Geochemistry, Geophysics, Geosystems

RESEARCH ARTICLE
10.1029/2020GC009349

**Key Points:**

- Pointwise divergence free velocity fields are necessary for accurate advection of tracers in incompressible velocity fields
- Partial differential equation (PDE)-constrained $l_2$ projection exactly conserves mass of tracer data
- High order time integration methods are unnecessary for low order spatial geodynamics numerical models

**Supporting Information:**

- Supporting Information S1

**Correspondence to:**
N. Sime,
nsime@carnegiescience.edu

**Citation:**
Sime, N., Maljaars, J. M., Wilson, C. R., & van Keken, P. E. (2021). An exactly mass conserving and pointwise divergence free velocity method: Application to compositional buoyancy driven flow problems in geodynamics. *Geochemistry, Geophysics, Geosystems*, *22*, e2020GC009349. https://doi.org/10.1029/2020GC009349

Received 6 AUG 2020
Accepted 5 FEB 2021



# An Exactly Mass Conserving and Pointwise Divergence Free Velocity Method: Application to Compositional Buoyancy Driven Flow Problems in Geodynamics

**Nathan Sime[1], Jakob M. Maljaars[2,3], Cian R. Wilson[1], and Peter E. van Keken[1]**

[1]Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC, USA, [2]Environmental Fluid Mechanics, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands, [3]S[&]T Corporation, Delft, The Netherlands

**Abstract** Tracer methods are widespread in computational geodynamics for modeling the advection of chemical data. However, they present certain numerical challenges, especially when used over long periods of simulation time. We address two of these in this work: the necessity for mass conservation of chemical composition fields and the need for the velocity field to be pointwise divergence free to avoid gaps in tracer coverage. We do this by implementing the hybrid discontinuous Galerkin (HDG) finite element (FE) method combined with a mass conserving constrained projection of the tracer data. To demonstrate the efficacy of this system, we compare it to other common FE formulations of the Stokes system and projections of the chemical composition. We provide a reference of the numerical properties and error convergence rates which should be observed by using these various discretization schemes. This serves as a tool for verification of existing or new implementations. We summarize these data in a reproduction of a published Rayleigh-Taylor instability benchmark, demonstrating the importance of careful choices of appropriate and compatible discretization methods for all aspects of geodynamics simulations.

## 1. Introduction

Melting at mid-oceanic ridges causes differentiation of the Earth's mantle with the formation of a distinct basaltic crust on top of a depleted peridotite layer. The recycling of this oceanic crust in subduction zones, where additional differentiation may occur, causes the continuous formation of chemical heterogeneity in the Earth's mantle which adds to compositional heterogeneity that may have been formed during Earth's formation and magma ocean solidification (e.g., Labrosse et al., 2007).

Mantle convection models that incorporate these differentiation processes can predict how any formed heterogeneity is mixed back in or retained in the mantle (Kellogg, 1992; Tackley, 2015; van Keken et al., 2002, 2014) and help decipher the processes that lead to the long term chemical evolution of the Earth as seen in geochemistry (Christensen & Hofmann, 1994; R. E. Jones et al., 2019; Samuel & Farnetani, 2003; Tucker et al., 2020; Xie & Tackley, 2004), the formation of the structure of the Earth's mantle as imaged by seismological methods (Ballmer et al., 2016; Bull et al., 2009; Haugland et al., 2018; T. D. Jones et al., 2020), and the Earth's thermal state and evolution (Li & McNamara, 2018; Nakagawa & Tackley, 2005; Zhong, 2006).

Implementing the chemical buoyancy term into the Stokes equations that follow from the conservation of momentum tends to be more challenging than that of thermal buoyancy because diffusion of relevant chemical species (such as silicon and oxygen) is many orders of magnitude smaller than thermal diffusion and can be assumed zero. In geodynamical applications several distinct methods have been used to represent the chemical buoyancy. These include (i) methods that delineate the boundary between two volumes of distinct chemical composition such as the marker chain method (Christensen & Yuen, 1984; Lin & van Keken, 2006; Schmeling, 1987) and volume or moment of fluid methods (Pilliod & Puckett, 2004; Robey & Puckett, 2019; Zalesak, 1979); (ii) tracer methods where individual tracers carry a relative proportion of the chemical buoyancy (Brandenburg et al., 2008; Christensen & Hofmann, 1994; Gerya & Yuen, 2003; O'Neill et al., 2006; Tackley & King, 2003); and (iii) representing the chemical density as a composition field and solving the advection-diffusion equation with low chemical diffusivity (e.g., Hansen & Yuen, 2000; Kellogg & King, 1993), which in the presence of thermal buoyancy leads to double-diffusive convection

(Turner, 1974). A comprehensive, if now slightly dated, comparison of these three methods is provided in van Keken et al. (1997).

Further complications occur if the compositionally distinct materials have different rheological behavior. In this case, the continuity in stress across the interface causes a discontinuity in strain rate which is not a trivial problem to solve. See Suckale et al. (2010) for an elegant solution using a combination of ghost fluids and the level-set method.

In the past decade or so, there has been significant progress improving methods for the accurate solution of thermochemical convection which are based on, among others, application of discontinuous Galerkin (DG) finite element (FE) methods (He et al., 2017), further use of level-set methods (Hillebrand et al., 2014; Samuel & Evonuk, 2010), and adaptive remeshing (Davies et al., 2007; Leng & Zhong, 2011). Modeling of thermochemical convection has also become more readily available to a larger group of researchers through community codes such as Advanced Solver for Problems in Earth's ConvecTion (as demonstrated by Gassmöller et al., 2018; and Robey & Puckett, 2019). The reliability of new methods has been demonstrated by careful benchmarking and error testing (e.g., Thielmann et al., 2014; Vynnytska et al., 2013).

In this study, we will focus on tracer methods. Their popularity arises primarily from their robustness when advecting sharp jumps in material properties in the absence of diffusion. In contrast, field-based methods under these circumstances are prone to instability arising as numerical oscillations or suffer from significant overdiffusion. The tracer methods are also more practical than the marker chain method for tracing chemical buoyancy over long time scales as typical mantle mixing leads to exponential growth of length of the marker chain (van Keken et al., 1997). Additionally, the implementation of basic tracer advection methods is straightforward in both two-dimensional and three-dimensional. Each position and chemical datum pair in a population of tracers is simply advected through a velocity field using discrete time integration.

Tracer methods are not without their issues however. Pointwise chemical data associated with the tracers must be transferred, or projected, onto a mesh-based composition field before they can be used to set material properties such as buoyancy or viscosity. Various methods have been proposed for this projection, many of which depend on treating the tracers as delta-function sources within the domain. In the Dirac delta source, or “Stokeslet” method (e.g., Christensen & Hofmann, 1994), direct integration of these delta functions with an associated mass per particle is used to derive the buoyancy source on the right hand side of the Stokes system. The so-called tracer ratio and absolute methods (Tackley & King, 2003) use integrals of the delta functions to count the number of particles of a particular type per cell. Weighting these counts by either a volume fraction per particle for the absolute method or by the total number of particles per cell for the ratio method results in a mesh-based composition field.

The properties of the projected composition field vary widely depending on the algorithm. Ideally, the field should be conserved both globally across the domain and locally within each cell of the mesh. Additionally, the error in the field representation should converge at an optimal rate as the number of cells in the mesh increases, provided some minimum number of particles per cell is maintained. The absolute and Dirac delta methods achieve global but not local conservation. The tracer ratio method is neither locally nor globally conservative (Trim et al., 2020) but it does have the advantage that the composition field remains bounded between set limits (Tackley & King, 2003). In general, however, the integration of Dirac delta functions does not satisfy the smoothness, or regularity, requirements for finite elements. This causes algorithms that depend on it to typically experience reduced error convergence rates even when using nonconforming methods (e.g., Houston & Wihler, 2012; Scott, 1973). When the composition field is used to drive or influence the flow, for example through the buoyancy and viscosity terms, respectively, this may in turn lead to suboptimal convergence of the velocity approximation.

During advection tracers may drift apart (e.g., Trim et al., 2020; Wang et al., 2015), leaving gaps in coverage that may cause projection algorithms to break down when cells lack particles. This often requires the introduction (and removal) of tracers (e.g., Moresi et al., 2003), adding to the complexity of the algorithm. In incompressible flows, however, the root cause of this problem is that the discrete velocity solution is generally only divergence free in some discrete sense and not at every point in the domain. Thus over long simulations two particles may diverge even if they are initially in close proximity to each other.

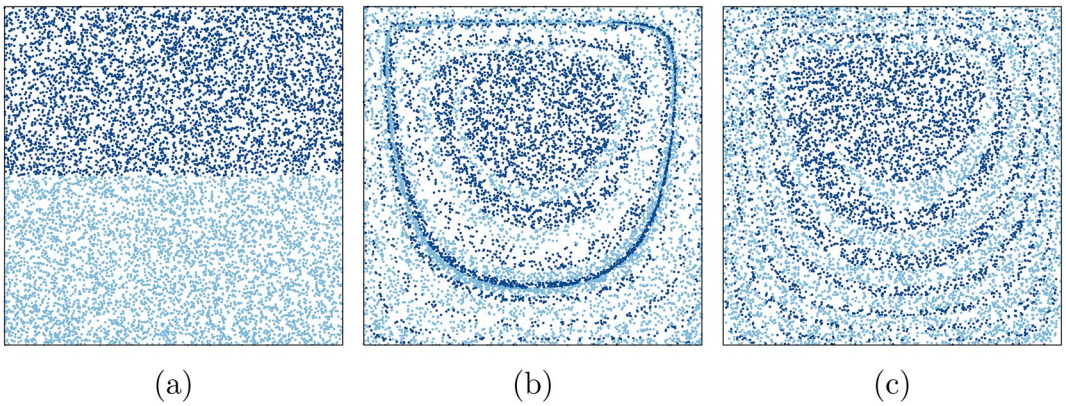


**Figure 1.** Example of tracers advected using a fourth order Runge-Kutta method in a forced cavity model. The geometry is a unit square and flow is forced by uniform horizontal velocity along the top boundary with free-slip boundaries elsewhere. The coloring of the tracers is provided solely as a visual aid and plays no other role in the simulation. (a) Initial configuration of the tracers. (b) Tracer distribution after 50 overturn times using a traditional Taylor-Hood finite element method. (c) as (b) but now with the new hybrid discontinuous Galerkin finite element method that guarantees the velocity field representation is pointwise divergence free. The mean number of particles per cell is 50 and the standard deviations from the mean in each case are: (a) 7.23, (b) 23.1, and (c) 8.01.

Various velocity field reconstruction techniques have been proposed that maintain uniform particle distributions in the incompressible limit (Jenny et al., 2001; McDermott & Pope, 2008; Wang et al., 2015). Importantly, Maljaars et al. (2018) demonstrated that particle drift is mitigated when the incompressibility constraint is accurately satisfied. This is demonstrated in Figure 1 where initially evenly distributed particles (Figure 1a) develop significant gaps as the driven lid cavity simulation progresses when the Stokes equations are discretized using the commonly used Taylor-Hood (TH) FE pair (Figure 1b). The use of a new hybrid discontinuous Galerkin (HDG) discretization that we will discuss further below guarantees that the incompressibility constraint is satisfied at every point in the domain. This maintains a consistent, even distribution of particles throughout the domain without the formation of gaps like those that occur with the TH discretization (Figure 1c).

The simple comparison in Figure 1 shows that, while stand-alone tracer advection algorithms are straightforward to implement, coupling them to a velocity solution may lead to significant artifacts. Here, we demonstrate a holistic approach to chemical advection by tracers in geodynamic models using finite elements on two-dimensional triangular meshes. This approach is straightforward to extend to three-dimensional tetrahedral meshes. Extension to elements with non-affine local-to-global mapping (e.g., general quadrilaterals/hexahedra and high order polyhedral cells) is not covered here as it would require different FE basis functions, chosen such that the divergence of velocity lies in the pressure space (see Section 5.3).

Our new approach requires us to reconsider the discretization of the entire system of equations. In the remainder of this study, we will demonstrate each stage individually: the advection of the tracers; the projection of their data to a chemical composition field in a conservative manner; and the appropriate spatial approximation of the velocity field such that it inherently satisfies the incompressibility constraint of the Stokes system of equations, overcoming spurious particle drift.

Advection of the composition field using tracers requires the solution of an ordinary differential equation (ODE) for each tracer in the domain. These ordinary differential equations (ODEs) may be discretized and numerically integrated using any appropriate method. In this work we will use (but are not limited to) flexible arbitrary order Runge-Kutta (RK) methods. We provide a brief overview of these methods in Section 3.

Before solving the Stokes system we must project the tracer data to a mesh-based composition field. In Section 4, we focus on the least squares projection, which preserves optimal error convergence during mesh refinement (Maljaars et al., 2018; Thielmann et al., 2014). However, it does not satisfy mass conservation so we exploit the partial differential equation (PDE)-constrained least squares projection (Maljaars et al., 2019,

**Table 1**
*Summary of Tracer Data Projection Methods. See Also Sections 1 and 4*

| This work | |
|---|---|
| $l_2$ projection | Minimizes square distance between tracer data ($X(t)$, $\Phi(t)$) and the composition field $\phi_h(\mathbf{x},t)$. Does not guarantee conservation. For smooth functions error converges at the optimal rate $\mathcal{O}(h^{k+1})$. When projecting into a DG FE space requires only local solves on each element |
| PDE-constrained $l_2$ projection | $l_2$ projection augmented by the PDE-constraint of the advection equation. HDG discretization of the variational problem exactly conserves the composition function. Requires a global solve over the domain |
| Not implemented here | |
| Dirac delta "Stokeslet" source | Direct application of point source data to the right hand side of the Stokes system. Does not satisfy regularity requirements of the FE scheme. Will reduce convergence rate of the error of the velocity FE approximation to $\mathcal{O}(h)$ (Scott, 1973; Houston & Wihler, 2012). Does not provide a field representation $\phi_h(\mathbf{x}, t)$ required by the viscosity model |
| Dirac delta projection | $L_2$ projection of tracer data to an FE space. Same issue applies in the projection of Dirac delta functions to a composition field whose FE approximation error measured in the $L_2$ norm will be suboptimal at a rate $\mathcal{O}(h)$. Requires a global solve over the domain if projecting into a continuous FE space |

2020), which allows us to enforce both global and local conservation properties on the composition field (Section 4.2). We provide a brief summary of tracer projection methods in Table 1.

A key benefit of this tracer projection approach is its treatment of tracer data as a pointwise approximation of a continuous field. Contrary to the "Stokeslet" method composed of Dirac delta sources, this approach allows us to both add and remove tracers mid-simulation without breaking local or global mass conservation when performing the PDE-constrained projection. This is not demonstrated here as the carefully selected velocity discretization, resulting lack of particle drift, and near uniform mesh resolution mean that all simulations are robust without the addition or removal of particles.

Numerous methods exist for the discretization of the Stokes system but we focus in this work on the FE method due to its ability to manage complex geometries, viscosity models, and tools for error analysis (Brenner & Scott, 2010). We summarize some of the more frequently used FE methods used in geodynamics in Table 2 but focus here on methods that produce pointwise divergence free velocity fields. For tracer advection in incompressible flows this has the essential property of preventing tracer "drift" as demonstrated in Figure 1c. Several methods exist with this property (e.g., Cockburn et al., 2007; Evans & Hughes, 2013; Guzmán & Neilan, 2014; Morgan & Scott, 2018; Scott & Vogelius, 1985). In this work, we exploit the new HDG method (Cockburn et al., 2010; Labeur & Wells, 2012; Rhebergen & Wells, 2018a) and compare its properties to the commonly used TH scheme.

After introducing the methods, we test them in Section 7 using a series of computational geophysics benchmarks from the literature. We examine the impact of each specific discretization method on the accuracy of the results with the goal of demonstrating how a holistic approach is necessary to maintain accuracy and consistency in a fully coupled tracer-based model.

Throughout this work, we focus on the equations underlying geophysical mantle convection models. However, we refer to Maljaars et al. (2018) regarding application to problems including inertia via the Navier-Stokes system and Maljaars et al. (2021) regarding the applicability of the proposed methods to various scientific models. This flexibility allows the use of the HDG method in a broad set of geophysical applications including those arising in oceanography and studies of core dynamics. While we demonstrate the applicability of this method to incompressible flows in this study, the method can also be applied to weakly compressible flows under the anelastic liquid approximation (e.g., Jarvis & Mckenzie, 1980). We plan to demonstrate this in a future submission.

## 2. Problem Definition and Geometry

Let $\Omega$ be the computational domain of spatial dimension $d = 2$ which accurately represents the problem geometry, for example a mantle convection cell. Furthermore, let $\partial\Omega$ be the boundary of this domain with outward pointing unit normal vector $\mathbf{n}$. A depiction of this discretization is shown in Figure 2. Given an

**Table 2**
*Summary of FE Discretization Schemes of the Stokes System Used in This work. See Also Section 5 and Section S4. We Refer to Boffi et al. (2013) and the References Cited Therein for More Information on FE pairs*

| | |
|---|---|
| Demonstrated in main text | |
| TH | Velocity and pressure FE approximations converge at the optimal rates $\mathcal{O}(h^{\mathsf{p}+1})$ and $\mathcal{O}(h^{\mathsf{p}})$ as measured in the $L_2$ norm. Globally satisfies the integral divergence free constraint $\int_\Omega \nabla\cdot\mathbf{u}_h d\mathbf{x} = 0$, however $\mathbf{u}_h$ is not pointwise divergence free |
| HDG | Scheme coupling FE functions defined on both the cells and the facets. Velocity and pressure $L_2$ error converges at optimal rates. Yields a pointwise divergence free velocity approximation $\nabla\cdot\mathbf{u}(\mathbf{x}) = 0 \forall \mathbf{x} \in \Omega$. Can be formulated such that local components of the problem are eliminated before a global solve by static condensation |
| Demonstrated in supporting information | |
| DG | Nonconforming FE scheme permitting discontinuities in the solution. Velocity and pressure $L_2$ error converges at optimal rates. Enforces the divergence free constraint such that $\int_\Omega \nabla\cdot\mathbf{u}_h d\mathbf{x} = 0$ and $\int_\kappa \nabla\cdot\mathbf{u}_h d\mathbf{x} = 0$, $\forall \kappa \in \mathcal{T}$. The resulting discontinuous velocity field $\mathbf{u}_h$ is incompatible with the PDE-constrained $l_2$ projection method. The degree of freedom count is typically much larger than conforming methods |
| MINI | Velocity FE space enriched by the cubic vector bubble element. Allows equal order approximation of velocity and pressure ($p \geq 1$). Enforces the divergence free constraint in the global sense $\int_\Omega \nabla\cdot\mathbf{u}_h d\mathbf{x} = 0$. |
| P2BDG1 | In some settings it is beneficial to have a discontinuous pressure approximation. This stable element pair combines the enriched velocity FE space of degree $p \geq 2$ of the MINI element with the $p-1$ degree FE DG space for the pressure. This element pair satisfies the divergence free constraint only in the global sense $\int_\Omega \nabla\cdot\mathbf{u}_h d\mathbf{x} = 0$ |
| P2DG0 | Also permits a discontinuous pressure approximation by combining the standard $p \geq 2$ vector FE space for the velocity and $p-2$ DG FE space for the pressure. Although this yields a decrease in the number of degrees of freedom, the convergence rate of the velocity solution as measured in the $L_2$ norm is suboptimal $\mathcal{O}(h^{\mathsf{p}})$ |

initial time $t = 0$ and final time $t_F$, we define the time interval $t \in \mathcal{I}_t \coloneqq (0, t_F]$. Subject to appropriate boundary and initial conditions we seek the velocity $\mathbf{u}$, pressure $p$, and chemical composition $\phi$ of the coupled problem in $\Omega \times \mathcal{I}_t$

$$-\nabla\cdot\left(2\eta\underline{\underline{\varepsilon}}(\mathbf{u})\right) + \nabla p = \mathbf{f}, \tag{1}$$

$$\nabla\cdot\mathbf{u} = 0, \tag{2}$$

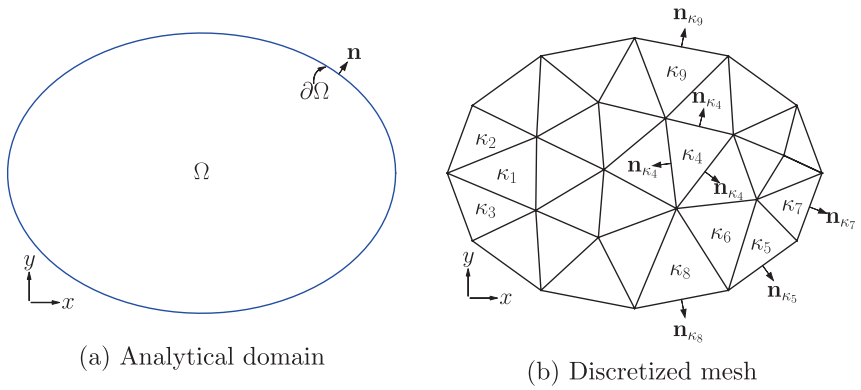


**Figure 2.** (a): an abstract domain $\Omega$ with boundary $\partial\Omega$ and outward pointing unit normal vector $\mathbf{n}$. (b): the discretization of $\Omega$ into a mesh comprised of tessellating triangles $\mathcal{T} = \{\kappa\}$. The first nine triangles are labeled. A subset of the outward pointing normals of the facets are also shown.

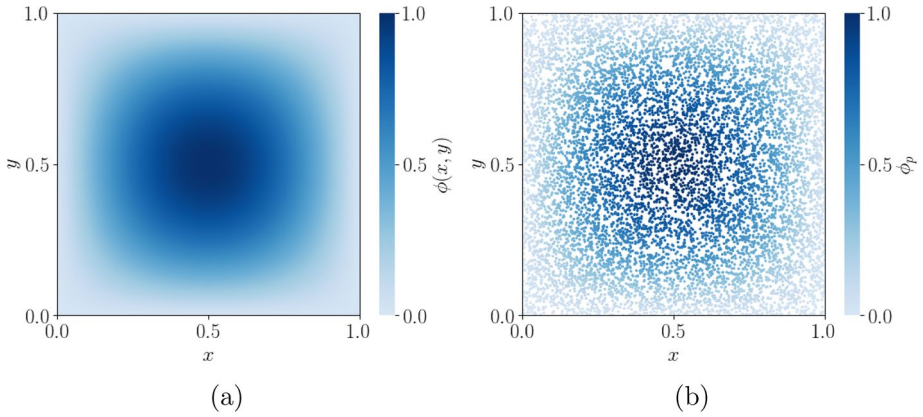


**Figure 3.** Discretization of a composition field by tracer data. (a): example analytical composition field $\phi(\mathbf{x}, t)$. (b): discretization of $\phi(\mathbf{x}, t)$ by pointwise interpolation to tracer data $\Phi(t) = \{\phi(\mathbf{x}_p, t)\}_{p=1}^{N_p}$ at positions $X(t)$.

$$\frac{\partial \phi}{\partial t} + \nabla \cdot (\mathbf{u}\phi) \quad = 0. \tag{3}$$

here, $\eta$ is the viscosity, $\underline{\underline{\varepsilon}}(\mathbf{u}) = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^{\top})$ is the rate of strain tensor and $\mathbf{f}$ is the momentum source. In this work we are primarily interested in the case where $\eta = \eta(\phi)$ and $\mathbf{f} = \mathbf{f}(\phi)$, however this is not a limitation. The definition of the viscosity and momentum source is prescribed by the underlying physical model.

To discretize the domain, we subdivide $\Omega$ into non-overlapping triangles which tessellate the geometry generating a conforming mesh. We write $\kappa$ to denote a triangle in the mesh $\mathcal{T}$ such that $\mathcal{T} := \{\kappa\}$. We refer to each $\kappa \in \mathcal{T}$ as an element or cell in the mesh which has measured diameter $h_\kappa$, the maximum distance between two points in the cell. We also employ the notation $h$ to indicate that a quantity has been discretized conforming to the underlying mesh. Furthermore, each cell $\kappa$ has a boundary $\partial\kappa$ with corresponding outward pointing unit normal vector $\mathbf{n}_\kappa$ (see Figure 2b).

## 3. Discretization of a Field by Tracer Data

Consider the advection equation stated in Equation 3. Instead of a field, $\phi(\mathbf{x}, t)$, we use a tracer method to make discrete pointwise approximations. In this setting, we generate $N_p$ tracers in the domain $\Omega$ such that each tracer has position $\mathbf{x}_p(t) \in \Omega$, $p = 1, ..., N_p$. We may therefore define the set of tracer positions $X$ inside the domain

$$X(t) := \left\{\mathbf{x}_p(t)\right\}_{p=1}^{N_p}. \tag{4}$$

Each tracer may carry arbitrary numerical data. For example, we assign the $p$th tracer's datum regarding the composition field value $\phi_p(t)$, $p = 1, ..., N_p$. So, we now define the composition data associated with the tracers

$$\Phi(t) := \left\{\phi_p(t)\right\}_{p=1}^{N_p}. \tag{5}$$

An example of the discretization of the field $\phi(\mathbf{x}, t)$ by the point coordinates $X(t)$ and the interpolated data $\Phi(t) = \{\phi_p(t) = \phi(\mathbf{x}_p(t), t)\}_{p=1}^{N_p}$ is shown in Figure 3.

It is important to note that in our field discretization implementation the tracer data have no associated physical mass or volume. It is the composition field itself that carries these physical data. Therefore, by

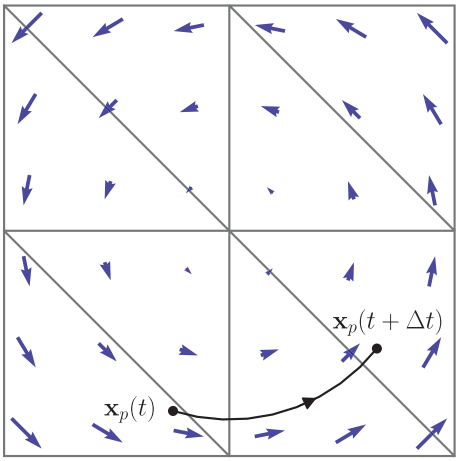


**Figure 4.** Example tracer advection through a velocity field. The blue arrows show a velocity field $\mathbf{u}_h(\mathbf{x})$ defined on an FE mesh. A tracer with position datum $\mathbf{x}_p(t)$ is advected to its new position at time step $\mathbf{x}_p(t+\Delta t)$.

this method we may add or remove tracers as necessary with no concern for breaking mass or volume conservation. However, conservation of the field itself requires consideration which we address later in Section 4.

### 3.1. Time Integration

The total derivatives of the tracer data subject to the advection equation read

$$\frac{d\phi_p}{dt} = 0, \tag{6}$$

$$\frac{d\mathbf{x}_p}{dt} = \mathbf{u}(\mathbf{x}_p,t), \quad p = 1,\ldots,N_p. \tag{7}$$

clearly, once the initial composition data $\Phi(t=0)$ have been prescribed, Equation 6 demands their values never change for $t>0$. However, the approximation of the evolution of the tracers' positions should be computed from Equation 7 by numerically integrating the $N_p$ independent ODEs. See Figure 4 for a depiction of numerical advection of a tracer through a velocity field.

The time derivative in Equation 7 must be integrated in order to compute the evolution of the tracer distribution. Precise time integration of the tracers is important to ensure a good approximation of the evolution of the composition field. In this work we will consider our implementation of the explicit RK time integration method. We write RK$\ell$, with $\ell \in \mathbb{N}$ indicating the order of the method we use. For detailed analysis of RK methods we refer to Butcher (1987).

Given an initial time $t_0 = 0$ and final time $t_F$, we discretize the simulation time interval $\mathcal{I}_t = [t_0,t_F]$ into the points

$$\mathcal{I}_{\Delta t} = \{t_0,t_1,\ldots,t_F\}, \quad t_0 < t_1 < \ldots < t_F, \tag{8}$$

where $\Delta t_n = t_{n+1} - t_n$. The way in which $\Delta t_n$ is chosen is a key factor in stability and accuracy of the RK method. Here, we make our choice based on a form of the Courant-Friedrichs-Lewy (CFL) criterion

$$\Delta t_n = C_{\mathrm{CFL}} \frac{h_{\min}}{\max_{x\in\Omega} \mathbf{u}_h(\mathbf{x},t_n)}, \tag{9}$$

where $C_{\mathrm{CFL}}$ is the constant CFL parameter and $h_{\min}$ is the smallest diameter measure of the mesh's cells. We refer to LeVeque (2007) for a comprehensive analysis of time integration methods.

### 3.2. Explicit Runge-Kutta Method Summary

The general form of the RK$\ell$ method is as follows: find $\mathbf{x}_{p,n+1}$, $p = 1, ..., N_p$ such that

$$\mathbf{x}_{p,n+1} = \mathbf{x}_{p,n} + \Delta t_n \sum_{i=1}^{\ell} b_i \mathbf{k}_{p,i} \tag{10}$$

where $\ell \geq 1$ and

$$\mathbf{k}_{p,i} = \mathbf{u}\left(\mathbf{x}_{p,n} + \Delta t_n \sum_{j=1}^{\ell} a_{ij}\mathbf{k}_{p,j}, t_n + \Delta t_n c_i\right). \tag{11}$$

**Table 3**
*The RK$\ell$ Methods Used in This Work*

| Method | Butcher tableau | Iteration |
|---|---|---|
| RK1 | $\begin{array}{cc} 0 & 0 \\ & 1 \end{array}$ | $\mathbf{k}_{p,1} = \mathbf{u}(\mathbf{x}_{p,n}, t_n)$<br>$\mathbf{x}_{p,n+1} = \mathbf{x}_{p,n} + \Delta t_n \mathbf{k}_{p,1}$ |
| RK2 | $\begin{array}{ccc} 0 & 0 & \\ 1 & 1 & 0 \\ & \frac{1}{2} & \frac{1}{2} \end{array}$ | $\mathbf{k}_{p,1} = \mathbf{u}(\mathbf{x}_{p,n}, t_n)$<br>$\mathbf{k}_{p,2} = \mathbf{u}(\mathbf{x}_{p,n} + \Delta t_n \mathbf{k}_{p,1}, t_n + \Delta t_n)$<br>$\mathbf{x}_{p,n+1} = \mathbf{x}_{p,n} + \frac{\Delta t_n}{2}\left(\mathbf{k}_{p,1} + \mathbf{k}_{p,2}\right)$ |
| RK3 | $\begin{array}{cccc} 0 & 0 & & \\ \frac{1}{2} & \frac{1}{2} & 0 & \\ \frac{3}{4} & 0 & \frac{3}{4} & 0 \\ & \frac{2}{9} & \frac{1}{3} & \frac{4}{9} \end{array}$ | $\mathbf{k}_{p,1} = \mathbf{u}(\mathbf{x}_{p,n}, t_n)$<br>$\mathbf{k}_{p,2} = \mathbf{u}(\mathbf{x}_{p,n} + \frac{1}{2}\Delta t_n \mathbf{k}_{p,1}, t_n + \frac{1}{2}\Delta t_n)$<br>$\mathbf{k}_{p,3} = \mathbf{u}(\mathbf{x}_{p,n} + \frac{3}{4}\Delta t_n \mathbf{k}_{p,2}, t_n + \frac{3}{4}\Delta t_n)$<br>$\mathbf{x}_{p,n+1} = \mathbf{x}_{p,n} + \Delta t_n\left(\frac{2}{9}\mathbf{k}_{p,1} + \frac{1}{3}\mathbf{k}_{p,2} + \frac{4}{9}\mathbf{k}_{p,3}\right)$ |
| RK4 | $\begin{array}{ccccc} 0 & 0 & & & \\ \frac{1}{2} & \frac{1}{2} & 0 & & \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & \\ 1 & 0 & 0 & 1 & 0 \\ & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array}$ | $\mathbf{k}_{p,1} = \mathbf{u}(\mathbf{x}_{p,n}, t_n)$<br>$\mathbf{k}_{p,2} = \mathbf{u}(\mathbf{x}_{p,n} + \frac{1}{2}\Delta t_n \mathbf{k}_{p,1}, t_n + \frac{1}{2}\Delta t_n)$<br>$\mathbf{k}_{p,3} = \mathbf{u}(\mathbf{x}_{p,n} + \frac{1}{2}\Delta t_n \mathbf{k}_{p,2}, t_n + \frac{1}{2}\Delta t_n)$<br>$\mathbf{k}_{p,4} = \mathbf{u}(\mathbf{x}_{p,n} + \Delta t_n \mathbf{k}_{p,3}, t_n + \Delta t_n)$<br>$\mathbf{x}_{p,n+1} = \mathbf{x}_{p,n} + \Delta t_n\left(\frac{1}{6}\left(\mathbf{k}_{p,1} + \mathbf{k}_{p,4}\right) + \frac{1}{3}\left(\mathbf{k}_{p,2} + \mathbf{k}_{p,3}\right)\right)$ |

The quantities $a_{ij}$, $b_i$, and $c_i$ are drawn from the so-called Butcher tableau (Butcher, 1987)

$$\begin{array}{ccccc} c_1 & a_{11} & & & \\ \vdots & \vdots & \ddots & & \\ c_\ell & a_{\ell 1} & \cdots & a_{\ell\ell} \\ & b_1 & \cdots & b_\ell \end{array} \tag{12}$$

Note that to achieve optimal convergence the RK$\ell$ method requires $\ell - 1$ additional computations of the Stokes FE solution, which may significantly increase the expense of higher order RK methods. We do not consider temporal interpolation of the velocity field to reduce this cost as it would introduce an additional source of numerical error and adversely affect the convergence rates we present. We note, however, that any higher order interpolant that aimed to balance expense and accuracy would also have to ensure that the divergence free properties of the velocity field are preserved. The RK$\ell$ methods implemented for this work are shown in Table 3.

## 4. Projection of Tracer Data to a Field

Let $\phi_h(\mathbf{x}, t)$ be the approximation of the true composition field $\phi(\mathbf{x}, t)$. With the tracer positions $X(t)$ and composition data $\Phi(t)$, we must define methods by which these data may be represented by $\phi_h(\mathbf{x}, t)$. We propose two methods here: the straightforward least squares ($l_2$) projection and the PDE-constrained $l_2$ projection. We summarize properties of these and other commonly used methods in Table 1.

### 4.1. Least Squares Projection

First we choose a finite dimensional space of functions $W^{h,k}$ which is composed of piecewise polynomials of degree $k$ into which the tracer data will be projected. We will specify choices for this space after we have

constructed the least squares projection operation. This $l_2$ projection minimizes the difference between the tracer data and its field projection in a least-squares sense. Given tracer positions $X(t)$ and composition data $\Phi(t)$, the $l_2$ projection operator corresponds to the following minimization problem: find $\phi_h(\mathbf{x}, t) \in W^{h,\mathsf{k}}$ which satisfies

$$\min_{\phi_h \in W^{h,\mathsf{k}}} \mathcal{J}(\phi_h) \coloneqq \sum_{p}^{N_p} \frac{1}{2}\Big(\phi_h(\mathbf{x}_p(t),t) - \phi_p(t)\Big)^2. \tag{13}$$

This corresponds to solving the following linear system: find $\phi_h(\mathbf{x}, t) \in W^{h,\mathsf{k}}$ such that

$$\sum_{p=1}^{N_p} \Big(\phi_h(\mathbf{x}_p(t),t) - \phi_p(t)\Big) w_h(\mathbf{x}_p(t)) = 0 \tag{14}$$

for all $w_h \in W^{h,\mathsf{k}}$.

We choose $W^{h,\mathsf{k}}$ to be the nonconforming FE space of discontinuous piecewise polynomials, that is the standard DG FE space. We therefore write $W^{h,\mathsf{k}}_{\mathrm{DG}}$ to emphasize the space's characteristic and define it as follows

$$\begin{aligned} W^{h,\mathsf{k}}_{\mathrm{DG}} \coloneqq \{v : v \text{ is square integrable in } \Omega \text{ and is composed} \\ \text{of piecewise polynomials of degree } \mathsf{k} \geq 0 \text{ defined on each } \kappa \in \mathcal{T}\}. \end{aligned} \tag{15}$$

This choice offers a number of benefits. We may better approximate the discontinuity between composition phases compared with standard continuous spaces. Furthermore the linear system (14) may be written as a summation of local (i.e., cellwise) equations, that can be solved independently for each cell $\kappa \in \mathcal{T}$, that is: find $\phi_h(\mathbf{x},t) \in W^{h,\mathsf{k}}_{\mathrm{DG}}$ such that

$$\sum_{\kappa \in \mathcal{T}} \sum_{p_\kappa=1}^{N_{p_\kappa}} \Big(\phi_h(\mathbf{x}_{p_\kappa}(t),t) - \phi_{p_\kappa}(t)\Big) w_h(\mathbf{x}_{p_\kappa}(t)) = 0 \tag{16}$$

for all $w_h \in W^{h,\mathsf{k}}_{\mathrm{DG}}$. Here, $N_{p_\kappa}$ is the number of tracers inside cell $\kappa$ and $p_\kappa = 1,\ldots,N_{p_\kappa}$ are the indices of the tracers which reside in the cell $\kappa$ exclusively. This local solution method is typically more computationally efficient than solving the global system that would be generated by a standard continuous FE space.

In some geodynamical models, we must constrain the composition function to reside within bounded limits. For example, $\phi_{\min} \leq \phi_h(\mathbf{x}, t) \leq \phi_{\max}\ \forall \mathbf{x} \in \Omega$ where $\phi_{\min}$ and $\phi_{\max}$ are minimum and maximum values permitted for $\phi_h$, respectively. This is pertinent in the case $k \geq 1$, considering numerical oscillations will yield overshoots and undershoots in the projected composition field. See Goldfarb and Idnani (1983) for a background on quadratic programming for such problems. In our implementation we make use of Quad-Prog++ to perform the box-constrained numerical optimization (Gaspero, 2020).

The $l_2$ projection scheme is not conservative. Hence, the composition field will incur a mass conservation error over time. To solve this issue we next exploit a PDE-constrained $l_2$ projection, recently introduced in Maljaars et al. (2019).

### 4.2. PDE-Constrained Projection

Consider again the minimization problem Equation 13. We augment this problem with the constraint of the advection Equation 3, so that the full problem reads: find $\phi_h(\mathbf{x},t) \in W^{h,\mathsf{k}}_{\mathrm{DG}}$ which satisfies

$$\min_{\phi_h \in W^{h,\mathsf{k}}_{\mathrm{DG}}} \mathcal{J}(\phi_h) \coloneqq \sum_{p}^{N_p} \frac{1}{2}\Big(\phi_h(\mathbf{x}_p(t),t) - \phi_p(t)\Big)^2 \tag{17a}$$

$$\text{subject to:} \quad \begin{aligned} \frac{\partial \phi_h}{\partial t} + \nabla \cdot (\mathbf{u}_h \phi_h) &= 0 \quad && \text{in } \Omega, \\ \phi_h &= \phi_{\text{in}} && \text{on } \partial\Omega_{\text{in}}, \end{aligned} \tag{17b}$$

where $\partial\Omega_{\text{in}} = \{\mathbf{x} \in \partial\Omega: \mathbf{u}_h(\mathbf{x})\cdot\mathbf{n}(\mathbf{x}) < 0\}$ is the region of the exterior boundary where the velocity field flows into the domain $\Omega$ and $\phi_{\text{in}}$ is the prescribed incoming composition data on the inflow boundary. We also write $\partial\Omega_{\text{out}} = \partial\Omega \backslash \partial\Omega_{\text{in}}$ which is the outflow portion of the exterior boundary.

The implication is that the minimization problem now carries information regarding conservation of $\phi_h(\mathbf{x}, t)$ as the simulation evolves. In fact, it can be shown that the HDG FE discretization of the system Equation 17 exactly conserves $\phi_h$ globally and locally (to machine precision). We exploit this exact mass conservation property in this work and refer to Maljaars et al. (2019) for the details of this analysis.

For now, we will state the FE problem which solves the system Equation 17. To formulate the HDG problem we require three FE spaces $W_{\text{DG}}^{h,\text{k}}$, $T^h$ and $\bar{W}_{\text{DG}}^{h,\text{k}}$. Here, $W_{\text{DG}}^{h,\text{k}}$ is the standard DG space of degree $k$, $T^h := W_{\text{DG}}^{h,0}$ is also a standard DG space of degree 0 (i.e., piecewise constants per cell[1] (This choice is made as no discernible advantage is observed when employing higher order spaces (Maljaars et al., 2019)) and $\bar{W}_{\text{DG}}^{h,\text{k}}$ is composed of piecewise polynomials of degree $k$ defined on the facets of the mesh and discontinuous between facets. By multiplying Equation 17b with a Lagrange multiplier $\lambda_h \in T^h$ and applying integration by parts, which introduces an unknown flux $\bar{\phi}_h \in \bar{W}_{\text{DG}}^{h,\text{k}}$ on interior facets, the optimization problem amounts to finding the stationary points of the Lagrangian functional:

$$\begin{aligned} \mathcal{L}(\phi_h, \bar{\phi}_h, \lambda_h) = & \sum_p^{N_p} \frac{1}{2}\Big(\phi_h(\mathbf{x}_p(t), t) - \phi_p(t)\Big)^2 \\ & + \int_\Omega \frac{\partial \phi_h}{\partial t}\lambda_h d\mathbf{x} - \sum_{\kappa\in\mathcal{T}} \int_\kappa \mathbf{u}_h \phi_h \cdot \nabla\lambda_h d\mathbf{x} \\ & + \sum_{\kappa\in\mathcal{T}} \int_{\partial\kappa\backslash\partial\Omega} \mathbf{u}_h \cdot \mathbf{n}_\kappa \bar{\phi}_h \lambda_h d\mathbf{s} + \int_{\partial\Omega_{\text{out}}} \mathbf{u}_h \cdot \mathbf{n}\phi_h\lambda_h d\mathbf{s} + \int_{\partial\Omega_{\text{in}}} \mathbf{u}_h \cdot \mathbf{n}\phi_{\text{in}}\lambda_h d\mathbf{s} \\ & + \sum_{\kappa\in\mathcal{T}} \int_{\partial\kappa} \frac{1}{2}\beta\Big(\bar{\phi}_h - \phi_h\Big)^2 d\mathbf{s} + \sum_{\kappa\in\mathcal{T}} \int_\kappa \frac{1}{2}\zeta(\nabla\phi_h)^2 d\mathbf{x}. \end{aligned} \tag{18}$$

Let us take a moment to examine the lines in Equation 18. The first should be familiar from Equation 17 arising from our desire to minimize the squared distance between the tracer values and the field approximation. The second and third lines are the HDG discretization of the weak formulation of the advection equation, that is, the constraint PDE in Equation 17. The first term in the final line arises from the HDG discretization where $\beta > 0$ is a parameter which prevents the problem from becoming singular when $\mathbf{u}_h\cdot\mathbf{n}_\kappa = 0$ by penalizing jumps between the field approximation defined on the cells $\phi_h$ and the facets $\bar{\phi}_h$. In all computations in this work we choose $\beta = 10^{-6}$ so that the regularization term governed by $\beta$ is smaller than the term resulting from the objective function in Equation 17a. The final term introduces a gradient penalty parameter $\zeta \geq 0$. For $\zeta > 0$, this term penalizes the overshoot and undershoot oscillations which occur when approximating a discontinuous function of degree $p \geq 1$ where a discontinuity is not aligned with the mesh.

The FE approximation $(\phi_h, \bar{\phi}_h, \lambda_h) \in W_{\text{DG}}^{h,\text{k}} \times \bar{W}_{\text{DG}}^{h,\text{k}} \times T^h$ at the stationary points of Equation 18 is computed by taking the Fréchet derivative and solving the resulting linear system (Maljaars et al., 2019), that is, find $(\phi_h, \bar{\phi}_h, \lambda_h) \in W_{\text{DG}}^{h,\text{k}} \times \bar{W}_{\text{DG}}^{h,\text{k}} \times T^h$ such that

$$\begin{aligned} & \sum_{p=1}^{N_p} \Big(\phi_h(\mathbf{x}_p(t), t) - \phi_p(t)\Big)\delta\phi_h(\mathbf{x}_p(t)) - \sum_{\kappa\in\mathcal{T}} \int_{\partial\kappa} \beta\Big(\bar{\phi}_h - \phi_h\Big)\delta\phi_h \, d\mathbf{s} + \int_\Omega \frac{\partial\delta\phi_h}{\partial t}\lambda_h \, d\mathbf{x} \\ & \quad - \sum_{\kappa\in\mathcal{T}} \int_\kappa \mathbf{u}_h \cdot \nabla\lambda_h\delta\phi_h \, d\mathbf{x} + \int_{\partial\Omega_{\text{out}}} \mathbf{u}_h \cdot \mathbf{n}\lambda_h\delta\phi_h \, d\mathbf{s} + \sum_{\kappa\in\mathcal{T}} \int_\kappa \zeta\nabla\phi_h \cdot \nabla\delta\phi_h \, d\mathbf{x} = 0, \end{aligned} \tag{19}$$

$$\begin{aligned} & \int_\Omega \frac{\partial\phi_h}{\partial t}\delta\lambda_h d\mathbf{x} - \sum_{\kappa\in\mathcal{T}} \int_\kappa \mathbf{u}_h\phi_h \cdot \nabla\delta\lambda_h d\mathbf{x} \\ & + \sum_{\kappa\in\mathcal{T}} \int_{\partial\kappa\backslash\partial\Omega} \mathbf{u}_h \cdot \mathbf{n}_\kappa \bar{\phi}_h \delta\lambda_h d\mathbf{s} + \int_{\partial\Omega_{\text{out}}} \mathbf{u}_h \cdot \mathbf{n}\phi_h\delta\lambda_h d\mathbf{s} + \int_{\partial\Omega_{\text{in}}} \mathbf{u}_h \cdot \mathbf{n}\phi_{\text{in}}\delta\lambda_h d\mathbf{s} \quad = 0, \end{aligned} \tag{20}$$

$$\sum_{\kappa\in\mathcal{T}}\int_{\partial\kappa\setminus\partial\Omega}\mathbf{u}_h\cdot\mathbf{n}_\kappa\lambda_h\delta\overline{\phi}_h d\mathbf{s}+\sum_{\kappa\in\mathcal{T}}\int_{\partial\kappa}\beta\left(\overline{\phi}_h-\phi_h\right)\delta\overline{\phi}_h d\mathbf{s}\quad=0, \tag{21}$$

for all $(\delta\phi_h,\delta\overline{\phi}_h,\delta\lambda_h)\in W_{\mathrm{DG}}^{h,\mathsf{k}}\times\overline{W}_{\mathrm{DG}}^{h,\mathsf{k}}\times T^h$. For numerical analysis details we refer to Maljaars et al. (2019), and for implementation and practical demonstration we refer to Maljaars et al. (2021). Although the algebra deriving this system appears daunting, the Unified Form Language component of the FEniCS project facilitates automatic symbolic differentiation of forms like Equation 18 (Alnæs et al., 2014). Furthermore, we exploit the concept of static condensation (also known as Guyan reduction, Guyan, 1965) by locally eliminating the unknowns $\phi_h$ and $\lambda_h$ in favor of $\overline{\phi}_h$ on each element during matrix assembly (see Section S2 for a brief overview and Maljaars et al. (2021) regarding practical implementation). Therefore we only need solve for $\overline{\phi}_h$ in a global sense, thereby vastly reducing size of the underlying linear system.

A key result of this formulation is that we achieve local and global conservation of $\phi_h$ (Maljaars et al., 2019). Consider the component of the Fréchet derivative in all directions $\delta\lambda_h$ Equation 20. To show global conservation we assign $\delta\lambda_h=1$ to yield

$$\int_\Omega\frac{\partial\phi_h}{\partial t}d\mathbf{x}+\sum_{\kappa\in\mathcal{T}}\int_{\partial\kappa\setminus\partial\Omega}\mathbf{u}_h\cdot\mathbf{n}_\kappa\overline{\phi}_h d\mathbf{s}+\int_{\partial\Omega_{\mathrm{out}}}\mathbf{u}_h\cdot\mathbf{n}\phi_h d\mathbf{s}+\int_{\partial\Omega_{\mathrm{in}}}\mathbf{u}_h\cdot\mathbf{n}\phi_{\mathrm{in}}d\mathbf{s}=0. \tag{22}$$

Since $\overline{\phi}$ is single valued across a facet, the flux at interior facets vanishes if $\mathbf{u}_h$ has continuous normal components across the facets (as in the case of the TH and HDG schemes), thus leaving us with the global conservation statement

$$\int_\Omega\frac{\partial\phi_h}{\partial t}d\mathbf{x}=-\int_{\partial\Omega_{\mathrm{out}}}\mathbf{u}_h\cdot\mathbf{n}\phi_h d\mathbf{s}-\int_{\partial\Omega_{\mathrm{in}}}\mathbf{u}_h\cdot\mathbf{n}\phi_{\mathrm{in}}d\mathbf{s}. \tag{23}$$

This equation shows that the net change in $\phi$ over the domain $\Omega$ equals the total ingoing flux over the exterior boundary of the domain, thereby proving global mass conservation. Note that using nonconforming FE methods such as standard DG FE to compute $\mathbf{u}_h$ will not inherently provide a $\mathbf{u}_h$ field with the necessary continuity requirements to obtain conservation of $\phi_h$.

To show local conservation we assign $\delta\lambda_h=1$ on $\kappa$ and $\delta\lambda_h=0$ on $\Omega\setminus\kappa$

$$\int_\kappa\frac{\partial\phi_h}{\partial t}d\mathbf{x}+\int_{\partial\kappa\setminus\partial\Omega}\mathbf{u}_h\cdot\mathbf{n}_\kappa\overline{\phi}_h d\mathbf{s}+\int_{\partial\kappa\cap\partial\Omega_{\mathrm{out}}}\mathbf{u}_h\cdot\mathbf{n}\phi_h d\mathbf{s}+\int_{\partial\kappa\cap\partial\Omega_{\mathrm{in}}}\mathbf{u}_h\cdot\mathbf{n}\phi_{\mathrm{in}}d\mathbf{s}=0. \tag{24}$$

The change in the composition field $\phi_h$ in a cell $\kappa$ is determined by the incoming and outgoing fluxes over the cell boundary $\partial\kappa$, thus demonstrating local conservation in terms of the numerical flux $\mathbf{u}_h\cdot\mathbf{n}\overline{\phi}_h$.

### 4.3. A $\theta$-Scheme (Which may be Independent of $\theta$)

Consider the semi-discrete formulation of the PDE-constrained $l_2$ projection Equations 19–21. We discretize the time derivative using a $\theta$ scheme such that

$$\phi_h^{m+\theta}=\theta\phi_h^{m+1}+(1-\theta)\phi_h^m \tag{25}$$

where $0\le\theta\le1$ such that the time derivative is approximated by

$$\frac{\partial\phi_h}{\partial t}\approx\frac{\phi_h^{m+1}-\phi_h^m}{\Delta t_m} \tag{26}$$

Furthermore, we note that

$$\frac{\partial\delta\phi_h}{\partial t}\approx\frac{\delta\phi_h^{m+1}-\delta\phi_h^m}{\Delta t_m}=\frac{\delta\phi_h^{m+1}}{\Delta t_m}, \tag{27}$$

$$\delta\phi_h^{m+\theta} = \theta\delta\phi_h^{m+1} + (1-\theta)\delta\phi_h^m = \theta\delta\phi_h^{m+1}, \tag{28}$$

since the Fréchet derivative of the known quantity $\delta\phi_h^m$ has no variation.

The discrete system may therefore be written

$$\begin{aligned}
&\sum_{p=1}^{N_p}\left(\phi_h^{m+1}(\mathbf{x}_p^{m+1}) - \phi_p\right)\delta\phi_h^{m+1}(\mathbf{x}_p^{m+1}) - \sum_{\kappa\in\mathcal{T}}\int_{\partial\kappa}\beta\left(\bar{\phi}_h^{m+1} - \phi_h^{m+1}\right)\delta\phi_h^{m+1}d\mathbf{s}\\
&\quad + \int_{\cdot}\frac{\delta\phi_h^{m+1}}{t_m}\lambda_h d\mathbf{x} - \theta\sum_{\kappa\in\mathcal{T}}\int_\kappa \mathbf{u}_h^{m+1}\cdot\nabla\lambda_h\delta\phi_h^{m+1}d\mathbf{x} + \theta\int_{\partial^{\cdot}_{\text{out}}}\mathbf{u}_h^{m+1}\cdot\mathbf{n}\lambda_h\delta\phi_h^{m+1}d\mathbf{s}\\
&\quad + \sum_{\kappa\in\mathcal{T}}\int_\kappa \zeta\nabla\phi_h^{m+1}\cdot\nabla\delta\phi_h^{m+1}d\mathbf{x} = 0,
\end{aligned} \tag{29}$$

$$\begin{aligned}
&\int_\Omega\frac{\phi_h^{m+1}-\phi_h^m}{t_m}\delta\lambda_h d\mathbf{x} - \sum_{\kappa\in\mathcal{T}}\int_\kappa(\mathbf{u}_h\phi_h)^{m+\theta}\cdot\nabla\delta\lambda_h d\mathbf{x}\\
&\quad + \sum_{\kappa\in\mathcal{T}}\int_{\partial\kappa\setminus\partial\Omega}\mathbf{u}_h^{m+1}\cdot\mathbf{n}_\kappa\bar{\phi}_h^{m+1}\delta\lambda_h d\mathbf{s} + \int_{\partial^{\cdot}_{\text{out}}}(\mathbf{u}_h\phi_h)^{m+\theta}\cdot\mathbf{n}\delta\lambda_h d\mathbf{s}\\
&\quad + \int_{\partial\Omega_{\text{in}}}(\mathbf{u}_h\phi_{\text{in}})^{m+\theta}\cdot\mathbf{n}\delta\lambda_h d\mathbf{s} = 0,
\end{aligned} \tag{30}$$

$$\sum_{\kappa\in\mathcal{T}}\int_{\partial\kappa\setminus\partial\Omega_{\text{in}}}\mathbf{u}_h^{m+1}\cdot\mathbf{n}_\kappa\lambda_h\delta\bar{\phi}_h^{m+1}d\mathbf{s} + \sum_{\kappa\in\mathcal{T}}\int_{\partial\kappa}\beta\left(\bar{\phi}_h^{m+1} - \phi_h^{m+1}\right)\delta\bar{\phi}_h^{m+1}d\mathbf{s} = 0. \tag{31}$$

Recall that we have made the choice that $T^h$ is composed of piecewise constant values on the facets of the mesh. Therefore, all terms including $\nabla\lambda_h$ and $\nabla\delta\lambda_h$ vanish. Furthermore, if $\mathbf{u}_h\cdot\mathbf{n} = 0$ on $\partial\Omega$ (as is common in geodynamical simulations), this leaves us with a formulation independent of $\theta$.

### 4.4. Number of Tracers per Cell

We must carefully consider the number of tracers per cell required to obtain an accurate reconstruction of $\phi_h(\mathbf{x}, t)$ from the particle data. The minimum data per cell to construct $\phi_h(\mathbf{x}, t)$ is one tracer per FE degree of freedom defined on the element. This is clear when examining Equation 16, since the linear system will be underdetermined should $N_{p_\kappa}$ be smaller than the local dimension of the approximating function space on $\kappa$, $\dim W_{\text{DG}}^{h,\mathsf{k}}(\kappa)$. In essence, we have the requirement that

$$N_{p_\kappa} \geq \dim W_{\text{DG}}^{h,\mathsf{k}}(\kappa), \quad \kappa\in\mathcal{T}. \tag{32}$$

Although this provides a minimum bound, in practice it is the authors' experience that choosing $N_{p_\kappa} \geq 15(\mathsf{k}+1)$ is robust over long simulation times.

## 5. Finite Elements for the Stokes System

In the previous sections, we established the methods by which the composition field will be approximated by tracer data. Specifically in Section 3, we established time integration methods for tracer positions and in Section 4 how we project the tracer data onto the composition field $\phi_h(\mathbf{x}, t)$. In this section, we show how the velocity approximation will be computed in the ODE Equation 7 by means of solving a Stokes system.

We discretize the momentum and continuity equations Equations 1 and 2 with the FE method, seeking approximations of the velocity and pressure, $\mathbf{u}_h$, and $p_h$ in the domain $\Omega$, respectively. This is subject to the boundary conditions $\mathbf{u}_h = \mathbf{u}_D$ on $\partial\Omega_D$ and $(2\eta\underline{\underline{\dot{\varepsilon}}}(\mathbf{u}) - p\underline{\underline{I}})\cdot\mathbf{n} = \mathbf{g}_N$ on $\partial\Omega_N$ where $\underline{\underline{I}}$ is the identity tensor. Here, $\partial\Omega_D$ and $\partial\Omega_N$, are the subdivision of $\partial\Omega$ into non-overlapping Dirichlet and Neumann components, respectively, where $\partial\Omega_D$ is not empty ($\partial\Omega = \partial\Omega_D\cup\partial\Omega_N$, $\partial\Omega_D\cap\partial\Omega_N = \emptyset$, and $\partial\Omega_D \neq \emptyset$).

There are numerous FE methods for approximating Equations 1 and 2. We will present the TH and HDG methods. A small subset of other methods is discussed in Section S4. Furthermore, for each method, we

summarize and state salient properties in Table 2. For extended detail see the cited monographs and references therein.

### 5.1. Taylor-Hood

The TH method (Taylor & Hood, 1973) is a commonly used FE method for discretizing the incompressible Stokes system. In this setting, we seek the velocity approximation $\mathbf{u}_h$ in the vector space $V^{h,\mathrm{p}}$ which comprises functions of dimension $d$ composed of piecewise polynomials of degree $p \geq 2$ defined on each element $\kappa \in \mathcal{T}$ and continuous in $\Omega$. This is the "typical" conforming and continuous FE vector space of functions. We further define $V^{h,\mathrm{p}}_{\mathrm{BC}} \coloneqq \{\mathbf{v} \in V^{h,\mathrm{p}} : \mathbf{v}\,|_{\partial\Omega_D} = \mathbf{u}_D\}$ and $V^{h,\mathrm{p}}_{\mathrm{BC}_0} \coloneqq \{\mathbf{v} \in V^{h,\mathrm{p}} : \mathbf{v}\,|_{\partial\Omega_D} = 0\}$ which are the continuous FE function spaces satisfying arbitrary Dirichlet boundary data $\mathbf{u}_D$ and homogeneous boundary data, respectively. We also seek the pressure approximation $p_h$ in $Q^{h,\mathrm{p}-1}$ which is the "typical" conforming and continuous scalar FE space of order $p-1$. The TH FE problem reads: find $(\mathbf{u}_h, p_h) \in V^{h,\mathrm{p}} \times Q^{h,\mathrm{p}-1}$ such that

$$\int_\Omega 2\eta \underline{\underline{\varepsilon}}(\mathbf{u}_h) : \underline{\underline{\varepsilon}}(\mathbf{v}_h) d\mathbf{x} - \int_\Omega p_h \nabla \cdot \mathbf{v}_h d\mathbf{x} = \int_\Omega \mathbf{f} \cdot \mathbf{v}_h d\mathbf{x} - \int_{\partial\Omega_N} \mathbf{g}_N \cdot \mathbf{v}_h d\mathbf{s} \tag{33}$$

$$-\int_\Omega q_h \nabla \cdot \mathbf{u}_h d\mathbf{x} = 0 \tag{34}$$

for all test functions $(\mathbf{v}_h, q_h) \in V^{h,\mathrm{p}}_{\mathrm{BC}_0} \times Q^{h,\mathrm{p}-1}$.

### 5.2. Hybrid Discontinuous Galerkin

Introduced in Cockburn and Gopalakrishnan (2009), Labeur and Wells (2007) and developed further in Cockburn and Sayas (2014), Cockburn et al. (2011), and Nguyen et al. (2010), HDG methods have recently gained significant interest (Cockburn et al., 2010; Labeur & Wells, 2012; Maljaars et al, 2018, 2019; Rhebergen & Wells, 2018a, 2020). One of the most attractive characteristics of the HDG method for the Stokes system is that it lends itself to a pointwise divergence free velocity field approximation. By this, we mean that

$$\nabla \cdot \mathbf{u}_h(\mathbf{x}, t) = 0 \quad \forall \mathbf{x} \in \Omega. \tag{35}$$

This is particularly important when using tracer methods in geodynamics where advection is dependent on pointwise evaluation of an incompressible velocity field.

The term "hybrid" is used to imply that the HDG method is composed of FE functions drawn from a pairing of FE spaces defined on the mesh cells and FE spaces defined on the mesh facets. We refer to Rhebergen and Wells (2020) for specific choices of these spaces, their analysis and properties. In this work we use the embedded discontinuous Galerkin HDG formulation drawing on the advantage of a pressure-robust[2] (The velocity error estimates do not depend on the pressure error (Rhebergen & Wells, 2020)., $H$(div)-conforming[3] (The divergence of the velocity is piecewise continuous (square integrable). and pointwise divergence free approximation with reduced degree of freedom count in the velocity component of the global linear system compared to other HDG methods (see Rhebergen & Wells, 2020).

The velocity and pressure approximations are comprised of two components:

1. $\mathbf{u}_h$ and $p_h$ are drawn from the vector FE space $V^{h,\mathrm{p}}_{\mathrm{DG}}$ and scalar FE space $Q^{h,\mathrm{p}-1}_{\mathrm{DG}}$, respectively, which are continuous on each element $\kappa \in \mathcal{T}$ and discontinuous in $\Omega$, that is, the standard DG spaces of degree $p$ and $p-1$, respectively

2. $\overline{\mathbf{u}}_h$ and $\overline{p}_h$ are drawn from $\overline{V}^{h,\mathrm{p}}$ and $\overline{Q}^{h,\mathrm{p}}_{\mathrm{DG}}$, respectively. Here $\overline{V}^{h,\mathrm{p}}$ is the vector space of piecewise polynomials of degree $p$ defined on the facets of the mesh and continuous between facets, and $\overline{Q}^{h,\mathrm{p}}_{\mathrm{DG}}$ is the scalar space composed of piecewise polynomials of degree $p$ defined on the facets of the mesh and discontinuous between facets

To enforce Dirichlet boundary conditions, we modify the vector space defined on the facets such that it satisfies the boundary data $\overline{\mathbf{u}}_h\,|_{\partial\Omega_D} = \mathbf{u}_D$, that is, $\overline{V}^{h,\mathrm{p}}_{\mathrm{BC}} \coloneqq \{\overline{\mathbf{v}} \in \overline{V}^{h,\mathrm{p}} : \overline{\mathbf{v}}\,|_{\partial\Omega_D} = \mathbf{u}_D\}$. Furthermore, we define $\overline{V}^{h,\mathrm{p}}_{\mathrm{BC}_0} \coloneqq \{\overline{\mathbf{v}} \in \overline{V}^{h,\mathrm{p}} : \overline{\mathbf{v}}\,|_{\partial\Omega_D} = 0\}$.

The HDG FE formulation is as follows: find $\left(\mathbf{u}_h, \overline{\mathbf{u}}_h, p_h, \overline{p}_h\right) \in V^{h,\mathrm{p}}_{\mathrm{DG}} \times \overline{V}^{h,\mathrm{p}}_{\mathrm{BC}} \times Q^{h,\mathrm{p}-1}_{\mathrm{DG}} \times \overline{Q}^{h,\mathrm{p}}_{\mathrm{DG}}$ such that

$$\begin{aligned} &-\sum_{\kappa\in\mathcal{T}} \int_\kappa \underline{\underline{\sigma}} : \nabla \mathbf{v}_h d\mathbf{x} + \sum_{\kappa\in\mathcal{T}} \int_{\partial\kappa} \widehat{\underline{\underline{\sigma}}}\mathbf{n}_\kappa \cdot \mathbf{v}_h d\mathbf{s} \\ &+ \sum_{\kappa\in\mathcal{T}} \int_{\partial\kappa} 2\eta\left(\overline{\mathbf{u}}_h - \mathbf{u}_h\right) \cdot \underline{\underline{\varepsilon}}(\mathbf{v}_h)\mathbf{n}_\kappa d\mathbf{s} = \int_\Omega \mathbf{f} \cdot \mathbf{v}_h d\mathbf{x}, \end{aligned} \tag{36}$$

$$\sum_{\kappa\in\mathcal{T}} \int_{\partial\kappa} \widehat{\underline{\underline{\sigma}}}\mathbf{n}_\kappa \cdot \overline{\mathbf{v}}_h d\mathbf{s} = \int_{\partial\Omega_N} \mathbf{g}_N \cdot \overline{\mathbf{v}}_h d\mathbf{s}, \tag{37}$$

$$\sum_{\kappa\in\mathcal{T}} \int_\kappa \mathbf{u}_h \cdot \nabla q_h d\mathbf{x} - \sum_{\kappa\in\mathcal{T}} \int_{\partial\kappa} \mathbf{u}_h \cdot \mathbf{n}_\kappa q_h d\mathbf{s} = 0, \tag{38}$$

$$\sum_{\kappa\in\mathcal{T}} \int_{\partial\kappa} \mathbf{u}_h \cdot \mathbf{n}_\kappa \overline{q}_h d\mathbf{s} - \int_{\partial\Omega} \overline{\mathbf{u}}_h \cdot \mathbf{n}\overline{q}_h d\mathbf{s} = 0, \tag{39}$$

for all $\left(\mathbf{v}_h, \overline{\mathbf{v}}_h, q_h, \overline{q}_h\right) \in V^{h,\mathrm{p}}_{\mathrm{DG}} \times \overline{V}^{h,\mathrm{p}}_{\mathrm{BC}_0} \times Q^{h,\mathrm{p}-1}_{\mathrm{DG}} \times \overline{Q}^{h,\mathrm{p}}_{\mathrm{DG}}$. Here,

$$\underline{\underline{\sigma}} = -2\eta\,\underline{\underline{\varepsilon}}(\mathbf{u}_h) + p_h \underline{\underline{I}}, \tag{40}$$

$$\widehat{\underline{\underline{\sigma}}} = -2\eta\,\underline{\underline{\varepsilon}}(\mathbf{u}_h) + \overline{p}_h \underline{\underline{I}} - \frac{\alpha}{h_\kappa} 2\eta\left(\overline{\mathbf{u}}_h - \mathbf{u}_h\right) \otimes \mathbf{n}_\kappa, \tag{41}$$

are the stress tensor and stress tensor numerical flux, respectively, where $h_\kappa$ is the measure of the cell diameter and $\alpha$ is a penalty parameter similar to those found in interior penalty methods chosen in this work to be $\alpha = 6p^2$ (as in Rhebergen & Wells, 2018a). Careful choice of $\alpha$ is required to ensure stability of the numerical method (e.g., Lew et al., 2004). The factor $p^2$ is chosen based on analysis (e.g., Prudhomme et al., 2000).

At first glance, the four equations appear to demand a large increase in computational expense when compared with the TH method. However, the local components, $\mathbf{u}_h$ and $p_h$ of Equations 36 and 38, may be eliminated in favor of the global components $\overline{\mathbf{u}}_h$ and $\overline{p}_h$ by the process of static condensation (see Section S2). Regarding preconditioning of this global system we refer to Rhebergen and Wells (2018b).

### 5.3. Pointwise Divergence Free Fields

In this section, we define a measure of the error in the incompressibility of our discretized velocity field $\mathbf{u}_h$. Moreover, we will briefly show why the HDG formulation yields a pointwise divergence free velocity approximation unlike the TH scheme.

We define the following error norm

$$\|\,\mathbf{u}_h\,\|_{\mathrm{div}} \coloneqq \|\,\nabla\cdot\mathbf{u}_h\,\|_{L_2(\Omega)} = \left(\int_\Omega \left(\nabla\cdot\mathbf{u}_h\right)^2 d\mathbf{x}\right)^{\frac{1}{2}} \tag{42}$$

which measures the square distance of the divergence of $\mathbf{u}_h$ from the desired divergence free value[4] ($L_2(\Omega)$ is the set of all real valued functions defined on $\Omega$ such that $\|\,v\,\|_{L_2(\Omega)} \coloneqq \left(\int_\Omega |\,v(x)\,|^2\,d\mathbf{x}\right)^{\frac{1}{2}} < \infty$).

By the TH formulation Equation 34, we have that $\int_\Omega q_h \nabla\cdot\mathbf{u}_h\,d\mathbf{x} = 0$ for all $q_h \in Q^{h,\mathrm{p}-1}$. In order to fulfill $\|\mathbf{u}_h\|_{\mathrm{div}} = 0$ we need to choose $q_h = \nabla\cdot\mathbf{u}_h$. However, this choice does not satisfy the regularity constraint

$q_h \in Q^{h,\mathrm{p}-1}$ given that $\nabla \cdot \mathbf{u}_h \in Q_{\mathrm{DG}}^{h,\mathrm{p}-1}$. In essence, $q_h$ and $\nabla\cdot\mathbf{u}_h$ are continuous and discontinuous across the elements $\kappa \in \mathcal{T}$, respectively, so that condition Equation 42 cannot be satisfied.

Now, consider the HDG formulation of the continuity Equation 38. Integrating by parts, we have $\sum_{\kappa\in\mathcal{T}} \int_\kappa q_h \nabla \cdot \mathbf{u}_h d\mathbf{x} = 0$ for all $q_h \in Q_{\mathrm{DG}}^{h,\mathrm{p}-1}$. In this setting, it is valid to make the choice $q_h = \nabla\cdot\mathbf{u}_h$ and the measure $\|\mathbf{u}_h\|_{\mathrm{div}} = 0$ is satisfied. One may further use this argument to show that $\overline{\mathbf{u}}_h$ is also pointwise divergence free (see Rhebergen & Wells, 2018a).

## 6. Implementation and Data Availability

The code developed to run all the numerical experiments in this work is available in the public repository Sime (2020). The dependencies of the code are briefly discussed in this section. For portable deployment in a Docker container we refer to Hale (2020) and Hale et al. (2017).

The core FE computations are performed using the components of the FEniCS project (Alnæs et al., 2015). The Lagrangian-Eulerian on Particles (LEoPart) library (Maljaars et al., 2021) is used in conjunction with FEniCS for managing tracer data, Runge-Kutta numerical integration, $l_2$ projection, PDE-constrained $l_2$ projection, and static condensation HDG system assembly. We further exploit the techniques described in Houston and Sime (2018) for the automatic generation of DG and HDG formulations. The PETSc (Balay et al., 2019a, 2019b) library is used for its data structures for sparse numerical linear algebra in addition to the multifrontal parallel solver MUltifrontal Massively Parallel Sparse direct Solver (Amestoy et al., 2000) for the direct factorization of matrices.

The components of the FEniCS project have been demonstrated to be scalable with parallel computation (e.g., Richardson et al., 2019). LEoPart leverages the inherent parallelism of the FEniCS project and Maljaars et al. (2021) demonstrate scalable performance of matrix assembly and particle advection. Regarding scalable solution of the HDG FE Stokes system reduced by static condensation using iterative methods we refer to the construction of an appropriate preconditioner discussed in Rhebergen and Wells (2018b).

## 7. Numerical Experiments

We have now established the fundamentals to be used in our tracer advection numerical experiments. We will examine our implementation of each aspect of the discretization of the incompressible Stokes Equations 1–3.

As a preliminary step, we verify our FE implementation of the Stokes system and advection of tracers by the RK$\ell$ method. In Section S5.1, we choose velocity and pressure solutions of the Stokes system a priori such that we may compute the error of their respective FE approximations. On a hierarchy of successively finer grids we observe this error to converge at optimal rates as predicted by FE analysis, verifying our implementation. In a similar fashion in Section S5.2, we verify our implementation of tracer advection by the RK$\ell$ method.

With this foundation, the projection of tracer data to a field by the $l_2$ and the PDE-constrained $l_2$ projection is verified in Section 7.1. Incorporation of these projection schemes into a coupled system where a solution is known a priori is examined in Section 7.2. With all of these methods verified, we proceed to solve the Rayleigh-Taylor problem benchmark showcased in van Keken et al. (1997) in Section 7.3. We provide more extensive experimental results in the Section S5.4, particularly applied to other frequently used FE schemes for the Stokes system as summarized in Equation 2.

Throughout this section, we write "bounded/penalized" projection to mean that we employ box-constrained optimization and the penalized gradient ($\zeta > 0$) methods in the case of $l_2$ and PDE-constrained $l_2$ projection, respectively. See Section 4.2 for details. Additionally, for each experiment, we select $C_{\mathrm{CFL}} = 1$.

### 7.1. Tracer Advection and Projection to a Field

With the Stokes system FE scheme and tracer advection verified (see Sections S5.1 and S5.2), we examine the projection of tracer data $X(t)$ and $\Phi(t)$ to a composition field representation. The manufactured exam-

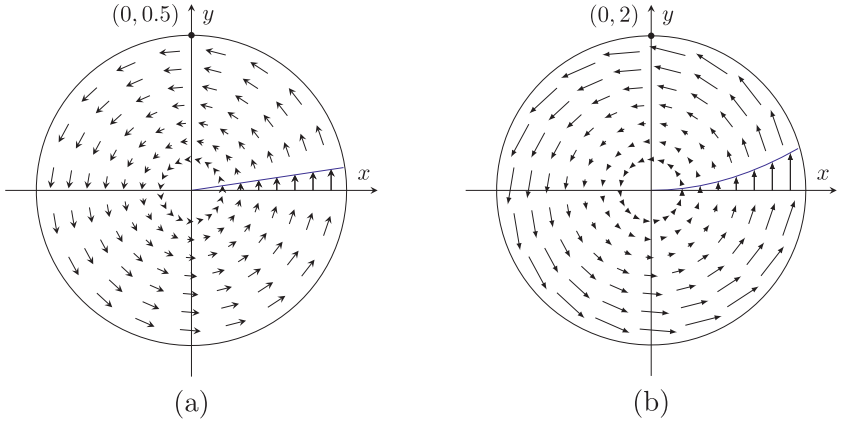


**Figure 5.** The domain $\Omega$ and manufactured velocity $\mathbf{u}(\mathbf{x}, t = 0)$ used in (a) Equation 43 and Section 7.1 and (b) Equation 47 and Section 7.2. The blue line emphasizes linear and quadratic growth in flow speed with radius.

ples in this section are inspired by those shown in Maljaars et al. (2019). Here, two different sources of error are to be considered:

1. The temporal approximation of the tracers' positions
2. The spatial approximation of the composition field representation

Let $\Omega$ be the unit disc of radius $r = 0.5$ centered at $r_0 = (0, 0)$. We define the rotational and time varying velocity field

$$\mathbf{u} = \begin{pmatrix} -y \\ x \end{pmatrix} \left( \cos^2 t + \frac{1}{2} \right). \tag{43}$$

This velocity field is shown in Figure 5 at $t = 0$. It can be shown that the period of one rotation of any tracer in this velocity field occurs at $t = 2\pi$. We examine two cases of a smooth and non-smooth composition field

$$\text{Smooth:} \qquad \phi(\mathbf{x}, t) = \sin(2\pi x_\theta)\sin(2\pi y_\theta), \tag{44}$$

$$\text{Non - smooth:} \qquad \phi(\mathbf{x}, t) = \begin{cases} 1 & \text{if } \theta_{\text{rot}} > 0, \\ 0 & \text{if } \theta_{\text{rot}} \leq 0, \end{cases} \tag{45}$$

where

$$\begin{aligned} \theta(\mathbf{x}, t) &= -\left( t + \frac{1}{4}\sin(2t) \right), \\ \begin{pmatrix} x_\theta \\ y_\theta \end{pmatrix} &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \\ \theta_{\text{rot}} &= \textbf{arctan2}(y_\theta, x_\theta), \end{aligned} \tag{46}$$

and arctan2$(\cdot, \cdot)$ is the standard two argument arctangent function (defined in Section S1). The initial conditions $\phi(\mathbf{x}, t = 0)$ in both cases are shown in Figure 6. The smooth case is designed to test the error convergence rate of the projected tracer data through time. The non-smooth case gives an indication of the performance of the projection methods when modeling chemical phase interfaces by a discontinuity in the composition function. For example those which may require bounded/penalized projection to yield physically relevant composition field approximations to mitigate the impact of overshoots and undershoots (see Sections 4.1 and 4.2).

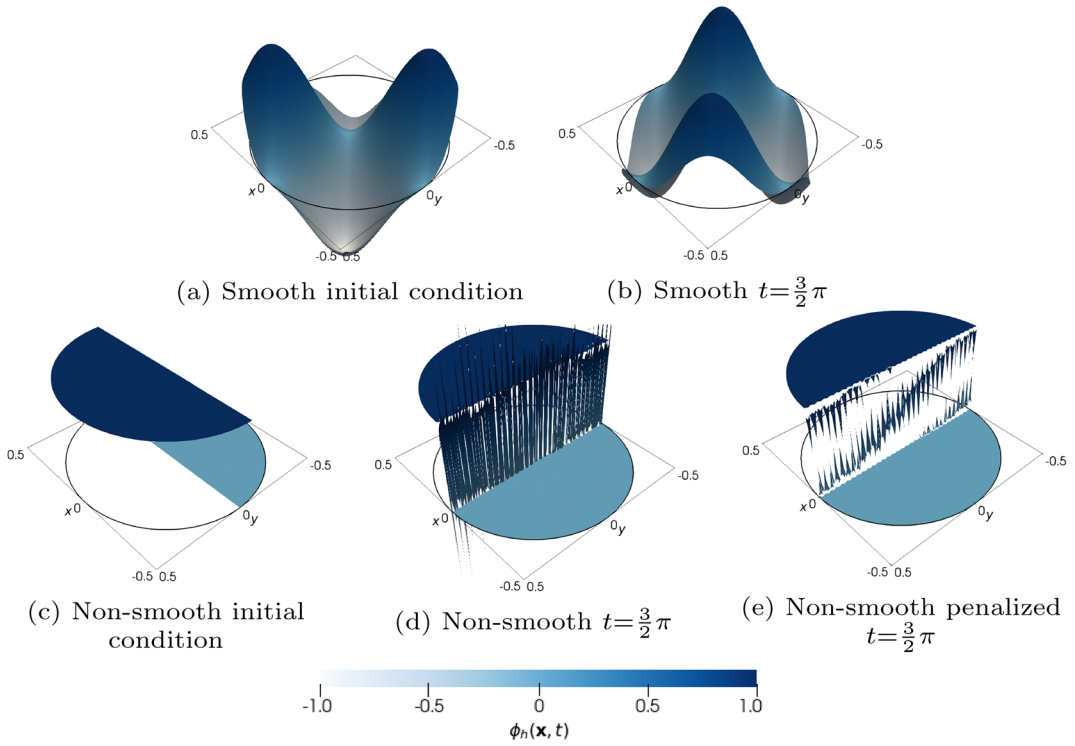


**Figure 6.** The smooth and non-smooth functions used in the manufactured field advection example in Section 7.1. Shown are the initial $\phi_h(\mathbf{x}, t = 0)$ and evolved $\phi_h(\mathbf{x}, t = \frac{3}{2}\pi)$ fields as computed by the PDE-constrained projection method where $k = 2$. Note in (d) the non-smooth case projection yields large overshoots and undershoots. (e) Using a penalized gradient method ($\zeta = 25$) reduces the severity of these overshoots and undershoots. The convergence of the error of these functions is quantified in Figure 7.

We measure two components of the error in this experiment:

1. $\| \phi - \phi_h \|_{L_2(\Omega)}$: The spatial error of the composition field approximation
2. $\epsilon_{\Delta\phi} = \int_\Omega (\phi_h(\mathbf{x}, t) - \phi_h(\mathbf{x}, 0)) d\mathbf{x}$: The mass conservation error

We compute the projected composition over a period of one rotation for the smooth, non-smooth, and non-smooth & bounded/penalized functions (see Figure 6 for examples using the PDE-constrained projection). In the bounded/penalized case, we use box-constrained optimization in the $l_2$ projection and we select $\zeta = 25$ in our PDE-constrained $l_2$ projection operator. Per mesh cell we generate 25 tracers by drawing their initial positions from a random uniform distribution defined in the cell's geometry.

The error through time for each of the smooth and non-smooth cases is shown in Figure 7. Here, we see in the smooth case that a comparison between the error using the $l_2$ and PDE-constrained $l_2$ projection methods is indistinguishable at this scale. In the case of $l_2$, projection method $\epsilon_{\Delta\phi}$ is converging. However, the difference in the mass conservation properties between the two methods is obvious as the PDE-constrained $l_2$ method conserves $\phi_h$ to machine precision.

In the non-smooth case, the composition field approximation error from the $l_2$ and PDE-constrained $l_2$ projection methods is very similar. However, the $k = 1$ and $k = 2$ high order approximations show no great benefit over the low order $k = 0$ approximation. This is expected as the interface discontinuity is not aligned with the mesh, yielding suboptimal convergence rates. See, for example, in Figures 7b1 and 7c1, the four "dips" in error corresponding to the interface discontinuity approximately lining up with the mesh constructed in our experiment. Despite this suboptimal convergence $\phi_h$ is conserved in the case of PDE-constrained $l_2$ projection by measurement of $\epsilon_{\Delta\phi}$. The $l_2$ projection shows no evidence of $\epsilon_{\Delta\phi}$ converging with increased approximation order.

In the non-smooth and bounded/penalized case, we see similar results to the non-smooth and unbounded/unpenalized case. However, we have that the errors in the approximation of the composition field are slightly larger when using both projection methods.

The convergence properties in the smooth initial condition case after one rotation at time $t = 2\pi$ are shown in Figure S6. Evidently we recover optimal rates and recover exact mass conservation $\epsilon_{\Delta\phi}$ when using the PDE-constrained projection. Not shown are the convergence properties of the non-smooth cases since the exact field is recovered at $t = 2\pi$ in even the lowest order FE approximation.

### 7.2. Coupled Manufactured Solution

We have demonstrated and verified: the spatial approximation of the Stokes subsystem FE solution in Section S5.1, the temporal approximation of the advected tracers' positions in Section S5.2, and the spatiotemporal approximation of the projection of the tracer values to a FE function representation in Section 7.1.

Now, we examine the fully coupled case where the tracers representing the composition data are advected based on the FE approximation of the Stokes subsystem, which in turn also depends on the projection of the tracer data onto an FE function space. Analytical solutions for such problems are extremely difficult to derive in the context of geodynamics simulations. Therefore, we manufacture the following problem.

Let $\Omega \coloneqq \{(x,y) : \sqrt{x^2 + y^2} < 2\}$ be the disc of radius 2. We compute the FE approximation of the Stokes system (Equations 1 and 2) in $\Omega$ where we prescribe the velocity and pressure fields

$$\begin{aligned} \mathbf{u}(\mathbf{x},t) &= \sqrt{x^2 + y^2}\begin{pmatrix} -y \\ x \end{pmatrix}\left(\cos^2 t + \frac{1}{2}\right), \\ p(\mathbf{x},t) &= 0. \end{aligned} \tag{47}$$

Furthermore, two composition functions are defined to be advected through the velocity field:

$$\text{Smooth:} \qquad \phi(\mathbf{x},t) = \exp\left(-\frac{(x_\theta - x_0)^2 + (y_\theta - y_0)^2}{2\xi}\right), \tag{48}$$

$$\text{Non - smooth:} \qquad \phi(\mathbf{x},t) = \begin{cases} 1 & \theta_{\text{rot}} > 0, \\ 0 & \theta_{\text{rot}} \le 0. \end{cases} \tag{49}$$

here, $\xi = 0.1$, $(x_0, y_0)$=(0.25, 0) is the initial center point of the Gaussian function and

$$\begin{aligned} \theta(\mathbf{x},t) &= -\sqrt{x^2 + y^2}\left(t + \frac{1}{4}\sin(2t)\right), \\ \begin{pmatrix} x_\theta \\ y_\theta \end{pmatrix} &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}, \\ \theta_{\text{rot}} &= \mathbf{arctan2}(y_\theta, x_\theta). \end{aligned} \tag{50}$$

The velocity field is shown in Figure 5b and the smooth and non-smooth composition functions are shown in Figure 8.

In the Stokes system, we set the viscosity and momentum source to

$$\eta = 1, \qquad \mathbf{f} = \mathsf{r} + \phi\hat{\mathbf{g}}, \tag{51}$$

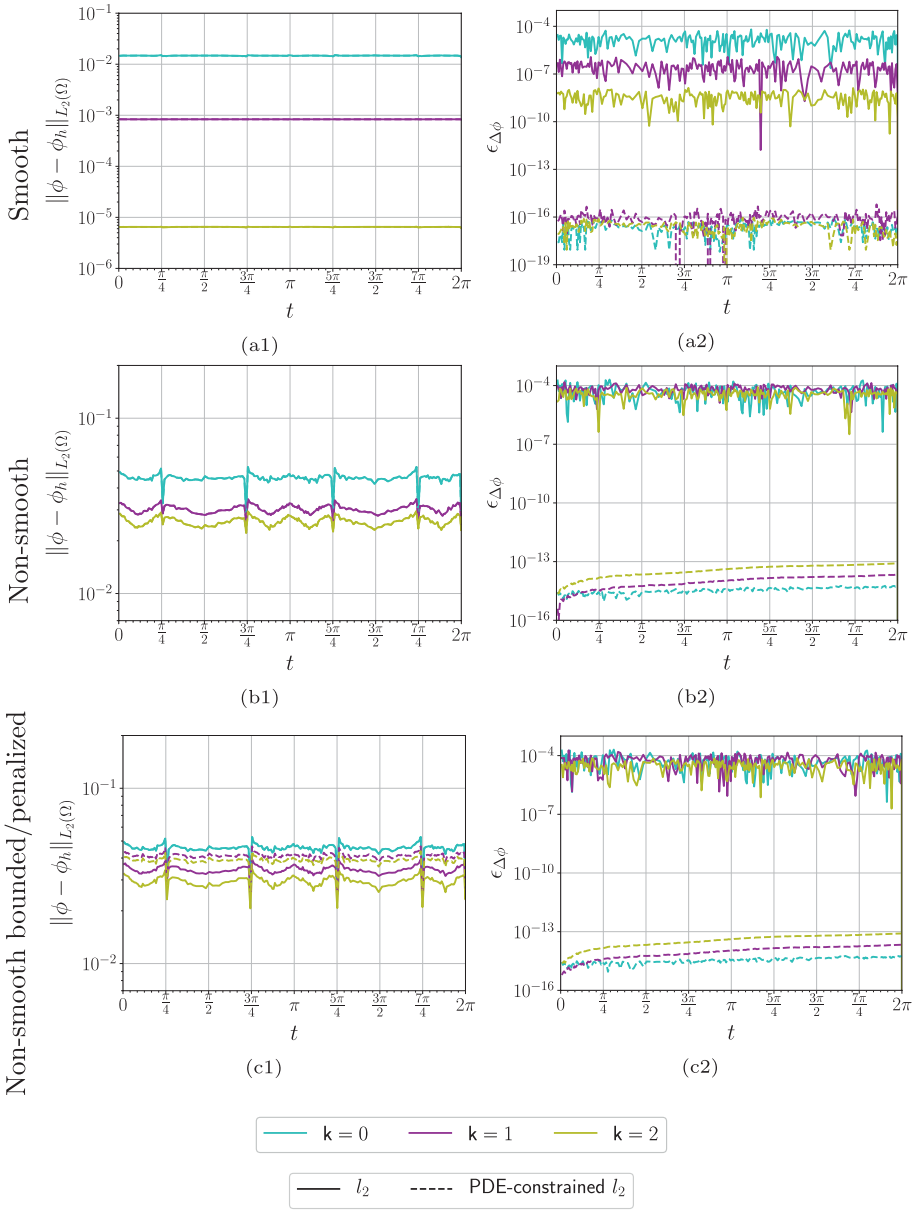


**Figure 7.** Computed error from the field advection manufactured solution experiment described in Section 7.1. The left column shows composition field approximation error measured in the $L_2$ norm with time. The right column shows and mass conservation error with time. The disc geometry is discretized with a mesh such that $\max_{\kappa\in\mathcal{T}} h_\kappa \approx \frac{1.42}{32}$ and $\min_{\kappa\in\mathcal{T}} h_\kappa = \frac{1}{32}$. In (a1) and (b1) the $L_2$ error difference between $l_2$ and PDE-constrained $l_2$ projection is indistinguishable at this scale. This is also true in (c1) for the $k = 0$ case. We show evidence for optimal convergence of the smooth function approximation in Figure S6. Projection of the non-smooth composition function is achieved with suboptimal convergence rates, to be expected from the inability to resolve the intraelement discontinuity. In (a2), (b2), and (c2), we achieve exact mass conservation by PDE-constrained projection (to machine precision), even when error convergence of the non-smooth composition field is suboptimal. An accumulation of machine precision error over simulation time is observed in the non-smooth cases.

where $\hat{\mathbf{g}}=(0,-1)^\top$ is the unit vector acting in the direction of gravity and

$$\mathsf{r} = \frac{3}{\sqrt{x^2+y^2}}\begin{pmatrix} y \\ -x \end{pmatrix}\left(\cos^2 t + \frac{1}{2}\right) - \phi\hat{\mathbf{g}} \tag{52}$$

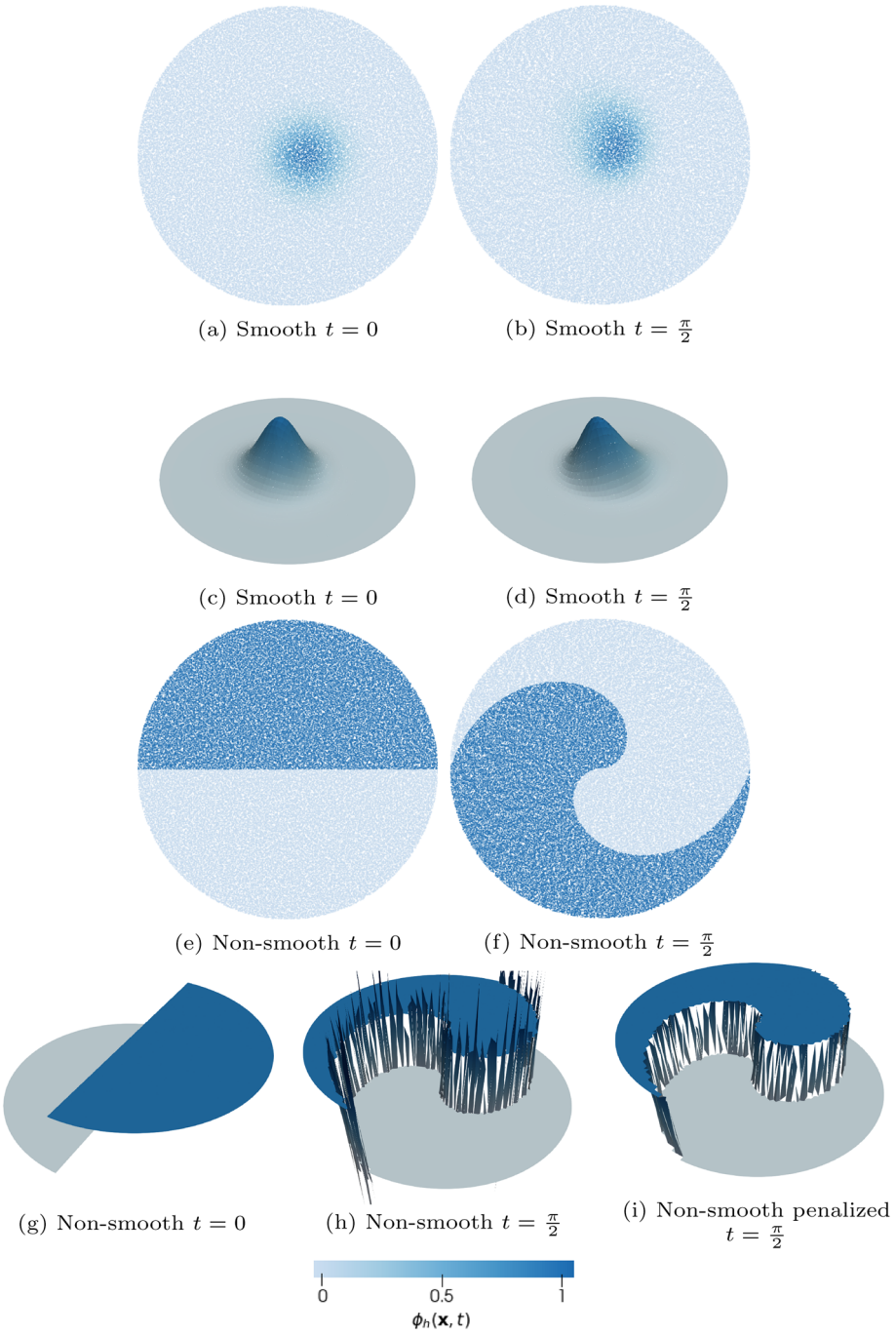


**Figure 8.** The smooth and non-smooth functions used in the manufactured solution example Equations 48 and 49 described in Section 7.2. Shown are the initial $t = 0$ and evolved $t = \frac{\pi}{2}$ particle distributions $X(t)$, $\Phi(t)$, and fields $\phi_h(\mathbf{x}, t)$ computed by PDE-constrained projection. Note in (h) the non-smooth case with no penalization of the gradient in the projection ($\zeta = 0$) yields large undershoots and overshoots. The convergence of the error of these functions is quantified in Figure 9.

is the conservation of momentum equation residual of the true solution (We use the true value of $\phi$ in the construction of the residual, $r$, prescribed by Equations 48 and 49).

We measure the following errors:

1. $\| \mathbf{u}(\mathbf{x},t) - \mathbf{u}_h(\mathbf{x},t) \|_{L_2(\Omega)}$: velocity field approximation error

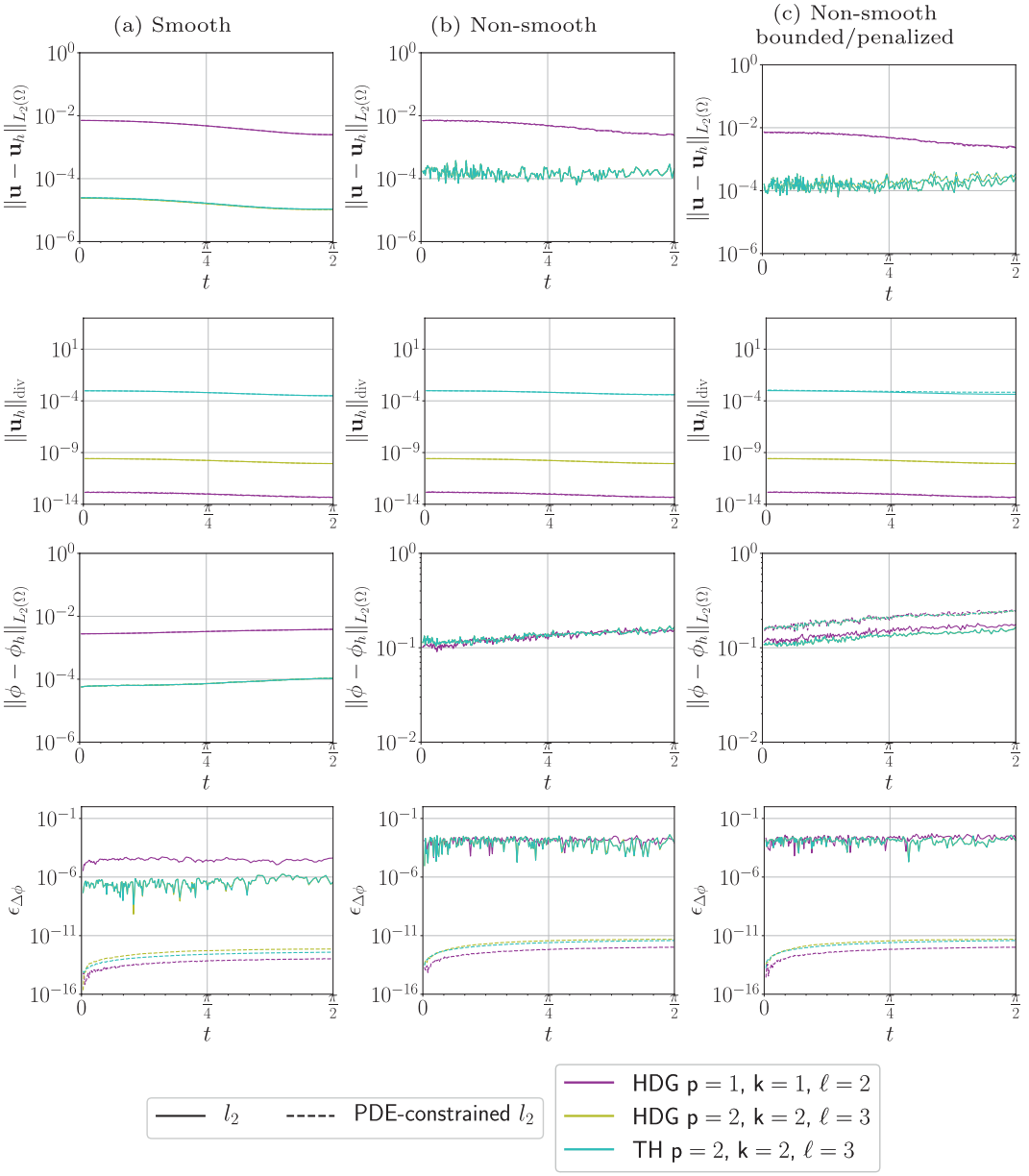


**Figure 9.** Coupled manufactured solution computed error. The disc geometry is discretized with a mesh such that $\max_{\kappa\in\mathcal{T}} h_\kappa \approx \frac{1.42}{32}$ and $\min_{\kappa\in\mathcal{T}} h_\kappa = \frac{1}{32}$. The velocity and composition field approximation errors using the third order HDG ($p = 2$) and TH methods are indistinguishable at this scale. From the data in columns (b) and (c), we observe no benefit using higher order methods to approximate a discontinuous composition field. In all cases, the PDE-constrained $l_2$ projection method preserves exact mass conservation and the HDG method provides pointwise divergence free velocity fields (to machine precision).

2. $\|\mathbf{u}_h(\mathbf{x}, t)\|_{\text{div}}$: divergence free constraint error
3. $\| \phi(\mathbf{x},t) - \phi_h(\mathbf{x},t) \|_{L_2(\Omega)}$: composition field approximation error
4. $\epsilon_{\Delta\phi} = \int_\Omega (\phi_h(\mathbf{x}, t) - \phi_h(\mathbf{x}, 0))\, d\mathbf{x}$: mass conservation error

Example solutions using the PDE-constrained projection at $t = \frac{\pi}{2}$, the error functionals computed in the simulation time domain $t \in \mathcal{I}_t = (0, \frac{\pi}{2}]$ and their convergence rates at final time $t = \frac{\pi}{2}$ are shown in Figures 8–10,

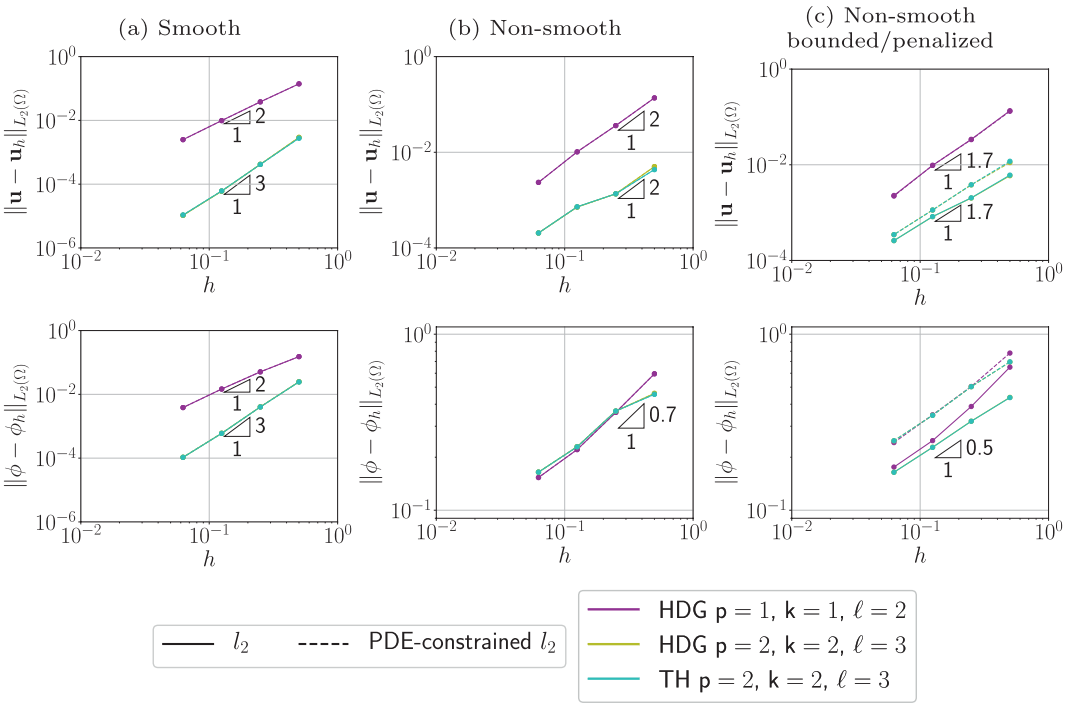


**Figure 10.** Coupled manufactured solution error functionals convergence rates measured at $t = \frac{\pi}{2}$. The cell measure here is selected such that $h = \min_{\kappa \in \mathcal{T}} h_\kappa$. The difference between the $p = 2$ HDG and TH methods, in addition to the $l_2$ and PDE-constrained $l_2$ projection, are indistinguishable at this scale. (a) Optimal convergence rates are observed in the approximation of the smooth composition field. (b) Inadequate representation of intraelement discontinuities causes convergence rates to be suboptimal in the non-smooth case. (c) Use of bounded/penalized projection to alleviate impact of undershoots in the composition field projection has a small detrimental impact on convergence rates with the benefit of a more robust solver for the Stokes system.

respectively. Here we demonstrate the second order HDG ($p = 1$, $k = 1$, and $\ell = 2$) and third order HDG and TH ($p = 2$, $k = 2$, and $\ell = 3$) methods. In each case the initial tracer configuration is generated with 25 tracers per mesh cell where $X(t = 0)$ is drawn from a random distribution defined on the cells' geometries.

In the case of the smooth composition field, we observe optimal convergence rates of the numerical methods. However, in the non-smooth case the inability to represent the intraelement discontinuity in the composition function yields suboptimal convergence rates. These suboptimal results are expected and follow from our findings in Section 7.1 and Figure 7.

### 7.3. Rayleigh-Taylor Instability

In the previous section, we demonstrated and validated: the exact conservation of the composition function; the pointwise divergence free velocity approximation; and the optimal convergence rates of the numerical approximation for smooth problems. We now turn to the Rayleigh-Taylor instability benchmark from van Keken et al. (1997) which is an example relevant to geodynamics simulations.

The Rayleigh-Taylor instability problem models the buoyancy driven evolution of a compositionally light material which initially resides below a compositionally denser layer. The chemical composition field is constructed such that $\phi = 0$ and $\phi = 1$ correspond to the immiscible light and dense materials, respectively. A key challenge lies in simulating the moving discontinuity at the interface between these immiscible materials.

In all of the numerical experiments in this section, we employ bounded/penalized projection methods of the tracer data. As in the previous section, the initial tracers' positions are drawn from a random distribution defined on the mesh cells' geometries where we generate 25 tracers per mesh cell.

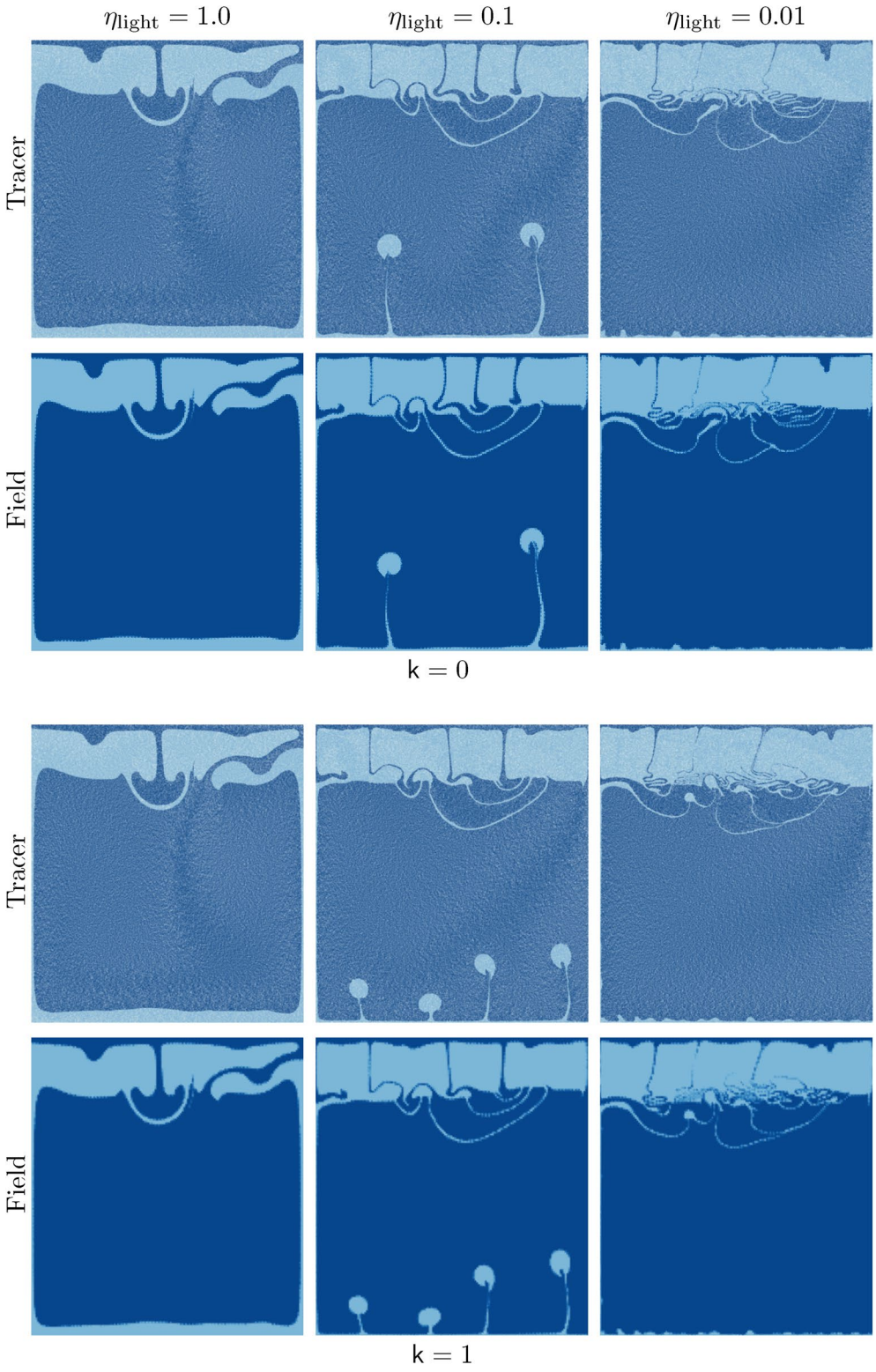


**Figure 11.** Snapshots of the Rayleigh-Taylor instability benchmark at $t = 1{,}500$ (cf. van Keken et al., 1997). Shown are the tracer distributions, tracer data and their PDE-constrained $l_2$ projection onto the composition field space $W_{\text{DG}}^{h,\text{k}}$ of degree $k \in \{0, 1\}$. Dark and light colors correspond to dense ($\phi$, $\phi_p = 1$) and light ($\phi$, $\phi_p = 0$) material, respectively. In all cases, the Stokes system is solved using the $p = 1$ HDG scheme on a mesh of $160 \times 160$ bisected quadrilaterals with 1,280,000 tracers and $\eta_{\text{dense}} = 1$. For more details see Section 7.3.2.

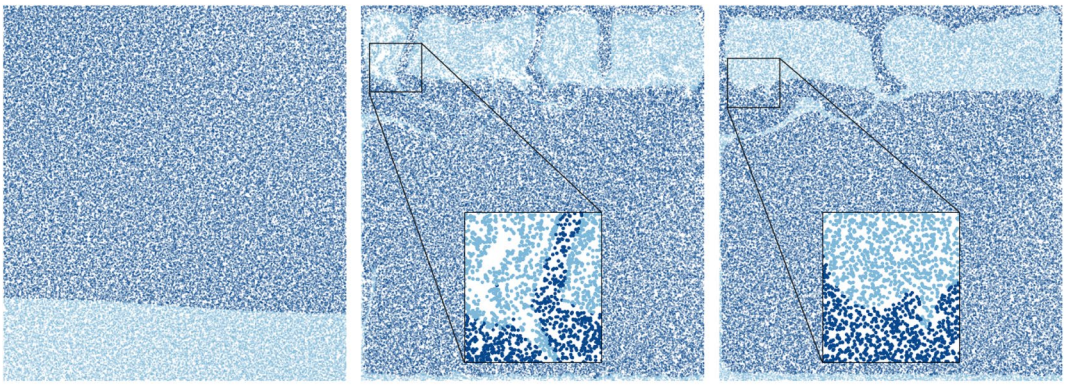


**Figure 12.** Tracer distribution in the Rayleigh-Taylor instability problem where $\eta_{\text{dense}} = 1$ and $\eta_{\text{light}} = 0.01$. Dark and light colors correspond to compositionally dense $\phi_p = 1$ and light $\phi_p = 0$, $p = 1, ..., N_p$, material, respectively. The mesh comprises $40 \times 40$ equally sized bisected quadrilaterals. Each triangle hosts 25 uniform randomly distributed tracers at its initial state giving rise to $N_p = 80{,}000$. (a) The initial condition Equation 55, (b) the TH solution at time $t \approx 634$ where there are large empty spaces between tracers and "bunching" close to the interface discontinuity, (c) the HDG solution at time $t \approx 634$ where the tracers have remained evenly distributed.

### 7.3.1. Problem Definition

The domain $\Omega$ is the rectangle with bottom left and top right corners positioned at (0, 0) and ($H = 1$, $L = 0.9142$), respectively. The chemical composition function is constructed such that

$$\phi(\mathbf{x},t) = \begin{cases} 0 & \forall x \in \text{light material}, \\ 1 & \forall x \in \text{dense material}. \end{cases} \tag{53}$$

the viscosity of the light and dense materials are the constant values $\eta_{\text{light}}$ and $\eta_{\text{dense}}$, respectively. We construct the global viscosity function from the composition function such that

$$\eta = \eta_{\text{light}} + \phi\left(\eta_{\text{dense}} - \eta_{\text{light}}\right). \tag{54}$$

the initial state of the composition field introduces a perturbation of the system from equilibrium

$$\phi(\mathbf{x},t=0) = \begin{cases} 0 & y < d_b + \dfrac{1}{50}\cos\left(\dfrac{\pi x}{L}\right), \\ 1 & \text{otherwise}. \end{cases} \tag{55}$$

here $d_b = 0.2$ is the thickness of the initial compositionally light (buoyant) material layer.

The velocity boundary conditions are set as follows:

1. No flow condition $u = (0,0)^{\mathsf{T}}$ on the top and bottom boundaries $y = 0$ and $y = H$
2. Free-slip condition $\mathbf{u}\cdot\mathbf{n} = 0$ and $\left((2\eta\,\underline{\underline{\varepsilon}}(\mathbf{u}) - p\underline{\underline{I}})\cdot\mathbf{n}\right)\cdot\tau = 0$ on the left ($x = 0$) and right ($x = L$) boundaries, respectively, where $\tau$ is a unit vector tangential to the boundary

The momentum source is prescribed such that $\mathbf{f} = \phi\hat{\mathbf{g}}$, where $\hat{\mathbf{g}}=(0,-1)^{\mathsf{T}}$ is the unit vector acting in the direction of gravity.

### 7.3.2. Computed Snapshots

The mesh of triangles is constructed from $160 \times 160$ equally spaced and bisected quadrilaterals. This mesh is then displaced to approximately align the facets with the discontinuity required by the initial condition

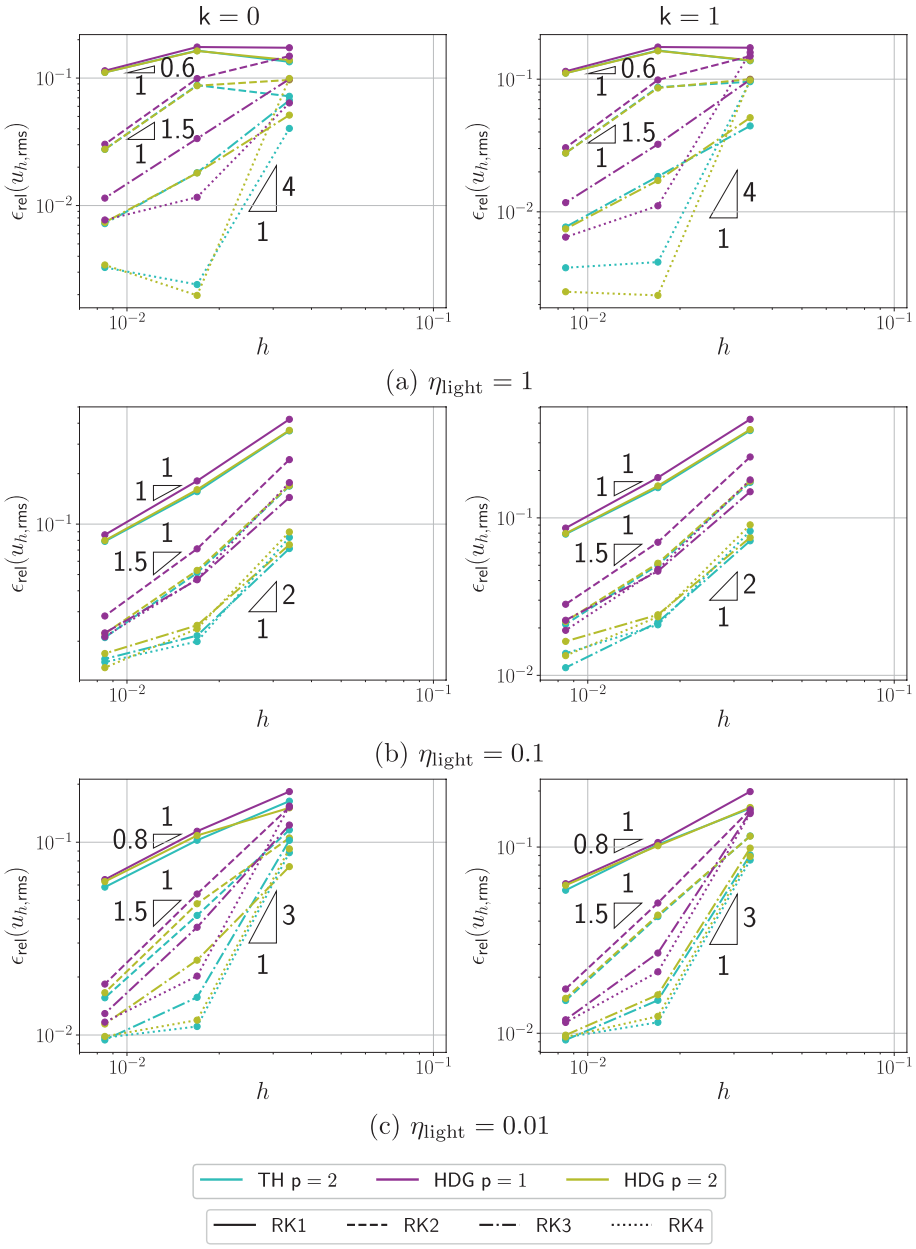


**Figure 13.** Measured relative error in $u_{\text{rms}}$ computed from the Rayleigh-Taylor instability benchmark. Here, we see we can at best achieve approximately second order error convergence rates. The expense of higher order RK methods ($\ell \geq 3$) and composition field function spaces $k \geq 1$ is not warranted unless intraelement discontinuities may be precisely resolved. PDE-constrained $l_2$ projection is used to generate the composition field representation of the tracer data. The left and right columns show cases where the composition field function belongs to the $k = 0$ and $k = 1$ DG FE spaces, respectively.

Equation 55. Each triangle is initially seeded with 25 tracers giving rise to a total of $N_p = 1{,}280{,}000$ tracers in the domain. The initial tracer data is assigned by pointwise interpolation of Equation 55, that is, $\Phi(t = 0) = \{\phi(\mathbf{x}_p(t = 0), t = 0)\}_{p=1}^{N_p}$.

In Figure 11, the Stokes system is discretized by the $p = 1$ HDG scheme. The composition field is constructed by exploiting PDE-constrained $l_2$ projection of the tracer data onto the $k \in \{0, 1\}$ DG FE space. The time integration of the particle advection is computed using the RK3 method, although this order of accuracy is superfluous given the second order Stokes and composition field approximation.

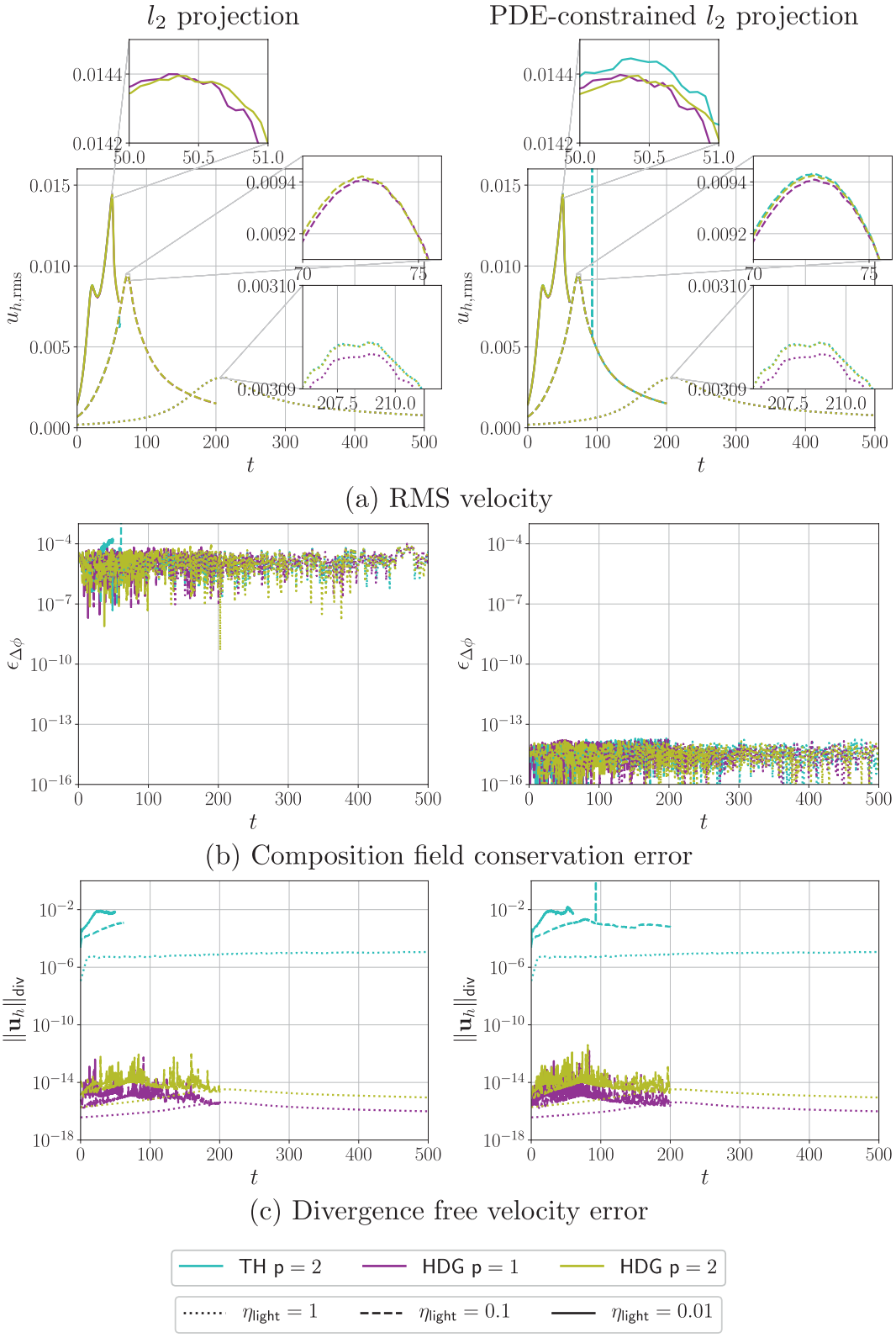


**Figure 14.** Rayleigh-Taylor instability benchmark data where $k = 0$. Using the TH and HDG methods combined with $l_2$ and PDE-constrained $l_2$ projection we compare the evolution of: (a) the root-mean-square velocity, (b) the mass conservation error and (c) the error in the divergence free velocity constraint. In all cases we see that in (b) PDE-constrained projection provides exact mass conservation (to machine precision). Furthermore in (c) the HDG FE method provides pointwise divergence free velocity field approximations. The TH method does not provide a robust solution framework prematurely ending our simulations (cf. Figure 12). See Section 7.3.5 for details.

In this first experiment, we examine the cases where $\eta_{\text{dense}} = 1$ and $\eta_{\text{light}} \in \{1, 0.1, 0.01\}$. We show the tracer distribution and projected composition field snapshots at time $t = 1{,}500$ in Figure 11 (cf. van Keken et al., 1997).

#### 7.3.3. The Impact of a Pointwise Divergence Free Field

A key message of this work is the benefit of pointwise divergence free velocity field approximations yielded from HDG FE solutions. The initial tracer configuration which interpolates the initial condition Equation 55 is shown in Figure 12. Also shown is the tracer distribution of the composition data computed from PDE-constrained $l_2$ projection method where the velocity field is approximated using either the TH or the HDG FE scheme at time $t \approx 634$ where $\eta_{\text{dense}} = 1$ and $\eta_{\text{light}} = 0.01$. This is the final time step after which the TH scheme fails due to the lack of sufficient tracer data in the mesh cells. This is evident from the large empty spaces in the tracer distribution using TH. We also see "bunching" of the tracers close to the interface discontinuity. Compare this with the pointwise divergence free velocity approximation of the HDG FE scheme, in which the tracers remain well distributed owing to this numerical property.

#### 7.3.4. Relative Error Convergence of Functionals

We examine three viscosity contrast scenarios where $\eta_{\text{dense}} = 1$ and $\eta_{\text{light}} \in \{1, 0.1, 0.01\}$. Based on our findings in Sections 7.1 and 7.2, we expect that sharp discontinuities in the composition field will have a negative impact on error convergence rates. The composition field appears in both the viscosity and momentum source terms in the Stokes system. Therefore, we pay close attention to the difference in convergence rates in the smooth and non-smooth viscosity cases, $\eta_{\text{light}} = 1$ and $\eta_{\text{light}} \in \{0.1, 0.01\}$, respectively.

The functional of interest is the root-mean-square velocity

$$u_{\text{rms}} = \sqrt{\frac{\int_\Omega \mathbf{u} \cdot \mathbf{u} d\mathbf{x}}{\int_\Omega d\mathbf{x}}}. \tag{56}$$

we measure $u_{\text{rms}}$ at final time $t_F$ where

$$t_F = \begin{cases} 100 & \eta_{\text{light}} = 1.0, \\ 50 & \eta_{\text{light}} = 0.1, \\ 10 & \eta_{\text{light}} = 0.01. \end{cases} \tag{57}$$

using the computed numerical data we evaluate the relative error changes between mesh refinement levels where $\mathcal{T}$ is a mesh consisting of $m \in \{20^2, 40^2, 80^2, 160^2\}$ bisected quadrilaterals, evenly dividing the domain into 2 $m$ triangles. Inside each triangle 25 tracers per cell are generated to form the initial condition Equation 55 such that there are 20,000, 80,000, 320,000 and 1,280,000 tracers in the mesh, respectively. In each case, we use the PDE-constrained $l_2$ projection to generate the composition field representation of the tracer data. The relative error in our computation of $u_{\text{rms}}$ is measured as follows

$$\epsilon_{\text{rel}}(u_{h,\text{rms}}) = \left| \frac{u_{h,\text{rms}} - u_{2h,\text{rms}}}{u_{h,\text{rms}}} \right|, \tag{58}$$

where $u_{h,\text{rms}}$ and $u_{2h,\text{rms}}$ correspond to fine and coarse approximations of $u_{\text{rms}}$ on meshes with cell diameter $h$ and $2h$, respectively, and $h = \min_{\kappa \in \mathcal{T}} h_\kappa$.

The computed root-mean-square velocity functional values with mesh (and implicitly time step by the CFL criterion) refinement are shown in Figure S7. The relative error is shown in Figure 13.

The convergence rates of the relative changes in error are suboptimal as expected from the inability to resolve intraelement discontinuities. The choice of $\ell \geq 2$ in the RK$\ell$ method plays only a small role in the approximation error. Although the RK4 method may appear to yield high order rates of convergence, in each case this is only featured in the $m = 80$ and $m = 40$ comparison. Examining the rates at $m = 160$ the

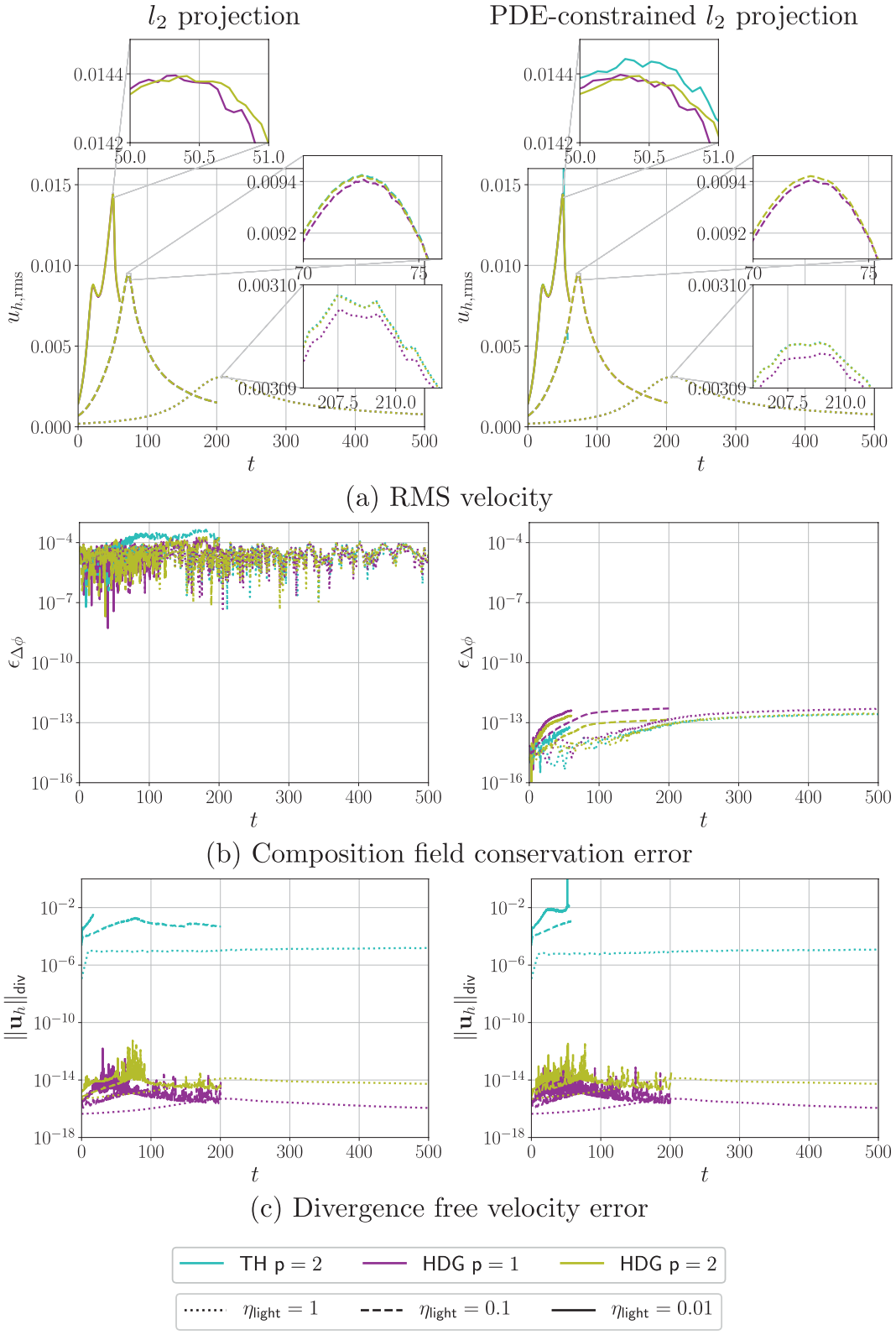


**Figure 15.** Rayleigh-Taylor instability benchmark data where $k = 1$. As in Figure 14, we compare the evolution of: (a) the root-mean-square velocity, (b) the mass conservation error and (c) the error in the divergence free velocity constraint in the TH and HDG methods. Again, we see that in (b) and (c) that PDE-constrained projection provides exact mass conservation and the HDG FE method provides pointwise divergence free velocity field approximations to machine precision, respectively. See Section 7.3.5 for details.

approximation error benefit of RK4 over its lower order counterparts is small considering its computational cost (three solutions of the Stokes equations).

Although the TH and HDG methods compare favorably at equivalent approximation order ($p = 2$), one must bear in mind that over long time periods the TH results will be severely affected by the lack of a pointwise divergence free field approximation (see Section 7.3.3 and Figure 12 for example). Evidently the higher order approximation of the composition field, $k = 1$, offers no obvious advantage over the $k = 0$ field representation. Although this is to be expected for the field representation of a discontinuity as examined in Sections 7.1 and 7.2.

### 7.3.5. Comparison of Stokes Finite Element Formulations

Here, we provide a reference of the various FE methods for the Stokes system when used in the Rayleigh-Taylor instability benchmark. The mesh is generated as described in the previous section. With each FE discretization scheme we are interested in the error incurred by the divergence free velocity approximation $\|\mathbf{u}_h\|_{\text{div}}$ as well as the mass conservation error through time $\varepsilon_{\Delta\phi}$. We employ the RK3 method for time integration in all experiments in this section.

The functional of interest is again the root-mean-square velocity $u_{\text{rms}}$ computed from the Stokes FE approximation equation Equation 56. We set $\eta_{\text{dense}} = 1$ and prescribe final time $t_F$ and $\eta_{\text{light}}$ according to

$$t_F = \begin{cases} 500 & \eta_{\text{light}} = 1, \\ 200 & \eta_{\text{light}} = 0.1, \\ 60 & \eta_{\text{light}} = 0.01. \end{cases} \tag{59}$$

these time intervals capture the period over which the bulk of the light material rises to the top of the domain showing large changes in $u_{\text{rms}}$. To see measured computations of $u_{\text{rms}}$ over larger time intervals we refer to van Keken et al. (1997) and Maljaars et al. (2021).

We show results computed using $p = 2$ TH and $p \in \{1, 2\}$ HDG FE schemes for the Stokes system. The cases where $k = 0$ and $k = 1$ are shown in Figures 14 and 15, respectively. Further results using other common FE schemes for the Stokes system (particularly those in Table 2) are shown in Section S5.4.

Some key observations are as follows:

1. Of the methods demonstrated, solely the HDG scheme offers pointwise divergence free velocity field approximations. One should expect that using a non-pointwise divergence free discretization would lead to failure over long time simulations as exhibited in Section 7.3.3. In fact, we find that even over short time simulations the TH scheme for the Stokes system is insufficient as a robust solver for the Rayleigh-Taylor benchmark with tracers. Our implementation fails to complete the TH simulations successfully due to insufficient tracer data in mesh cells, even in the isoviscous $\eta_{\text{light}} = 1$ case
2. In the TH scheme, larger viscosity contrasts incur larger $\|\mathbf{u}\|_{\text{div}}$ error. This should be expected as rapid changes in viscosity will be difficult to resolve for a high order conforming FE method
3. A comparison of the $l_2$ and PDE-constrained $l_2$ projection methods appear to yield only subtle differences in the evaluation of $u_{\text{rms}}$. However, using the non-conserving $l_2$ projection method breaks the physical law demanded by the conservation model
4. Projection of the tracer data into the $k = 1$ space appears to offer little to no benefit over the $k = 0$ space. This is expected as discussed in Section 7.1 for non-smooth composition fields

## 8. Conclusion

Using the HDG Stokes and PDE-constrained $l_2$ tracer projection numerical scheme, we have the following attractive properties:

1. Exact mass conservation of the composition field
2. A pointwise divergence free velocity field which mitigates tracer dispersion
3. Optimal convergence rates of the approximation error as measured in the $L_2$ norm provided regularity is satisfied

## Data Availability Statement

The code developed to run all the numerical experiments in this work and a list of its dependencies is available in the public repository Sime (2020) (see Section 6 for details).

**Acknowledgments**
The authors thank Sia Ghelichkhan and two anonymous reviewers for constructive reviews that helped improve the manuscript. This research study was partly supported by NSF-EAR-CSE-DI grant 1664642. N. Sime gratefully acknowledges the support of Carnegie Institution for Science President's Fellowship.

## References

Alnæs, M. S., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., et al. (2015). The FEniCS Project Version 1.5. *Archive of Numerical Software*, *3*(100), 9–23.

Alnæs, M. S., Logg, A., Ølgaard, K. B., Rognes, M. E., & Wells, G. N. (2014). Unified form language: A domain-specific language for weak formulations of partial differential equations. *ACM Transactions on Mathematical Software*, *40*(2), 9:1–9:37.

Amestoy, P. R., Duff, I. S., & L'Excellent, J.-Y. (2000). Multifrontal parallel distributed symmetric and unsymmetric solvers. *Computer Methods in Applied Mechanics and Engineering*, *184*, 501–520.

Balay, S., Abhyankar, S., Adams, M. F., Brown, J., Brune, P., Buschelman, K., et al. (2019). *PETSc users manual (Technical Report No. ANL-95/11 - Revision 3.12)*. Argonne National Laboratory. https://www.mcs.anl.gov/petsc

Balay, S., Abhyankar, S., Adams, M. F., Brown, J., Brune, P., Buschelman, K., et al. (2019). *PETSc Web page*. https://www.mcs.anl.gov/petsc

Ballmer, M. D., Schumacher, L., Lekic, V., Thomas, C., & Ito, G. (2016). Compositional layering within the large low shear-wave velocity provinces in the lower mantle. *Geochemistry, Geophysics, Geosystems*, *17*, 5056–5077. https://doi.org/10.1002/2016GC006605

Boffi, D., Brezzi, F., & Fortin, M. (2013). Mixed finite element methods and applications (Vol. 44). New York, NY: Springer Series in Computational Mathematics.

Brandenburg, J. P., Hauri, E. H., van Keken, P. E., & Ballentine, C. J. (2008). A multiple-system study of the geochemical evolution of the mantle with force-balanced plates and thermochemical effects. *Earth and Planetary Science Letters*, *276*, 1–13.

Brenner, S. C., & Scott, L. R. (2010). *The mathematical theory of finite element methods* (3rd ed.). New York, NY: Springer.

Bull, A., McNamara, A., & Ritsema, J. (2009). Synthetic tomography of plume clusters and thermochemical piles. *Earth and Planetary Science Letters*, *278*, 152–162.

Butcher, J. C. (1987). *The numerical analysis of ordinary differential equations: Runge-Kutta and general linear methods*. Chichester: John Wiley & Sons.

Christensen, U. R., & Hofmann, A. W. (1994). Segregation of subducted oceanic crust in the convecting mantle. *Journal of Geophysical Research*, *99*, 19867–19884. https://doi.org/10.1029/93JB03403

Christensen, U. R., & Yuen, D. A. (1984). The interaction of a subducting lithospheric slab with a chemical or phase boundary. *Journal of Geophysical Research*, 4389–4402. https://doi.org/10.1029/JB089iB06p04389

Cockburn, B., & Gopalakrishnan, J. (2009). The derivation of hybridizable discontinuous Galerkin methods for Stokes flow. *SIAM Journal on Numerical Analysis*, *47*(2), 1092–1125.

Cockburn, B., Gopalakrishnan, J., Nguyen, N., Peraire, J., & Sayas, F.-J. (2011). Analysis of HDG methods for stokes flow. *Mathematics of Computation*, *80*(274), 723–760.

Cockburn, B., Kanschat, G., & Schötzau, D. (2007). A note on discontinuous Galerkin divergence-free solutions of the Navier-Stokes equations. *Journal of Scientific Computing*, *31*(1–2), 61–73.

Cockburn, B., Nguyen, N. C., & Peraire, J. (2010). A comparison of HDG methods for Stokes flow. *Journal of Scientific Computing*, *45*(1–3), 215–237.

Cockburn, B., & Sayas, F.-J. (2014). Divergence-conforming HDG methods for Stokes flows. *Mathematics of Computation*, *83*(288), 1571–1598.

Davies, D., Davies, J., Hassan, O., Morgan, K., & Nithiarasu, P. (2007). Investigations into the applicability of adaptive finite element methods to two-dimensional infinite Prandtl number thermal and thermochemical convection. *Geochemistry, Geophysics, Geosystems*, *8*, Q05010. https://doi.org/10.1029/2006GC001470

Evans, J. A., & Hughes, T. J. R. (2013). Isogeometric divergence-conforming B-splines for the steady Navier-Stokes equations. *Mathematical Models and Methods in Applied Sciences*, *23*(08), 1421–1478.

Gaspero, L. D. (2020). *QuadProg++*. https://github.com/liuq/QuadProgpp

Gassmöller, R., Lokavarapu, H., Heien, E., Puckett, E. G., & Bangerth, W. (2018). Flexible and scalable particle-in-cell methods with adaptive mesh refinement for geodynamic computations. *Geochemistry, Geophysics, Geosystems*, *19*, 3596–3604. https://doi.org/10.1029/2018GC007508

Gerya, T. V., & Yuen, D. A. (2003). 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.

Goldfarb, D., & Idnani, A. (1983). A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming*, *27*(1), 1–33.

Guyan, R. J. (1965). Reduction of stiffness and mass matrices. *AIAA Journal*, *3*(2), 380.

Guzmán, J., & Neilan, M. (2014). Conforming and divergence-free Stokes elements on general triangular meshes. *Mathematics of Computation*, *83*(285), 15–36.

Hale, J. S. (2020). *FEniCS in Docker*. https://fenics.readthedocs.io/projects/containers/en/latest/

Hale, J. S., Li, L., Richardson, C. N., & Wells, G. N. (2017). Containers for portable, productive, and performant scientific computing. *Computing in Science & Engineering*, *19*(6), 40–50.

Hansen, U., & Yuen, D. (2000). Extended-Boussinesq thermal-chemical convection with moving heat sources and variable viscosity. *Earth and Planetary Science Letters*, *176*, 400–411.

Haugland, S. M., Ritsema, J., van Keken, P. E., & Nissen-Meyer, T. (2018). Analysis of PKP scattering using mantle mixing simulations and axisymmetric 3D waveforms. *Physics of the Earth and Planetary Interiors*, *276*, 226–233.

He, Y., Puckett, E. G., & Billen, M. I. (2017). A discontinuous Galerkin method with a bound preserving limiter for the advection of non-diffusive fields in solid Earth geodynamics. *Physics of the Earth and Planetary Interiors*, *263*, 23–37.

Hillebrand, B., Theuilot, C., Geenen, T., van den Berg, A., & Spakman, W. (2014). Using the level set method in geodynamical modeling of multi-material flows and Earth's free surface. *Solid Earth*, *5*, 1087–1098.

Houston, P., & Sime, N. (2018). Automatic symbolic computation for discontinuous Galerkin finite element methods. *SIAM Journal on Scientific Computing*, *40*(3), C327–C357.

Houston, P., & Wihler, T. P. (2012). Discontinuous Galerkin methods for problems with Dirac delta source. *ESAIM: Mathematical Modelling and Numerical Analysis*, *46*(6), 1467–1483.
Jarvis, G. T., & Mckenzie, D. P. (1980). Convection in a compressible fluid with infinite prandtl number. *Journal of Fluid Mechanics*, *96*(3), 515–583.
Jenny, P., Pope, S., Muradoglu, M., & Caughey, D. (2001). A hybrid algorithm for the joint PDF equation of turbulent reactive flows. *Journal of Computational Physics*, *166*(2), 218–252.
Jones, R. E., van Keken, P. E., Hauri, E. H., Tucker, J. M., Vervoort, J., & Ballentine, C. J. (2019). Origins of the terrestrial Hf-Nd mantle array: Evidence from a combined geodynamical-geochemical approach. *Earth and Planetary Science Letters*, *518*, 26–39.
Jones, T. D., Maguire, R. R., van Keken, P. E., Ritsema, J., & Koelemeijer, P. (2020). Subducted oceanic crust as the origin of seismically slow lower-mantle structures. *Progress in Earth and Planetary Science*, *7*, 17.
Kellogg, L. H. (1992). Mixing in the mantle. *Annual Review of Earth and Planetary Sciences*, *20*, 365–388.
Kellogg, L. H., & King, S. D. (1993). Effect of mantle plumes on the growth of D" by reaction between the core and mantle. *Geophysical Research Letters*, *20*, 379–382. https://doi.org/10.1029/93GL00045
Labeur, R. J., & Wells, G. N. (2007). A Galerkin interface stabilization method for the advection–diffusion and incompressible Navier–Stokes equations. *Computer Methods in Applied Mechanics and Engineering*, *196*(49), 4985–5000.
Labeur, R. J., & Wells, G. N. (2012). Energy stable and momentum conserving hybrid finite element method for the incompressible Navier Stokes equations. *SIAM Journal on Scientific Computing*, *34*(2), A889–A913.
Labrosse, S., Hernlund, J., & Coltice, N. (2007). A crystallizing dense magma ocean at the base of the Earth's mantle. *Nature*, *450*, 866–869.
Leng, W., & Zhong, S. (2011). Implementation and application of adaptive mesh refinement for thermochemical mantle convection studies. *Geochemistry, Geophysics, Geosystems*, *12*, Q04006. https://doi.org/10.1029/2010GC003425
LeVeque, R. J. (2007). Finite difference methods for ordinary and partial differential equations: Steady-state and time-dependent problems (Vol. 98). Philadelphia, PA: SIAM.
Lew, A., Neff, P., Sulsky, D., & Ortiz, M. (2004). Optimal BV estimates for a discontinuous Galerkin method for linear elasticity. *Applied Mathematics Research eXpress*, *2004*(3), 73–106.
Li, M., & McNamara, A. K. (2018). The influence of deep mantle compositional heterogeneity on Earth's thermal evolution. *Earth and Planetary Science Letters*, *500*, 86–96.
Lin, S.-C., & van Keken, P. E. (2006). Dynamics of thermochemical plumes: 1. Plume formation and entrainment of a dense layer. *Geochemistry, Geophysics, Geosystems*, *7*, Q02006. https://doi.org/10.1029/2005GC001071
Maljaars, J. M., Labeur, R. J., & Mller, M. (2018). A hybridized discontinuous Galerkin framework for high-order particle mesh operator splitting of the incompressible Navier–Stokes equations. *Journal of Computational Physics*, *358*, 150–172.
Maljaars, J. M., Labeur, R. J., Trask, N., & Sulsky, D. (2019). Conservative, high-order particle mesh scheme with applications to advection-dominated flows. *Computer Methods in Applied Mechanics and Engineering*, *348*, 443–465.
Maljaars, J. M., Labeur, R. J., Trask, N., & Sulsky, D. (2020). Optimization based particle-mesh algorithm for high-order and conservative scalar transport. In E. van Brummelen, A. Corsini, S. Perotto, & G. Rozza (Eds.), *Lecture notes in computational science and engineering* (Vol. *132*, pp. 265–275). Cham: Springer.
Maljaars, J. M., Richardson, C. N., & Sime, N. (2021). LEoPart: A particle library for FEniCS. *Computers & Mathematics with Applications*, *81*, 289–315.
McDermott, R., & Pope, S. (2008). The parabolic edge reconstruction method (PERM) for Lagrangian particle advection. *Journal of Computational Physics*, *227*(11), 5447–5491.
Moresi, L., Dufour, F., & Mhlhaus, H.-B. (2003). A Lagrangian integration point finite element method for large deformation modeling of viscoelastic geomaterials. *Journal of Computational Physics*, *184*(2), 476–497.
Morgan, H., & Scott, L. R. (2018). Toward a unified finite element method for the Stokes equations. *SIAM Journal on Scientific Computing*, *40*(1), A130–A141.
Nakagawa, T., & Tackley, P. J. (2005). Deep mantle heat flow and thermal evolution of the Earth's core in thermochemical multiphase models of mantle convection. *Geochemistry, Geophysics, Geosystems*, *6*, Q08003. https://doi.org/10.1029/2005GC000967
Nguyen, N. C., Peraire, J., & Cockburn, B. (2010). A hybridizable discontinuous Galerkin method for Stokes flow. *Computer Methods in Applied Mechanics and Engineering*, *199*(9–12), 582–597.
O'Neill, C., Moresi, L., Müller, D., Albert, R., & Dufour, F. (2006). Ellipsis 3D: A particle-in-cell finite element hybrid code for modeling mantle convection and lithospheric deformation. *Computers & Geosciences*, *32*, 1769–1779.
Pilliod, J. E., & Puckett, E. G. (2004). Second-order accurate volume-of-fluid algorithms for tracking material interfaces. *Journal of Computational Physics*, *199*(2), 465–502.
Prudhomme, S., Pascal, F., Oden, J. T., & Romkes, A. (2000). *Review of a priori error estimation for discontinuous Galerkin methods (Technical Report)*. Austin, TX: TICAM, University of Texas at Austin.
Rhebergen, S., & Wells, G. N. (2018). A hybridizable discontinuous Galerkin method for the Navier–Stokes equations with pointwise divergence-free velocity field. *Journal of Scientific Computing*, *76*(3), 1484–1501.
Rhebergen, S., & Wells, G. N. (2018). Preconditioning of a hybridized discontinuous Galerkin finite element method for the Stokes equations. *Journal of Scientific Computing*, *77*(3), 1936–1952.
Rhebergen, S., & Wells, G. N. (2020). An embedded hybridized discontinuous Galerkin finite element method for the Stokes equations. *Computer Methods in Applied Mechanics and Engineering*, *358*, 112619.
Richardson, C. N., Sime, N., & Wells, G. N. (2019). Scalable computation of thermomechanical turbomachinery problems. *Finite Elements in Analysis and Design*, *155*, 32–42.
Robey, J. M., & Puckett, E. G. (2019). Implementation of a volume-of-fluid method in a finite element code with applications to thermochemical convection in a density stratified fluid in the earth's mantle. *Computers & Fluids*, *190*, 217–253.
Samuel, H., & Evonuk, M. (2010). Modeling advection in geophysical flows with particle level sets. *Geochemistry, Geophysics, Geosystems*, *11*, Q08020. https://doi.org/10.1029/2010GC003081
Samuel, H., & Farnetani, C. G. (2003). Thermochemical convection and helium concentrations in mantle plumes. *Earth and Planetary Science Letters*, *207*, 39–56.
Schmeling, H. (1987). On the relation between initial conditions and late stages of Rayleigh-Taylor instabilities. *Tectonophysics*, *133*, 65–80.
Scott, L. R. (1973). Finite element convergence for singular data. *Numerische Mathematik*, *21*, 317–327.
Scott, L. R., & Vogelius, M. (1985). Norm estimates for a maximal right inverse of the divergence operator in spaces of piecewise polynomials. *ESAIM: Mathematical Modelling and Numerical Analysis*, *19*(1), 111–143.
Sime, N. (2020). *GeoPart: Particles in geodynamics supporting code*. https://bitbucket.org/nate-sime/geopart

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. https://doi.org/10.1029/2009JB006916

Tackley, P. J. (2015). Mantle geochemical geodynamics. In G. Schubert & D. Bercovici (Eds.), *Treatise on geophysics, "Mantle Dynamics"* (2nd ed.), (Vol. *7*, pp. 521–585). Amsterdam: Elsevier.

Tackley, P. J., & King, S. D. (2003). Testing the tracer ratio method for modeling active compositional fields in mantle convection simulations. *Geochemistry, Geophysics, Geosystems*, *4*(4). https://doi.org/10.1029/2001GC000214

Taylor, C., & Hood, P. (1973). A numerical solution of the Navier-Stokes equations using the finite element technique. *Computers & Fluids*, *1*(1), 73–100.

Thielmann, M., May, D., & Kaus, B. (2014). Discretization errors in the hybrid finite element particle-in-cell method. *Pure and Applied Geophysics*, *171*, 2165–2184.

Trim, S. J., Lowman, J. P., & Butler, S. L. (2020). Improving mass conservation with the tracer ratio method: Application to thermochemical mantle flows. *Geochemistry, Geophysics, Geosystems*, *21*(2). https://doi.org/10.1029/2019GC008799

Tucker, J. M., van Keken, P. E., Jones, R. E., & Ballentine, C. J. (2020). A role for subducted oceanic crust in generating the depleted morb mantle. *Geochemistry, Geophysics, Geosystems*, *21*, ee2020GC009148. https://doi.org/10.1029/2020GC009148

Turner, J. (1974). Double-diffusive phenomena. *Annual Review of Fluid Mechanics*, *6*, 37–54.

van Keken, P. E., Ballentine, C., & Hauri, E. H. (2014). Convective mixing in the Earth's mantle. In H. Holland, K. Turekian & R. Carlson (Eds.), *Treatise on geochemistry, "The mantle and core"* (2nd ed.), (Vols. *3*, pp. 509–525). Amsterdam: Elsevier.

van Keken, P. E., Hauri, E. H., & Ballentine, C. J. (2002). Mantle mixing: The generation, preservation, and destruction of chemical heterogeneity. *Annual Review of Earth and Planetary Sciences*, *30*, 493–525.

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*, 22477–22495. https://doi.org/10.1029/97JB01353

Vynnytska, L., Rognes, M. E., & Clark, S. R. (2013). Benchmarking FEniCS for mantle convection simulations. *Computers & Geosciences*, *50*, 95–105.

Wang, H., Agrusta, R., & van Hunen, J. (2015). Advantages of a conservative velocity interpolation (cvi) scheme for particle-in-cell methods with application in geodynamic modeling. *Geochemistry, Geophysics, Geosystems*, *16*(6), 2015–2023. https://doi.org/10.1002/2015GC005824

Xie, S., & Tackley, P. J. (2004). Evolution of helium and argon isotopes in a convecting mantle. *Physics of the Earth and Planetary Interiors*, *146*, 417–439.

Zalesak, S. T. (1979). Fully multidimensional flux-corrected transport algorithms for fluids. *Journal of Computational Physics*, *31*(3), 335–362.

Zhong, S. (2006). Constraints on thermochemical convection of the mantle from plume heat flux, plume excess temperature, and upper mantle temperature. *Journal of Geophysical Research*, *111*, B04409. https://doi.org/10.1029/2005JB003972