A Discontinuous Galerkin Method with a Bound Preserving Limiter for the Advection of non-Diffusive Fields in Solid Earth Geodynamics

Ying He and Elbridge Gerry Puckett¹

Department of Mathematics, U. C. Davis, Davis, CA 95616, USA

Magali I. Billen

Department of Earth and Planetary Science, U. C. Davis, Davis, CA 95616, USA

Abstract

Mineral composition has a strong effect on the properties of rocks (e.g., viscosity, density, thermal conductivity) and is an essentially non-diffusive property in the context of large-scale mantle convection. Due to the non-diffusive nature and the origin of compositionally distinct regions in the Earth (e.g., crustal layers, chemical piles) the boundaries between distinct regions can be nearly discontinuous. Existing methods for tracking rock composition in mantle convection simulations are either computationally expensive, smear sharp boundaries or are not easily adapted to tracking multiple compositional fields. Here we present a discontinuous Galerkin method with a bound preserving limiter and a second order Runge-Kutta, strong stability-preserving time discretization method for the advection of non-diffusive fields (abbreviated as DG-BP). Using the falling box benchmark problem, we show that our method is accurate, stable and is computationally cost-effective. Compared to a finite element method using an artificial entropy viscosity approach the DG-BP method maintains much sharper compositional boundaries. In addition, when used with adaptive mesh refinement the DG-BP method results in fewer degrees of freedom due to smaller regions of mesh refinement in the neighborhood of the discontinuity. We also find that,

Email address: yinghe@math.ucdavis.edu (Ying He and Elbridge Gerry Puckett)

¹Corresponding Author

with the DG-BP method, the Stokes solver converges for a larger viscosity ratio compared to the finite element method: this is likely because the discontinuous solution has smaller viscosity gradients within individual elements. The DG-BP method provides many of the capabilities required for solid Earth geodynamics problems including efficiently tracking multiple non-diffusive compositional fields through all the physical processes expected within the mantle (e.g., subduction, mixing, stretching, entrainment) in both 2-D and 3-D computations. At the present time, the DG-BP method has only been implemented for incompressible flow. However, the approach presented here can be extended to the compressible formulation by using a weighted, essentially non-oscillatory limiter.

Keywords: Geodynamics, Composition Advection, Discontinuous Galerkin Method, Stokes Equation, Bound Preserving Limiter, Strong Stability Preserving (SSP) Runge-Kutta

1. Introduction

Study of the deformation and evolution of the solid Earth relies on computational geodynamics to explore a range of scientific questions from the origin of tectonic plates (Tackley, 2000) and large low seismic velocity provinces imaged
5 in the lower mantle (Karason and van der Hilst, 2000; McNamara and Zhong, 2004), to the processes governing subduction dynamics (Gurnis and Hager, 1988; Christensen, 1996), spreading at mid-ocean ridges (Gerya, 2010), and long term mixing of geochemical signatures in the deep mantle (Christensen, 1989; Kellogg, 1991; Kellogg et al., 1999). Because compositionally-distinct regions exist
10 within both tectonic plates and the mantle, the importance of modeling the complete thermo-chemical evolution of these systems has long been recognized and incorporated into computational studies (e.g., Davies, 1977; Hoffman and McKenzie, 1985). However, modeling the time evolution of compositional fields is challenging because these fields are non-diffusive, and thus require methods
15 that can track sharp boundaries. In addition, while there are finite element

methods that are well suited for solving the advection-diffusion equation for the diffusive temperature field, these same methods can be unstable or lead to unwanted smearing of boundaries when applied to a non-diffusive variable (Zhong et al., 2007).

20 Several approaches employed by the solid Earth geodynamics community to address the computational challenges of advecting a non-diffusive compositional field include (Zhong et al., 2007): 1) a small, but non-zero artificial diffusivity used to stabilize the standard methods (Kellogg, 1991; Kellogg and King, 1993), 2) a non-linear shock capture technique used to remove over/under-shoots and
 25 sharpen the compositional boundary (Lenardic and Kaula, 1993), and 3) tracer particles used instead of a continuous field (Christensen, 1989; Tackley and King, 2003; Gerya and Yuen, 2003a; McNamara and Zhong, 2004; O’Neill et al., 2006). Each of these approaches has benefits and costs, most notably that field-based methods (i.e., a continuous variable similar to temperature) tend to smooth
 30 sharp boundaries, while tracer methods are generally more expensive computationally. Before describing these methods and considering their strengths and weaknesses, it is useful first to consider two examples of how compositionally distinct regions enter into geodynamic simulations.

1.1. Non-Diffusive Fields in the Geodynamical Simulations

35 Two examples of geodynamic processes are illustrated in Figure 1: subduction of a tectonic plate and upwelling of a compositional *pile* at the base of the mantle. In both cases there are compositionally distinct regions (i.e., layers, blobs) and the properties of these materials (e.g., density, viscosity, radiogenic heating) may depend on the composition. Therefore, any approximations or
 40 errors in the advection of the compositional field will contribute to approximations or errors in the solution of the Stokes and the thermal advection-diffusion equations.

Subduction Zone. In the case of subduction, the main compositional structure is that of layering within the tectonic plates including sediments, crust, a residuum

Figure 1: Illustration of compositional variability in A) an upper mantle subduction zone, and B) the lower mantle near a compositional pile. Compositional variability exists as layers with both sharp (crust/lithosphere) and gradational (lithosphere/mantle) boundaries within the tectonic plates and at the base of the mantle, but these layers are stretched, broken, mixed and entrained by mantle flow. Multiple distinct compositional types need to be accurately tracked throughout this process.

45 left-over after melting and the mantle (Figure 1a). The subducting plate forms at a spreading center due to melting of the mantle with an approximately pyrolytic composition. Melting creates a thin crustal layer (5–10 km) overlying a residuum layer (15–25 km thick) composed of primarily olivine (e.g., harzburgite composition). The boundary between the crust and residuum layer is sharp
50 (a few hundred meters), while the boundary between the residuum and the pyrolytic mantle is gradational over tens of kilometers (Fischer et al., 2010). Similarly, the continental or overriding plate has a thicker crust (30–70 km) and residuum layer (up to ≈ 150 km), and the composition of these layers differ from each other and from these same layers in the oceanic plate.

55 While the mantle below tectonic plates does have heterogeneities in composition or geochemical signatures, these are not usually of primary interest for these types of subduction studies and therefore this layer is treated as compositionally homogeneous. One exception is the tracking of composition related to movement of volatile components, which can cause changes in the mineralogy

60 and melting within the mantle wedge (Gerya and Yuen, 2003b). For this system, tracking of composition includes tracking all the layers within the tectonic plates, as well as the entrainment and mixing of material into the mantle wedge above the subducting plate.

Compositional Piles. Large regions with low seismic velocity in the lower mantle have been interpreted as irregular layers of compositionally distinct material (e.g., Kellogg et al., 1999; Zhao et al., 2015). Simulations involving this kind of compositionally-distinct region often includes modeling the evolution of the whole mantle over the age of the Earth. In these simulations the focus is on how piles are moved around and deformed through interaction with the large-scale mantle flow as well as entrainment of parts of this layer into rising thermal plumes that form on or near the piles (Kellogg et al., 1999; McNamara and Zhong, 2005; Lin and van Keken, 2005; Tackley, 2012). Entrainment of this material into hot rising plumes is inferred from the geochemical signatures measured in rocks erupted at the Earth’s surface (e.g., Hofmann, 1997). Seismic observations suggest that the boundaries of these piles are sharp in some regions (e.g., Garnero and McNamara, 2008). In addition the possible preservation of these regions over > 4.0 billion years suggest that sharp boundaries are maintained by distinct material properties.

1.2. Previous Methods For Tracking Non-Diffusive Fields

80 The examples described above illustrate that for a range of geodynamical studies it is necessary to accurately track large continuous regions with sharp boundaries, as well as, properly treat the entrainment and mixing of these components over long time-scales. In the geodynamics literature two approaches have been used to track compositional variations within mantle convection simulations: field methods and tracer methods. In early models using field methods, a small artificial diffusivity was used (Kellogg, 1991; Kellogg and King, 1993; Christensen, 1989), characterized by the Lewis number ($Le = 1/\kappa$, where κ is the diffusivity). The use of a compositional field (rather than tracers) has

the advantage of using less memory and in some cases using the same solver
90 method as that used for the thermal field. For these early models, resolution
was severely limited by computational resources, so this approximation was rea-
sonable. However, as computational power increased and understanding related
to compositional structures in the mantle improved, so too has the need for
more accurate modeling of truly non-diffusive fields.

95 One of the first improvements was to adapt a two-step, non-linear shock cap-
ture method to mantle convection problems (Lenardic and Kaula, 1993). This
method removes dispersion errors by applying a global filter to remove any values
that are above or below the defined minimum and maximum of the field. This
method has been applied within several computational codes using advection of
100 compositional fields (e.g., Farnetani and Richards (1995); Moresi and Lenardic
(1997); Tackley and King (2003)). More recently Kronbichler et al. (2012) im-
plemented the entropy-viscosity (ν_k) stabilization scheme originally proposed
in Guermond et al. (2011) to obtain correct solutions to problems with high
Péclet number, such as mantle convection. In these problems, the standard
105 finite element discretizations introduce spurious oscillations around steep gra-
dients (Donea and Huerta, 2005). The entropy-viscosity stabilization removes
such oscillations, but tends to smooth and broaden initially sharp boundaries.

The use of tracer particles to track compositional fields in geodynamical
simulations began with applying the marker chain method to track the bound-
110 ary between two compositions (e.g., Woidt, 1978; Christensen, 1982). In two
dimensional (2-D) simulations, this results in a chain of roughly evenly-spaced
particles, and is a cost effective tracer method as it uses few tracers to track only
the boundary at high resolution. However, the number of tracers can quickly in-
crease for models with fine grids, or with mixing of the material, which can lead
115 to exponential growth in the length of the boundary (van Keken et al., 1997). It
is also difficult to extend the marker chain method to three dimensional (3-D)
models.

More recently, the level set method has been applied in geodynamical prob-
lems to track the interface between two compositions (Suckale et al., 2010;

120 Samuel and Evonuk, 2010). Unlike the marker-chain method, which tracks
the boundary between two compositions, the level-set method tracks a surface
whose zero contour is the boundary between the two compositions. Level set
methods have the advantage of working well in both 2-D and 3-D, and easily
tracking surfaces that fold and break apart. One challenge in implementing
125 level-set is tracking multiple (> 2) compositional fields.

More general tracer methods have become a popular approach for including
multiple compositional fields into geodynamical solutions, and offer the unique
advantage of being able to track the history of strain of a volume of material,
which can be related to observable seismic properties (e.g., seismic anisotropy).
130 Tracer particles have been added to existing eulerian, finite element method
(FEM) and finite difference (FD) simulation codes (McNamara and Zhong, 2004;
Tackley and King, 2003; Gerya and Yuen, 2003a) or more explicitly included
through Lagrangian particle-in-cell (PIC) methods, in which the tracers become
the integration points within an element on the Eulerian grid (O'Neill et al.,
135 2006).

One advantage of using tracers rather than a field is that there is no ex-
plicit diffusion or dispersion at the boundary between two compositional fields.
Instead, the solution to the Stokes equation is applied to advect the tracers,
and then the tracer properties are used to update the material properties for
140 the next time step. However, this last step can effectively lead to smoothing of
sharp boundaries due to interpolation from the tracers to the grid. Lagrangian
PIC methods avoid this smoothing by making the tracers the integration points.
The main limitation of tracers is the increased computational overhead. In par-
ticular, the tracer ratio method, in which tracers are used to represent all com-
145 positional fields present, requires only about 5 tracers per cell in 2D, compared
to > 40 tracers per cell if no tracers are used for the *background* material (van
Keken et al., 1997). However, the total number of tracers is still large for 2D
models and increases by an order of magnitude for 3-D simulations.

1.3. Discontinuous Galerkin (DG) Method with a Bound Preserving Limiter

150 To improve the solution for the advection of non-diffusive fields in geodynamics applications, we look to other disciplines for possible methods. The aim is to find a stable, efficient approach to accurately advect compositional fields while preserving the sharp discontinuities that exist in these problems. One such approach is the Discontinuous Galerkin method, which by-passes the continuous shape functions used to represent quantities in the FEM to allow for
155 discontinuities. Here, we present a new method of addressing the stability of the non-diffusive field advection problem in geodynamics simulations. Specifically, we apply the Discontinuous Galerkin (DG) method (Cockburn and Shu, 1998a), which stabilizes the advection equation by a monotone upwind flux, and
160 a bound preserving (BP) limiter (Zhang and Shu, 2010; Zhang et al., 2013).

The DG method was first introduced to solve the hyperbolic neutron transport equation (Reed and Hill, 1973). Later, Runge-Kutta Discontinuous Galerkin (RKDG) methods were developed for hyperbolic conservation laws (Cockburn and Shu, 1989; Cockburn et al., 1990; Cockburn and Shu, 1998b). More recently, a uniformly high order accurate DG scheme, which preserves positivity
165 of density and pressure for the compressible Euler equations of the gas dynamics was constructed (Zhang and Shu, 2010). Subsequently, a second order accurate DG scheme, which satisfies a strict maximum principle for general nonlinear convection diffusion equations on unstructured triangular meshes, was designed
170 (Zhang et al., 2013). For pure advection problems, one should note that there is no difference between the DG method and the LDG method, although we aim to extend this work to the time-dependent temperature advection-diffusion equation using the LDG method in future work.

The DG method has recently been applied to simulations of mantle convection using the FEniCS framework to study the time-dependent isothermal compositional convection (Rayleigh-Taylor instability) benchmark problem (Vynnytska et al., 2013). In this study, the incompressible Stokes equations are discretized using standard finite element method with Taylor-Hood elements
175 (Taylor and Hood, 1973). The discretization of the compositional equation is

180 carried out by an implicit Euler scheme in time, the DG method in space, and
a shock capture filtering method is used to correct for numerical dispersion
(Lenardic and Kaula, 1993). Their numerical examples demonstrated that the
discontinuous Galerkin scheme is an effective alternative for tracking compo-
sitional interfaces without significant numerical diffusion or dispersion errors.
185 However as can be seen in the figures presented, the shock-capture filtering
methods does not preserve any bound property and significant errors in the
compositional field still exist near the boundaries.

1.4. Overview of Work Presented

Here we present the DG method and with a bound preserving limiter (BP
190 limiter) to solve the compositional equation for a falling box benchmark prob-
lem, and compare our numerical results with a FEM approach with entropy
viscosity stabilization. The aim of the work presented here is not only tracking
compositional interfaces without significant numerical diffusion or dispersion
errors, but also to preserve the global bound of the composition. The global
195 bound on the composition is an important physical property, as any overshoot or
undershoot of the composition will be strongly amplified by the compositionally-
dependent viscosity and will therefore directly impact the solution to the Stokes
problem.

To the best of our knowledge, there is no related work on DG, LDG or RKDG
200 methods that has been applied to solid Earth geodynamics modeling problems
with the BP limiter, and in particular as to the equations of mantle convection.
Therefore, we present the computational scheme required for applying of the
DG method with BP limiter within a FEM mantle convection simulation code.
Specifically, we first describe the governing equations for our model problem and
205 formulate its weak form after the time discretization using a strong stability-
preserving (SSP) second order time discretization method (Sec. 2.1). Next, we
present the spatial discretization by the FEM, the stabilization techniques and
the preconditioner for the Stokes Problem (Secs 2.2–2.4). Then, we present the
DG discretization for the composition problem (Sec. 2.5), and the BP limiter

210 used to stabilize the DG scheme (Sec. 2.6). Finally, we review the Adaptive
mesh refinement (AMR) technique used in the work (Sec. 2.7).

To validate our numerical schemes and to demonstrate their efficiency and
stability we present some numerical results using a benchmark problem called
the falling box (Gerya and Yuen, 2003a) described in Section 3.1. We com-
215 pare the results and performance of the DG-BP method with a FEM using a
viscosity-entropy scheme (Secs. 3.2–3.4). We find that the DG method with a
BP limiter provides stable solutions without artificial smearing at compositional
boundaries and is computationally more efficient. We also show that choice of
time integration scheme is key to obtaining non-oscillatory results with the BP
220 limiter.

2. Methods

The model problem we consider is described by the equations of conservation
of momentum and mass (the Stokes Equation) for incompressible flow, and the
advection of a non-diffusive quantity, C :

$$-\nabla \cdot (2\eta(C)\epsilon(\mathbf{u})) + \nabla p = \rho(C)\mathbf{g}, \quad (1a)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (1b)$$

$$\frac{\partial C}{\partial t} + \mathbf{u} \cdot \nabla C = 0, \quad (1c)$$

where $\mathbf{u}$ is the velocity, p is the pressure, C is the composition, $\epsilon(\mathbf{u}) = \frac{1}{2}(\nabla\mathbf{u} +$
 $(\nabla\mathbf{u})^T)$ is the strain-rate tensor, $\rho(C) = \rho_0 + (\rho_1 - \rho_0)C$ is the composition-
dependent density, and $\eta(C) = (\frac{\eta_1}{\eta_0})^C \eta_0$ the composition-dependent viscosity
225 with positive constant reference densities ($\rho_0 \leq \rho_1$) and viscosities ($\eta_0 \leq \eta_1$).
These equations are solved with no-flux boundary conditions: $\mathbf{u} \cdot \mathbf{n} = 0$ and
 $\partial C / \partial \mathbf{n} = 0$,

One should notice that equations (1b)-(1c) imply

$$\frac{\partial C}{\partial t} + \nabla \cdot (\mathbf{u}C) = 0. \quad (2)$$

which is the conservative form of the equation for advection of the compositional
field C and is the preferred form to use with the DG method.

To illustrate the time discretization scheme used for the model problem we begin with the first order Forward-Euler time discretization for the composition advection equation. Given the initial time t_0 , time step size Δt and initial composition field C^0 , at current time $t_k = t_0 + k \cdot \Delta t$, we solve for current velocity $\mathbf{u}^k$ and pressure p^k from the Stokes equation (3a)-(3b) using the composition field C^k from the current time-step k . We also solve for the composition field C^{k+1} for the next time-step $k+1$ from the advection equation (3c) using the calculated current velocity $\mathbf{u}^k$, as shown below.

$$-\nabla \cdot (2\eta(C^k)\epsilon(\mathbf{u}^k)) + \nabla p^k = \rho(C^k)\mathbf{g}, \quad (3a)$$

$$\nabla \cdot \mathbf{u}^k = 0, \quad (3b)$$

$$\frac{C^{k+1} - C^k}{\Delta t} + \nabla \cdot (\mathbf{u}^k C^k) = 0. \quad (3c)$$

For higher order time discretization, when the FEM is used, one can apply the BDF2 or Crank-Nicolson methods. However, if the DG method is used, then a Strong Stability-Preserving (**SSP**) high order time discretization method (Gottlieb, 2005; Gottlieb et al., 2001) must be applied in order to maintain the bound preserving property from the limiter.

If the forward Euler method is stable under the time constriction $\Delta t < L$, then the SSP high order time discretization method satisfies the same stability under the time step restriction with a constant coefficient c given by

$$\Delta t < cL.$$

A simple optimal second order two stage **SSP** Runge-Kutta (RK) method with a constant $c = 1$ is given by:

$$C^{k+1/2} = C^k + \Delta t \nabla \cdot (\mathbf{u}^k C^k), \quad (4a)$$

$$C^{k+1} = \frac{1}{2}C^k + \frac{1}{2}C^{k+1/2} + \frac{1}{2}\Delta t \nabla \cdot (\mathbf{u}^{k+\frac{1}{2},*} C^{k+1/2}), \quad (4b)$$

where $\mathbf{u}^{k+\frac{1}{2},*}$ can be obtained either by solving equations (3a)(3b) with $C^{k+1/2}$ or extrapolation by using numerical solutions at previous time steps.

Alternatively, one can use the second order SSP three-step method with a constant $c = \frac{1}{2}$ given by:

$$C^{k+1} = \frac{3}{4}C^k + \frac{3}{2}\Delta t \nabla \cdot (\mathbf{u}^{\mathbf{k}} C^k) + \frac{1}{4}C^{k-2}. \quad (5)$$

One should be aware that the given second order SSP three-step method (5) assumes that the time step size is a constant in time. However, the second order two stage **SSP** Runge-Kutta (RK) method (4a)-(4b) can still be applicable for
240 variable time step size. We will show results for the DG method and FEM using both the RK and three-step time discretization schemes.

2.2. Spatial Discretization for the Stokes Problem

For the spatial discretization, the FEM is applied for Equations (3a) and
245 (3b), where we need to solve a discrete saddle point Stokes problem. For the composition advection problem (3c) both the FEM and DG method are applied in order to compare their performance.

After the spatial discretization of equations (3a) and (3b) by a FEM similar to the approach used in Kronbichler et al. (2012), we obtain the following matrix system

$$\begin{bmatrix} A_k & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \mathbf{U}_k \\ \bar{P}_k \end{bmatrix} = \begin{bmatrix} \mathbf{F}_k \\ 0 \end{bmatrix} \quad (6)$$

where

$$(A_k)_{ij} = (2\eta(C^k)\epsilon(\psi_j^{\mathbf{u}}), \epsilon(\psi_i^{\mathbf{u}})), \quad (7a)$$

$$B_{ij} = -(\psi_j^{\mathbf{p}}, \nabla \cdot \psi_i^{\mathbf{u}}), \quad (7b)$$

$$(\mathbf{F}_k)_j = (\rho(C^k)\mathbf{g}, \psi_j^{\mathbf{u}}), \quad (7c)$$

$$\mathbf{U}_k = \sum_j U_j^k \psi_j^{\mathbf{u}}, \quad (7d)$$

$$\bar{P}_k = \sum_j P_j^k \psi_j^{\mathbf{p}} \quad (7e)$$

and $\{\psi_j^{\mathbf{u}}\}$ and $\{\psi_j^{\mathbf{p}}\}$ are the two sets of basis functions of the finite element space, and $\{U_j^k\}_j$ and $\{P_j^k\}_j$ are the two sets of basis expansion coefficients

used for velocity and pressure, respectively. The matrix A_k itself is symmetric positive definite, however the larger block matrix

$$\mathbf{A} = \begin{bmatrix} A_k & B^T \\ B & 0 \end{bmatrix}$$

is indefinite due to the zero matrix located in the (2,2) block. Therefore, the Conjugate Gradient (CG) method can not be directly applied, but rather an iterative method is required. Furthermore, a good preconditioner is necessary to accelerate the convergence if an iterative method is used to solve the above matrix system (Eq. 6).

2.3. Preconditioner for the Stokes Problem

We can solve the system of equations given by equation (6) by using the Schur complement of the matrix block $\mathbf{A}$

$$BA_k^{-1}B^TP_k = BA_k^{-1}\mathbf{F}_k, \quad (8)$$

$$A_k\mathbf{U}_k = \mathbf{F}_k - B^T\bar{P}_k, \quad (9)$$

or

$$\begin{bmatrix} A_k & B^T \\ 0 & BA_k^{-1}B^T \end{bmatrix} \begin{bmatrix} U_k \\ \bar{P}_k \end{bmatrix} = \begin{bmatrix} F_k \\ BA_k^{-1}\mathbf{F}_k \end{bmatrix}. \quad (10)$$

Let $S_k = BA_k^{-1}B^T$ be the Schur complement, then a good right preconditioner is

$$\mathbf{Y} = \begin{bmatrix} A_k & B^T \\ 0 & -S_k \end{bmatrix} \quad (11)$$

and its inverse matrix is

$$\mathbf{Y}^{-1} = \begin{bmatrix} A_k^{-1} & A_k^{-1}B^TS_k^{-1} \\ 0 & -S_k^{-1} \end{bmatrix}. \quad (12)$$

We can easily verify that

$$\mathbf{A}\mathbf{Y}^{-1} = \begin{bmatrix} I & 0 \\ BA_k^{-1} & I \end{bmatrix}, \quad (13)$$

which is much better preconditioned than the matrix block $\mathbf{A}$. We refer the
 255 interested reader to May and Moresi (2008) for a more comprehensive study of
 the preconditioner of this type of problem, and particularly for problems arising
 in the geodynamics community for modeling mantle convection.

However in practice, we do not directly apply the preconditioner $\mathbf{Y}$ in (11),
 as it requires one to form the matrix vector products $A_k^{-1}x$ and $S_k^{-1}x$, which are
 260 equivalent to solving the matrix systems $A_k y = x$ and $S_k y = BA_k^{-1}B^T y = x$.
 These are computationally expensive to solve inside each GMRES iteration.
 Instead, we use an approximate preconditioner.

In Kronbichler et al. (2012), two types of approximate preconditioners $\widetilde{\mathbf{Y}}^{-1}$
 of preconditioner $\mathbf{Y}^{-1}$ are studied and implemented, where $\widetilde{\mathbf{Y}}^{-1}$ is in the format
 of

$$\widetilde{\mathbf{Y}}^{-1} = \begin{bmatrix} \widetilde{A}_k^{-1} & \widetilde{A}_k^{-1} B^T \widetilde{S}_k^{-1} \\ 0 & -\widetilde{S}_k^{-1} \end{bmatrix}. \quad (14)$$

The authors firstly propose to use $\widetilde{A}_k^{-1}$ and $-\widetilde{S}_k^{-1}$ as blocks in the preconditioner $\widetilde{\mathbf{Y}}^{-1}$, approximating the inverse matrices A_k^{-1} and $-S_k^{-1}$, where

$$(\widetilde{A}_k)_{ij} = (2\eta(C^k)\nabla(\Phi_j), \nabla(\Phi_i)), \quad (15)$$

$$(\widetilde{S}_k)_{ij} = (\eta^{-1}(C^k)\phi_j, \phi_i). \quad (16)$$

If η is a constant, then roughly we can consider the Schur complement $S_k = BA_k^{-1}B^T$ as an operator and assume the operators commute, which leads to

$$\begin{aligned} S_k &= BA_k^{-1}B^T = -\nabla \cdot (-\nabla \cdot (\eta\nabla))^{-1} \nabla \\ &= \nabla \cdot (\nabla \eta \cdot \nabla + \eta \Delta)^{-1} \nabla \approx \frac{1}{\eta} \Delta (\Delta)^{-1} = \frac{1}{\eta} \mathbf{I}. \end{aligned} \quad (17)$$

Therefore, $\widetilde{S}_k$ defined in equation (16) is a good approximation to S_k if $\eta(C^k)$ is
 a constant or in the case with only a small jump variation, as they are spectrally
 265 closed (Benzi et al., 2005).

However, if $\eta(C^k)$ has a big jump variation, then we can not ignore the effect
 of the term $\nabla \eta(C^k) \cdot \nabla$ in the derivation of equation (17). In this case $\widetilde{S}_k^{-1}$ is

not a good preconditioner, especially when C^k is obtained by the FEM. The numerical results discussed in a later section will demonstrate this observation.

270 Applying the right preconditioner $\widetilde{\mathbf{Y}}$ to equation (10) is equivalent to implementing the following two steps,

1. Solve b from $A\widetilde{\mathbf{Y}}^{-1}b = f$,
2. Solve x from $\widetilde{\mathbf{Y}}x = b$.

If we denote $x = (x_1, x_2)$ and $b = (b_1, b_2)$, then in Kronbichler et al. (2012), the
275 second step is implemented as below:

1. Solve $x_2 = -\widetilde{S}_k^{-1}b_2$ by performing an inexact CG solver with ILU preconditioner of the system $\widetilde{S}_kx_2 = -b_2$
2. Compute $y_1 = b_1 - B^Tx_2$.
3. Solve $x_1 = \widetilde{A}_k^{-1}y_1$ by performing an inexact CG solver for the system
280 $\widetilde{A}_kx_1 = y_1$ with ML-AMG preconditioner based on the matrix $\widetilde{A}_k$.

In the third step as shown above, we refer to the preconditioner as an ***X preconditioner*** and $\widetilde{A}_k^{-1}$ is chosen to be only a single, cheap V-cycle with $\widetilde{A}_k$ at the expense of more outer iterations.

Notice that in equation (7a), if the viscosity parameters $\eta_1 = \eta_0$, i.e., the
285 viscosity is a constant $\eta = \eta_0$, then it is independent of composition. Therefore, the matrix $A_k = A_{k-1} = \dots = A_0$ is time-independent, which is the most efficient and easiest case, as we do not need to recompute the matrix at every time-step and the matrix system is even better conditioned. However, in this study, our interests are more in those cases with viscosity variations on the
290 order of 10 and 100. Consequently, the matrix A_k needs to be recomputed at every time-step, and therefore for future studies with larger viscosity variations, seeking a good preconditioner to reduce the time cost will be important. Future work will explore better preconditioners, but that is beyond the scope of the present study. Instead, we refer the interested read to May and Moresi (2008),
295 which presents several different types of preconditioners with their pros and cons, and demonstrates the numerical performance of those preconditioners with examples.

2.4. FEM Stabilization of the Composition Advection Problem

For the purposes of testing and comparing the DG method developed here we also compute solutions using an FEM with entropy-viscosity stabilization (Kronbichler et al., 2012) for the composition advection (equation 3c). In this case the resulting matrix system is

$$MC^{k+1} = MC^k + \Delta t \cdot G^k \quad (18)$$

where $M_{ij} = (\psi_j, \psi_i)_\Omega$ is the mass matrix, and the right hand side vector G^k is calculated using previous time-step solutions C^k and $\mathbf{u}^k$ by the formula

$$G_j^k = (\mathbf{u}^k \nabla C^k, \psi_j)_\Omega + (\nu_k \nabla C^k, \nabla \psi_j)_\Omega \quad (19)$$

where ν_k is the entropy viscosity function, which only adds artificial diffusion

300 where the local Péclet number is large and the solution is non-smooth.

2.5. DG Discretization of the Composition Advection Problem

For the DG method the spatial discretization is different from the FEM because it must allow for discontinuities between elements. We first consider the theoretical basis for this method and then explain how it is implemented.

We denote our computational domain as $\Omega = \cup_{e=1}^E \Omega^e$, where Ω^e represents non-overlapping body-conforming quadrilateral elements. We consider a local approximate solution $u^e(x, y)$ that has the local nodal representation

$$u^e(x, y) = \sum_{i,j=0}^N u_{ij}^e \psi_{ij}^e(\xi(x), \eta(y)), \quad (20)$$

305 where $\{(x_i, y_j)\}_{i,j=0}^N$ denotes the set of grid points on Ω^e . The basis coefficients u_{ij}^e are the nodal values $u^e(x_i, y_j)$ on Ω^e , and the basis $\psi_{ij}(\xi, \eta) = \ell_i(\xi)\ell_j(\eta)$, or simply ψ_{ij} , has a tensor product form of multi-dimensional N th-order Lagrange interpolation polynomials based on $N + 1$ grid points in each direction. Grid points can be equa-distant (FEM), or Gauss-Legendre-Lobatto (GLL) quadrature grid points (i.e., the Spectral Element Method).

310

We map each physical coordinate $(x, y) \in \Omega^e$ onto the reference domain $(\xi, \eta) \in I = [-1, 1]^2$ through the affine mapping and formulate the computational scheme on the reference domain. Let us define a finite-dimensional approximation space $V_N \subset H^1(\Omega)$ such that $V_N = \text{span}\{\psi_{ij}^e(\xi, \eta)\}_{i=0, j=0, e=1}^{N, N, E}$. With this approximation space, one can represent the global solution as the direct sum of local piecewise polynomials

$$u(x, y) \simeq u_h(x, y) = \bigoplus_{e=1}^E u_h^e(x, y), \quad (21)$$

where $u_h^e(x, y)$ is defined on each element Ω_e by equation (20). At the overlapping interfaces the DG space functions do not require any continuity, which is the main difference compared with continuous Galerkin approach.

When the DG method is applied (Cockburn and Shu, 1998a; Hesthaven and Warburton, 2008), after the spatial discretization of equation (3c), for the weak formulation one must find $C^{k+1} \in V_N$ such that

$$(C^{k+1}, w)_{\Omega_e} = (C^k, w)_{\Omega_e} + \Delta t (C^k \mathbf{u}^k, \nabla w)_{\Omega_e} - \Delta t (\mathbf{u}^k \cdot \mathbf{n} C^{k,*}, w)_{\partial\Omega_e}, \quad (22)$$

for any $w \in V_N$, where an upwind flux is used for $C^{k,*}$ (Cockburn and Shu, 1998a; Hesthaven and Warburton, 2008) and is defined as

$$C^{k,*} = \begin{cases} C^{k,-}, & \text{if } \mathbf{u}^k \cdot \mathbf{n} > 0 \\ C^{k,+}, & \text{if } \mathbf{u}^k \cdot \mathbf{n} < 0 \end{cases} \quad (23)$$

and $C^{k,-}$ = local/interior solution on Ω_e , $C^{k,+}$ = neighbor/exterior solution of Ω_e . Although, equation (22) appears to be only defined locally on each element Ω_e , it actually also depends on the adjacent solutions through the flux term $C^{k,*}$, which is defined at each element interface using two side values. Generally, the flux term $C^{k,*}$ is the most difficult part to determine when one wants to design a DG scheme, as it is the key point used to ensure the scheme's stability.

If we consider the following DG numerical expansions in V_N

$$C^{k+1}(x, y) = \bigoplus_{e=1}^E \sum_{i,j=0}^N C_{ij}^{k+1,e} \psi_{ij}^e(\xi(x), \eta(y)), \quad (24)$$

$$C^k(x, y) = \bigoplus_{e=1}^E \sum_{i,j=0}^N C_{ij}^{k,e} \psi_{ij}^e(\xi(x), \eta(y)), \quad (25)$$

$$\mathbf{u}^k(x, y) = \bigoplus_{e=1}^E \sum_{i,j=0}^N \mathbf{u}_{ij}^{k,e} \psi_{ij}^e(\xi(x), \eta(y)), \quad (26)$$

and the mass matrix $M = \text{diag}(M^1, \dots, M^E)$ where

$$(M^e)_{\hat{i}\hat{j}, i\hat{j}} = (\psi_{i\hat{j}}^e(\xi(x), \eta(y)), \psi_{i\hat{j}}^e(\xi(x), \eta(y)))_{\Omega_e} \quad (27)$$

and the right hand side vector $G^{k,*} = (G^{k,1}, \dots, G^{k,E})^T$ where

$$\begin{aligned} (G^{k,e})_{\hat{i}\hat{j}} &= (C^k \mathbf{u}^k, \nabla \psi_{i\hat{j}}^e(\xi(x), \eta(y)))_{\Omega_e} \\ &\quad - (\mathbf{u}^k \cdot \mathbf{n} C^{k,*}, \psi_{i\hat{j}}^e(\xi(x), \eta(y)))_{\partial\Omega_e} \end{aligned} \quad (28)$$

then the matrix system of equation is given by

$$MC^{k+1} = MC^k + \Delta t \cdot G^{k,*}. \quad (29)$$

320 2.6. Adding a Bound-Preserving (BP) Limiter

Following an idea similar to that presented by Zhang and Shu (2010), after each time-step a BP limiter is applied to the newly obtained numerical solution C^{k+1} in a post-processing procedure. Therefore, we modify the scheme in equation (29) as follows

$$MC^{k+1} = M\tilde{C}^k + \Delta t \cdot G^{k,*}, \quad (30a)$$

$$\tilde{C}^{k+1} = \text{BPLimiter}(C^{k+1}) \quad (30b)$$

where $\text{BPLimiter}(\cdot)$ represents the BP limiter applied to the DG solution of (30a). At time step k , if the DG solution is bounded by $m \leq C^k \leq M$, then after applying the limiter, the solution C^{k+1} is guaranteed to satisfy the same upper and lower bounds $m \leq C^{k+1} \leq M$.

325 In previous studies three key ingredients required for constructing a high
order maximum-principle-satisfying DG scheme were introduced (Zhang and
Shu, 2010; Zhang et al., 2013):

1. The composition cell average at the next time step by forward Euler time
discretization scheme is a monotone function with respect to certain point
330 values (Gauss-Lobatto points for the one-dimensional case), under a suit-
able CFL condition.
2. A simple scaling limiter modifies the DG polynomial $p(x)$ on cell I_e into
 $\tilde{p}(x)$ such that $\tilde{p}(x) \in [m, M]$ at these special points without changing
its cell average. Moreover, it can be proven that the modified polynomial
335 $\tilde{p}(x)$ is also a high order approximation just as $p(x)$. Thus we have $\bar{C}^{k+1} \in$
 $[m, M]$ if all the degrees of freedom at time level k in the right hand side
of equation (29) are replaced by using those of modified polynomials $\tilde{p}(x)$.
3. The forward Euler method is replaced by a strong stability preserving
(SSP) high order time discretization scheme, which is a convex combina-
340 tion of forward Euler and thus, will preserve the bounds.

In our work, we follow these steps as a post-processing to the numerical solution
 C^{k+1} in (30a) to obtain a modified solution $\tilde{C}^{k+1}$ such that its DG polynomials
at some special points are bound preserved and $\bar{\tilde{C}}^{k+1} = \bar{C}^{k+1} \in [m, M]$, which
is then used again in the next time step to ensure the cell average $\bar{C}^{k+2}$ is also
345 bound preserved.

2.7. Adaptive Mesh Refinement

Adaptive Mesh Refinement (AMR) is used to refine and, or coarsen the grid
size subject to user defined criteria such temperature or strain-rate gradient
across a cell (Berger and Oliger, 1984; Verfürth, 1996). This method has been
350 broadly applied in the area of scientific computing, such as in the field of geo-
dynamics, where large 3D simulation is usually required. The advantage of the
AMR technique is that it provides the same overall accuracy as a uniformly re-
fined mesh with the same minimal mesh size, but numerical cost is significantly

reduced by increasing mesh size where lower resolution is sufficient. In our work,
355 all the numerical tests use the Kelly error indicator (Kelly et al., 1983) as the
error estimator for the criterion of mesh refinement, which is implemented in
deal.II (Bangerth et al., 2015). This criterion has been found to be simple, yet
universally useful in adaptively refining meshes.

The accuracy of the FEM and the DG method is mostly determined by the
360 overall minimal mesh size if the same order of degree of polynomial basis is used.
With the same uniformly refined mesh, the cost of the DG method is much
more expensive than the FEM method when using a lower order polynomial
basis (degree $k \leq 3$). This is because the DG method uses extra degrees of
freedom between the element interfaces. However, if we use the AMR method
365 to partition the mesh with the same global minimal mesh size, the need for more
degrees of freedom is not obviously true any more because fewer elements may
be needed to resolve the discontinuous boundary. Indeed, using our numerical
examples we will show that for problems with a discontinuity the cost of the
DG method is much cheaper than the FEM.

370 **3. Numerical Results**

3.1. Falling Box Benchmark Problem

In this section, we apply the DG method with BP limiter to the sinking hard
box benchmark problem first used by Gerya and Yuen (2003a) to demonstrate
the capabilities of a characteristics-based marker-in-cell method simulation code,
375 I2ELVIS. The benchmark problem is defined on a 2-D Cartesian domain, which
is 500 km $\times$ 500 km. A small block (100 km $\times$ 100 km) is placed with its top
edge at 50 km below the top of the domain and centered horizontally (Fig. 2-a).
The viscosity contrast between the box is varied as a way of testing the codes
ability to give accurate and stable solutions when there is a sharp contrast in
380 material properties.

We present numerical results for three examples. In all of the models the

η_1/η_0	# Cells	# Dofs($\mathbf{u} + \mathbf{p}, C$)	Over (C, η)	Under (C, η)
FEM				
1	20404	(190543, 84677)	0.018%	0.100%
10	17020	(158827, 70581)	0.034%	0.072%
100	NA	NA	NA	NA
DG with BP Limiter, SSP RK2				
1	7759	(77384, 69831)	0.014%	0.022%
10	6271	(62628, 56439)	0.050%	0.015%
100	5353	(53403, 48177)	0.174%	0.000%
DG with BP Limiter, SSP 2nd order multi-step				
1	7711	(76871, 69399)	0.014%	0.026%
10	6184	(61854, 55656)	0.050%	0.000%
100	5317	(52998, 47853)	0.178%	0.003%

Table 1: Comparison of Numerical Test Results. Results are presented at time step 5000 for the cases $\eta_1/\eta_0 = 1, 10$, and at time step 4000 for the case $\eta_1/\eta_0 = 100$. For each viscosity ratio, we list the number cells resulting from the AMR, the number of degrees of freedom for the Stokes system ($\mathbf{u} + \mathbf{p}$) and the advection solver (C) and the overshoot/undershoot of the compositional field.

following parameters are held fixed:

$$\begin{aligned}
\mathbf{g}' &= (0, 9.8) \text{ m/s}^2, & \text{grav. acceleration} \\
\tilde{a} &= 500 \text{ km} & \text{domain height and width} \\
\eta_0 &= 10^{21} \text{ Pa} \cdot \text{s} & \text{background viscosity} \\
\rho_0 &= 3200 \text{ kg/m}^3, & \text{background density} \\
\rho_1 &= 3300 \text{ kg/m}^3, & \text{small box density}
\end{aligned} \tag{31}$$

while the viscosity of the small box, η_1 , is increased from 10^{21} Pa·s to 10^{23} Pa·s for each test, i.e. the viscosity contrast between the small, hard block, and the softer surroundings is increased from 1 to 10 to 100. The initial location and dimension of the small box is defined using the composition field, which is given as

$$C_0(x, 0) = \begin{cases} 1, & \text{if } (x, y) \in [\frac{3}{8}\tilde{a}, \frac{5}{8}\tilde{a}] \times [\frac{5}{8}\tilde{a}, \frac{7}{8}\tilde{a}] \\ 0, & \text{otherwise} \end{cases}. \tag{32}$$

In all tests we use the same maximal refinement level $l = 10$, global minimal mesh size $h = (\frac{1}{2})^{l-1}\tilde{a}$. To compare the numerical results at the same final time, the time-step size is determined by the CFL constraint ($\Delta t = 0.05 \frac{h}{\min\{\gamma 10^{-4}, \|\mathbf{v}\|_\infty\}}$), where $\gamma = 2, 1.5, 1$ for the viscosity ratio 1, 10 and 100 respectively. However, in general, $\Delta t = 0.1 \frac{h}{\|\mathbf{v}\|_\infty}$ is sufficient for the FEM and the DG-SSP-RK2 methods. To stabilize the numerical scheme, the DG method uses the upwind flux and the BP limiter, as well as the SSP time discretization schemes in section 2.1. For comparison we also run the benchmark problems using the FEM with the entropy-viscosity stabilization of the composition advection equation with explicit BDF2 time discretization scheme. All these tests are implemented using the deal.II C++ open source library (Bangerth et al., 2015).

The falling box benchmark is a good test case for our DG-BP method for several reasons. First, it is a simple model problem to implement. Second, it has a strong discontinuity, which makes the numerical simulation much more challenging. Third, the deformation of the box is different for different viscosity ratios providing a straightforward approach for examining how the scheme

performs for different shapes of compositional variation. For comparison, the numerical results for the same benchmark problem studied in Gerya and Yuen (2003a) and are shown in Fig. 2. It is important to note that these figures only show the final positions of the tracer particles. In order to get a composition field, one needs to do a linear interpolation from the tracers particles to the mesh. A similar sinker problem is also studied in May and Moresi (2008), where a purely Stokes problem is considered for variable jump viscosities.

3.2. Comparison of DG With and Without a BP Limiter

For the problems that satisfy the conservation laws, the DG and FVM methods have been demonstrated to have better numerical stability and numerical convergence than the FEM. A first order FVM can maintain a numerical maximal principle as the analytic solution does. However, without the help of a limiter, the numerical scheme derived by the DG method can not satisfy the maximal principle or the bound preserving property.

To illustrate the importance of the BP limiter, we show the numerical results for the DG case using the second order SSP Runge-Kutta time discretization algorithm with and without a BP limiter in Fig. 3. For this test, we consider the viscosity variation to be constant, i.e. $\eta_1/\eta_0 = 1$. Thus, only the composition field and density have a discontinuity. For the refinement at timestep 5000, the DG method needs more refinement near the discontinuous interface without applying the BP limiter (Fig. 3A left, # cells: 9589) than that with applying a BP limiter (Fig. 3A right, # cells: 7759). We zoom in on the compositional field in the top and bottom parts of the inner box in Fig. 3B (left: no limiter; right: BP limiter). The DG without the BP limiter results show a broad numerical boundary layer. Furthermore, the overshoot (undershoot) without the BP limiter (Fig. 3B left) is over 24% (16%). A horizontal profile (Fig. 3C) across top part of the inner box confirms that the DG-BP method is much better at preserving sharp compositional boundary without overshoot/undershoot. Therefore, for the remainder of the tests, we always apply the BP limiter to the DG method.

Figure 2: Numerical results of the falling box benchmark from Gerya and Yuen (2003a) (Reproduced with permission – pending). The location of the box is tracked and illustrated by the location of the tracers particles. A) Initial set-up, B) Case $\eta_1/\eta_0 = 1$ after 9.886 Myr, C) Case $\eta_1/\eta_0 = 10$ after 15.446 Myr, and D) Case $\eta_1/\eta_0 = 100$ after 19.623 Myr. The initial set-up used in this study is identical to that shown in (A) but uses a compositional field instead of tracer particles.

Figure 3: Comparison of DG with and without the BP limiter. A) Composition shown on AMR mesh at timestep 5000 for DG without limiter (left) and with limiter (right), B) Enlarged-view Comparison of Compositional Field for Top and Bottom of Falling Box, C) Horizontal profile across top. Location of profiles are shown in top sub-figure of (B).

3.3. Comparison of FEM and DG Method

To compare the numerical performance between the FEM and DG method, we take three examples with the parameters given in (31) and initial condition given in (32), but with each test using a different viscosity ratio of $\eta_1/\eta_0 = 1, 10$, and 100.

The composition field and AMR mesh for the DG-BP method and the FEM are shown in Fig. 4, Fig. 5 and Fig. 6 for increasing values of η_1/η_0 . As we can see from the figures of the composition field, the DG-BP method and FEM achieve similar accuracy and preserve the bound (between 0 and 1) within acceptable accuracy (all less than 0.2%). However, from the figures showing the mesh resulting from AMR, we can see that the DG-BP method uses fewer cells and therefore, fewer degrees of freedom for the Stokes solver (see also Table 1). Therefore, the computational cost of the DG-BP method is much smaller than the FEM. It is also worth noting that the DG-BP method continues to be stable and to preserve the sharp compositional boundary as the viscosity contrast is increased to $\eta_1/\eta_0 = 100$. In contrast, the FEM method fails to converge for the higher viscosity ratio case. One reason for this difference in convergence behavior may be that for our model problem, the FEM is more sensitive to the choice of the preconditioner for the Stokes solver than the DG-BP method, as

Figure 4: Comparison of Compositional Field for Case $\eta_1/\eta_0 = 1$. Composition shown on AMR mesh at time-step 5000 for, A) FEM and C) DG method with BP limiter. B) Compositional field at initial time (top) and time-step 5000 (bottom) for FEM (left) and DG-BP (right).

Figure 5: Comparison of Compositional Field for Case $\eta_1/\eta_0 = 10$. Composition shown on AMR mesh at time-step 5000 for, A) FEM and C) DG method with BP limiter. B) Compositional field at initial time (top) and time-step 5000 (bottom) for FEM (left) and DG-BP (right).

Figure 6: Case $\eta_1/\eta_0 = 100$ for DG method with BP limiter (FEM tests failed to converge). A) Composition shown on AMR mesh at time-step 4000. B) Compositional field at time-step 4000.

here we only apply a very simple and general preconditioner.

We zoom in on the compositional field in the top part of the inner box for the first two cases ($\eta_1/\eta_0 = 1, 10$) in Fig. 7. The FEM results show a smooth, broad (smeared out) numerical boundary layer. The smoothing effect extends into the tail forming above the deformed corners of the box in the case with $\eta_1/\eta_0 = 1$ and has the effect of truncating the length of the full compositional anomaly in the tail. The DG-BP method preserves much sharper boundaries and shows that the full compositional anomaly extends to the very narrow tip of the tail.

The importance of using a second order SSP time discretization scheme is also demonstrated by comparing to the results for the DG method using a first order forward Euler time discretization scheme (Fig. 7 C, F). The lower order time discretization scheme causes the boundary layer to smear out and causes oscillations within this smeared boundary layer. Therefore, the second order SSP Runge-Kutta time discretization scheme is necessary to preserve sharp, non-oscillatory compositional boundary when applying the DG method with the BP limiter.

Figure 7: Enlarged-view Comparison of Compositional Field for Top of Falling Box. Case with $\eta_1/\eta_0 = 1$ with A) FEM, B) DG with BP limiter and second order SSP RK2 time-step algorithm, and C) DG with BP limiter and 1st order forward Euler algorithm. Case with $\eta_1/\eta_0 = 10$ with D) FEM, E) DG with BP limiter and second order SSP RK2 time-step algorithm, and F) DG with BP limiter and 1st order forward Euler algorithm. Arrows point to tail (top) and corner (bottom) features discussed in the text.

4. Discussion

465 The numerical tests confirm that the DG-BP method is more effective than FEM at preserving sharp compositional boundaries (Figure 8). While the FEM, using the viscosity-entropy stabilization, smears the boundaries of the falling box over a large number of elements, the DG-BP method maintains a sharp boundary over only a few elements. The numerical tests also show that the DG-
 470 BP method uses fewer degrees of freedom than the FEM when implemented with the AMR scheme. This is because a smaller region of highly refined elements is needed at the sharp boundary maintained in the DG-BP method, while this region expands to cover most of the falling box and its surroundings for the FEM.

475 We also find that while the FEM is not able to converge for a viscosity contrast of 100 while the DG method does converge. One reason for the improved convergence behavior with the DG method is the smaller gradient of viscosity within the interior of each cell, which occurs because the viscosity can be discontinuous like the composition on which it depends. In contrast, the gradient of

Figure 8: Profiles of Compositional field across falling block for case with $\eta_1/\eta_0 = 10$. A) Vertical profile bisecting falling block: DG – blue, FEM – red. B) Horizontal profile across bottom edge. C) Horizontal profile across top. Location of profiles are shown in (D) for DG result and (E) for FEM result. Arrows point to locations where the differences between the two results clearly illustrate the strengths (sharp compositional boundaries for DG method) and weaknesses (significant smoothing for FEM method) of each method.

viscosity from the FEM must be smooth, and therefore, there are larger gradients in viscosity within individual cells. Because the simple preconditioner used here ignores gradients in viscosity, large viscosity gradients can cause failure of convergence.

One of the main goals for implementing the DG-BP method was to maintain sharp boundaries with small overshoot and undershoot errors ($< 0.01 - 0.1\%$ depending on viscosity jump) and the results show that the DG-BP method applied with a SSP second order Runge Kutta scheme does achieve this goal. However, it is also clear that for both the FEM and DG-BP method the size of the overshoot error depends on the viscosity contrast (Table 1). To understand the source of this error we re-ran the DG-BP models using a uniform mesh with an element size corresponding to the finest resolution in the AMR models. In this case, the overshoot error is smaller than in the FEM test results. In addition, we ran tests where the AMR method was only allowed to refine and not coarsen the element size: in this case the error is the same as that for the uniform mesh. Therefore, we conclude that the AMR algorithm plays a role in

introducing errors in the DG-BP method.

Further examining the problem specification, we find that the original source for this error is in fact the choice of (P_2, P_1) Taylor-Hood (or Hood-Taylor) elements (second order spatial element basis for velocity field and first order spacial
500 element basis for pressure) for the Stokes problem. For this choice of elements, the flow solution is not exactly point-wise divergence free ($\nabla \cdot \mathbf{u} \neq \mathbf{0}$) and will introduce a small error ($\mathcal{O}(h)$). Application of the BP limiter requires that the flow is point-wise divergence free in order to keep the composition within the upper and lower bounds ($C \in [0, 1]$). Because this requirement is violated by
505 the choice of (P_2, P_1) elements, BP limiter is not able to keep the composition values strictly between the bounds. When the AMR algorithm is applied the error introduced by the element type is magnified. This source of error can be eliminated by choosing an element type that will give a point-wise divergence free solution for the Stokes Equation, such as (P_2, P_1^{-1}) elements, with a second
510 order spatial element basis for the velocity field and a discontinuous first order spacial element basis for the pressure.

4.1. Limitations of the DG Method with BP limiter

The numerical results clearly demonstrate the advantages of the DG-BP method, however, there are some limitations to the DG-BP method as presented
515 here, which will need to be improved in the future.

First, the requirement for a divergence free flow solution for use of the BP limiter also means that the method presented here can only be used for the incompressible (Boussinesq or Extended Boussinesq) form of the conservation of mass and momentum. This limitation can be better understood by considering the form of the conservation equation used in the DG-BP scheme. Equation (1c) is the non-conservative form of the conservation equation,

$$\frac{\partial C}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{C} = \mathbf{0}, \quad (33)$$

which implies that the compositional field satisfies the global bound $[0, 1]$ at each time-step. If the material is incompressible, then the above equation is

equivalent to the conservative form,

$$\frac{\partial C}{\partial t} + \nabla \cdot (C\mathbf{u}) = \mathbf{0}, \quad (34)$$

in which the second term is divergence of a flux term $f(C) := C\mathbf{u}$. This term can be written in a form that gives a weakly bound preserved average compositional field on each cell, which is needed to apply the BP limiter,

$$\bar{C}^{k+1} = \bar{C}^k - \frac{\Delta t}{|E|} \left(\int_{\partial E^{out}} f(\tilde{C}^{k,*}) ds - \int_{\partial E^{in}} f(\tilde{C}^{k,*}) ds \right), \quad (35)$$

$$\tilde{C}^{k+1} = \text{BPlimiter}(C^{k+1}) \in [0, 1], \quad (36)$$

$$\bar{\tilde{C}}^{k+1} = \bar{C}^{k+1} \in [0, 1]. \quad (37)$$

Note, in this case, the average compositional field on each cell at the current time step only depends on the cell average at the previous time step and the flux terms on the cell interface.

For the compressible form of the conservation equation, the non-conservative form is the same as above, and so the compositional field still satisfies the global bound $[0, 1]$ at each time-step. However, the local conservative form of the equation is,

$$\frac{\partial(\rho C)}{\partial t} + \nabla \cdot (\rho C\mathbf{u}) = \mathbf{0}. \quad (38)$$

In this case the flux term is $\rho C\mathbf{u}$ and not $C\mathbf{u}$. However, for a compressible
520 flow there is no local bound on the term ρC because the density within the cell can also change due to thermal or pressure induced expansion or contraction. However, it may be possible to stabilize the DG method with compressible flow using a weighted essentially non-oscillatory (WENO) limiter (Qiu and Shu, 2005; Zhu et al., 2008), which limits spurious oscillations of the numerical solution in
525 the presence of strong discontinuities.

Second, for the problems with smooth solutions, the FEM is much more computationally efficient than the DG-BP method due to the number of degree of freedoms required to achieve the same order of accuracy. Therefore, for practical considerations, the combination of applying DG-BP method to solve
530 the advection equation and applying FEM to solve Stokes equation is a good method only for cases with sharp boundaries or discontinuities.

Third, for theoretical considerations, the question of existence of a bound preserving limiter for any implicit time discretization scheme using DG is still open. However, numerically one can attempt to use the bound preserving limiter for an implicit time discretization scheme and expect improved numerical results.

4.2. Implications for Geodynamics Simulations

While we have demonstrated the advantages of the DG-BP method compared to the FEM, the question remains as to whether this is a preferred method for mantle convection simulations, and if so, under what conditions. Here we address those questions that we consider essential for the kinds of models described in the introduction.

Tracking Multiple Compositional Fields. The DG-BP method can be used to track multiple compositional fields in the same model without any modification from that presented here. When there are multiple compositional fields and there is no reaction between them, one can still assume the similar maximum and minimum bound for each individual compositional field and each field also satisfies the advection equation (3c) independently.

Long-term Stretching, Mixing and Folding. We have shown that the DG-BP method works well for a single sharp boundary, but we have not demonstrated how the method works when a layer, such as the crust, is mixed into the mantle over a billion year time-scale causing stretching, thinning, shearing, and folding of the initial crustal layer. Because this method is discontinuous, as features are stretched and break up in to smaller pieces, the boundaries between these fields will continue to be tracked and remain sharp. However, the accuracy of capturing the details of the mixing behavior is limited by the smallest mesh size, and therefore future benchmark tests are needed to determine how the computational cost (#DOFs) increases as compositional anomalies are mixed into finer scales. The ability to maintain sharp boundaries through time in a simulation is likely to impact conclusions related to entrainment of compositional signals from boundary layers into rising thermal plumes. The results

indicated that coupling the method to AMR will limit the computational cost as mixing proceeds, but eventually mixing can lead to compositional heterogeneity at all scales.

565 **Application to Diffusive Fields.** The DG-BP method may be applicable to advection-diffusive fields, such as temperature, if the energy equation can be rewritten to satisfy the scalar conservation laws. However, because the temperature field will have more smoothly varying gradients, the FEM is likely to be a better choice in terms of computational cost and efficiency.

570 **Use in both 2-D and 3-D Simulations.** The DG-BP method can be used in both 2-D and 3-D simulation. In addition, because it uses a field variable, in 3-D models it will require fewer degrees of freedom compared to tracers methods, providing an effective cost-efficient alternative.

5. Conclusions

575 Advection of compositional fields is a key component of mantle convection studies that requires tracking sharp boundaries between fields as they are subducted, mixed into the mantle and entrained in upwellings. However, as described in the introduction, existing methods are either computationally expensive, smooth over sharp boundaries or are not easily adapted to tracking
580 multiple compositional fields. The numerical tests demonstrate that the DG method with BP limiter is a stable, accurate and cost-effective approach to solving advection of a non-diffusive compositional field coupled to the Stokes equations. It does require that a higher order time discretization method is utilized, which requires holding 2–3 previous solution of the velocity field in
585 memory. However, when implemented with an AMR algorithm it uses fewer degrees of freedom than the FEM with a viscosity-entropy stabilization method. The DG method with BP limiter provides many of the capabilities needed for solid Earth geodynamics problems, which have not been found in any one of the previous approaches. Specifically, it can be used to track multiple compositional
590 fields through all the physical processes expected within the mantle (e.g.,

subduction, mixing, stretching, entrainment) in both 2-D and 3-D simulations. In 3-D simulations in particular, it will require fewer computational resources than tracer methods, and will not lead to smearing of boundaries that occurs as information from tracers are interpolated onto the mesh. As presented here
595 the BP limiter can not be used with the compressible form of the problem. Future work will explore other methods for stabilizing the DG method for the composition-advection equation with compressible flow, such as the weighted essentially non-oscillatory (WENO) scheme (Qiu and Shu, 2005; Zhu et al., 2008).

600 Acknowledgements

We would like to thank Jon Robey for discussions on level-sets and the time discretization approach, and John Naliboff for his input on the manuscript. This work was supported by the National Science Foundation (grant number 1440811).

605 References

- Bangerth, W., Heister, T., Heltai, L., Kanschä, G., Kronbichler, M., Maier, M., Turcksin, B., 2015. The `deal.II` library, version 8.3. preprint.
- Benzi, M., Golub, G. H., Liesen, J., 2005. Numerical solution of saddle point problems. *Acta numerica* 14 (1), 1–137.
- 610 Berger, M. J., Oliger, J., 1984. Adaptive mesh refinement for hyperbolic partial differential equations. *Journal of Computational Physics* 53 (3), 484–512.
- Christensen, U., 1982. Phase boundaries in finite amplitude mantle convection. *Geophysical Journal of the Royal Astronomical Society* 68, 487–497.
- Christensen, U. R., 1989. Mixing by time-dependent convection. *Earth and Plan-*
615 *etary Science Letters* 95, 382–394.

- Christensen, U. R., 1996. The influence of trench migration on slab penetration into the lower mantle. *Earth and Planet. Sci. Lett.* 140, 27–39.
- Cockburn, B., Hou, S., Shu, C.-W., 1990. The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. IV. The
620 multidimensional case. *Mathematics of Computation* 54 (190), 545–581.
- Cockburn, B., Shu, C.-W., 1989. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. II. General framework. *Mathematics of computation* 52 (186), 411–435.
- Cockburn, B., Shu, C.-W., 1998a. The local discontinuous Galerkin method
625 for time-dependent convection-diffusion systems. *SIAM Journal on Numerical Analysis* 35 (6), 2440–2463.
- Cockburn, B., Shu, C.-W., 1998b. The Runge–Kutta discontinuous Galerkin method for conservation laws V: Multidimensional systems. *Journal of Computational Physics* 141 (2), 199–224.
- 630 Davies, G. F., 1977. Whole-mantle convection and plate tectonics. *Geophysical Journal of the Royal Astronomical Society* 49 (459–486).
- Donea, J., Huerta, A., 2005. Steady transport problems, 33–78.
- Farnetani, D. G., Richards, M. A., 1995. Thermal entrainment and melting in mantle plumes. *Earth and Planetary Science Letters* 136, 251–267.
- 635 Fischer, K. M., Ford, H. A., Abt, D. L., Rychert, C. A., 2010. The lithosphere–asthenosphere boundary. *Annual Reviews of Earth and Planetary Science* 38, 551–575.
- Garnero, E. J., McNamara, A. K., 2008. Structure and dynamics of Earth’s lower mantle. *Science* 320, 626–628.
- 640 Gerya, T., 2010. Dynamical instability produces transform faults at mid-ocean ridges. *Science* 329 (1047).

- Gerya, T. V., Yuen, D. A., 2003a. Characteristics-based marker-in-cell method with conservative finite-differences schemes for modeling geological flows with strongly variable transport properties. *Physics of the Earth and Planetary Interiors* 140, 293–318.
- Gerya, T. V., Yuen, D. A., 2003b. Rayleigh-Taylor instabilities from hydration and melting propel 'cold plumes' at subduction zones. *Earth and Planetary Science Letters* 212.
- Gottlieb, S., 2005. On high order strong stability preserving Runge–Kutta and multi step time discretizations. *Journal of Scientific Computing* 25 (1), 105–128.
- Gottlieb, S., Shu, C.-W., Tadmor, E., 2001. Strong stability-preserving high-order time discretization methods. *SIAM review* 43 (1), 89–112.
- Guermond, J.-L., Pasquetti, R., Popov, B., 2011. Entropy viscosity method for nonlinear conservation laws. *Journal of Computational Physics* 230 (11), special issue High Order Methods for {CFD} Problems.
- Gurnis, M., Hager, B. H., 1988. Controls of the structure of subducted slabs. *Nature* 335 (22), 317–321.
- Hesthaven, J. S., Warburton, T., 2008. Nodal discontinuous Galerkin methods: algorithms, analysis, and applications. Vol. 54. Springer.
- Hoffman, N. R. A., McKenzie, D. P., 1985. The destruction of geochemical heterogeneities by differential fluid motions during mantle convection. *Geophysical Journal International* 82 (2), 163–206.
- Hofmann, A. W., 1997. Mantle geochemistry: the message from oceanic volcanism. *Nature* 385, 219–229.
- Karason, H., van der Hilst, R. D., 2000. Improving global tomography models of P-wavespeed I. incorporation of differential travel times for refracted and

- diffracted core phases (PKP, Pdiff). *Journal of Geophysical Research* 106, 6569–6587.
- 670 Kellogg, L. H., 1991. Interaction of plumes with compositional boundary at 670 km. *Geophysical Research Letters* 18 (5), 8650868.
- Kellogg, L. H., Hager, B. H., van der Hilst, R. D., 1999. Compositional stratification in the deep mantle. *Science* 283, 1881–1884.
- Kellogg, L. H., King, S. D., 1993. Effect of mantle plumes on the growth of
675 D” by reaction between the core and mantle. *Geophysical Research Letters* 29 (5), 379–382.
- Kelly, D., Gago, D. S., Zienkiewicz, O., Babuska, I., et al., 1983. A posteriori error analysis and adaptive processes in the finite element method: Part i—error analysis. *International journal for numerical methods in engineering*
680 19 (11), 1593–1619.
- Kronbichler, M., Heister, T., Bangerth, W., 2012. High accuracy mantle convection simulation through modern numerical methods. *Geophysical Journal International* 191 (1), 12–29, cited By 14.
- Lenardic, A., Kaula, W. M., 1993. A numerical treatment of geodynamic viscous flow problems involving the advection of material interfaces. *Journal of*
685 *Geophysical Research* 98 (B5), 8243–8260.
- Lin, S.-C., van Keken, P. E., 2005. Multiple volcanic episodes of flood basalts caused by thermochemical mantle plumes. *Nature* 436, 250–252.
- May, D. A., Moresi, L., 2008. Preconditioned iterative methods for Stokes flow
690 problems arising in computational geodynamics. *Physics of the Earth and Planetary Interiors* 171 (1), 33–47.
- McNamara, A. K., Zhong, S., 2004. Thermochemical structures within a spherical mantle: Superplumes or piles? *Journal of Geophysical Research* 109 (B07402).

- 695 McNamara, A. K., Zhong, S., 2005. Thermochemical structures beneath Africa and the Pacific Ocean. *Nature* 437, 1136–1139.
- Moresi, L.-N., Lenardic, A., 1997. Three-dimensional numerical simulations of crustal deformation and subcontinental mantle convection. *Earth and Planetary Science Letters* 150, 233–243.
- 700 O'Neill, C., Moresi, L., Müller, D., Albert, R., Dufour, F., 2006. Ellipsis 3D: A particle-in-cell finite-element hybrid code for modelling mantle convection and lithospheric deformation. *Computers & Geosciences* 32 (10), 1769–1779.
- Qiu, J., Shu, C.-W., 2005. Runge–kutta discontinuous galerkin method using weno limiters. *SIAM Journal on Scientific Computing* 26 (3), 907–929.
- 705 Reed, W., Hill, T., Oct 1973. Triangular mesh methods for the neutron transport equation.
- Samuel, H., Evonuk, M., 2010. Modeling advection in geophysical flows with particle level sets. *Geochemistry, Geophysics and Geosystems* 11 (8), Q08020.
- Suckale, J., Nave, J.-C., Hager, B. H., 2010. It takes three to tango: 1. Simulating buoyancy-driven flow in the presence of large viscosity contrasts. *Journal of Geophysical Research* 115 (B07409).
- 710 Tackley, P. J., June 2000. Mantle convection and plate tectonics: Toward an integrated physical and chemical theory. *Science* 288, 2002–2007.
- Tackley, P. J., 2012. Dynamics and evolution of the deep mantle resulting from thermal, chemical, phase and melting effects. *Earth-Science Reviews* 110, 1–25.
- 715 Tackley, P. J., King, S. D., 2003. Testing the tracer ratio method for modeling active compositional fields in mantle convection simulations. *Geochemistry, Geophysics and Geosystems* (G³) 4 (4), 8302.
- 720 Taylor, C., Hood, P., 1973. A numerical solution of the navier-stokes equations using the finite element technique. *Computers & Fluids* 1 (1), 73–100.

- van Keken, P. E., King, S. D., Schmeling, H., Christensen, U. R., Neumeister, D., Doin, M.-P., 1997. A comparison of methods for the modeling of thermochemical convection. *Journal of Geophysical Research* 102 (B10), 22477–22495.
- Verfürth, R., 1996. A review of a posteriori error estimation and adaptive mesh-refinement techniques. John Wiley & Sons Inc.
- Vynnytska, L., Rognes, M., Clark, S., 2013. Benchmarking fenics for mantle convection simulations. *Computers and Geosciences* 50, 95 – 105, benchmark problems, datasets and methodologies for the computational geosciences.
- URL <http://www.sciencedirect.com/science/article/pii/S0098300412001689>
- Woidt, W.-D., 1978. Finite element calculations applied to slat dome analysis. *Tectonophysics* 50, 369–386.
- Zhang, X., Shu, C.-W., Nov. 2010. On positivity-preserving high order discontinuous Galerkin schemes for compressible Euler equations on rectangular meshes. *J. Comput. Phys.* 229 (23), 8918–8934.
- Zhang, Y., Zhang, X., Shu, C.-W., Feb. 2013. Maximum-principle-satisfying second order discontinuous Galerkin schemes for convection-diffusion equations on triangular meshes. *J. Comput. Phys.* 234, 295–316.
- Zhao, C., Garnero, E. J., McNamara, A. K., Schmerr, N., Carlson, R. W., 2015. Seismic evidence for a chemically distinct thermochemical reservoir in Earth’s deep mantle beneath Hawaii. *Earth and Planetary Science Letters* 426, 143–153.
- Zhong, S. J., Yuen, D. A., Moresi, L. N., 2007. Numerical methods for mantle convection. In: *Treatise on Geophysics*. Elsevier, Ch. 7.05, pp. 227–252.
- Zhu, J., Qiu, J., Shu, C.-W., Dumbser, M., 2008. Runge–kutta discontinuous galerkin method using weno limiters ii: unstructured meshes. *Journal of Computational Physics* 227 (9), 4330–4353.