-adjoint.

In summary, the complete form of the adjoint magnetic induction system reads

$$\begin{aligned} &L_1^\dagger(\mathbf{B}^\dagger)+L_2^\dagger(\mathbf{B}^\dagger)+L_\alpha^\dagger(\mathbf{B}^\dagger)\\ &\quad=-\nabla p^\dagger+R_m[(\nabla\times\mathbf{B}^\dagger)\times\mathbf{u}+\alpha(\nabla\times\mathbf{B}^\dagger)]+\nabla^2\mathbf{B}^\dagger\\ &\qquad+R_m[\hat{\mathbf{r}}\,(\mathbf{u}\cdot\mathbf{B}^\dagger)\delta(r-1)+\alpha(\mathbf{B}^\dagger\times\hat{\mathbf{r}})\delta(r-1)], \end{aligned}$$

where L_α is the linear operator for the α-effect term [2].

Having the adjoint operator of the induction equation in hand, we may write the adjoint system of Eq. (13) as follows:

$$\begin{aligned} &\nabla_{\mathbf{B}_0}\chi^2=-\mathbf{B}_0^\dagger,\\ &-\frac{\partial\mathbf{B}^\dagger}{\partial t}=R_m[(\nabla\times\mathbf{B}^\dagger)\times\mathbf{u}+\hat{\mathbf{r}}\,(\mathbf{u}\cdot\mathbf{B}^\dagger)\delta(r-1)]+\nabla^2\mathbf{B}^\dagger\\ &\qquad+R_m[\alpha(\nabla\times\mathbf{B}^\dagger)+\alpha(\mathbf{B}^\dagger\times\hat{\mathbf{r}})\delta(r-1)]-\nabla p^\dagger\\ &\qquad-O^\dagger[O(\mathbf{B})-y], \end{aligned}\tag{24}$$

where O stands for the observation operator, y stands for the observations, the boundary condition of $\mathbf{B}^\dagger$ stays the same as $\mathbf{B}$, namely, an electrically insulating boundary condition (15), and $p^\dagger$ vanishes at $r=1$ and satisfies $\nabla^2 p^\dagger=R_m\nabla\cdot[(\nabla\times\mathbf{B}^\dagger)\times\mathbf{u}]$.

B. Example 2: The Hall-effect dynamo model and its adjoint

The Hall-effect model of magnetic field evolution of a neutron star was first proposed by Jonse [46], where the nondimensionalized governing equation reads [47]

$$\frac{\partial\mathbf{B}}{\partial t}=R_B\nabla\times[(\mathbf{u}-\nabla\times\mathbf{B})\times\mathbf{B}]+\nabla^2\mathbf{B}.\tag{25}$$

The Hall parameter R_B [48] measures the ratio of the Hall-effect time scale $\tau_H=\frac{B}{en_eR^2}$ against the magnetic diffusion time scale $\tau_D=R^2/\eta$, $R_B=\frac{\tau_H}{\tau_D}=\frac{B}{en_e\mu_0\eta}$ in SI units, where R is the radius of the star. The whole Hall system is nondimensionalized using τ_D for the time scale. In the definition of R_B, B is the typical magnetic field strength, μ_0 is the magnetic permeability, e is the charge of the electron, n_e is the electron's number density, and η is the magnetic diffusivity. In order to focus on the Hall effect, we set $\mathbf{u}=0$.

We choose a vertical boundary condition on the magnetic field, i.e., $B_\theta=B_\phi=0$ at $r=1$. It can be shown that the Hall-effect term does not create or annihilate energy and the energy of the whole system in the unit sphere will monotonically decay. Integrating by parts, we have the energy of the system in the unit sphere,

$$\begin{aligned} &\frac{1}{2}\frac{d}{dt}\int_V\mathbf{B}^2\,dV=-\int_V R_B\,\mathbf{B}\cdot\nabla\times[(\nabla\times\mathbf{B})\times\mathbf{B}]\\ &-\mathbf{B}\cdot\nabla^2\mathbf{B}\,dV=-\int_V(\nabla\times\mathbf{B})^2\,dV<0. \end{aligned}\tag{26}$$

For this boundary condition, the radial basis functions $I_n^l(r)$ and $\Psi_n^l(r)$ of the poloidal and toroidal magnetic fields have to satisfy

$$\frac{dI_n^l(r)}{dr}=0\quad\text{and}\quad\Psi_n^l(r)=0\quad\text{at}\quad r=1.\tag{27}$$

In Appendix A, we define the spectral basis functions $I_n^l(r)$ and $\Psi_n^l(r)$ for the whole sphere satisfying these boundary conditions.

In light of the discussions of Sec. III A, we are in a position to quickly derive the adjoint of Eq. (25). The misfit functional χ^2 constrained by the Hall-effect model and the divergence-free condition of the magnetic field can be written as

$$\begin{aligned} \chi^2=&\frac{1}{2}\int_{t=0}^{\tau}\langle O(\mathbf{B})-y,O(\mathbf{B})-y\rangle dt+\int_{t=0}^{\tau}\langle p^\dagger,\nabla\cdot\mathbf{B}\rangle dt\\ &+\int_{t=0}^{\tau}\left\langle\mathbf{B}^\dagger,\frac{\partial\mathbf{B}}{\partial t}+R_B\nabla\times[(\nabla\times\mathbf{B})\times\mathbf{B}]-\nabla^2\mathbf{B}\right\rangle dt, \end{aligned}$$

where the inner product $\langle\ \rangle$ is defined as the volume integral over the unit sphere, the time integral is over the whole observation time window $t\in[0,\tau]$, and the Lagrange multiplier $p^\dagger$ is the adjoint pressure. Taking the total derivative of χ^2 with respect to $\mathbf{B}_0$, integrating by parts and imposing the terminal condition that $\mathbf{B}^\dagger=0$ at $t=\tau$ and a zero boundary condition on $p^\dagger$ ($p^\dagger|_{r=1}=0$), one finds

$$\begin{aligned} &\nabla_{\mathbf{B}_0}\chi^2=-\mathbf{B}_0^\dagger,\\ &0=\int_{t=0}^{\tau}\left\langle D\mathbf{B},-\frac{\partial\mathbf{B}^\dagger}{\partial t}+R_B(\nabla\times\mathbf{B}^\dagger)\times(\nabla\times\mathbf{B})\right.\\ &\qquad+R_B\nabla\times[\mathbf{B}\times(\nabla\times\mathbf{B}^\dagger)]-\nabla^2\mathbf{B}^\dagger+\nabla p^\dagger\\ &\qquad\left.+O^\dagger[O(\mathbf{B})-y]\right\rangle dt+\mathsf{S}_1+\mathsf{S}_2+\mathsf{S}_3, \end{aligned}$$

where the three boundary terms are

$$\begin{aligned} \mathsf{S}_1&=R_B\int_{t=0}^{\tau}\int_\Sigma\{[(\nabla\times D\mathbf{B})\times\mathbf{B}]\times\mathbf{B}^\dagger\}\cdot d\boldsymbol{\Sigma}\,dt,\\ \mathsf{S}_2&=R_B\int_{t=0}^{\tau}\int_\Sigma\{[\mathbf{B}\times(\nabla\times\mathbf{B}^\dagger)]\times D\mathbf{B}\}\cdot d\boldsymbol{\Sigma}\,dt,\\ \mathsf{S}_3&=R_B\int_{t=0}^{\tau}\int_\Sigma\{[(\nabla\times\mathbf{B})\times D\mathbf{B}]\times\mathbf{B}^\dagger\}\cdot d\boldsymbol{\Sigma}\,dt. \end{aligned}\tag{28}$$

The integrands of the three boundary terms are either orthogonal to $\mathbf{B}$ or $\mathbf{B}^\dagger$, thus they do not have radial components and the boundary terms vanish. The adjoint system can be represented as

$$\begin{aligned} &\nabla_{\mathbf{B}_0}\chi^2=-\mathbf{B}_0^\dagger,\\ &-\frac{\partial\mathbf{B}^\dagger}{\partial t}=-R_B\{(\nabla\times\mathbf{B}^\dagger)\times(\nabla\times\mathbf{B})+\nabla\times[\mathbf{B}\times(\nabla\times\mathbf{B}^\dagger)]\}\\ &\qquad+\nabla^2\mathbf{B}^\dagger-\nabla p^\dagger-O^\dagger[O(\mathbf{B})-y], \end{aligned}\tag{29}$$

where

$$\begin{aligned} &p^\dagger|_{r=1}=0,\\ &\nabla^2p^\dagger=-R_B\nabla\cdot[(\nabla\times\mathbf{B}^\dagger)\times(\nabla\times\mathbf{B})], \end{aligned}$$

and the boundary condition on $\mathbf{B}^\dagger$ stays the same as that for $\mathbf{B}$, namely, Eq. (27).

C. Numerical method

We discretize the spatial part of the magnetic induction term $\nabla\times(\mathbf{u}\times\mathbf{B})$, the Hall-effect term $\nabla\times[(\nabla\times\mathbf{B})\times\mathbf{B}]$, and their adjoints using the Galerkin method [33]. Each poloidal and toroidal radial basis function satisfies the electrically insulating boundary condition for the kinematic dynamo and

a vertical boundary condition for the Hall-effect problem. A similar approach can be found in [35]. The Laplace operator ∇^2 is discretized using the same poloidal and toroidal scalars at the same resolution.

As we discussed in Sec. II B 2, if the continuous adjoint system is derived and discretized in the same Hilbert space, the continuous and discrete approaches are numerically equivalent. In our analysis, we derive the adjoints for the kinematic dynamo and Hall-effect problems using the energy norm. Hence if we discretize these two systems using the same energy norm, the continuous and discrete adjoint are equivalent. In a spherical geometry, the incompressible flow $\mathbf{u}$ can also be expanded in poloidal and toroidal fields,

$$\mathbf{u} = \sum_{l,m} \nabla \times \nabla \times \left[s_l^m(r) Y_l^m(\theta,\phi)\hat{\mathbf{r}}\right] + \nabla \times \left[t_l^m(r) Y_l^m(\theta,\phi)\hat{\mathbf{r}}\right],$$

where s_l^m and t_l^m are poloidal and toroidal scalars, respectively. We first check the numerical accuracy of the adjoint algorithm for computing $-\nabla p^\dagger + (\nabla \times \mathbf{B}^\dagger) \times \mathbf{u} + \hat{\mathbf{r}}\ (\mathbf{u}\cdot\mathbf{B}^\dagger)\delta(r-1)$, where the flow is incompressible and vanishes at the surface of the sphere ($r = 1$) and we take the poloidal and toroidal scalars for the representation of the flow $\mathbf{u}$ as [49]

$$s_2^0(r) = c_2 r^3 (1-r^2)^3, \qquad t_1^0(r) = c_1 r^2 (1-r^2). \tag{30}$$

The coefficients are $c_1 = 8.107\,929\,179\,422\,066$ and $c_2 = 1.193\,271\,237\,996\,972$ in fully normalized real spherical harmonics. We demonstrate the equivalence of the discrete and continuous approaches to this problem by computing the matrix representation $\mathcal{M}$ of $\nabla \times (\mathbf{u} \times \mathbf{B})$ using the algorithm in Appendix B2a and by computing the matrix Λ of $-\nabla p^\dagger + (\nabla \times \mathbf{B}^\dagger) \times \mathbf{u} + \hat{\mathbf{r}}\ (\mathbf{u}\cdot\mathbf{B}^\dagger)\delta(r-1)$ using Appendix B2b, where analytically $\Lambda = \mathcal{M}^\dagger$. Notice that the flow is axisymmetric, thus $\mathcal{M}$ and Λ are block diagonal with each block representing a different spectral order m. We compute the largest numerical error of each nonzero entry of $\mathcal{M}^\dagger$ and Λ, where the relative error is $E = \max|\frac{M_{j,i}-\Lambda_{i,j}}{M_{j,i}}|$. For the highest resolutions tested, $N_{\max} = L_{\max} = 30$ and $m = 0$, the numerical algorithm exhibits machine precision, namely, the relative error E is found to be less than 10^{-13}. Notice that the spectral transform in the ϕ direction is carried out via fast Fourier transform using FFTW [50], which is a well known stable and fast algorithm, thus there is no need to test the nonaxisymmetric modes other than $m = 0$.

The time stepping is carried out via a Crank-Nicolson scheme, where the diffusion term is treated implicitly and the magnetic induction and Hall-effect terms are treated explicitly for both forward and adjoint problems. For example, the induction equation in time is discretized as

$$\mathbf{B}_{i+1} = \left(1 - \frac{\Delta t}{2}\nabla^2\right)^{-1} \times \left[\mathbf{B}_i + \Delta t R_m \nabla \times (\mathbf{u} \times \mathbf{B}_i) + \frac{\Delta t}{2}\nabla^2 \mathbf{B}_i\right].$$

The minimization of the misfit is carried out using a limited memory quasi-Newton method (L-BFGS). Approximating the misfit χ^2 up to the quadratic term, we have

$$\chi^2(\mathbf{P_0} + \Delta\mathbf{P_0}) \approx \chi^2(\mathbf{P_0}) + \nabla_{\mathbf{P}_0}\chi^2(\mathbf{P_0})\cdot\Delta\mathbf{P_0} + \beta\tfrac{1}{2}\Delta\mathbf{P_0}\cdot\mathcal{H}\cdot\Delta\mathbf{P_0},$$

where $\mathcal{H}$ is known as the Hessian and the parameter β is chosen to satisfy the Wolfe conditions [39]. Then the Newton step $\Delta\mathbf{P}_0$ satisfies

$$\nabla_{\mathbf{P}_0}\chi^2(\mathbf{P}_0 + \Delta\mathbf{P}_0) = \nabla_{\mathbf{P}_0}\chi^2(\mathbf{P}_0) + \beta\mathcal{H}\cdot\Delta\mathbf{P}_0 = 0. \tag{31}$$

Hence the improvement of the initial condition is $\Delta\mathbf{P}_0 = -\beta\mathcal{H}^{-1}\nabla_{\mathbf{P}_0}\chi^2(\mathbf{P}_0)$, where the inverse Hessian $\mathcal{H}^{-1}$ is estimated and gradually improved in the minimization step. In our numerical study, we have not directed much effort to the optimal choice of parameters (e.g., β) in Eq. (31). We simply use the algorithm to demonstrate gradual improvement of the estimation of the initial condition. Our Galerkin method ensures that the correct boundary conditions are always adhered to. The numerical algorithm is based on Nocedal [51] and the software package is acquired from [52].

IV. VALIDATION OF THE ADJOINT SYSTEMS

A. Synthetic data generation

In the following examples we demonstrate the ability of our algorithms to correctly recover the initial conditions of some physical systems. To do so we generate synthetic data under two scenarios. In the first, which we refer to as two-dimensional (2D) observations, we generate data corresponding to the value of the spectral coefficients of the poloidal scalar $\mathbf{S}_l^m$ on the core surface for all degrees and orders of spherical harmonic up to the model resolution at regular time intervals. This corresponds exactly to the real-life situation of boundary value observations. In the second scenario we supply values of the spectral coefficients of the poloidal scalar ${}_n\mathbf{S}_l^m$ at regular time intervals. This corresponds to full knowledge of the poloidal field within the core (hence we term this 3D observations), but there is no information concerning the toroidal field. We use this test to discover whether we have sufficient sensitivity to find the toroidal field in the core.

We perform the closed-loop testing by first defining the true initial condition of the kinematic dynamo and Hall-effect system. Starting from the truth, the dynamical system generates a trajectory $\mathbf{B}(\mathbf{B}_0,t)$ in phase space, where $\mathbf{B}_0$ is the initial condition of $\mathbf{B}$. Although we carry out our computations in terms of nondimensionalized variables, we will report our results here in dimensional units (years), pertinent to the Earth. The conversion requires specification of the magnetic diffusivity and size of the conducting core; we take $\eta = 1.5\ \mathrm{m^2\ s^{-1}}$ and $r = 3500$ km. The lowest decay modes for a core with insulating or vertical boundary conditions have eigenvalues $\lambda = \pi^2$ and $\lambda = 7.5$, respectively; the corresponding decay times $\frac{r^2}{\lambda\eta}$ are 26 000 and 34 000 years. Neglecting this difference of 30%, we take the decay time of the slowest mode to be 30 000 years in both cases. We choose different combinations of observation time and different observation techniques for study, where the observation time window is either 7000 years or 30 000 years and the observation technique is either 2D or 3D. We define the discrete misfit as

$$\chi^2 = \frac{1}{2}\sum_i [\mathcal{O}\cdot\mathbf{a}_i - \mathbf{y}_i]^T \cdot [\mathcal{O}\cdot\mathbf{a}_i - \mathbf{y}_i], \tag{32}$$

where $\mathbf{a}_i$ is the coefficient list of the poloidal magnetic field and the observed data at the CMB ($r = 1$) in spectral space (l,m) at

time t_i, $\mathbf{y}_i$ is the list of observations at t_i, and $\mathcal{O}$ is the discretized observation operator. $\mathcal{O}$ splits in the spherical degree l and m and hence is a block-diagonal matrix for both scenarios. For 2D observations, each diagonal matrix is actually a row vector, $\mathcal{O}_{l,m} = [\frac{l(l+1)\Phi_1^l}{r^2}, \frac{l(l+1)\Phi_2^l}{r^2}, \ldots, \frac{l(l+1)\Phi_n^l}{r^2}]_{r=1}$, and for 3D observations, each $\mathcal{O}_{l,m}$ is also a diagonal matrix, $\mathcal{O}_{l,m} = \text{diag}[\Phi_1^l, \Phi_2^l, \ldots, \Phi_n^l]_{r=1}$, where Φ_n^l is the poloidal radial basis.

B. Example 1—Kinematic dynamo numerical results

We choose Eq. (30) as the velocity of the kinematic dynamo. Notice the flow is axisymmetric, thus the magnetic field decouples in spherical harmonic order m. When solving axisymmetric induction problems, the spatial computation complexity increases as the cube of the spatial resolution. For fully 3D problems, the computational complexity scales as spatial resolution to the power of 4 [35]. Furthermore, the time step is a decreasing function of spatial degree, $\max[N_{\max}, L_{\max}]$. We set the true initial condition of $\mathbf{B}$ to be $a_{(1,1,1c/s)} = 1$ and $b_{(1,1,1c/s)} = 1$, where $\mathbf{a}_{(n,l,m_{c/s})}$ and $\mathbf{b}_{(n,l,m_{c/s})}$ are poloidal and toroidal coefficients, and set the starting estimate of the initial condition to be $\mathbf{B}_0^{(0)} = 0$. Limited by single CPU computing power, we choose the spatial resolution for the kinematic dynamo to be $N_{\max} = L_{\max} = 15$ for the $m = 1$ component (since the solution is entirely $m = 1$) and the magnetic Reynolds number to be $R_m = 50$. The optimal time step is about 10^{-4}, found experimentally.

We observe the radial component of the magnetic field everywhere exactly on the CMB ($r = 1$) for each spherical harmonic degree l and order m up to the model resolution every 100 years for 30 000 years.

The initial condition of the magnetic field can be retrieved well. Figure 1 illustrates the true initial condition (dashed black) and the rebuilt initial condition (solid gray) of the magnetic field at the 500th iteration. The misfit drops by 11 orders of magnitude in 500 iterations (Fig. 2), where χ^2 is renormalized by its value at the first iteration. Notice that there is no information concerning the toroidal field in the observed data. The toroidal field is retrieved, due to the inductive coupling [40], where the poloidal velocity interacts with the toroidal magnetic field and creates the poloidal magnetic field.

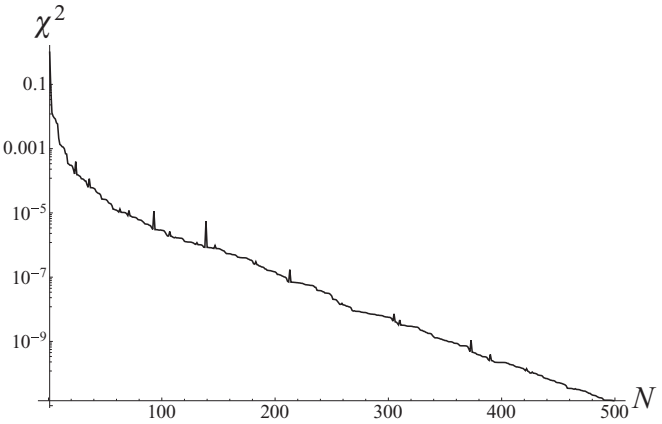

FIG. 2. For the test case of a kinematic dynamo problem for $R_m = R'_m = 50$, the reduction of the misfit χ^2 as a function of the number of iterations N, where χ^2 is normalized by its value at χ^2 of the first iteration.

In the 4DVar framework, we implicitly assume that the dynamical equations are a perfect description of the underlying physics. However, we can illustrate the sensitivity to model error by altering the value of R_m from its correct value. To illustrate this we try to assimilate using an incorrect value for R_m, namely, $R'_m = 45$. Figure 3 illustrates the true initial condition (dashed black lines) and the rebuilt initial condition (solid gray lines) of the magnetic field at the 200th iteration. The misfit χ^2 drops by five orders of magnitude in about 200 iterations (Fig. 4) and saturates at this level, where χ^2 is renormalized by its value at the first iteration. We are still able to retrieve the poloidal field well, since the observations are directly of the poloidal magnetic field and $R'_m = 45$ is just 10% different from the true value $R_m = 50$. However, due to the lack of direct observations of the toroidal field, the retrieved toroidal field differs somewhat from the truth.

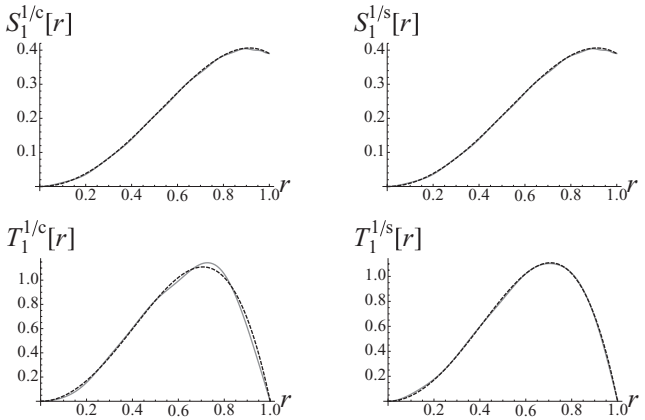

FIG. 1. The test case of a kinematic dynamo for $R_m = R'_m = 50$: the true versus the rebuilt initial condition for an axisymmetric kinematic dynamo represented by the poloidal and toroidal radial scalars of the $m = 1$ component, where the dashed black lines stand for the true initial condition and the solid gray lines stand for the rebuilt initial condition after 500 iterations.

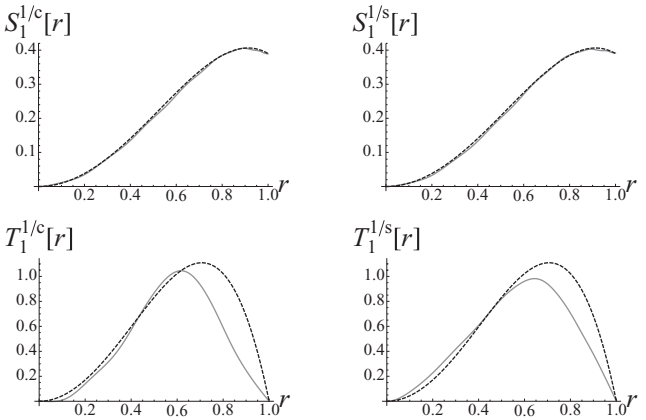

FIG. 3. The test case of a kinematic dynamo for $R_m = 50$ and $R'_m = 45$: the true versus the rebuilt initial condition for an axisymmetric kinematic dynamo represented by the poloidal and toroidal radial scalars of the $m = 1$ component, where the dashed black lines stand for the true initial condition and the solid gray lines stand for the rebuilt initial condition after 200 iterations.

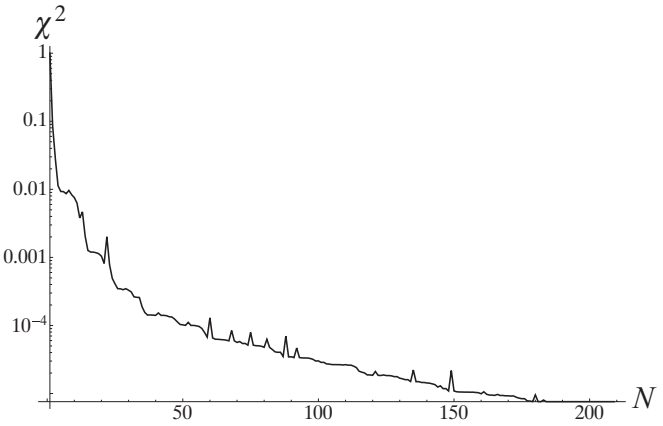

FIG. 4. For the test case of a kinematic dynamo problem for $R_m = 50$ and $R'_m = 45$, the reduction of the misfit χ^2 as a function of the number of iterations N, where χ^2 is normalized by its value at χ^2 of the first iteration. The misfit saturates at the 10^{-5} level in around 200 iterations.

It is perhaps interesting to remark that both the initial condition $\mathbf{B}_0$ and the control parameter R_m can be retrieved using the variational data assimilation framework. The mathematical derivation is similar to the pure initial-value problem that we treat extensively here, where the only difference is that the Gâteaux differential of χ^2 is not only with respect to the prediction of the initial condition $\mathbf{B}_0$ but also to the prediction of R_m. However, it is not the main scope of this paper, and we skip such mathematical derivations.

C. Example 2—Hall-effect problem

Eschewing for the present time a parallel approach, we use a purely serial code and choose a modest truncation of $N_{\max} = L_{\max} = m_{\max} = 5$ as the spatial resolution, the maximum Hall parameter is $R_B = 20$, and the initial condition is set as in Table I. The optimal time step in this setup is discovered to be about 10^{-4} for $R_B = 5$ and to be about 10^{-5} for $R_B = 20$ by experiment. In the numerical experiments, we carry out three test cases (see Table II), using two different observation techniques.

Figures 5–10 illustrate the reconstructed initial conditions compared with the true initial conditions and the reduction of χ^2 for these three cases, where the dashed black lines stand for the true state, and the solid gray lines are for the

TABLE I. The initial condition of the Hall-effect problem, where $\{a_{(n,l,m_{c,s})}\}$ and $\{b_{(n,l,m_{c,s})}\}$ are the spectral coefficients of the initial magnetic field $\mathbf{B}_0$.

$a_{(1,1,0)} = 0.4$	$b_{(1,1,0)} = 0.4$
$a_{(1,1,1/c)} = -0.25$	$b_{(1,1,1/c)} = -0.25$
$a_{(1,1,1/s)} = 0.25$	$b_{(1,1,1/s)} = 0.15$
$a_{(1,2,0)} = 0.35$	$b_{(1,2,0)} = 0.35$
$a_{(1,2,1/c)} = -0.1$	$b_{(1,2,1/s)} = -0.1$
$a_{(1,2,1/s)} = 0.15$	$b_{(1,2,1/s)} = 0.05$
$a_{(1,2,2/c)} = -0.05$	$b_{(1,2,2/c)} = 0.15$
$a_{(1,2,2/s)} = 0.1$	$b_{(1,2,2/s)} = 0.1$

TABLE II. Three test case setups for the Hall-effect problems, where $\Delta\tau$ is the time interval between two observations and τ is the time window. R_B is the Hall parameter and "obs" defines the observations to be either 2D (at $r = 1$) or 3D. The e-folding time of the Hall-effect problem with the vertical boundary condition is about 30 000 years.

	Obs	R_B	τ (yr)	$\Delta\tau$ (yr)	Figures
Case 1	3D	5	7000	15	5, 6
Case 2	2D	5	30000	60	7, 8
Case 3	2D	20	7000	15	9–11

rebuilt state. The 3D observation strategy is more accurate and efficient for rebuilding the initial condition, especially for the toroidal fields. The rebuilt poloidal fields overlap with the true states in Figs. 5, 7, and 9, due to the direct observations on poloidal fields. However, for the toroidal part, the 3D observation (Fig. 6) performs better than the 2D observation strategy (Fig. 8), because of the full knowledge of the poloidal field in space and time, which leads to stronger convexity of χ^2 for 3D observations. But with the help of stronger advection (nonlinear term), one is able to find out the nondirectly observable field (toroidal part) even in a shorter time window. The Hall effect (convection) in test case 3 is four times as strong as that for test case 2. Comparing these two cases, we just need a quarter of the observation time window (7000 years) for test case 3 to retrieve the poloidal and toroidal scalar (Fig. 10) as accurately as in test case 2 (Fig. 8) over 30 000 years. Furthermore, the efficiency of the inversion increases when increasing the nonlinearity (Figs. 6 and 8), and for 500 iterations, χ^2 of case 2 with $R_B = 5$ drops seven orders of magnitude, and in contrast, χ^2 of case 3 with $R_B = 20$ drops eight orders. Our ability to retrieve the initial condition is shown for case 3 in physical, rather than spectral space, in Fig. 11; the recovery of the $|\mathbf{B}|$ isosurface is excellent.

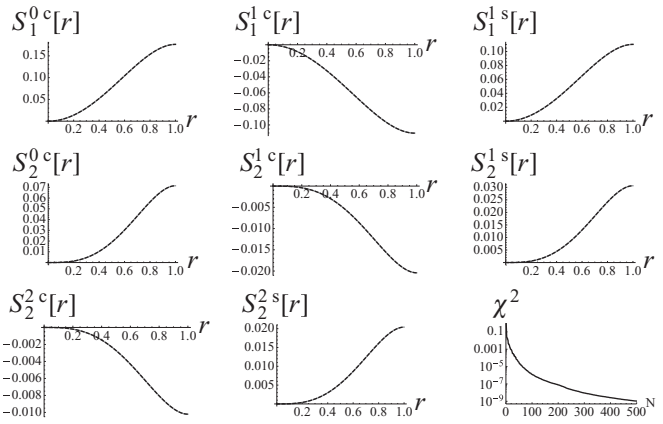

FIG. 5. For the Hall-effect problem, case 1: The rebuilt poloidal initial condition (in solid gray) against the true state (in dashed black) and the reduction of the misfit as a function of the iteration number N for the Hall-effect problem for 500 equally spaced 3D observations in 7000 years with $R_B = 5$.

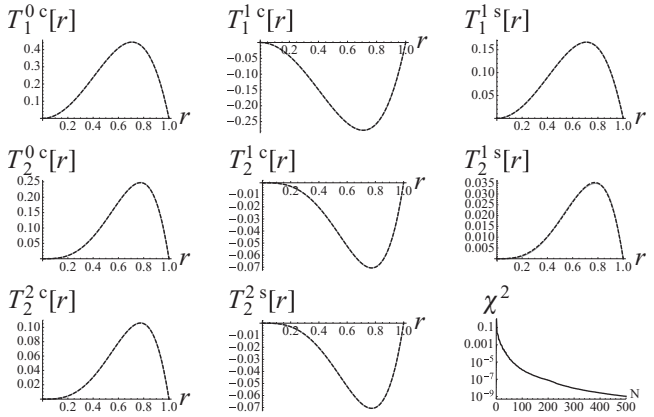

FIG. 6. For the Hall-effect problem, case 1: The rebuilt toroidal initial condition (in solid gray) against the true state (in dashed black) and the reduction of the misfit as a function of the iteration number N for the Hall-effect problem for 500 equally spaced 3D observations in 7000 years with $R_B = 5$.

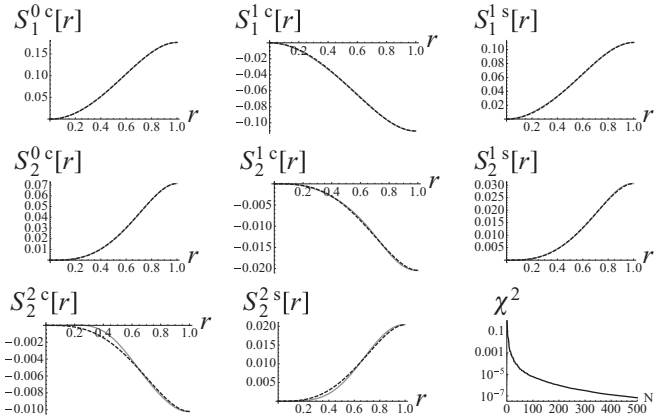

FIG. 7. For the Hall-effect problem, case 2: The rebuilt poloidal initial condition (in solid gray) against the true state (in dashed black) and the reduction of the misfit as a function of the iteration number N for the Hall-effect problem for 500 equally spaced 2D observations in 30 000 years with $R_B = 5$.

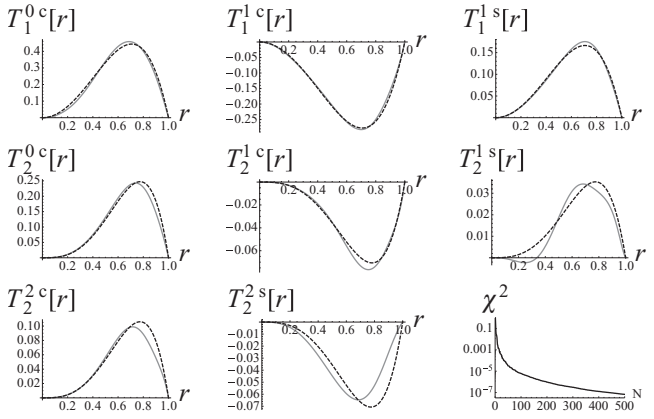

FIG. 8. For the Hall-effect problem, case 2: The rebuilt toroidal initial condition (in solid gray) against the true state (in dashed black) and the reduction of the misfit as a function of the iteration number N for the Hall-effect problem for 500 equally spaced 2D observations in 30 000 years with $R_B = 5$.

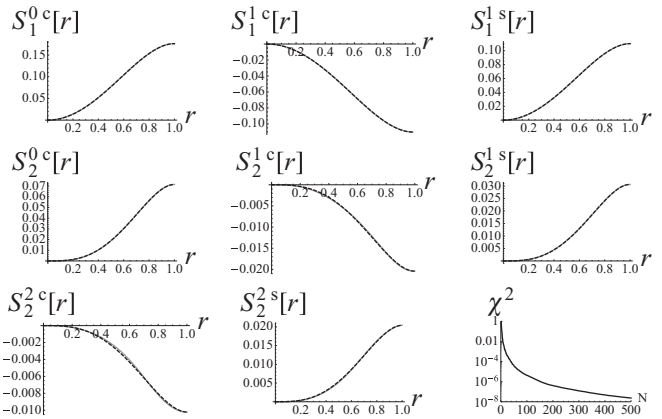

FIG. 9. For the Hall-effect problem, case 3: The rebuilt poloidal initial condition (in solid gray) against the true state (in dashed black) and the reduction of the misfit as a function of the iteration number N for the Hall-effect problem for 500 equally spaced 2D observations in 7000 years with $R_B = 20$.

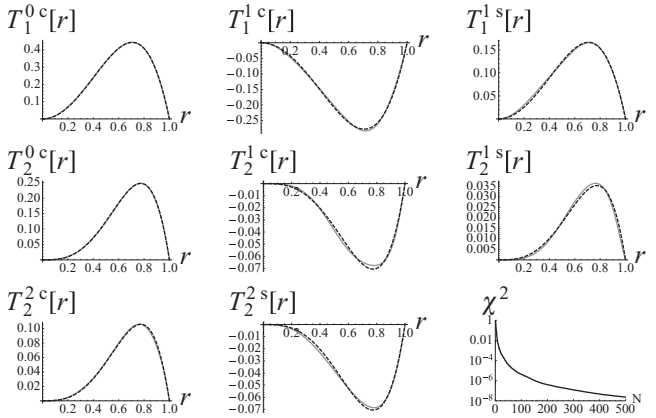

FIG. 10. For the Hall-effect problem, case 3: The rebuilt toroidal initial condition (in solid gray) against the true state (in dashed black) and the reduction of the misfit as a function of the iteration number N for the Hall-effect problem for 500 equally spaced 2D observations in 7000 years with $R_B = 20$.

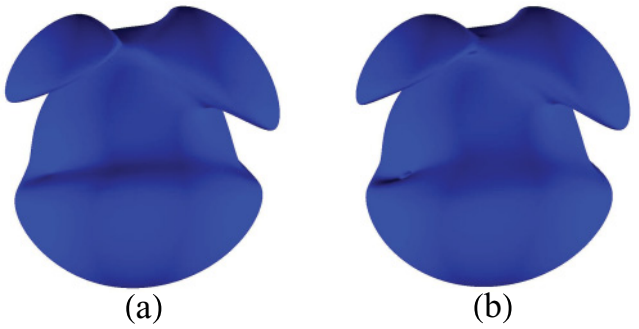

FIG. 11. (Color online) Isosurfaces for the Hall-effect problem, case 3. Shown is the $|\mathbf{B}| = 0.45$ isosurface for (a) the true initial state and (b) the recovered initial state from 2D observations.

V. CONTINUOUS ADJOINT OF THE BOUSSINESQ INCOMPRESSIBLE DYNAMO

The geodynamo system is governed by three coupled equations for velocity field $\mathbf{u}$, magnetic field $\mathbf{B}$, and temperature field T_c. The flow is assumed incompressible and vanishes at the CMB, $r = 1$. The Earth's mantle is a weak electrical conductor compared with the Earth's core, hence we consider the mantle as an electrical insulator and the magnetic field satisfies Eq. (15) at the CMB. We set the temperature field T_c as a constant at the CMB, hence the temperature field perturbation of the Earth's core, T, can be set as $T = 0$ at the CMB. The geodynamo system can be written as

$$\mathrm{Eq}_1 := \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{u} + 2\boldsymbol{\Omega}\times\mathbf{u} + \frac{1}{\rho}\nabla p - \frac{1}{\rho}\mathbf{J}\times\mathbf{B} - \nu\nabla^2\mathbf{u} - \alpha g T\hat{\mathbf{r}} = 0, \tag{33}$$

$$\mathrm{Eq}_2 := \frac{\partial \mathbf{B}}{\partial t} - \nabla\times(\mathbf{u}\times\mathbf{B}) - \eta\nabla^2\mathbf{B} = 0, \tag{34}$$

$$\mathrm{Eq}_3 := \frac{\partial T}{\partial t} + \mathbf{u}\cdot\nabla T - \kappa\nabla^2 T - h = 0, \tag{35}$$

where $\mathbf{J}$ is the electrical current density, ρ is the mass density of the core, ν is the kinematic viscosity, $\boldsymbol{\Omega}$ is the angular velocity, η is the magnetic diffusivity and κ is the thermal diffusivity, g is the gravitational acceleration, α is the thermal expansion coefficient, $\alpha g T\hat{\mathbf{r}}$ is the buoyancy force, and h is the internal heating. Applying the definition of the misfit and introducing the PDE constraints of the dynamo system and divergence-free conditions of $\mathbf{u}$ and $\mathbf{B}$, where adjoint fields are denoted as $\mathbf{u}^\dagger$, $\mathbf{B}^\dagger$, and $T^\dagger$, we have

$$\begin{aligned}\chi^2 = &\int_{t=0}^{\tau}\frac{1}{2}\langle O(\mathbf{B}) - y, O(\mathbf{B}) - y\rangle dt \\ &+ \int_{t=0}^{\tau}[\langle \mathbf{u}^\dagger, \mathrm{Eq}_1\rangle + \langle \mathbf{B}^\dagger, \mathrm{Eq}_2\rangle + \langle T^\dagger, \mathrm{Eq}_3\rangle]dt \\ &+ \int_{t=0}^{\tau}[\langle p_1^\dagger, \nabla\cdot\mathbf{u}\rangle + \langle p_2^\dagger, \nabla\cdot\mathbf{B}\rangle]dt,\end{aligned}$$

where the inner product $\langle\ \rangle$ is defined in Eq. (2) and the integration weight w is chosen as $w = 1$, $\mathbf{u}^\dagger$ vanishes at the CMB, $\mathbf{B}^\dagger$ satisfies the same insulating boundary condition (15) as $\mathbf{B}$, the boundary condition for $T^\dagger$ is $T^\dagger = 0$ at $r = 1$, and $p_1^\dagger$ and $p_2^\dagger$ are the adjoint pressure terms for $\mathbf{u}$ and $\mathbf{B}$. Taking the total derivative with respect to $\mathbf{u}_0$, $\mathbf{B}_0$, and T_0, integrating by parts and imposing the terminal condition that $\mathbf{u}^\dagger = 0$, $\mathbf{B}^\dagger = 0$, and $T^\dagger = 0$ and the zero conditions for $p_1^\dagger$ and $p_2^\dagger$, we have

$$\nabla_{\mathbf{u}_0}\chi^2 = -\mathbf{u}_0^\dagger,$$

$$\nabla_{\mathbf{B}_0}\chi^2 = -\mathbf{B}_0^\dagger,$$

$$\nabla_{T_0}\chi^2 = -T_0^\dagger,$$

provided that $\mathbf{u}^\dagger$, $\mathbf{B}^\dagger$, and $T^\dagger$ satisfy the adjoint system

$$\begin{aligned}0 = &-\frac{\partial \mathbf{u}^\dagger}{\partial t} + \nabla\times(\mathbf{u}\times\mathbf{u}^\dagger) + \mathbf{u}^\dagger\times(\nabla\times\mathbf{u}) + 2\mathbf{u}^\dagger\times\boldsymbol{\Omega} \\ &+ \nabla p_1^\dagger - \mathbf{B}\times(\nabla\times\mathbf{B}^\dagger) - \nu\nabla^2\mathbf{u}^\dagger + T\nabla T^\dagger, \\ 0 = &-\frac{\partial \mathbf{B}^\dagger}{\partial t} - (\nabla\times\mathbf{B}^\dagger)\times\mathbf{u} + \nabla p_2^\dagger \\ &- \frac{1}{\rho\mu_0}[\nabla\times(\mathbf{B}\times\mathbf{u}^\dagger) + \mathbf{u}^\dagger\times(\nabla\times\mathbf{B})] \\ &- \eta\nabla^2\mathbf{B}^\dagger + O^\dagger[O\mathbf{B} - y], \\ 0 = &-\frac{\partial T^\dagger}{\partial t} - \mathbf{u}\cdot\nabla T^\dagger - \alpha g\mathbf{u}^\dagger\cdot\hat{\mathbf{r}} - \kappa\nabla^2 T^\dagger,\end{aligned} \tag{36}$$

where both $\mathbf{u}^\dagger$ and $T^\dagger$ vanish at $r = 1$, $\mathbf{B}^\dagger$ satisfies the insulating boundary condition (15), and the adjoint pressure terms $p_1^\dagger$ and $p_2^\dagger$ with the zero boundary condition $p_1^\dagger|_{r=1} = p_2^\dagger|_{r=1} = 0$ satisfy

$$\begin{aligned}\nabla^2 p_1^\dagger = &\nabla\cdot[-\mathbf{u}^\dagger\times(\nabla\times\mathbf{u}) - 2\mathbf{u}^\dagger\times\boldsymbol{\Omega} \\ &+ \mathbf{B}\times(\nabla\times\mathbf{B}^\dagger) - T\nabla T^\dagger], \\ \nabla^2 p_2^\dagger = &\nabla\cdot\left[(\nabla\times\mathbf{B}^\dagger)\times\mathbf{u} + \frac{1}{\rho\mu_0}\mathbf{u}^\dagger\times(\nabla\times\mathbf{B})\right].\end{aligned}$$

It is perhaps of interest to consider the geodynamo system satisfying different velocity boundary conditions, e.g., stress-free [21], and different temperature boundary conditions [3], and to consider how our derivation could be implemented in the situation where the setting is a spherical shell rather than a full sphere. This situation is the most frequently considered situation because of physical relevance for the Earth (possessing, as it does, a solid inner core). By introducing adjoint pressure terms and boundary driving terms, adjoints of the Navier-Stokes equation, magnetic induction equation, and temperature equation can be derived using the same techniques as those used to obtain Eq. (36). However, deriving an adjoint for the Laplace operator for different temperature boundary conditions is slightly different and the detailed derivations can be found in Appendix C.

VI. DISCUSSION

In our variational data assimilation problem, the downhill direction for minimizing the misfit functional χ^2 is computed by solving the adjoint system. Our optimal approach is obtained by solving the continuous system. Our focus in this work has been to develop the adjoint geodynamo system and to develop the corresponding numerical algorithms for solving the geodynamo initial-value problem.

Interestingly, the adjoint dynamo system exhibits different analytical properties than the forward problem. Hence it requires different numerical algorithms for its solution. We apply our analysis and numerical algorithm to solve a kinematic dynamo and the Hall-effect initial-value problem and have found very encouraging results.

The numerical algorithms described in Appendix B for solving the forward dynamo problem and its adjoint are compatible with most geodynamo codes. Solving the adjoint dynamo system requires an extra Poisson solver. In a spherical geometry and using a spherical harmonic expansion, the 3D Poisson equation becomes a 1D ordinary differential equation with a zero boundary condition at $r = 1$.

For Earth's dynamo system, the velocity, magnetic, and temperature fields are highly nonlinearly coupled together and generate complex magnetic field variations in space and

time. Although the toroidal magnetic field is hidden 3000 km beneath the surface of the Earth, as we demonstrated in the Hall-effect problem, one is still able to rebuild the initial condition of the toroidal magnetic field. This is due to the mechanism in which a poloidal velocity field interacts with a toroidal magnetic field to create a poloidal magnetic field. In cases 2 and 3 of the Hall-effect problem, we use 2D observations of the radial component of the magnetic field at the CMB to rebuild the poloidal and toroidal initial conditions. In case 2, the convection due to the Hall effect is five times faster than the magnetic diffusion and we need about one diffusion time, or 30 000 years of magnetic data, to retrieve the initial condition of the magnetic field. In contrast, for case 3, the Hall effect is four times as fast as that for case 2 and we just need $1/4$ of a magnetic diffusion time for inversion. For Earth, the magnetic Reynolds number $R_m = UL/\eta$ is about 500, where U is the typical velocity assumed to be 20 km/yr [53], L is the radius of Earth's core, and η is the magnetic diffusivity. This means that, for Earth's dynamo, it takes about 100 yr, L/U, for the poloidal velocity field to push the toroidal magnetic field at the top of the inner core to the CMB, which is 25 times faster than the largest Hall-effect convection rate in our model, and therefore proportionately less data are required. Despite the geodynamo being a much more nonlinear dynamical system than the Hall-effect problem, we believe that historical magnetic data spanning several convection times (several hundred years) are of a suitable duration for geomagnetic data assimilation. Using these data, we expect to quantify the initial condition of the velocity, magnetic, and temperature fields. However, it still remains as an open question whether these several hundred years of data are sensitive enough to determine the initial condition of the core fields or not. Nevertheless, as we observed in the inversion of the kinematic dynamo model, without any knowledge of the toroidal magnetic field, the toroidal part of the rebuilt initial condition is somewhat sensitive to R_m. Presumably, for the geodynamo problem, the retrieved toroidal part of the magnetic field is sensitive not only to R_m but is also sensitive to the other control parameters, such as the Ekman number (the ratio of viscous force to Coriolis force) and the Rayleigh number (the ratio of the buoyancy force to the viscous force). To answer these questions, we need to implement the adjoint dynamo model for further numerical studies.

Central to our derivation is the use of an inner product defined in terms of volume integrals. In the case of a shell one must remain with this definition, however, there is no longer an obvious radial spectral basis $\Phi_i(r)$ with which to work. The orthogonality of the basis over the interval $[r_i, r_o]$ (where r_i and r_o are inner and outer core radii, respectively) would require

$$\int_{r_i}^{r_o} r^2 \Phi_n(r)\Phi_m(r)\, dr = \delta_{nm},$$

and no set of orthogonal polynomials immediately presents itself. One can of course use the Gram-Schmidt process to create such a basis, most likely complete with satisfaction of the boundary conditions, but the process does not appear entirely straightforward. In our initial study, we discretize the adjoint system in the same Hilbert space as we analytically derive the adjoint system. Hence, the continuous approach and discrete approach are numerically equivalent. However, the solutions of the PDEs should not depend on the numerical scheme. If the forward and the adjoint scheme are accurate enough, one could solve the forward and the adjoint system using a different method or different discretizations. This is the subject of future work.

ACKNOWLEDGMENTS

We are very grateful to J. Tromp who listened patiently to our problems and offered generous advice and encouragement. We also thank A. Fournier, J. Velimsky, and A. Fichtner for several stimulating discussions and suggestions. This work is partially supported by SNF Grant No. 200021-112293, ERC Grant No. 247303 ("MFECE"), and Grant No. NE/G014043/1.

APPENDIX A: POLOIDAL AND TOROIDAL FIELDS DEFINED USING THE ENERGY NORM

We choose the toroidal and poloidal radial basis functions, such that each mode is orthogonal to the others in the energy norm,

$$\begin{aligned}
\int_V \mathbf{\Phi}_{(n_1,l_1,m_1)} \cdot \mathbf{\Phi}_{(n_2,l_2,m_2)} dV &= \delta_{n_1,n_2}\delta_{l_1,l_2}\delta_{m_1,m_2},\\
\int_V \mathbf{\Psi}_{(n_1,l_1,m_1)} \cdot \mathbf{\Psi}_{(n_2,l_2,m_2)} dV &= \delta_{n_1,n_2}\delta_{l_1,l_2}\delta_{m_1,m_2},\\
\int_V \mathbf{\Phi}_{(n_1,l_1,m_1)} \cdot \mathbf{\Psi}_{(n_2,l_2,m_2)} dV &= 0,
\end{aligned} \tag{A1}$$

where $dV = r^2 \sin\theta\, d\theta\, d\phi\, dr$ is the volume element in spherical coordinate, $\mathbf{\Phi}_{(n,l,m)}$ and $\mathbf{\Psi}_{(n,l,m)}$ are defined as

$$\begin{aligned}
\mathbf{\Phi}_{(n,l,m)} &= \nabla \times \nabla \times (\Phi_n^l Y_l^m \hat{\mathbf{r}})\\
\mathbf{\Psi}_{(n,l,m)} &= \nabla \times (\Psi_n^l Y_l^m \hat{\mathbf{r}}),
\end{aligned}$$

and Φ_n^l and Ψ_n^l are the radial basis. Integrating in θ and ϕ, Eq. (A1) can be simplified as radial integrals for each l and m,

$$\begin{aligned}
\delta_{n_1,n_2} &= l(l+1)\int_{r=0}^{1}\left[\frac{l(l+1)}{r^2}\Phi_{n_1}^l\Phi_{n_2}^l + \frac{\partial\Phi_{n_1}^l}{\partial r}\frac{\partial\Phi_{n_2}^l}{\partial r}\right]dr,\\
\delta_{n_1,n_2} &= l(l+1)\int_{r=0}^{1}\Psi_{n_1}^l\Psi_{n_2}^l\, dr.
\end{aligned} \tag{A2}$$

The unnormalized poloidal and toroidal radial basis functions satisfying the insulating boundary conditions (15) are

$$\Phi_n^l \propto c_0 P_n^{(0,l+1/2)}(2r^2-1) + c_1 P_{n-1}^{(0,l+1/2)}(2r^2-1) + c_2 \tag{A3}$$

and

$$\Psi_n^l(r) \propto r^{l+1}(1-r^2)P_{n-1}^{(2,l+1/2)}(2r^2-1), \tag{A4}$$

and the unnormalized poloidal radial basis for the vertical boundary condition (27) is

$$I_n^l \propto r^{l+1}\left[d_0 P_n^{(0,l+1/2)}(2r^2-1) + d_1 P_{n-1}^{(0,l+1/2)}(2r^2-1)\right], \tag{A5}$$

where

$$\begin{aligned}
c_0 &= -2n^2(l+1) - n(l+1)(2l-1) - l(2l+1),\\
c_1 &= 2(l+1)n^2 + (2l+3)(l+1)n + (2l+1)^2,
\end{aligned}$$

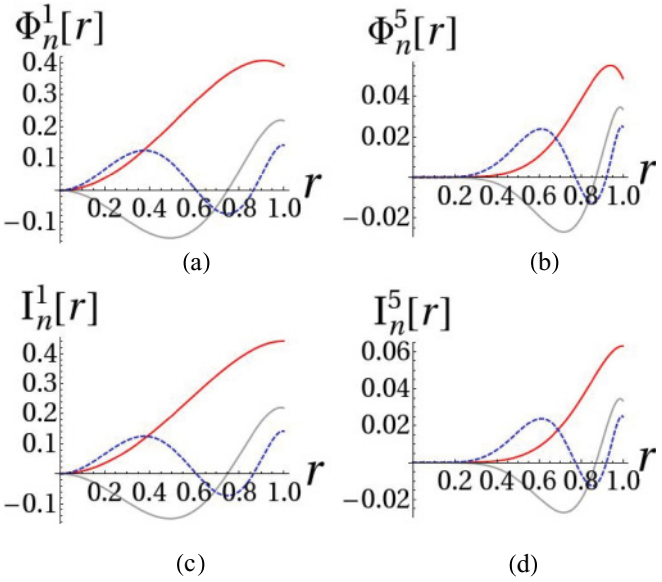

FIG. 12. (Color online) Plots (a) and (b) illustrate the first three radial basis functions of the poloidal field satisfying the insulating boundary condition for $l = 1$ and $l = 5$, and (c) and (d) illustrate the first five radial basis functions of the poloidal field satisfying the vertical boundary condition for $l = 1$ and $l = 5$, in the order of red, gray, and dashed blue for $n = 1,2,3$.

$$c_2 = 3nl + l(2l+1),$$
$$d_0 = -(2n-1)(n+l),$$
$$d_1 = (2n+1)(n+l+1),$$

and $P_n^{(0,l+1/2)}$ is the nth order Jacobi polynomial. The poloidal radial basis functions Φ_n^l and I_n^l are newly derived using the method described in [54–56], where in Fig. 12, one can find the plots of the first three poloidal radial bases Φ_n^l and I_n^l for $l = 1$ and $l = 5$. The toroidal radial basis Ψ_n^l is the same as we introduced in [35].

APPENDIX B: ALGORITHMS FOR COMPUTING THE FORWARD AND ADJOINT TERMS

1. Algorithm for discretizing forward magnetic induction

In order to discretize the induction term $\nabla \times (\mathbf{u} \times \mathbf{B})$ we use the pseudospectral method. For the horizontal transform in θ and ϕ components, we use the standard method [57]. Define

$$\mathbf{H} = \left(\Delta_r, \frac{\Delta_\theta}{\sin\theta}, \frac{\Delta_\phi}{\sin\theta}\right) \quad \text{and} \quad \Delta = \mathbf{u} \times \mathbf{B}. \tag{B1}$$

The poloidal Γ_l^m and toroidal scalar Π_l^m of the induction term can be computed as follows:

$$\Gamma_l^m = \frac{r^2}{l(l+1)}[\mathbf{r} \cdot \nabla \times (\mathbf{u} \times \mathbf{B})]_{(l,m)},$$
$$\Pi_l^m = \frac{r^2}{l(l+1)}[\mathbf{r} \cdot \nabla \times \nabla \times (\mathbf{u} \times \mathbf{B})]_{(l,m)}.$$

In spectral (l,m) space, the poloidal and toroidal scalars are equivalent to

$$\Gamma_l^m(r) = \frac{c_l^m r}{l}[H_\phi]_{(l-1,m)}(r) - \frac{c_{l+1}^m r}{l+1}[H_\phi]_{(l+1,m)}(r) - \frac{r}{l[l+1]}\left[\frac{\partial H_\theta}{\partial \phi}\right]_{(l,m)}(r) \tag{B2}$$

and

$$\Pi_l^m(r) = [H_r]_{(l,m)}(r) + \frac{\partial}{\partial r}\left[r\Theta_l^m(r)\right], \tag{B3}$$

as shown by Glatzmaier [57], where c_l^m are normalization coefficients and

$$\Theta_l^m(r) = \frac{c_l^m r}{l}[H_\theta]_{(l-1,m)}(r) - \frac{c_{l+1}^m r}{l+1}[H_\theta]_{(l+1,m)}(r) + \frac{r}{l[l+1]}\left[\frac{\partial H_\phi}{\partial \phi}\right]_{(l,m)}(r). \tag{B4}$$

Having computed Γ_l^m and Π_l^m, one can use the orthogonality relation in Eq. (A2) to further transform Γ_l^m and Π_l^m into the spectral coefficients $a_{(n,l,m)}$ and $b_{(n,l,m)}$.

2. Algorithm for computing the adjoint magnetic induction term

a. Algorithm for computing $(\nabla \times \mathbf{B}^\dagger) \times \mathbf{u} - \nabla p^\dagger$

For simplicity, in our discussion here we assume $R_m = 1$ but the extension for $R_m \neq 1$ is obvious. Computing the adjoint induction term must be performed in a different manner to the computation of the forward problem. Notice that $(\nabla \times \mathbf{B}^\dagger) \times \mathbf{u} - \nabla p^\dagger$ is solenoidal, and the boundary condition on $p^\dagger$ is $p^\dagger = 0$ at $r = 1$. This means that if we can solve for $p^\dagger$ first and subtract $\nabla p^\dagger$ from $(\nabla \times \mathbf{B}^\dagger) \times \mathbf{u}$, the remaining term is the adjoint field. Mathematically speaking, the poloidal scalar $[\Gamma^\dagger]_l^m(r)$ of the adjoint field takes the following form:

$$[\Gamma^\dagger]_l^m(r) = \frac{r^2}{l(l+1)}\{\hat{\mathbf{r}} \cdot [(\nabla \times \mathbf{B}^\dagger) \times \mathbf{u} - \nabla p^\dagger]\}_{(l,m)} \tag{B5}$$

and for toroidal scalar $[\Pi^\dagger]_l^m(r)$, one has

$$[\Pi^\dagger]_l^m(r) = \frac{r^2}{l(l+1)}\{\hat{\mathbf{r}} \cdot \nabla \times [(\nabla \times \mathbf{B}^\dagger) \times \mathbf{u} - \nabla p^\dagger]\}_{(l,m)} = \frac{r^2}{l(l+1)}\{\hat{\mathbf{r}} \cdot \nabla \times [(\nabla \times \mathbf{B}^\dagger) \times \mathbf{u}]\}_{(l,m)}. \tag{B6}$$

Further projecting $[\Gamma^\dagger]_l^m(r)$ and $[\Pi^\dagger]_l^m(r)$ onto Eqs. (A3) and (A4) via Eq. (A2), one can get the spectral coefficients of the adjoint induction term $a_{(n,l,m)}^\dagger$ and $b_{(n,l,m)}^\dagger$. Notice that mathematically the adjoint toroidal transform (B6) is the same as the forward poloidal transform, thus one can use Eq. (B2) in computation. The only problem that remains is how to compute $p^\dagger$ for Eq. (B5). One way to do this is to solve a Poisson equation by taking the divergence of the adjoint induction term, namely,

$$\nabla^2 p^\dagger = \nabla \cdot [(\nabla \times \mathbf{B}^\dagger) \times \mathbf{u}]. \tag{B7}$$

This equation can be easily solved via a Galerkin method. Similarly to Eq. (B1), we define

$$\mathbf{H}^\dagger = \left(\Delta_r^\dagger, \frac{\Delta_\theta^\dagger}{\sin\theta}, \frac{\Delta_\phi^\dagger}{\sin\theta}\right) \quad \text{and} \quad \Delta^\dagger = (\nabla \times \mathbf{B}^\dagger) \times \mathbf{u}. \tag{B8}$$

The right-hand side of Eq. (B7) is

$$\left[\frac{\partial H_r^\dagger}{\partial r}\right]_{l,m} + \frac{1}{r}\Big(2[H_r^\dagger]_{l,m} + (l+1)c_l^m[H_\theta^\dagger]_{l-1,m} - lc_{l+1}^m[H_\theta^\dagger]_{l+1,m} + \left[\frac{\partial H_\phi^\dagger}{\partial\phi}\right]_{l,m}\Big) \tag{B9}$$

in spherical harmonics (l,m). Hence, for Eq. (B7), one just needs to solve an ordinary differential equation in r for each spherical harmonic degree l and order m for $p^\dagger$. Since $p^\dagger = 0$ at $r = 1$, one can modify the toroidal $\mathbf{B}$ radial basis by reducing l to $l-1$ due to spherical continuity [28], where an unnormalized basis set is $\Psi_n^{l-1} = r^l(1-r^2)P_{n-1}^{(2,l-3/2)}(2r^2-1)$.

b. Algorithm for computing $(\mathbf{u}\cdot\mathbf{B}^\dagger)\hat{\mathbf{r}}\delta(r-1)$

Notice that the boundary driving is in the radial direction, thus this only influences the poloidal field. In the Galerkin method, we have to project the term $(\mathbf{u}\cdot\mathbf{B}^\dagger)\hat{\mathbf{r}}\delta(r-1)$ onto the spectral space (n,l,m) and we have

$$\int_V {}_n\mathbf{S}_l^m \cdot (\mathbf{u}\cdot\mathbf{B}^\dagger)\hat{\mathbf{r}}\delta(r-1)dV = \left[l(l+1)\Phi_n^l\right][\mathbf{u}\cdot\mathbf{B}^\dagger]_{(l,m)}\big|_{r=1}, \tag{B10}$$

where Φ_n^l is the radial basis function of the poloidal scalar, $\frac{l(l+1)}{r^2}\Phi_n^l$ is the radial component of ${}_n\mathbf{S}_l^m$ and $[\mathbf{u}\cdot\mathbf{B}^\dagger]_{(r,l,m)}$ is the radial function in spherical harmonic space (l,m).

APPENDIX C: ADJOINT TEMPERATURE DIFFUSION TERM FOR DIFFERENT BOUNDARY CONDITION

In spherical geometry, the temperature field can be decomposed as $T_c = \sum_{l,m} f_l^m(r)Y_l^m(\theta,\phi)$, where f_l^m is the radial scalar function and Y_l^m is a spherical harmonic. Hence the Laplace operator decouples in spherical degree l and order m.

As we discussed in Sec. II, if the temperature field satisfies the linear boundary condition $K(T_c) = c$ for both homogeneous and inhomogeneous cases, the adjoint field $T^\dagger$ has to satisfy $K(T^\dagger) = 0$ for the adjoint data assimilation of the initial-value problem. For the temperature field T_c satisfying either (i), a constant flux boundary condition or (ii), a heterogeneous boundary condition, one can represent T_c as temperature perturbation T superimposed on a background temperature T_b: (i) For a constant flux boundary condition, $\nabla T_c = c\hat{\mathbf{r}}$, where c is a constant, the temperature field T_c can be written as $T_c = T + cr$, where $\nabla T = 0$ and $\nabla T_b = c\hat{\mathbf{r}}$. (ii) For a heterogeneous condition, $T_c = \sum_{l,m} f_l^m Y_l^m$, where $f_l^m(r=1) = d_l^m$ are known constants, the temperature field T_c can be written as $T_c = T + T_b$, where $T(r=1) = 0$ and $T_b = \sum_{l,m} d_l^m r^l Y_l^m$.

For both cases, we have $\nabla^2 T_b = 0$ and the adjoint Laplace operator obeys

$$\int_V T^\dagger \nabla^2 T\, dV = \int_V T\nabla^2 T^\dagger\, dV + \mathsf{S}_1 + \mathsf{S}_2.$$

Denote $d\Sigma$ as the surface element. The two boundary terms S_1 and S_2 are zero,

$$\mathsf{S}_1 = \int_\Sigma T^\dagger \nabla T \cdot d\boldsymbol{\Sigma} = 0 \quad \text{and} \quad \mathsf{S}_2 = -\int_\Sigma T\nabla T^\dagger \cdot \boldsymbol{d}\Sigma = 0,$$

since for case (i), we have $\nabla T = \nabla T^\dagger = 0$ at $r = 1$ and for case (ii), we have $T = T^\dagger = 0$ at $r = 1$. Thus for both cases, the Laplace operator for the temperature perturbation is self-adjoint, $(\nabla^2)^\dagger = \nabla^2$.

[1] J. Larmor, Elect. Rev. **85**, 512 (1919).
[2] H. K. Moffatt, *Magnetic Field Generation in Electrically Conducting Fluids* (Cambridge University Press, UK, 1983).
[3] A. Sakuraba and P. H. Roberts, Nature **2**, 802 (2009).
[4] T. Miyagoshi, A. Kageyama, and T. Sato, Nature (London) **463**, 793 (2010).
[5] M. Dumberry and C. C. Finlay, Earth Planet. Sci. Lett. **254**, 146 (2007).
[6] A. Jackson, A. R. Jonkers, and R. M. Walker, Philos. Trans. R. Soc. London, Ser. A **358**, 957 (2000).
[7] M. Korte, A. Genevey, C. G. Constable, U. Frank, and E. Schnepp, Geochem. Geophys. Geosyst. **6**, 32 (2005).
[8] M. Kono, in *Treatise on Geophysics, Geomagnetism*, Vol. 5 (Elsevier, New York, 2007).
[9] A. Tarantola, Geophysics **49**, 1259 (1984).
[10] J. Tromp, C. Tape, and Q. Lui, Geophys. J. Int. **160**, 195 (2005).
[11] O. Talagrand and P. Courtier, Q. J. R. Meteorol. Soc. **113**, 1311 (1987).
[12] C. C. T. Pringle and R. R. Kerswell, Phys. Rev. Lett. **105**, 154502 (2010).
[13] A. Fournier, C. Eymin, and T. Alboussiere, Nonlinear Processes Geophys. **14**, 163 (2007).
[14] E. Canet, A. Fournier, and D. Jault, J. Geophys. Res. **114**, B11101 (2009).
[15] N. Gillet, D. Jault, E. Canet, and A. Fournier, Nature (London) **465**, 74 (2010).
[16] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter* (Springer, Berlin, 2006).
[17] Z. Sun, A. Tangborn, and W. Kuang, Nonlinear Processes Geophys. **14**, 181 (2007).
[18] D. Liu, A. Tangborn, and W. Kuang, J. Geophys. Res. **112**, B08103 (2007).
[19] A. Fournier, G. Hulot, D. Jault, W. Kuang, A. Tangborn, N. Gillet, E. Canet, J. Aubert, and F. Lhuillier, Space. Sci. Rev. **155**, 247 (2010).
[20] W. Kuang and A. Tangborn, *IAGA Special Book Series*, edited by M. Mandea and M. Korte, Vol. 5 (Springer, New York, 2010).
[21] W. Kuang and J. Bloxham, J. Comput. Phys. **153**, 51 (1999).
[22] W. Kuang, A. Tangborn, W. Jiang, D. Liu, Z. Sun, J. Bloxham, and Z. Wei, Commun. Comput. Phys. **3**, 85 (2008).

[23] W. Kuang, A. Tangborn, Z. Wei, and T. Sabaka, Geophys. J. Int. **179**, 1458 (2009).
[24] W. Kuang, Z. Wei, R. Holme, and A. Tangborn, Earth, Planets Space **62**, 775 (2010).
[25] C. C. Finlay *et al.*, Geophys. J. Int. **183**, 1216 (2010).
[26] A. Fichtner, H.-P. Bunge, and H. Igel, Phys. Earth, Planet. Inter. **157**, 86 (2006).
[27] D. Gubbins and N. Roberts, Geophys. J. R. Astron. Soc. **73**, 675 (1983).
[28] G. Backus, R. Parker, and C. Constable, *Foundations of Geomagnetism* (Cambridge University Press, Cambridge, England, 1996).
[29] R. L. Parker, *Geophysical Inverse Theory* (Princeton University Press, Princeton, NJ, 1994).
[30] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications* (SIAM, Philadelphia, PA, 2008).
[31] J. P. Aubin, *Applied Functional Analysis* (Wiley, New York, 2000).
[32] R. Reed and S. Barry, *Methods of Modern Mathematical Physics, 1: Functional Analysis* (revised and enlarged ed.) (Academic, New York, 1980).
[33] J. P. Boyd, *Chebyshev and Fourier Spectral Methods*, 2nd ed. (Dover, New York, 2001).
[34] T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson, *Introduction to Algorithms*, 2nd ed. (McGraw-Hill, New York, 2001).
[35] K. Li, P. W. Livermore, and A. Jackson, J. Comput. Phys. **229**, 8666 (2010).
[36] C. Jones, P. Boronski, A. Brun, G. Glatzmaier, T. Gastine, M. Miesch, and J. Wicht, Icarus **216**, 120 (2011).
[37] O. Talagrand, in *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*, edited by A. Griewank and F. Corliss (SIAM, Philadelphia, PA, 1991), pp. 169–180.
[38] R. Giering and T. Kaminski, ACM Trans. Math. Softw. **24**, 437 (1998).
[39] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. (Springer, New York, 2006).
[40] E. C. Bullard and H. Gellman, Philos. Trans. R. Soc. London, Ser. A **247**, 213 (1954).
[41] T. Namikawa and S. Matsushita, Geophys. J. R. Astron. Soc. **19**, 395 (1970).
[42] G. W. Hanson and A. B. Yakovlev, *Operator Theory for Electromagnetics* (Springer, New York, 2001).
[43] P. H. Roberts, J. Math. Anal. Appl. **1**, 195 (1960).
[44] R. D. Gibson and P. H. Roberts, in *Magnetism and the Cosmos*, edited by W. R. Hindmarsh, F. J. Lowes, P. H. Roberts, and S. K. Runcorn (Oliver and Boyd, Edinburgh/London, 1966), pp. 108–120.
[45] M. Kono and P. H. Roberts, J. Geomagn. Geoelectr. **43**, 839 (1991).
[46] B. P. Jones, Mon. Not. R. Astr. Soc. **233**, 875 (1988).
[47] P. Goldreich and A. Reisenegger, Astrophys. J. **395**, 250 (1992).
[48] R. Hollerbach and G. Rüdiger, Mon. Not. R. Astron. Soc. **337**, 216 (2002).
[49] P. W. Livermore and A. Jackson, Geophys. Astrophys. Fluid Dyn. **99**, 467 (2005).
[50] See, http://www.fftw.org/.
[51] J. Nocedal, Math. Comput. **35**, 773 (1980).
[52] See, http://users.eecs.northwestern.edu/nocedal/lbfgs.html.
[53] J. Bloxham and A. Jackson, Rev. Geophys. **29**, 97 (1991).
[54] P. W. Livermore, http://escholarship.org/uc/item/9vk1c6cm/, 2009.
[55] P. W. Livermore and G. Ierley, Numerical Algorithms **54**, 533 (2009).
[56] P. W. Livermore, J. Comput. Phys. **229**, 2046 (2010).
[57] G. A. Glatzmaier, J. Comput. Phys. **55**, 461 (1984).