Chapter 2

Adjoint Method in Mantle Convection[1]

2.1 Theoretical Basis of the Adjoint Method

The adjoint method for data assimilation is a gradient-based inversion, which is particularly useful for inverting nonlinear dynamic systems. Derivation of the adjoint method for an evolving system is based on perturbation theory, where a mismatch in the model output against observation is attributed to an error in the model input, with their relation approximated as a first order derivative (gradient) of the least-squared mismatch with respect to the input. To formulate the concept mathematically, consider an initial value problem in which all the governing equations and boundary conditions are perfectly known and numerical errors are negligible. Any mismatch in the prediction should be attributed to errors in the initial condition (i.e., model input). This relation can be simply expressed as $dJ = (\partial J / \partial \vec{a}) d\vec{a}$, where J is a scalar cost function, which defines the mismatch of prediction from data and $\vec{a}$ is the initial variable that potentially carries error. If an explicit form of the expression $\partial J / \partial \vec{a}$ can be obtained, then the perturbation (i.e., error) $d\vec{a}$ associated with the initial condition can be retrieved.

---

[1] This chapter is based on: 1) Liu L. and M. Gurnis (2008), *Simultaneous Inversion of Mantle Properties and Initial Conditions Using an Adjoint of Mantle Convection*, J. of Geophy. Res., 113, B08405, doi:10.1029/2008JB005594. 2) Liu, L. and M. Gurnis (2010), *Adjoint method and its application in mantle convection*, Earth Science Frontiers (Chinese with English abstract), in press.

Specifically, we define the cost function J as a function of temperature T

$$J = \iint_{t,V} (T_p - T_d)^2 \, dvdt \tag{1}$$

where T_p is the predicted temperature, T_d is the actual temperature (with the subscript d denoting data), t is time, and V is volume. In mantle convection, T_d is available only at the present day t_1, so effectively J is a delta function in the time domain with a nonzero value at t_1.

The governing equations for mantle convection, assuming an incompressible and Newtonian fluid, are

$$\nabla \cdot \vec{u} = 0 \tag{2}$$

$$\nabla P + \nabla \cdot (\eta \nabla \vec{u}) = \rho_m \alpha \Delta T \vec{g} \tag{3}$$

$$\frac{\partial T}{\partial t} + \vec{u} \cdot \nabla T = \kappa \nabla^2 T \tag{4}$$

where $\vec{u}$ is velocity, P dynamic pressure, η dynamic viscosity, ρ_m ambient mantle density, α coefficient of thermal expansion, ΔT temperature anomaly, $\vec{g}$ gravitational acceleration, T temperature, and κ thermal diffusivity.

If we assume $T(t_0)$ is the only variable that brings error into our prediction, our goal, in order to retrieve this quantity, is to obtain the expression $\partial J / \partial T(t_0)$, where t_0 refers to the initial time. This expression can be obtained through a constraint condition of the energy equation by introducing the adjoint variable λ (a Lagrangian multiplier) that forms a Lagrangian function L

$$L = J + \iint_{t,V} \lambda \left(\frac{\partial T}{\partial t} + \vec{u} \cdot \nabla T - \kappa \nabla^2 T \right) dvdt \tag{5}$$

A perturbation in L corresponds to perturbations in J and T. Subsequently, we will use δ to denote the perturbed quantities.

$$\delta L = \delta J + \iint_{t,V} \lambda(\frac{\partial \delta T}{\partial t} + \vec{u} \cdot \nabla \delta T - \kappa \nabla^2 \delta T) dv dt \qquad (5')$$

In principle, the velocity u should also contribute to this perturbation since it depends on T (see Eq. 3 and Bunge *et al*. [2003]), but we choose to neglect the velocity dependence in Eq. (5'). This is because, first, a full differentiation of Eq. (5) leads to a coupled system of the adjoint and forward models that is numerically changing to implement [Bunge *et al*., 2003]; second, inaccuracy from omission of the u dependence in Eq. (5') is diminished by the variational approach to the single temperature-adjoint solution through iterative schemes we will describe. By applying integration by parts over time and space to Eq. (5') with prescribed boundary conditions, we obtain

$$\delta L = \delta J + \int_V (\lambda \delta T) dv \Big|_{t_0}^{t_1} - \iint_{t,V} \delta T(\frac{\partial \lambda}{\partial t} + \vec{u} \cdot \nabla \lambda + \kappa \nabla^2 \lambda) dv dt \qquad (5'')$$

When $\delta L \to 0$, all terms associated with δT should vanish (λ is nonzero only at t_0 and t_1). Therefore, for any instant of time between t_0 and t_1, we have

$$\frac{\delta J}{\delta T} = \iint_{t,V} (\frac{\partial \lambda}{\partial t} + \vec{u} \cdot \nabla \lambda + \kappa \nabla^2 \lambda) dv dt \qquad (6)$$

This is called the adjoint equation or adjoint operator.

In practice, $\delta J / \delta T$ is nonzero only at the final time (t_1) in a mantle convection model when the mismatch between prediction and data is made. In effect, the term $\delta J / \delta T$ represents the residual temperature field at the final time. Therefore, the adjoint energy

equation (Eq. 6) is the same as the forward energy equation (Eq. 4) except for the diffusion term that has an opposite sign. This difference also means Eq. (6) is numerically unstable in describing a forward-time evolution, but ideal for representing a backward-time process. If we consider *t* as always being forward in time while substituting Eq. (1), then the differential form of Eq. (6) becomes

$$\frac{\partial \lambda}{\partial(-t)} - \vec{u}\cdot\nabla\lambda - \kappa\nabla^2\lambda - 2(T_p - T_d)_{t_1} = 0 \qquad (6')$$

where $2(T_p - T_d)_{t_1}$ is a forcing term at t_1. So, with $T_p - T_d$ describing the residual field at the final instant of time (which also provides a state for the system to start with), Eq. (6') represents a backward-in-time advection-diffusion process.

It can be seen that with Eq. (6), Eq. (5") can be reduced to

$$\delta L = \int_V (\lambda \delta T) dv \Big|_{t_0}^{t_1} \qquad (7)$$

Alternatively,

$$\frac{\partial L}{\partial T(t_0)} = -\int_V \lambda(t_0) dv \qquad (7')$$

Eq. (7') indicates that the gradient of the Lagrangian function *L* with respect to the initial temperature can be explicitly expressed as the adjoint quantity at the initial time. Since the Lagrantian function is an augmented (constrained) cost function, as can be seen from Eq. (5) where the zero valued constraint (Eq. 4) is prescribed at both t_0 and t_1, we can conclude that the same gradient relation as shown in Eq. (7') also holds for the cost function *J*.

$$\frac{\partial J}{\partial T(t_0)} = -\int_V \lambda(t_0) dv \qquad (7'')$$

This equation eventually allows for the following numerical algorithm to be reached.

For more references, this adjoint of the energy equation has been derived for meteorological [Sun *et al.*, 1991; Sirkes and Tziperman, 1997] and mantle convection problems [Bunge *et al.*, 2003; Ismail-Zadeh *et al.*, 2004].

In order to reverse a nonlinear process like mantle convection, iterative solvers are inevitable. We interleaved the backward adjoint calculation with a forward solution of the energy and momentum equations within an iterative procedure similar to that proposed by Bunge *et al.* [2003].

Our convention for subscript refers to time (0 for initial; 1 for present) while those for superscripts refer to the number of iterations. The number of iterations is determined by the accuracy to which we desire our prediction to satisfy data. Specifically these are described in the steps followed:

(i) Solve the forward problem with all three governing equations (Eq. 2 to 4) with initial condition T_0^n ($0 \le n \le N$, where *n* is the iteration number) and predict T_1^n . The first initial condition T_0^0 is potentially arbitrary. Store the velocity field for all time steps.

(ii) Compute the mismatch J^n and its gradient $\partial J^n / \partial T_1^n$; solve the adjoint energy equation (Eq. 7) with the velocity stored in (i) from t_1 to t_0 and obtain λ_0^n .

(iii) Update the initial field: $T_0^{n+1} = T_0^n - \alpha(n) \cdot \lambda_0^n$, where $\alpha(n)$ is a damping factor (defined as in Ismail-Zadeh *et al*. [2004] except that we took a simple form assuming $\alpha(n)$ only depends on *n*), with n_0 an adjustable integer

$$\alpha(n) = 1/(n + n_0) \tag{8}$$

In general, the coefficient α can also be a constant with values no more than 0.5, in order to reduce overshoots.

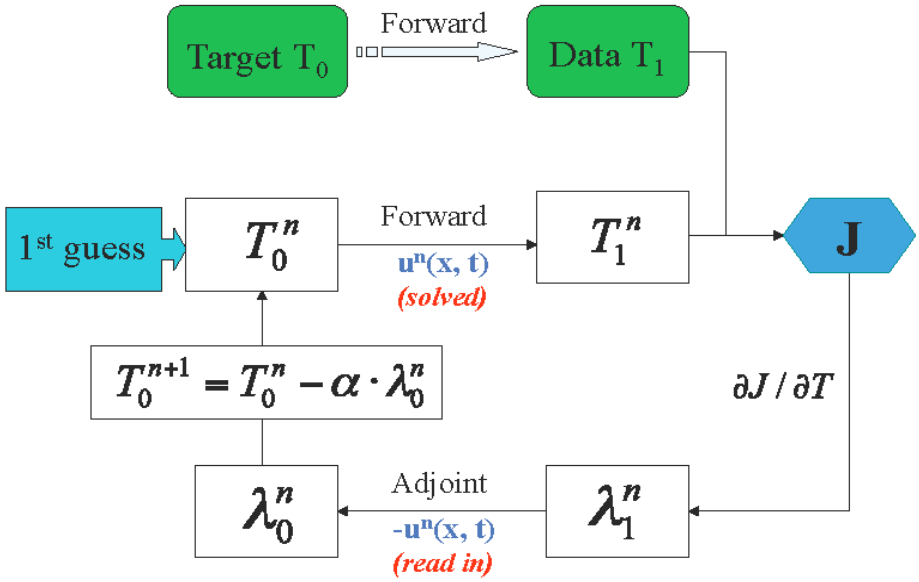

Figure 1 Illustration of the forward- adjoint iterative solver. T_0 and T_1 represent the reference initial and final states, respectively. u^n is velocity for the ^{n}th iteration, which is solved during the forward run and read in from storage during the adjoint run.

Figure 1 illustrates the whole workflow of this iterative solver for a synthetic test. We assume that the true mantle temperature in both the past and present are known, which we call reference states. This past mantle state is the solution (target) we seek by applying

the adjoint method, and the present state is the function we try to match with the forward predictions. The reference states are generated by a forward run that solves the normal convection equations (Eqs. 2, 3, 4). The forward and adjoint iterations follow the procedures described above. Iterations stop upon convergence or when the mismatch is small enough.

2.2 Numerical Implementation of the Adjoint Method

2.2.1 Solving 1D Linear Problems

Before we apply the adjoint method to more complex problems, we first design a very simple example: to invert a 1D kinematic thermal-diffusion problem, through which we illustrate the workings of the adjoint method. This simulation is carried out with a code based on the finite-element method (FEM) written in the programming language Matlab.

Imagine the reference initial condition is a thermal upwelling, e.g., a mantle plume, situated in the lower mantle sometime in the geological past (Fig. 2A), which is also the target solution we seek by inverting the present-day "observed" structure of this upwelling. For simplicity, only the forward energy equation (Eq. 4) and its adjoint operator (Eq. 7) are solved, with a prescribed velocity field. Without having to solve the advection term in Eq. (4), this is essentially a linear problem.

As Fig. 2B illustrates, the first guess of the initial condition has little correlation with the reference, so does the first prediction. This creates the largest mismatch between prediction and observation (i.e., residual temperature) among all iterations (Fig. 2B and C). The first correction to the initial condition, generated by advecting this large residual back to the initial time, is also the largest among all iterations. The initial condition, after five updates, becomes much closer to the reference while the mismatch decreases. After 15 iterations, the solution and target converge with a small mismatch. The convergence process is also reflected through a global root-mean-squared (RMS) residual (Fig. 2 C) for the initial and final states that decay exponentially with iterations.

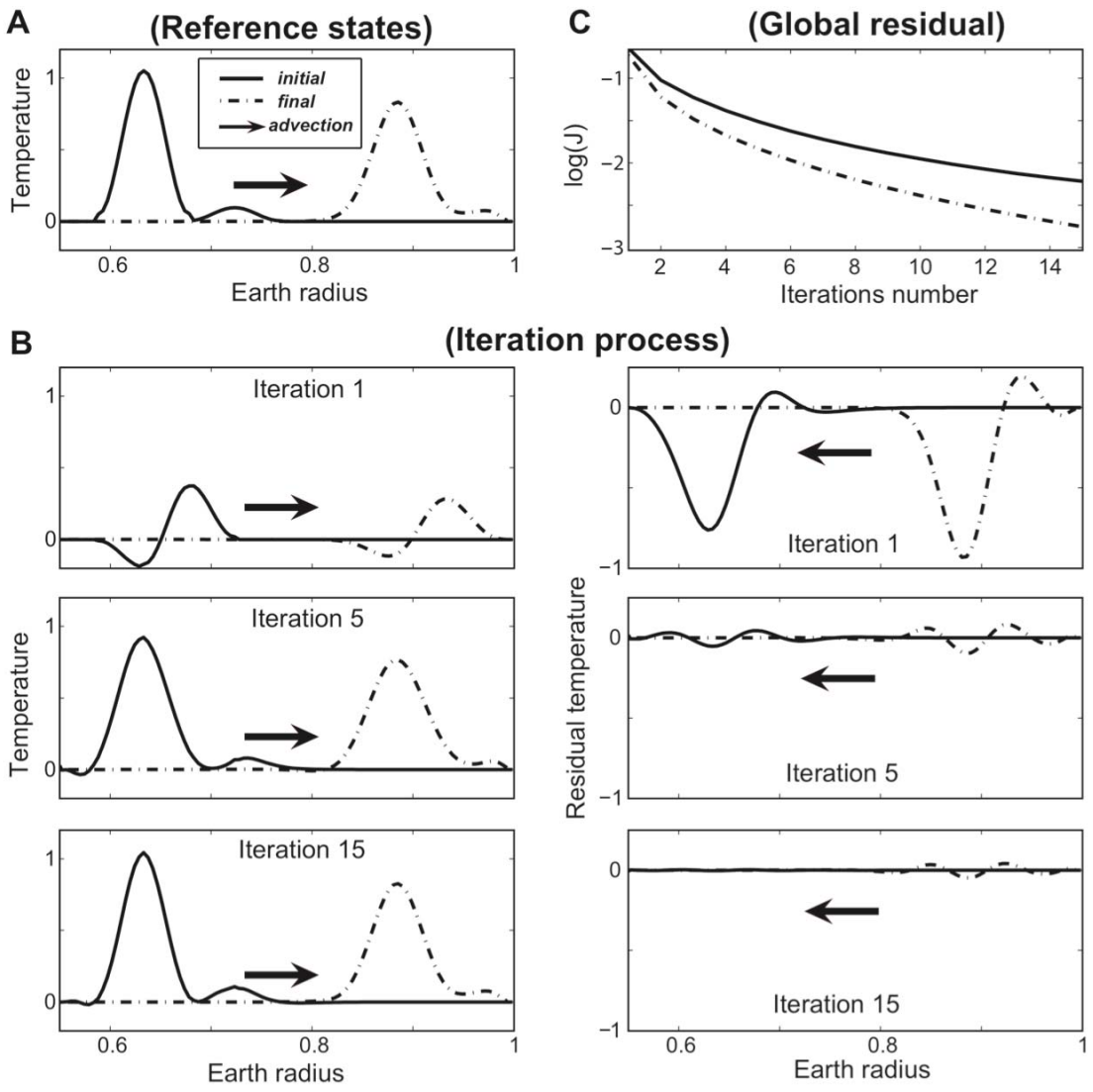

Figure 2 Adjoint inversion of a 1D linear advection-diffusion process. A. Reference initial (solid) and final (dashed) states, the bold arrow indicates the direction of the prescribed velocity. Both axes are dimensionless. B. The iterative procedure solving the initial condition. Left column shows the forward runs, while the right shows the adjoint runs. Three different iterations are shown. C. Global root-mean-squared residuals from both the initial and final states, as a function of iterations. Note the vertical axis is on a log scale.

Additional tests with different first guesses show that the solution is independent of prior information about the true initial condition. This indicates the adjoint method convergences to the true solution unconditionally for this linear inverse problem, consistent with inversion theory [Tarantola, 2005].

2.2.2 Solving 3D Nonlinear Problems

For spherical problems, we have implemented the adjoint algorithm into software using the finite element and designed specifically for mantle convection, *CitcomS* [Zhong *et al*., 2000; Tan, *et al*., 2006]. The version of *CitcomS* used here solves the equations within a spherical geometry and scales well on large parallel computers. Our changes were made to *CitcomS* version 2.1.0 obtained from the Computational Infrastructure for Geodynamics (https://geodynamics.org).

Upon implementation of the adjoint method within *CitcomS*, we hope to obtain a good solution to the initial condition while reducing the computational cost as well. Two sets of numerical experiments are designed that used the forward-adjoint looping to estimate initial conditions. The first set has a uniform viscosity (η=1), a constant ambient mantle temperature, and a Rayleigh number of 1.0×10^8. The second set of experiments has a layered viscosity, a top thermal boundary layer, and a higher Rayleigh number.

The model domain is: colatitude $\theta \in [1.27, 1.87]$, longitude $\varphi \in [0.0, 0.6]$ (both in radians), and radius (normalized by outer radius of the earth) $r \in [0.55, 1]$. Boundary conditions for the forward model are: $\vec{n} \cdot \vec{u} = 0$ and $\vec{n} \cdot \nabla \vec{u}_{tg} = 0$ on all boundaries, where $\vec{n}$

is the outer normal vector, $\vec{u}$ the velocity vector, and $\vec{u}_{tg}$ the tangential velocity; the surface and core-mantle boundary (CMB) are isothermal, while the sidewalls have zero heat flux, $\vec{n} \cdot \nabla T = 0$. The adjoint model has zero adjoint temperature on the surface and CMB, and zero adjoint heat flux on the sidewalls. The Rayleigh number is

$$Ra = \frac{\rho_m g \alpha R_o^3 \Delta T}{\eta_o \kappa} \tag{9}$$

where ΔT is the temperature drop from CMB to surface. Time is non-dimensionalized, with the actual time t related to the model time t' by

$$t = t' R_o^2 / \kappa \tag{10}$$

All symbols with their dimensional values are listed in Table 1. Hereafter, all physical quantities are normalized with their dimensional values, if not noted separately.

2.2.2.1. Models within a single layer

In the first set (Set I) of experiments, the reference states include an initial condition (Fig. 3A) that has a spherical hot anomaly in the lower part of mantle (with a maximum temperature increase of $\Delta T = 0.3$ at the center and a Gaussian temperature profile across the center). The final condition was produced by running the model forward for 9 Myr (Fig. 3B). These two reference states are the targets we attempted to predict with the adjoint method. All models are computed on a 33×33×33 grid. We assumed $n_0 = 1$ in Eq. (8), for the forward-adjoint looping.

Table 1: Summary of Model Parameters

Parameter	Symbol	Value
Radius of the earth	R_o	6371 km
Gravitational acceleration	g	9.81 $m\ s$
Reference mantle density	ρ_m	3300 $kg\ m^{-3}$
Reference viscosity	η_o	10^{21} Pas
Coefficient of thermal expansion	κ	$10^{-6}\ m^2\ s^{-1}$
Thermal expansity	α	3x10
Super-adiabatic temperature drop from CMB to surface	ΔT	393˚C (Set I); 1965˚C (Set II)

All iterations were started with different first guesses to the target initial condition (each of these guesses constituted different cases, A1–A6, with "A" denoting adjoint method). Either we assumed a uniform temperature, a temperature that was a function of the actual initial condition, or generated an estimate through a simple backward integration of the governing equations (hereafter, we refer to this state of the system the SBI). The SBI was obtained by integrating the governing equations from t_1 to t_0 while reversing the sign of gravity from the forward calculation. The initial guesses were arranged in order of how close they are to the target initial condition (Table 2). Specifically, Case A1 had a nearly isothermal condition with a tiny perturbation. Case A2 had an anomaly with the same center as the target, but with a smaller volume (1/8×) and hence buoyancy. Cases A3 had the same shape and buoyancy compared to the target, but the center was shifted upward by 500 km.

Case A4 also had the same shape and buoyancy as the target but its center was shifted horizontally by 400 km. Case A5 had the same center but the anomaly had a larger volume (2.4×) and hence buoyancy. Case A6 used the SBI first guess to obtain the first guess.

Table 2: Description of the thermal anomaly structures in the reference initial state and various first guesses

	Geometry	**Center** (θ, φ, r)	**Diameter**[¶] (Dimensionless)	**Magnitude** (ΔT)
Reference	Sphere	(1.57, 0.3, 0.7)	0.3	0.3
Case A1	Sphere	(1.57, 0.3, 0.7)	0.3	0.001
Case A2	Sphere	(1.57, 0.3, 0.7)	0.15	0.5
Case A3	Sphere	(1.57, 0.3, 0.78)	0.3	0.3
Case A4	Sphere	(1.57, 0.39, 0.7)	0.3	0.3
Case A5	Sphere	(1.57, 0.3, 0.7)	0.4	0.3
Case A6	SBI[§]	--	--	--

Note: SBI[§] means simple backward integration, i.e., reverse the sign of gravity and run the forward model from the present-day mantle structure for the same amount of time.

[¶]: Diameter of the spherical anomaly, normalized by R_o, radius of the earth.

For comparison, we ran all cases for 50 iterations while tracking the recovered initial conditions. These retrieved initial conditions were noticeably different, and the recoveries in

Case A4 to A6 (Fig. 3F–H) were better than those in A1 to A3 (Fig. 3C–E). Case A6 gave the best recovery (Fig. 3H). The root-mean-squared (RMS) residuals between the recovered initial conditions and the target initial (Fig. 4A), and those between the predicted and target final (Fig. 4B) decreased with the number of iterations. The terminal (at $n = 50$) residuals for both the initial and final states (Fig. 4A, B), decreased from Case A1 to A6 as the first guess more closely reflected the target initial condition. The SBI first guess (A6), especially, started the first iteration with residuals far smaller than the others and the residuals with the final state remained small in comparison to the other cases (Fig. 4A, B). The rate of convergence was also dependent on the initial guess: the closer the first guess to the target initial condition, the faster the convergence (Fig. 4A, B). The case based on an SBI first guess was one of the fastest converging cases and required the least number of iterations to converge. If the final solution is achieved when the slope of the residual between predicted and target final decreases to below a specified small value, then solving for the initial condition with the SBI first guess is almost an order of magnitude faster than the others. The convergence of Case A1, with the nearly isothermal initial condition, is far smaller than A6 using the SBI first guess and much of this difference arises from the organization of the forward-adjoint looping. For Case A1, the adjoint temperature at t_1 is nearly the negative of the final temperature, in other words, almost possessing the same buoyancy used in the strict reverse calculation (SBI first guess). However, when the adjoint temperature in A1 is advected from t1 to t0, the stored velocity field from the forward calculation is used, but this velocity field is quite different from the actual. The SBI first

guess overcomes this limitation by using the velocity field from the actual backward calculation.

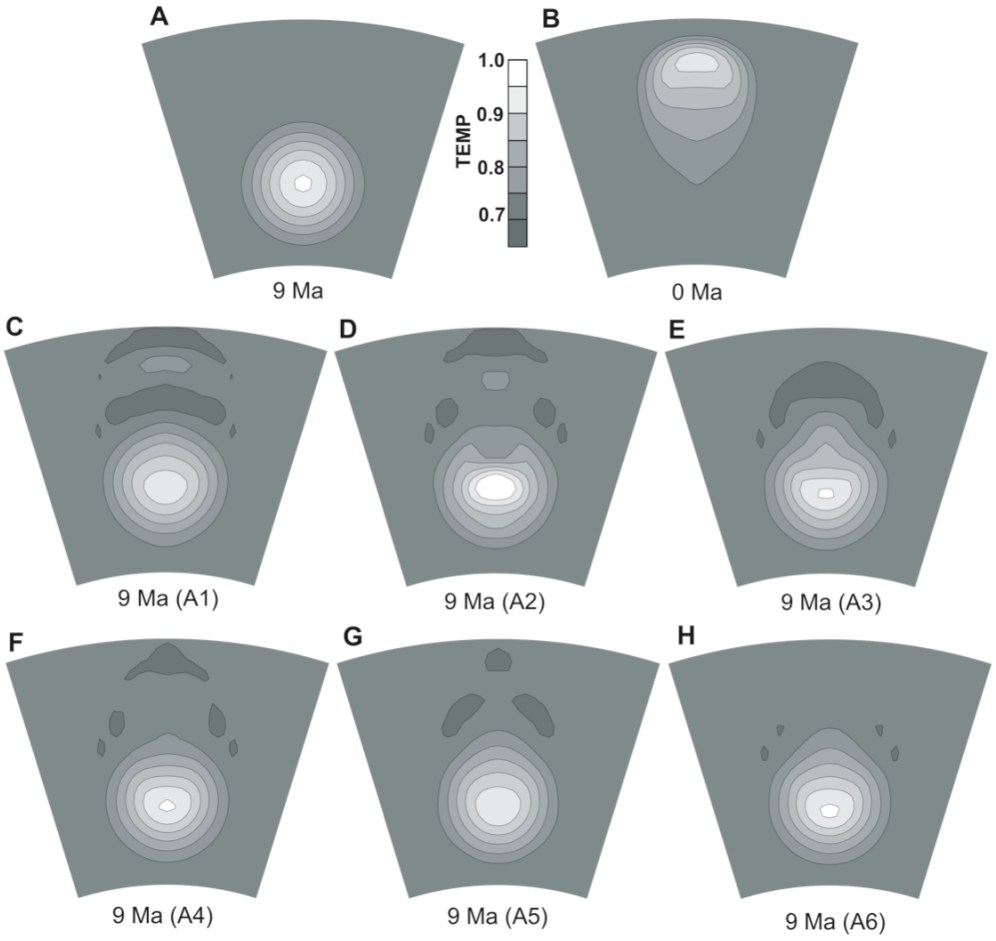

Figure 3 Three-dimensional forward-adjoint models (with a 33×33×33 mesh) for a mantle with a single-layer viscosity and uniform background temperature. Shown is temperature for vertical cross-sections along lines of latitude through center of the domain. Reference thermal states at 9 Ma (A) and present (B). (C to H) Retrieved initial states at 9 Ma using six different initial guesses (Case A1−A6, Table 2). For all cases, 50 forward-adjoint iterations were used.

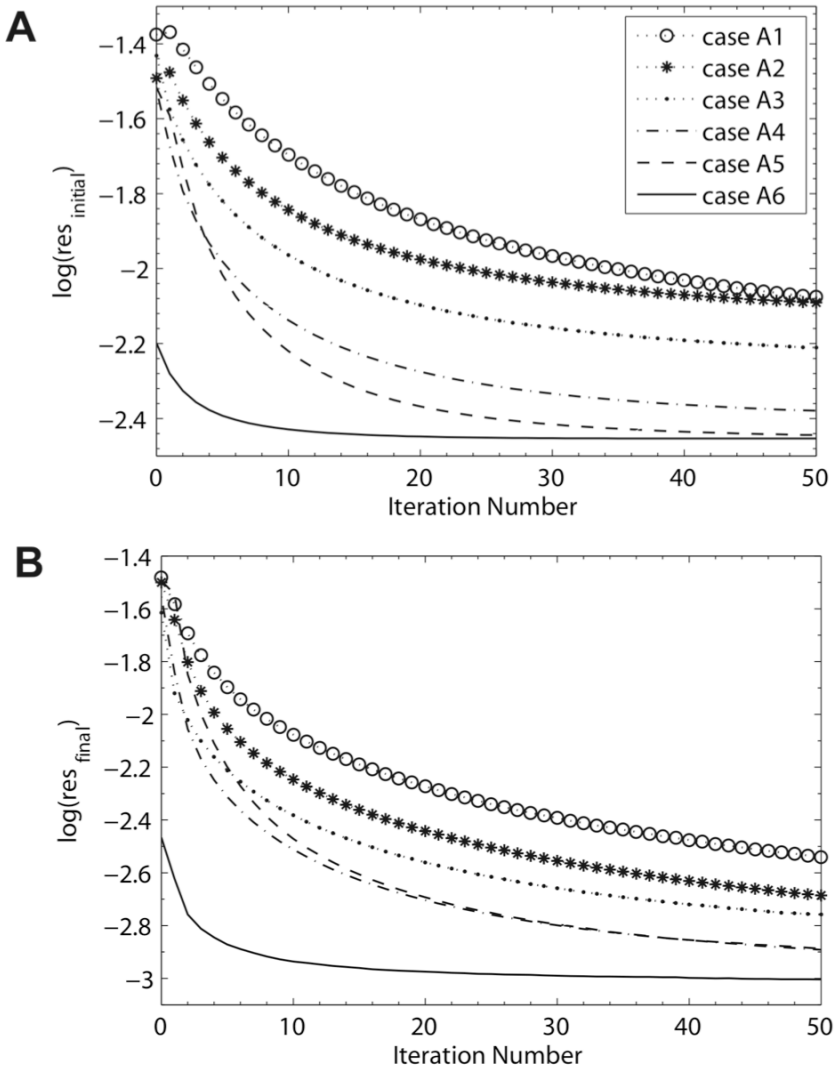

Figure 4 Convergence of the models shown in Fig. 3. (A) Root-mean-squared (RMS) residuals of recovered initial conditions with respect to the reference initial versus iteration. (B) RMS residuals of the predicted final conditions with respect to the reference final versus iteration.

Since the solutions are dependent on the first guess, finding the optimal one is important to decrease the computational cost while obtaining a reasonable solution. Because the SBI first guess gives the best solution to the initial condition, both in terms of the terminal residual and the rate of convergence, we consider this to be a useful means to

obtain an optimal first guess. Another advantage of obtaining the first guess via the SBI is that it requires no *a priori* information of the solution. Algorithmically, it is also easy to obtain.

2.2.2.2. Models with thermal boundary layers and depth- and temperature-dependent viscosities

The second set (Set II) of experiments is geophysically more realistic with a top thermal boundary layer (TBL) representing the lithosphere and a four-layer mantle with temperature-dependent viscosity. The TBL has an error function temperature profile typical of 40 Ma oceanic lithosphere. The viscosities for lithosphere, asthenosphere, transition zone, and the lower mantle, without temperature-dependence, are 10, 1, 10, and 40, respectively. The temperature dependence of viscosity is

$$\eta_T = \eta_o \times \exp(\frac{1}{T + 0.3} - 1) \tag{11}$$

where η_T is temperature-dependent viscosity and η_o is the depth-dependent prefactor. This results in an order of magnitude decrease in viscosity from T=0 to 1. Compared to Set I, we used a higher Rayleigh number at 5.0×10^8.

The target initial condition has the same thermal anomaly as that in model Set I, only that it has a TBL on top (Fig. 5A). The target final condition (Fig. 5B) is 52 Myr after the anomaly had risen through the asthenosphere and eroded the lithosphere. We named six different cases as AL1 to AL6 (L denoting lithosphere). The first guesses in Case AL1 to AL5 had the same anomaly structures as described in Table 2, and they all had the same

TBL as in the target initial. Case A6 is with the SBI first guess. Comparatively, these first guesses in AL1–AL5 had more information on the target initial than those in A1–A5, because we assumed the correct TBL in these guess. On the other hand, Case AL6 (using the SBI first guess) had less information on the initial condition because the TBL had to be entirely recovered with the forward-adjoint looping. All models were realized with a 49×49×49 mesh with an under resolved lithosphere spanned with just two mesh points.

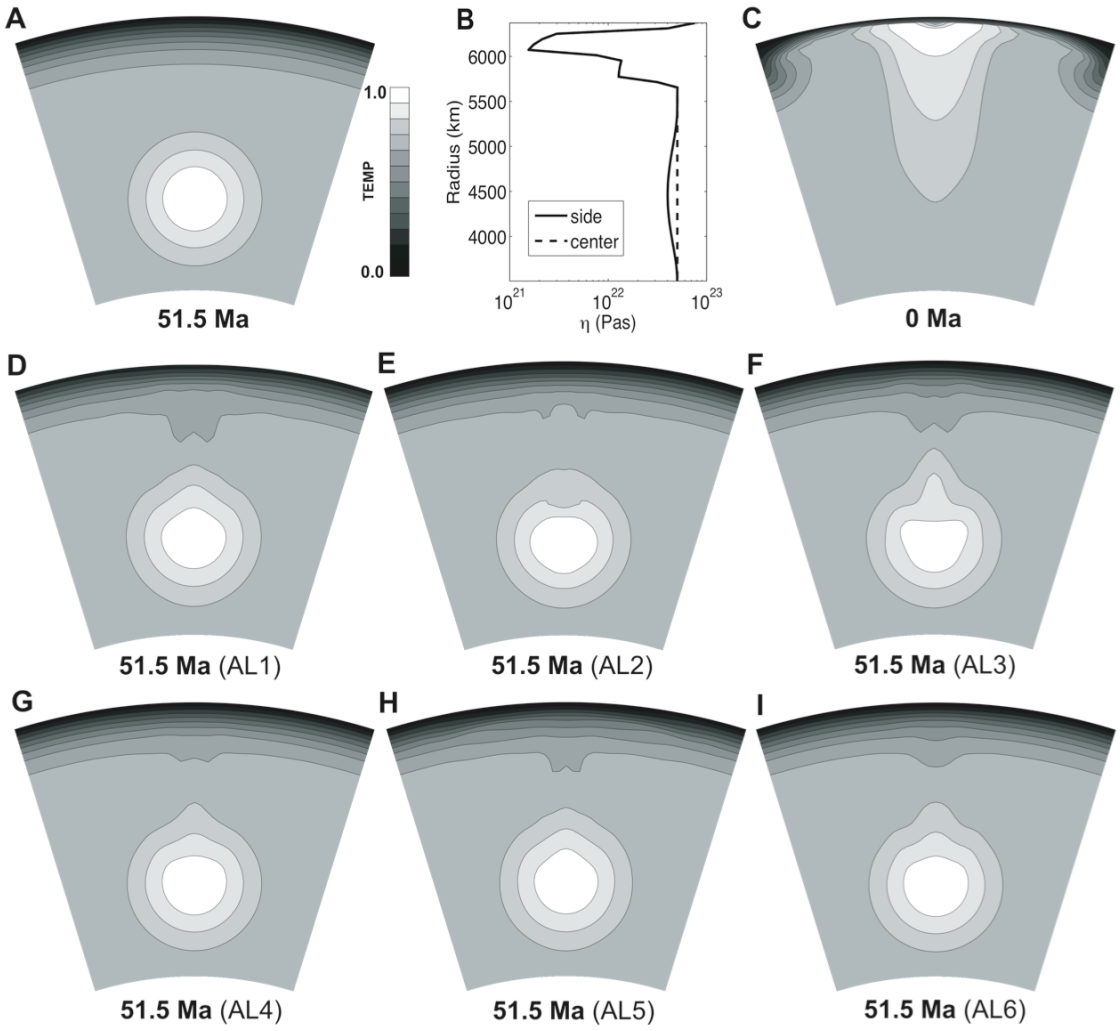

Figure 5 Three-dimension forward-adjoint models (with a 49×49×49 mesh) for a model with a radially stratified viscosity and a top thermal boundary layer. Shown is temperature through vertical cross-sections. Reference states at 51.5 Ma (A) and present (C). (B) Radial viscosity profiles in the reference initial condition. (D to I) Retrieved initial states from six different initial guesses (AL1–AL6) after 50 iterations.

Since these models are more complex, and thus more nonlinear, than models in Set I, a smaller damping factors α with $n_o = 2$ is adopted, in order to avoid overcorrection in the iterative process. We integrated the forward and adjoint equations for 50 iterations to obtain the solutions (Fig. 5C–H). Since the temperature field includes a TBL and a lower mantle anomaly, a small residual would entail recovering both well. The comparison of recoveries in these cases is not as obvious as that of only a rising Stokes sphere (Set I). Case AL5 with the closest initial guess also accumulated substantial errors through the nonlinear interaction between the rising spherical anomaly and the thermal erosion of the lithosphere. Case AL4 and AL6 both gave good recoveries with the smallest residuals between recovered and target initial condition (Fig. 6A), and AL6 among all cases had smallest residual between predict and target final condition (Fig. 6B). The SBI first guess (AL6) also led to fastest convergence, and most of the residuals were reduced within the first 10 iterations. From both the residuals and rate of convergence, we concluded that the SBI still gives the best initial guess even in such a complex model.

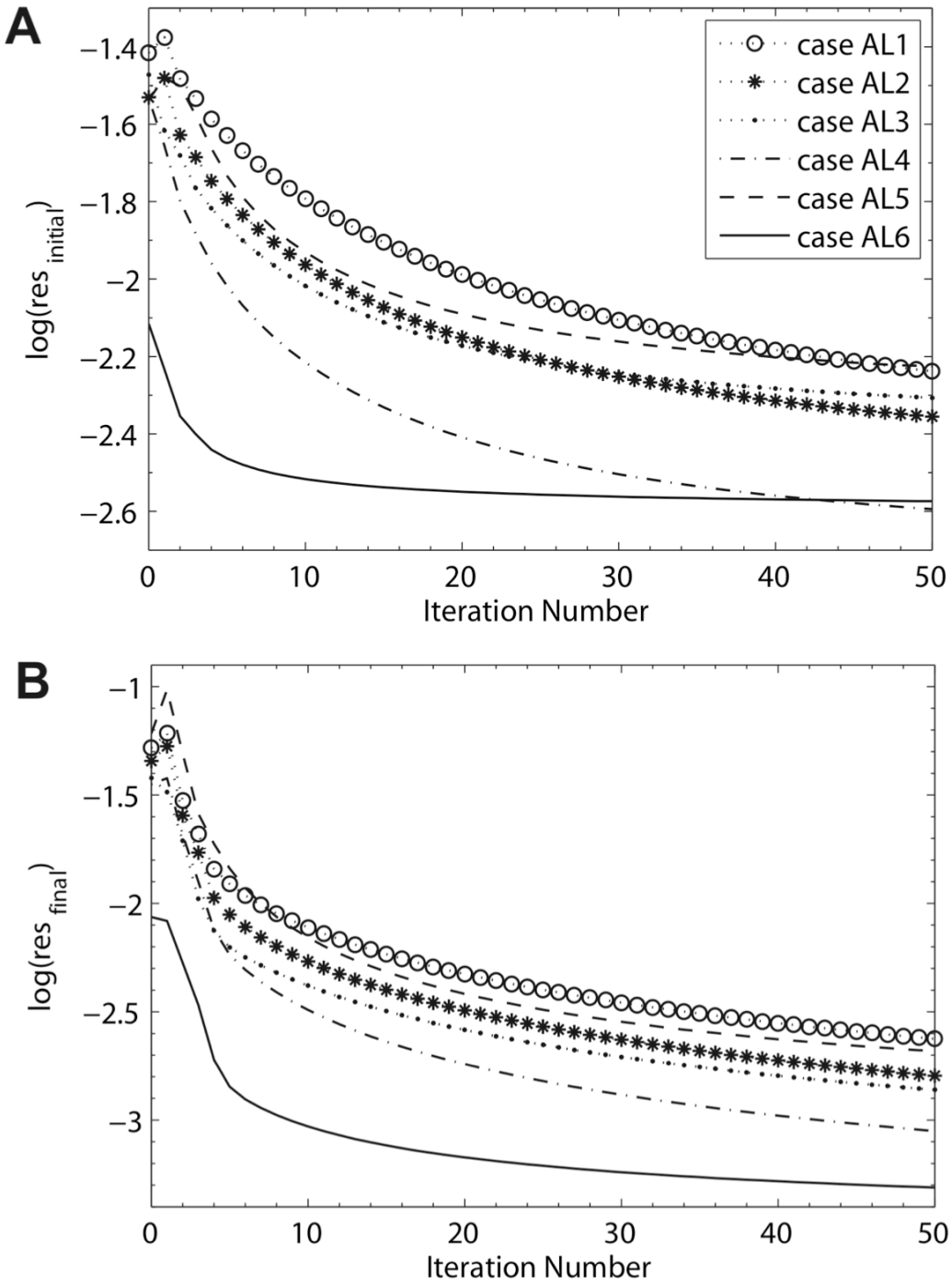

Figure 6 Same as Fig. 4 except for the models shown in Fig. 5.

Other experiments showed that by increasing n_o, hence decreasing α, we could decrease the terminal residuals of the initial condition in AL1 to AL5 upon convergence, some of which (AL4 and AL5) could be even smaller than in AL6, indicating a better recovery, but the terminal residual of the final condition in AL6 was always the smallest. However, for these tests, we had to increase the number of iterations to obtain the same amount of reduction of residuals; in other words, we reduced the rate of convergence substantially in A1 to A5, while A6 always had the fastest convergence and smallest residuals during most of the iterations. This indicates that the SBI first guess always produces good solutions with the least computational cost

2.2.3 Discussion

Inferring initial conditions with adjoint methods for mantle convection seems inherently ambiguous compared to atmospheric circulation problems where direct constraints on initial conditions from measurements in the system interior are used. Using a technique similar to that in Bunge *et al*. [2003], we first inferred initial conditions via the looping between forward and adjoint calculations to minimize the difference between a prediction and the final state of the mantle (a state that can be determined from seismic tomography). An optimal convergence requires some constraint on the initial condition. Starting the first forward calculation with an isothermal mantle was less efficient than with an initial guess obtained by the simple backward integration (SBI) of the convection equations that starts with present-day structure. Even when a Stokes sphere interacts with

and distorts a thermal boundary layer, where diffusion is important, the SBI first guess continues to provide a good estimate for the initial condition.

The adjoint method is an iterative gradient method that solves for a linearized problem. For the final solution to reach the global minimum in the residual space, the trial solution in the first iteration must be close to the true solution. Since the SBI initial guess makes use of present-day mantle information, this inverse of mantle convection approximates the true solution to first order. Therefore, the SBI initial guess guarantees a good solution with the adjoint method, as long as the model has not been run so long that diffusion at boundary layers dominates the problem. However, the approximation of initial conditions based on SBI will face difficulty when the anomalies reach a thermal boundary layer (TBL) and gradually diffuse away, which means an SBI estimate will not provide the same amount of buoyancy force. This is the natural limit for the adjoint method [Ismail-Zadeh *et al*., 2004].

The SBI initial guess is close to optimal for most mantle convection problems because advection dominates thermal diffusion with typical Peclet numbers ~10^3. To best approximate the true solution, an initial guess must capture its total buoyancy and geometry that we demonstrated with several numerical experiments in which either the buoyancy was underestimated or the initial location was incorrect. In these cases, the trial solutions all have large initial errors that must be iteratively removed with forward-adjoint looping. An idealized case with the correct initial location and buoyancy that is close to the actual initial condition recovers the initial condition nearly as well as with the SBI first guess. Since the SBI first guess involves the solution of the three conservation equations (Eq. 2–4), we

obtain a condition that has almost the same total buoyancy as that in the true solution and with its geometry defined through the coupled solution of flow and advection; this initial guess will, of course, lead to a good solution.

Seismology has revealed that the mantle has both low and high velocity regions that putatively represent a complex combination of thermal and chemical anomalies [Masters *et al*., 2000; Ishii and Tromp, 1999; Ni *et al*., 2002]. In these real cases where mantle anomalies have irregular geometry and amplitude, arbitrary initial guesses can hardly capture the true solution in the first place, and the SBI initial guess will be especially beneficial in retrieving a reasonable representation of the true initial condition.

However, it is worthwhile to point out that the tests performed in this chapter all assume that mantle properties, including the viscosity distribution, constitutive relation, and mantle density anomalies are perfectly known. In other words, these are idealized situations, which do not exist in the earth. For geophysical problems, many other unknowns need to be solved. Besides the dynamic properties like mantle viscosity and density anomaly, chemical composition and its temporal variation are other questions requiring solution. The numerical experiments shown in this chapter all treat mantle density anomalies as being thermal in origin, and possible chemical heterogeneities are not considered. In the upcoming chapters, I will try to address several of these issues.