



Geochemistry, Geophysics, Geosystems

RESEARCH ARTICLE

10.1002/2015GC006228

Key Points:

- Incompressible Stokes with viscoplastic rheologies is highly nonlinear and requires accurate nonlinear solvers
- Pressure-dependent rheologies (e.g., Drucker-Prager) can be ill-posed causing Newton solvers to fail
- Developing physical, tractable, and solvable geodynamic models of crustal failure remains an open challenge

Supporting Information:

- Supporting Information S1
- Figure S1
- Figure S2

Correspondence to:

M. Spiegelman,
mspiegel@ideo.columbia.edu

Citation:

Spiegelman, M., D. A. May, and C. R. Wilson (2016), On the solvability of incompressible Stokes with viscoplastic rheologies in geodynamics, *Geochem. Geophys. Geosyst.*, 17, 2213–2238, doi:10.1002/2015GC006228.

Received 14 DEC 2015

Accepted 3 MAY 2016

Accepted article online 9 MAY 2016

Published online 21 JUN 2016

On the solvability of incompressible Stokes with viscoplastic rheologies in geodynamics

Marc Spiegelman^{1,2}, Dave A. May³, and Cian R. Wilson¹

¹Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, USA, ²Department of Applied Physics and Applied Math, Columbia University, New York, New York, USA, ³Department of Earth Sciences, Institute of Geophysics, ETH Zürich, Zürich, Switzerland

Abstract Plasticity/failure is an essential ingredient in geodynamics models as earth materials cannot sustain unbounded stresses. However, many questions remain as to appropriate models of plasticity as well as effective solvers for these strongly nonlinear systems. Here we present some simplified model problems designed to elucidate many of the issues involved for the description and solution of *viscoplastic* problems as currently used in geodynamic modeling. We consider compression and extension of a viscoplastic layer overlying an isoviscous layer and introduce a single plastic yield criterion which includes the most commonly used viscoplasticity models: von Mises, depth-dependent von Mises, and Drucker-Prager. We show that for all rheologies considered, successive substitution schemes (aka Picard iteration) often stall at large values of the nonlinear residual, producing spurious solutions. However, combined Picard-Newton schemes can be effective for rheologies that are independent of the dynamic pressure. Difficulties arise when solving incompressible Stokes problems for rheologies that depend on the dynamic pressure such as Drucker-Prager viscoplasticity. Analysis suggests that incompressible Stokes can become ill-posed when the dependence of the deviatoric stress tensor on dynamic pressure (i.e., $|\partial\tau/\partial p'|$) becomes large. We demonstrate empirically that, in these cases, Newton solvers can fail by introducing spurious shear bands and discuss the consequence of interpreting the results of nonconverged computations. Even for problems where solvers converge, Drucker-Prager viscoplasticity can produce dynamic pressures that deviate significantly from lithostatic and both the velocity and pressure fields should be evaluated to determine whether solutions are geologically reasonable.

1. Introduction

The dynamics of coupled brittle-ductile systems on geological time scales is central to our understanding of a range of important tectonic processes on Earth, including mountain belt formation, subduction initiation, and feedbacks between surface processes and deformation in the mantle-lithosphere.

Our understanding of the rheology of terrestrial rocks, has, to a large extent, been shaped by laboratory experiments [Byerlee, 1978; Ranalli, 1995; Scholz, 2002]. The first-order temperature, pressure, and strain-rate dependence of the ductile flow laws associated with olivine, quartzite, and other Earth materials is well established [e.g., Karato and Wu, 1993]. Consequently, Arrhenius-type flow laws based on experimentally determined parameters are heavily used by the computational geodynamics community for both mantle-scale and regional lithospheric-scale modeling. At low temperatures however, these rheologies predict extremely high viscosities, which at geologically reasonable strain rates will generate commensurately high stresses.

Of course, rocks do not possess an infinite yield strength. Numerous physical mechanisms exist to limit the strength of rocks, including low-temperature plasticity [Evans and Goetze, 1979; Kameyama et al., 1999; Karato, 2011] and brittle fracture [Anderson, 1905; Byerlee, 1978]. However, a key challenge remains to develop appropriate constitutive relationships which describe the “brittle behavior” of rocks that can be upscaled and applied to describe the behavior of macroscale (km-scale) faults and shear zones. For the purpose of regional and global-scale computational geodynamic simulations, the latter is of paramount importance, and much effort within the geodynamics community has been devoted to develop such continuum-based models. While many yield criteria exist within the broader solid mechanics community (to describe the behavior of granular material, soils, etc.), the most commonly used yield criteria within geodynamics are

stress-limiters based on von Mises [von Mises, 1913], or the Drucker-Prager [Drucker and Prager, 1952; Vermeer and de Borst, 1984] model.

When deciding on an appropriate plasticity model, two fundamental questions arise: (i) Does the chosen constitutive law produce brittle structures (in some metric) which reflect reality, e.g., as observed in the field and or in the laboratory?; (ii) Is the chosen constitutive behavior computationally tractable/solvable? Attempts have been made to validate Drucker-Prager based models using analog modeling studies involving sandbox experiments which mimic the compression of brittle wedges [Buiter et al., 2006]. More attention has been given to trying to understand the geometry of shear zones formed by Drucker-Prager models (e.g., shear band angle, shear band thickness, symmetric versus asymmetric structures) and how these factors relate to boundary conditions and rheological properties of the underlying brittle material [Lemiale et al., 2008; Kaus, 2010; Buiter, 2012; Huismans and Beaumont, 2003; Huismans et al., 2005; Huismans and Beaumont, 2008]. To date, few studies have thoroughly explored the issue of “solvability” raised above. This issue is important as plasticity models introduce strong nonlinearities into the underlying problems. In general, developing robust and efficient nonlinear solvers is far more delicate, than if the problem is linear (or weakly nonlinear). The first details of a potential solvability issue connected with Drucker-Prager viscoplasticity appeared in Lemiale et al. [2008], where it was evident that the Drucker-Prager rheology was difficult to converge unless strain-softening was introduced. In Popov and Sobolev [2008], it was clearly illustrated that the shear band thickness is dependent on the accuracy of the nonlinear solve.

The purpose of this paper is to clarify and illustrate key computational issues arising from the most commonly used formulations of viscoplasticity used in geodynamic models. Through a series of simplified model problems, similar to those in the existing literature, we first demonstrate the importance of monitoring the full nonlinear convergence of these problems and introduce some useful algorithms for accurate solution of incompressible Stokes flow with viscoplastic rheologies. We then use these algorithms to compare the behavior of three commonly used plasticity models, von Mises, depth-dependent von Mises, and Drucker-Prager, in well resolved and converged solutions. In particular, we demonstrate computational problems that arise from combining dynamic pressure-dependent rheologies such as Drucker-Prager viscoplasticity with incompressible Stokes where the dynamic pressure enforces the incompressibility constraint. We show that when this pressure is fed back into the rheology, the problems become highly nonlinear, can develop potentially nonphysical solutions or have no solutions at all as the problem is ill-posed. Moreover, we discuss the structure of nonconverged solutions to provide guidance in interpreting existing and future plasticity formulations which are relevant to describing the brittle behavior of rocks on Earth.

2. Governing Equations

All of the models presented solve some form of Stokes equations for the viscoplastic deformation of an incompressible solid under a gravitational body force in a domain Ω :

$$-\nabla \cdot \boldsymbol{\tau} + \nabla p = \rho \mathbf{g}, \quad (1)$$

$$-\nabla \cdot \mathbf{v} = 0, \quad (2)$$

where \mathbf{v} is the velocity, p is the total pressure, ρ is the material density, and \mathbf{g} is the acceleration due to gravity and $\boldsymbol{\tau}$ is the deviatoric stress tensor. For the isotropic media we consider, the deviatoric stress is

$$\boldsymbol{\tau} = 2\eta(\mathbf{v}, p)\dot{\boldsymbol{\epsilon}}, \quad (3)$$

where η is the shear viscosity. We assume the shear viscosity can include multiple deformation mechanisms, as well as plastic limiters and thus in general, will be a nonlinear function of the velocity and pressure. In general, the deviatoric strain-rate tensor is given by

$$\dot{\boldsymbol{\epsilon}} = \frac{1}{2} [\nabla \mathbf{v} + \nabla \mathbf{v}^T] - \frac{1}{3} \nabla \cdot \mathbf{v} \mathbf{I}. \quad (4)$$

although the last term vanishes in the case of incompressible flows.

The system defined by equations (1) and (2) is closed with boundary conditions on velocity (Dirichlet) or stress (homogenous Neumann)

$$\mathbf{v} = \bar{\mathbf{v}} \quad \text{for } \mathbf{x} \in \Gamma_D, \quad (5)$$

$$\boldsymbol{\sigma} \cdot \mathbf{n} = \mathbf{0} \quad \text{for } \mathbf{x} \in \Gamma_N, \quad (6)$$

where $\boldsymbol{\sigma} = \boldsymbol{\tau} - p\mathbf{I}$ is the total stress, \mathbf{n} is the outward pointing normal to the domain boundary $\partial\Omega$. Γ_D, Γ_N denote the regions along the domain boundary where the Dirichlet and Neumann boundary conditions are applied, respectively, subject to the restrictions that $\Gamma_D \cap \Gamma_N = \emptyset$ and $\Gamma_D \cup \Gamma_N = \partial\Omega$.

2.1. Plasticity

To close equations (1) and (2) requires a constitutive relation for η which can be a function of \mathbf{v}, p (as well as temperature, composition etc.). Classical ductile flow laws for Earth materials such as diffusion or dislocation creep often have an Arrhenius temperature dependence which predicts extremely high strength materials at low temperatures. When such flow laws are used in geodynamic models, the strong materials can lead to unrealistic stresses if forced at geological strain rates. Alternatively, if the available stresses are bounded, e.g., by buoyancy forces, the very strong materials can lead to negligible deformation such as stagnant lid convection [Moresi and Solomatov, 1995; Solomatov and Moresi, 1996]. For this reason, it is important to include weakening mechanisms, or material failure models to limit stresses.

Scalar viscoplasticity is one form of weakening mechanism that is relatively easy to implement. The basic assumption is that we can find an effective plastic viscosity η_p such that the deviatoric stress tensor

$$\boldsymbol{\tau} = 2\eta_p \dot{\boldsymbol{\epsilon}} \quad (7)$$

is bounded by some yield stress. In general, η_p is a fourth-order tensor relating stress to strain-rate, however, in geodynamics, it is generally assumed that η_p is isotropic and scalar (i.e., the strain-rate tensor and deviatoric stress tensor share a common coordinate system). In this case, we can compute the second invariant of both sides to define

$$\eta_p = \frac{\tau_{II}}{2\dot{\epsilon}_{II}}, \quad (8)$$

where $\dot{\epsilon}_{II}$ and τ_{II} are the square root of the second invariant of the deviatoric strain-rate and deviatoric stress tensor, respectively;

$$\dot{\epsilon}_{II} = \sqrt{\frac{1}{2} \dot{\boldsymbol{\epsilon}} : \dot{\boldsymbol{\epsilon}}}, \quad \tau_{II} = \sqrt{\frac{1}{2} \boldsymbol{\tau} : \boldsymbol{\tau}}. \quad (9)$$

Furthermore, if we assume that the second invariant of the stress tensor is bounded by some yield criterion i.e., $\tau_{II} = Y$, then we find

$$\eta_p = \frac{Y}{2\dot{\epsilon}_{II}} \quad (10)$$

2.1.1. Viscoplasticity in Geodynamics

The previous description of plasticity is somewhat heuristic (but leads to most of the simple plasticity models used in geodynamics). These relations can alternatively be derived using the plasticity formalism developed within the engineering literature. To illustrate the relationship with classical plasticity, we begin by assuming that the total strain-rate $\dot{\boldsymbol{\epsilon}}$ can be decomposed into a purely viscous and purely plastic component:

$$\dot{\boldsymbol{\epsilon}} = \dot{\boldsymbol{\epsilon}}_v + \dot{\boldsymbol{\epsilon}}_p. \quad (11)$$

The plastic strain rate $\dot{\boldsymbol{\epsilon}}_p$ is defined in terms of a scalar plastic multiplier $\dot{\lambda}$ and a scalar plastic potential g according to

$$\dot{\boldsymbol{\epsilon}}_p = \dot{\lambda} \frac{\partial g}{\partial \boldsymbol{\tau}}. \quad (12)$$

The use of incompressible materials mandates that we use a plastic potential g which is not a function of the pressure p —in the remainder of this paper we will assume $g = \tau_{II}$.

To derive an expression for the plastic viscosity η_p , we assume that deviatoric stresses arise purely from the viscous component of the strain rate, that is

$$\tau = 2\eta \dot{\epsilon}_v. \quad (13)$$

Substituting equations (12), (11) into equation (13), and noting that $\partial g / \partial \tau = \frac{1}{2} \tau_{II}^{-1} \tau$ yields

$$\tau = 2\eta \left(\dot{\epsilon} - \dot{\lambda} \frac{\tau}{2\tau_{II}} \right), \quad (14)$$

or rearranging for τ :

$$\tau = \left(\frac{2\eta}{1 + \dot{\lambda} \eta \tau_{II}^{-1}} \right) \dot{\epsilon}. \quad (15)$$

The yield surface, $f(\sigma)$ is a scalar function defining the failure (yield) state of a material. Yield surfaces are assumed to be of the following form

$$f(\sigma) = \tau_{II} - Y(\sigma),$$

where we will refer to the scalar $Y(\cdot)$, as the yield criterion. For materials which are not yielding $f < 0$, while at yield $f = 0$. Under the condition of yielding ($f > 0$), it is common practice in geodynamics to define the plastic multiplier $\dot{\lambda}$ which exactly satisfies $f = 0$, or equivalently $\tau_{II} = Y$ [Lemiale et al., 2008]. Invoking the assumption of equation (8) applied to equation (15), we obtain

$$\tau_{II} = \left(\frac{2\eta}{1 + \dot{\lambda} \eta \tau_{II}^{-1}} \right) \dot{\epsilon}_{II}.$$

Under the condition that at yield $\tau_{II} = Y$, we obtain the following definition of $\dot{\lambda}$:

$$\dot{\lambda} = 2\dot{\epsilon}_{II} - \frac{Y}{\eta},$$

and finally we can express the effective viscoplastic viscosity as

$$\eta_p = \frac{Y}{2\dot{\epsilon}_{II}}, \quad (16)$$

which we note is identical to equation (10). We note that the derivation for $\dot{\lambda}$ did not impose that upon yield, the stress must remain at the yield surface for all time.

2.1.2. Yield Criteria

The various viscoplasticity models used in geodynamics, simply make different assumptions about the functional form of Y . For example, a von Mises plasticity model (henceforth denoted as VM) assumes

$$Y^{VM} = C, \quad (17)$$

where C is a constant cohesion independent of pressure. In contrast, Drucker-Prager (henceforth denoted as DP) introduces a *friction angle* φ and defines the yield surface according to

$$Y^{DP} = C \cos(\varphi) + \sin(\varphi)p, \quad (18)$$

which depends on the confining pressure p . The pressure dependence causes the material to be stronger at higher pressures. In the context of Stokes equation, the confining pressure is assumed to be given by the total pressure defined in equation (1) [Kaus, 2010].

Approximations are often made to Drucker-Prager in which the pressure p is assumed to be the lithostatic pressure (as opposed to the total pressure). This implicitly assumes that the normal stress is independent of the dynamic pressure. As such the effective yield criterion is depth-dependent and the yield criterion is given by:

$$Y^{DDM} = C \cos(\varphi) + \sin(\varphi)p_{\text{lith}}, \quad (19)$$

where p_{lith} is the lithostatic pressure [e.g., Moresi and Solomatov, 1998; Tackley, 2000; Stein et al., 2004]. We will refer to this model as depth-dependent von Mises (DDM).

These three yield criterion can be generalized by the following definition:

$$\gamma^{\alpha DP} = A + B(p_{\text{lith}} + \alpha p'), \quad (20)$$

where $p = p_{\text{lith}} + p'$, p' is the dynamic pressure and $\alpha \in [0, 1]$. Throughout the remainder of this paper, we will refer to equation (20) as the α -Drucker-Prager model.

With specific choices for A , B , and α , we can recover all plasticity models introduced. In particular:

1. von Mises: $A = C$, $B = 0$
2. depth-dependent von Mises: $A = C \cos(\varphi)$, $B = \sin(\varphi)$, $\alpha = 0$
3. Drucker-Prager: $A = C \cos(\varphi)$, $B = \sin(\varphi)$, $\alpha = 1$

We also note that the α plasticity model also permits ad hoc mixtures of different yield criteria by using $0 < \alpha < 1$. This aspect of the formulation will be discussed in more detail in section 5.3.

From equation (16), the effective plastic viscosity for the α -Drucker-Prager model is

$$\eta_p = \frac{A + B(p_{\text{lith}} + \alpha p')}{2\dot{\epsilon}_{II}}. \quad (21)$$

3. Development of a Representative Model

To evaluate the solvability of the proposed viscoplastic rheologies (VM, DDM, DP) with an incompressible Stokes formulation, we have designed a simplified problem which is representative of a collisional setting. In order to develop this model, we have made several important decisions which we discuss in more detail in the following subsections.

3.1. Pressure Split Formulation

To make the role of dynamic pressure and lithostatic pressure in these rheologies more precise, in equation (20), we chose to decompose the total pressure p as

$$p = p_{\text{lith}} + p', \quad (22)$$

where $p_{\text{lith}} = \rho_0 g d$ is a *reference* lithostatic pressure at depth d below some datum, ρ_0 is a reference density, and p' is the dynamic pressure due to viscous deformation and residual buoyancy.

Given this decomposition, we can rewrite incompressible Stokes in terms of velocity and dynamic pressure by substituting equation (22) into equation (1) to obtain

$$-\nabla \cdot 2\eta \dot{\epsilon} + \nabla p' = (\rho - \rho_0) \mathbf{g} = \mathbf{f}_m. \quad (23)$$

We note that the conservation of mass (equation (2)) remains unchanged. This pressure decomposition is purely formal and does not change the behavior of solutions although care should be taken when solving the pressure split equations when the free surface is not horizontal [e.g., Kramer *et al.*, 2012]. Further details related to this point are provided in Appendix A.

3.2. Viscosity Mixing Model

Equations (16) or (21) are only valid when the material is at failure and an additional subtlety remains how to combine the plastic rheology with classical ductile flow laws. The simplest approach is to just use the weakest mechanisms via

$$\eta = \min(\eta_1, \eta_2, \dots, \eta_N, \eta_p), \quad (24)$$

where η_1, \dots, η_N are the viscosities computed from individual *ductile* flow laws (e.g., dislocation creep, diffusion creep) and η_p is the effective plastic viscosity. Equation (24), however, is not differentiable which can make it harder to analyze, or use within nonlinear Newton solvers.

If we regard the viscoplastic rheology as another flow law defining an isotropic viscosity which is simultaneously active together with the ductile flow laws, then the strain rates from each individual mechanism are additive (e.g., a generalized Maxwell body consistent with equation (11)), and the total strain rate tensor can be expressed as

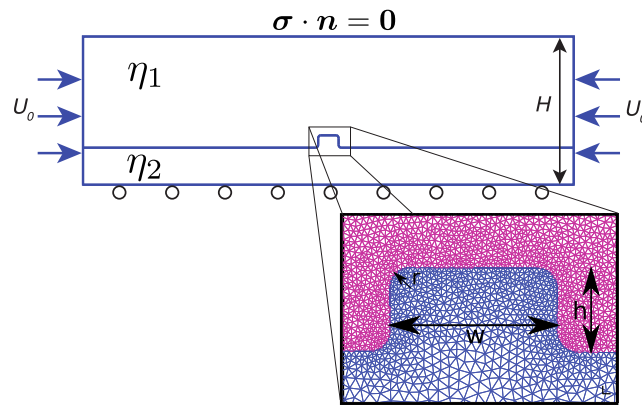


Figure 1. Geometry and boundary conditions for the simple compressional notch test. Inset shows close-up view of the rounded notch along with the typical unstructured mesh resolution.

$$\dot{\epsilon} = \frac{\tau}{2\eta_1} + \dots + \frac{\tau}{2\eta_N} + \frac{\tau}{2\eta_p}. \quad (25)$$

Moreover, as each viscosity is a scalar quantity, we can define a stress tensor τ with a single scalar effective viscosity such that $\dot{\epsilon} = \frac{\tau}{2\eta_{\text{eff}}}$ where

$$\eta_{\text{eff}} = \left[\frac{1}{\eta_1} + \dots + \frac{1}{\eta_N} + \frac{1}{\eta_p} \right]^{-1}. \quad (26)$$

This style of “effective viscosity” approach for viscoplasticity formulations has been used in large-scale geodynamic simulations of convection [e.g., Trompert and Hansen, 1998; Stein et al., 2004; Moresi et al., 2007; van Heck and Tackley, 2008; Tosi et al., 2015].

In our representative model, we consider a viscoplastic material with a single ductile flow law defined by a Newtonian (constant) viscosity η_1 . Hence the effective viscosity is given by

$$\eta_{\text{eff}} = \left[\frac{1}{\eta_1} + \frac{2\dot{\epsilon}_{\parallel}}{A + B(\rho_{\text{lith}} + \alpha p')} \right]^{-1}. \quad (27)$$

Throughout the remainder of paper, η_1 will be referred to as either the “background”, or “reference” viscosity. By writing the effective viscosity in terms of η_p

$$\eta_{\text{eff}} = \frac{\eta_1 \eta_p}{\eta_1 + \eta_p}, \quad (28)$$

it is clear that in the limit that the background viscosity is considerably greater than the viscoplastic viscosity ($\eta_1 \gg \eta_p$), then the effective rheology becomes purely plastic and independent of the background viscosity. Appendix C provides analysis of a viscoplastic layer in simple shear that shows how the purely plastic limit can be ill-posed and how the inverse viscosity mixing model regularizes the problem.

3.3. Geometry

Figure 1 shows the model geometry and boundary conditions for our representative model. The overall domain is 4×1 with depth $H = 30$ km. It has a viscoplastic upper layer (“crustal” region) with a rheology given by equation (27) with constant background viscosity η_1 . The lower layer (“asthenospheric” region) is isoviscous with viscosity η_2 .

A small notch is introduced within the upper layer to initiate stress localization. In these models, the notch has been deliberately designed to have smooth (nonreentrant) corners to avoid introducing singularities within the strain rate field. Removing geometric singularities results in a strain rate which is finite and bounded—thus we can numerically resolve $\dot{\epsilon}_{\parallel}$. If the notch possessed sharp corners, the computed strain rate invariant at the corner of the notch would continually increase under mesh refinement. Given the form of the effective viscosity (see equation (27)), this would have the effect of driving $\eta_p \rightarrow 0$ and would thus introduce a mesh dependence in the resulting shear band thickness. A mesh-dependent solution makes the question of “solubility” impossible to address. The inset within Figure 1 shows a close-up of the notch with a representative view of the unstructured mesh used in this region. The model geometry and rheology prescribed in our representative models defines a simple idealization of the Earth’s lithosphere, similar to that used by others [e.g., Huismans and Beaumont, 2003; Moresi et al., 2007].

Boundary conditions are a constant horizontal velocity $\mathbf{v} \cdot \mathbf{n} = -U_0$ on the left and right boundaries, a free-slip bottom boundary with $\mathbf{v} \cdot \mathbf{n} = 0$ on the bottom and a free-stress condition $\boldsymbol{\sigma} \cdot \mathbf{n} = \mathbf{0}$ on the top boundary. Table 1 lists the parameters and values used in this paper.

Table 1. Parameters Used in the Compressional Notch Test

Name	Description	Values
η_1	background viscosity of upper viscoplastic layer	10^{23} – 5×10^{24} Pa s
η_2	Viscosity of the lower layer and notch (isoviscous)	10^{21} Pa s
η_0	Reference viscosity for scaling	10^{22} Pa s
H	Domain depth (and scale length)	30 km
H_1	Upper layer depth	$0.75H$
H_2	Lower layer depth	$0.25H$
L	Domain width	$4H$
w	Notch width	$1/6H$
h	Notch height	$1/12H$
r	Radius of curvature of notch corners	$0.02H$
U_0	Horizontal velocity (compression)	0.0025–0.0125 m/yr
C	Cohesion	10^8 Pa
φ	Friction angle	0 – 30°
ρ	Density (both layers)	2700 kg m^{-3}
g	Acceleration due to gravity	9.81 m s^{-2}

3.4. Time Independence

In the absence of a damage model, or other time-dependent softening mechanisms, the incompressible viscoplastic model defined is instantaneous. That is, no temporal evolution is required for the development of shear bands. This is in contrast with elastoplastic or elasto-viscoplastic problems. Given that our intention is to quantify, the “solvability” of viscoplastic formulations for incompressible media with a simplified model, we restrict our discussion in this work to the instantaneous problem. We do not introduce softening of the cohesion or friction angle as a function of accumulated plastic strain.

4. Numerical Methods

4.1. Nonlinear Solvers

Equations (2),(23) together with equation (27) and boundary conditions form a strongly nonlinear system of PDE’s where the primary nonlinearity arises through the strain rate and pressure dependence of the viscosity. To solve these problems numerically, we utilize iterative nonlinear solvers with well understood and monitored convergence criteria. For the purpose of introducing the different nonlinear solvers, in the context of incompressible viscoplastic Stokes problems, we use the following definitions:

1. The velocity and pressure solution will be denoted as a single vector, $\mathbf{x}=(\mathbf{v},p')$.
2. The differential operators defined by the left-hand sides of equation (23) and equation (2) will be denoted via $\mathcal{L}(\cdot)$.
3. The right-hand sides of equation (23) and equation (2) will be denoted via a single vector $\hat{\mathbf{f}}=((\rho-\rho_0)\mathbf{g},0)$

4.1.1. Successive Substitution

Successive substitution (often referred to as Picard iteration), solves a sequence of linearized problems using the velocity and pressure at the previous iteration to approximate the viscosity. Given an initial guess for the coupled Stokes problem to be \mathbf{x}_0 , we summarize the successive substitution procedure in Algorithm 1.

Algorithm 1 Successive Substitution

```

Require:  $\mathbf{x}_0$ 
for  $i=1,2,\dots$  do {loop until converged}
    Solve  $\mathcal{L}(\mathbf{x}_{i-1})\mathbf{x}_i=\hat{\mathbf{f}}$  for  $\mathbf{x}_i$ 
end for

```

The issue with successive substitution is how best to define an appropriate stopping condition for the iteration that can measure the degree of accuracy of the current solution estimate, \mathbf{x}_i .

In some numerical simulations [e.g., Lemiale et al., 2008; Kaus, 2010], nonlinear iterations are terminated when changes in the solution between iterations, $\Delta\mathbf{x}=\mathbf{x}_i-\mathbf{x}_{i-1}$ become small in some norm. However, as we demonstrate in section 5.1, this does not guarantee a converged solution as many nonlinear iterative

methods can converge slowly or stall. A more reliable measure of convergence for a well scaled nonlinear problem is to calculate the nonlinear residual at iteration i

$$\mathbf{F}(\mathbf{x}_i) = \mathcal{L}(\mathbf{x}_i)\mathbf{x}_i - \hat{\mathbf{f}} \quad (29)$$

and stop when $\|\mathbf{F}(\mathbf{x}_i)\| \leq \text{tol}$ in some norm. Here we use the discrete 2-norm $\|\mathbf{F}(\mathbf{x})\|_2 = \sqrt{\mathbf{F} \cdot \mathbf{F}}$ and define our stopping criteria as

$$\|\mathbf{F}(\mathbf{x}_i)\|_2 \leq \max(\text{rel_tol}\|\mathbf{F}(\mathbf{x}_0)\|_2, \text{abs_tol}) \quad (30)$$

where rel_tol and abs_tol are relative and absolute tolerances. Unless otherwise specified, all the problems in this paper use $\text{rel_tol} = 10^{-7}$ and $\text{abs_tol} = 10^{-11}$.

4.1.2. Approximate Newton Iteration

As written, successive substitution can be difficult to monitor for nonlinear convergence as it requires an additional residual calculation. However, it is straightforward to rewrite the exact same problem in terms of a residual and a correction that allows for automatic convergence testing as well as more control in the iterative method. The principal idea is that at every iteration, we can write our new solution as

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \delta\mathbf{x} \quad (31)$$

where $\delta\mathbf{x}$ is the subsequent correction. Substituting equation (31) into

$$\mathcal{L}(\mathbf{x}_{i-1})\mathbf{x}_i = \hat{\mathbf{f}} \quad (32)$$

and rearranging, yields the following defect-correction problem

$$\mathcal{L}(\mathbf{x}_{i-1})\delta\mathbf{x} = \hat{\mathbf{f}} - \mathcal{L}(\mathbf{x}_{i-1})\mathbf{x}_{i-1} = -\mathbf{F}(\mathbf{x}_{i-1}) \quad (33)$$

which is then solved for correction $\delta\mathbf{x}$. This formulation of the nonlinear problem leads to a more general iterative strategy which is summarized in Algorithm 2. Here we have introduced the parameter β which serves as a globalization parameter that controls how much of the correction gets added to the previous solution. Successive substitution is equivalent to setting $\beta = 1$, however, in more sophisticated schemes β can be adjusted to maintain monotonicity in the reduction of the residual. In addition, Algorithm 2 is less sensitive to rounding error because the RHS of equation (33) approaches $\mathbf{0}$ as the problem converges, and each nonlinear iteration starts with an initial guess of $\delta\mathbf{x} = \mathbf{0}$. More importantly, the nonlinear residual is naturally calculated as part of this algorithm allowing straightforward convergence monitoring.

Algorithm 2: Approximate Newton

Require: \mathbf{x}_0

Initialize Iteration: $i = 0$

set $\mathbf{F}(\mathbf{x}_0) = \mathcal{L}(\mathbf{x}_0)\mathbf{x}_0 - \hat{\mathbf{f}}$

while $\|\mathbf{F}(\mathbf{x}_i)\|_2 > \text{tol}$ **do**

Solve $\mathcal{L}(\mathbf{x}_{i-1})\delta\mathbf{x} = -\mathbf{F}(\mathbf{x}_{i-1})$ for $\delta\mathbf{x}$

Set $\mathbf{x}_i = \mathbf{x}_{i-1} + \beta\delta\mathbf{x}$

end while

Unfortunately, for many problems Algorithms 1 or 2 can show very slow convergence and possibly even stalling where $\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i$ without $\|\mathbf{F}(\mathbf{x}_i)\|_2 \rightarrow 0$ (see below). However, with a small change in the operator, this algorithm can be made much more efficient.

4.1.3. Newton Iteration

The previous algorithm has the exact same structure as a Newton iteration with the only exception that the linear operator in the inner solve is calculated from the original quasi-linear problem rather than using the full Jacobian for the problem (and thus the name approximate Newton). To see this, we briefly review

Newton's method which attempts to solve the fully nonlinear problem $\mathbf{F}(\mathbf{x}) = \mathbf{0}$. Again, it starts with a guess \mathbf{x}_{i-1} such that $\mathbf{F}(\mathbf{x}_{i-1}) \neq \mathbf{0}$ and seek a correction $\delta\mathbf{x}$ such that $\mathbf{F}(\mathbf{x}_{i-1} + \delta\mathbf{x}) = \mathbf{0}$. Taylor expansion gives

$$\mathbf{F}(\mathbf{x}_{i-1} + \delta\mathbf{x}) = \mathbf{F}(\mathbf{x}_{i-1}) + \mathbf{J}(\mathbf{x}_{i-1})\delta\mathbf{x} + O(\delta\mathbf{x}^2) \quad (34)$$

where

$$\mathbf{J}(\mathbf{x}) = \frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}} \quad (35)$$

is the *Jacobian* or functional derivative of the nonlinear residual with respect to the solution \mathbf{x} . If we truncate equation (34) to neglect the higher-order terms and assume $\delta\mathbf{x}$ is an exact correction, we arrive at Newton's algorithm (with globalization) which is shown in Algorithm 3.

Algorithm 3 Newton

Require: \mathbf{x}_0

Initialize Iteration: $i = 0$, set $\mathbf{F}(\mathbf{x}_0) = \mathcal{L}(\mathbf{x}_0)\mathbf{x}_0 - \hat{\mathbf{f}}$

while $\|\mathbf{F}(\mathbf{x}_i)\|_2 > \text{tol}$ **do**

Solve $\mathbf{J}(\mathbf{x}_{i-1})\delta\mathbf{x} = -\mathbf{F}(\mathbf{x}_{i-1})$ for $\delta\mathbf{x}$

Set $\mathbf{x}_i = \mathbf{x}_{i-1} + \beta\delta\mathbf{x}$

end while

Comparison with the previous section shows that the algorithms are identical except for the choice of Jacobian \mathbf{J} . If the problem is well posed and the Jacobian not singular, then it is possible for Newton's algorithm to demonstrate quadratic convergence for initial guesses that are close to the solution. In general, raw Newton cannot be guaranteed to converge, however, with sufficient controls such as line search or trust region (which are supplied with any good nonlinear solver package), it can provide much better and rapid solutions with small residuals. We will compare all of these strategies (as well as hybrid strategies) on our various model problems in subsequent sections.

4.2. Spatial Discretization: Finite Element Method

We emphasize that these different nonlinear solver strategies are independent of the specific form of discretization and can be used with finite element, finite volume, and finite difference methods. Here, however, we solve equations (2) and (23) using the Finite Element method which, with appropriate software, can make calculation of non-linear residuals and exact Jacobians precise. Specific details and weak forms are given in Appendix A.

4.3. Software Implementation

To readily explore a range of plasticity models and solver strategies, we use the open source software package TerraFERMA (The Transparent Finite Element Rapid Model Assembler) [Wilson and Spiegelman, 2015]. TerraFERMA leverages three open source computational libraries to provide a framework for rapid composition and exploration of coupled multiphysics models. The first library is from the FEniCS project [Logg, 2012], which provides a high-level language (UFL) for the description and assembly of weak forms as well as automatic differentiation of forms for calculating exact Jacobians of complicated nonlinear residuals. Given a high-level description of the forms, the FEniCS Form Compiler (FFC) autogenerates compilable C++ code for assembly of linear algebraic objects. The second library, PETSc [Balay et al., 2012a, 2012b], provides a wide range of scaleable linear and nonlinear solvers and preconditioners for the solution and monitoring of algebraic systems. In particular, we make considerable use of PETSc SNES objects (Scalable Nonlinear Equation Solver) for solving and monitoring convergence of both approximate and exact Newton schemes. Finally, we use the SPuD library [Ham et al., 2009] which is an application-neutral options handling

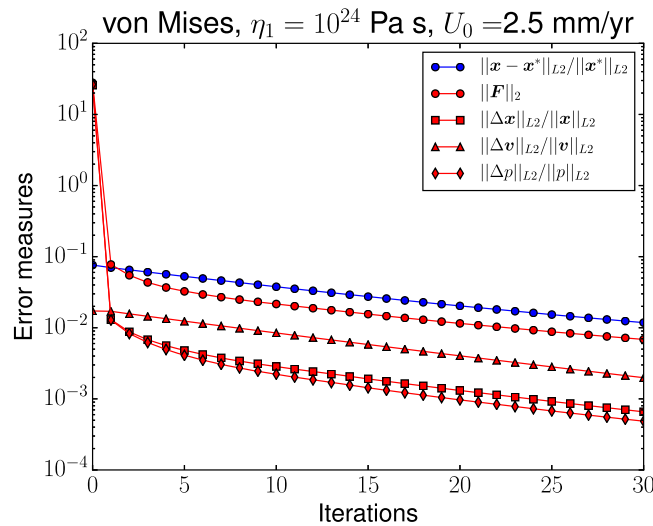


Figure 2. Convergence behavior of successive substitution (approximate Newton) for the notch test with an upper layer viscosity $\eta_1 = 10^{24}$ Pa s, compression velocity $U_0 = 2.5$ mm/yr, and a von Mises plasticity model $A = C$, $B = 0$. The red curves show reduction in the discrete 2-norm of the full nonlinear residual ($\|F\|_2$) as well as measures of the relative change in solution between iterations as defined in the text. The blue curve shows the relative error $\|x - x^*\|_{L2}/\|x^*\|_{L2}$ with respect to a converged solution x^* with nonlinear residual $\|F\|_2 \sim 10^{-10}$ (i.e., Figure 4b). Note that even when the velocity solution is changing by less than 10^{-3} per iteration, both the relative error and the nonlinear residual are $\sim 10^{-2}$.

nonlinearity into the solution. This model also provides a useful illustration of the convergence behavior of the various solver strategies. Figure 2 shows typical convergence behavior for successive substitution (or approximate Newton) for an upper layer reference viscosity $\eta_1 = 10^{24}$ Pa s and a compressional velocity $U_0 = 2.5$ mm/yr. The initial state is calculated from the isoviscous solution without plasticity. We then take 30 iterations of approximate Newton (Algorithm 2). We use a sparse-direct solver (UMFPACK) [Davis, 2004] for the Stokes system for each linear solve so that the linear residuals, $\mathcal{L}(x_{i-1})x_i - \hat{f}$ are of machine precision at each iteration. The figure shows several measures used to gauge convergence. Three of the measures are similar to those used in the literature [e.g., Kaus, 2010; Lemiale et al., 2008] based on relative changes in the solution from iteration to iteration. For example

$$\|\Delta v\|_{L2}/\|v\|_{L2} = \left[\frac{\int_{\Omega} (v_i - v_{i-1}) \cdot (v_i - v_{i-1}) dV}{\int_{\Omega} v_i \cdot v_i dV} \right]^{1/2} \quad (36)$$

is a measure of the relative change in the velocity field from iteration $i - 1$ to i which is reduced to $\sim 2 \times 10^{-3}$ by the final iteration. Here we define the errors in terms of the $L2$ -norm for functions which is easily calculated using TerraFERMA. Similarly $\|\Delta p\|_{L2}/\|p\|_{L2}$ shows relative changes in the pressure field while $\|\Delta x\|_{L2}/\|x\|_{L2}$ are changes in the entire solution $x = (v, p)$. Both of these measures are reduced to $< 10^{-3}$ which is comparable to some of the stopping criteria reported in the literature (e.g., Lemiale et al. [2008] used 10^{-3} and Kaus [2010] used 10^{-4}). However, while the solution is changing slowly, the discrete 2-norm of the nonlinear residual, is still relatively large with $\|F\|_2 \sim 10^{-2}$. Of course, the magnitude of the nonlinear residual will depend on how the problem is scaled and we find that it is useful to scale variables such that they are roughly $O(1)$ when solving. For this problem, we scale velocities by U_0 , viscosities by $\eta_0 = 10^{22}$ Pa s, and stresses/pressures by $\eta_0 U_0 / H$ where H is the layer depth. With this scaling, we find that $\|F\|_2$ is a good approximation to the norm of the relative error $e_{rel} = \|x - x^*\|_{L2}/\|x^*\|_{L2}$ where x^* is a well-resolved discrete solution with small nonlinear residual. For this problem, we use the solution shown in Figure 4b which is a solution of this problem with nonlinear residual $\|F\|_2 \sim 10^{-11}$.

We now consider solving the same model problem using a Newton based solver. The convergence of Newton's method is known to be highly dependent on the initial solution vector. While we expect to obtain

system to manage all of the scientific and computational choices that make up a model into a single, hierarchical options file that contains both compile time and run time options. Given a .tfml (TerraFERMA markup language) file, TerraFERMA compiles, and runs a custom executable for a specific model that can be readily shared, reproduced and modified by anyone with the underlying software. All .tfml files used in models in this paper, together with meshes and "simulation harness" files for conducting parameter sweeps and plotting figures are available as supporting information in an accompanying git repository (<http://bitbucket.org/mspieg/plasticitymodels>).

5. Results

5.1. von Mises Plasticity

We begin with von Mises plasticity ($B = 0$) as it is the simplest of the plasticity models and introduces the least

quadratic convergence with respect to $\|F\|_2$, this is only possible if the initial vector is “close” to the true solution. To obtain a reliable starting vector x_0 , we perform a number of Picard iterations and then switch to Newton’s method. We summarize our approach in Algorithm 4.

Algorithm 4 Combined Picard-Newton iteration

Initialize: Set $\eta = \eta_1$ and solve for x_0

Solve using approximate Newton to $\|F\|_2 < \max(10^{-5}\|F(x_0)\|_2, 10^{-3})$ or $i = 30$

Solve with Newton until $\|F\|_2 < \max(10^{-7}\|F(x_0)\|_2, 10^{-11})$

Figure 3 shows the convergence of $\|F\|_2$ and e_{rel} using the nonlinear solver strategy described in Algorithm 4. While the first 30 approximate Newton (Picard) iterations reduce the nonlinear residual to only $\sim 10^{-2}$, another 10 Newton iterations reduce it to 10^{-11} . Note, that to reduce the relative error to $< 10^{-6}$ requires a nonlinear residual reduction to $\|F\|_2 \approx 10^{-6}$.

To illustrate the difference between a nonconverged and converged solution, Figure 4 compares the solutions for $\dot{\epsilon}_{II}$ from nonconverged successive substitution (4a) to a converged solution using Newton (4b). The structure and intensity of the shear bands are noticeably different with almost a factor of 2 difference in strain rate within the shear bands. These results suggest that any discussion of the structure of viscoplastic shear bands requires a demonstration of convergence in the nonlinear residual.

Finally, Figure 5a shows the convergence behavior of Algorithm 4 for the von Mises plasticity model using a wider range of upper layer viscosities and compression velocities. While stronger or more rapidly deformed layers tend to take more computational work, particularly in the approximate Newton phase, this algorithm demonstrates good convergence behavior for all parameters considered.

5.2. Depth-Dependent von Mises Plasticity

von Mises plasticity is relatively straightforward to solve accurately and efficiently with a good Newton solver. However, as a model of plasticity it allows yielding to occur at all depths in the Earth. Depth-

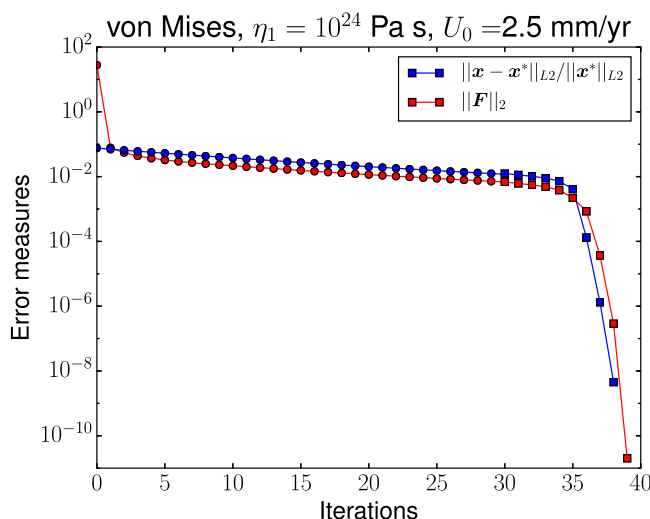


Figure 3. Convergence behavior of the combined Picard-Newton iterations for the same problem shown in Figure 2. Initial iterations using an approximate Newton scheme are indicated via circles ($i=0, \dots, 30$) and full Newton steps are indicated via squares. After an initial period of slow convergence due to line search restrictions, the solution shows classical quadratic convergence. Also for this scaling of the problem, the nonlinear residual is a reasonably good estimator of the relative-error (blue curve).

dependent von Mises plasticity approximates Byerlee like failure in the crust by allowing the yield stress to increase with depth. This model has the same nonlinearity as von Mises but is only a good approximation if the dynamic pressure is small (an assumption we will show is not necessarily valid for strong layers).

Figure 5b shows the convergence behavior for this model ($A = C \cos(\varphi)$, $B = \sin(\varphi)$, $\varphi = 30^\circ$, $\alpha = 0$) for the same range of layer strengths and compression velocities as Figure 5a. The convergence behavior is approximately the same for von Mises and depth-dependent von Mises except for the strongest/fastest deforming layers where the convergence rate of both approximate Newton and Newton iterations degrade, although all models converge eventually to a solution with $\|F\|_2 \leq 10^{-10}$.

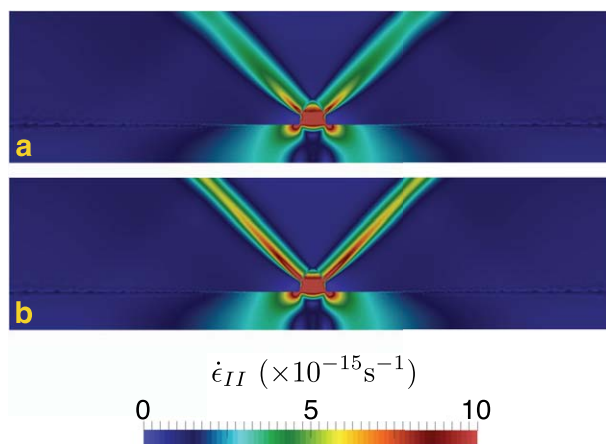


Figure 4. Comparison of strain-rate invariant fields $\dot{\epsilon}_{II}$ between (a) a poorly converged solution and (b) a well-converged solution for the notch-test with von Mises plasticity. (top) Results from successive substitution with poor convergence after 30 iterations ($\|\mathbf{x} - \mathbf{x}^*\|_{L2} / \|\mathbf{x}^*\|_{L2} \approx \|\mathbf{F}\|_2 \sim 10^{-2}$). (bottom) The result of combined Picard-Newton iterations with $\|\mathbf{F}\|_2 \sim 10^{-11}$ (and $\mathbf{x} = \mathbf{x}^*$). For these problems, poor convergence leads to more weakly defined shear bands.

the two runs (second column in Figure 6). Whereas von Mises plasticity reduces the overall viscosity of the upper layer to $\sim 5 \times 10^{22}$ Pa s outside of the shear zones (and $\sim 5 \times 10^{21}$ Pa s within the shear zones), depth-dependent von Mises plasticity allows the lower part of the plastic layer to achieve viscosities closer to the intrinsic layer viscosity $\eta_1 = 10^{24}$ Pa s. The viscosity inside the shear bands are higher, leading to a slightly thicker shear zone. Somewhat surprisingly, however, these shear bands are more effective at reducing the strain rate in the central uplifted block, which also increases its strength, leading to a more rigid behavior in the central block.

It is important to note that the change in behavior is only due to the inclusion of the lithostatic pressure into the effective viscosity. However, this change also has consequences for the dynamic pressure. The third column of Figure 6 compares p' for the same runs and shows a pronounced overpressure of ~ 200 – 500 MPa in the viscoplastic layer, with a particularly high-pressure region at the lower corner of the uplifted block. This behavior is a natural consequence of Stokes equation, where the dynamic pressure acts to enforce the incompressibility condition. In particular, strong regions under compression will often have large positive dynamic pressures (which can be comparable to lithostatic) to maintain incompressibility.

5.3. Drucker-Prager Plasticity

Including the dynamic pressure in the viscosity via a Drucker-Prager model, makes the problem significantly more nonlinear and can lead to both computational and geodynamical issues. Figure 5c shows convergence plots for full Drucker-Prager plasticity ($A = C \cos(\varphi)$, $B = \sin(\varphi)$, $\varphi = 30^\circ$, $\alpha = 1$) for the same parameters as in Figures 5a and 5b. Convergence is comparable for weak/slowly deforming layers, however the nonlinear residual is observed to stagnate for viscosities $\eta_1 \geq 10^{24}$ Pa s and velocities $U_0 > 2.5$ mm/yr. Figure 6 (third row) compares the strain rate, viscosity, and dynamic pressure for the strongest Drucker-Prager plasticity run that converges ($\eta_1 = 10^{24}$ Pa s, $U_0 = 2.5$ mm/yr) to those for the other two plasticity models using the same parameters. When full pressure dependence is included in this problem, the shear bands show a sharp curvature in angle near the notch but are weaker in the upper part of the layer. Examination of the viscosity and pressure fields suggest that this behavior stems from the increased viscosity in the lower corner of the uplifted block which simply causes the shear bands to deflect into the weaker regions around it. The “angle” of the shear bands has little to do with the intrinsic friction angle, and simply reflects the more complex viscosity structure due to an effective viscosity that also depends on p' . More seriously, the increasing viscosity near the base of the layer, further increases the dynamic pressure near the bottom of the layer and in the uplifted wedge to over 500 MPa (and in some regions as high as 1 GPa, the colorbar has been clipped for better comparison of the models) which is of the same order as lithostatic pressures. The primary effect of including full Drucker-Prager plasticity is to actually make the layer harder to fail (and

Figure 6 compares the shear band structure, effective viscosity, and dynamic pressure p' for von Mises and depth-dependent von Mises (as well as Drucker-Prager) plasticity models. All solutions shown here use $\eta_1 = 10^{24}$ Pa s and $U_0 = 2.5$ mm/yr and are well converged with nonlinear residuals $\|\mathbf{F}\|_2 < 10^{-10}$. We note that $\eta_1 = 10^{24}$ Pa s, $U_0 = 2.5$ mm/yr are the parameter choices which result in the highest stresses where all three models still converge.

Comparing von Mises to depth-dependent von Mises models, the two solutions are very similar with slightly thicker shear bands for depth-dependent von Mises. In particular, there is no significant change in shear band angle although the nominal “friction angle” has changed from $\varphi = 0^\circ$ to $\varphi = 30^\circ$. More significant is the change in effective viscosity between

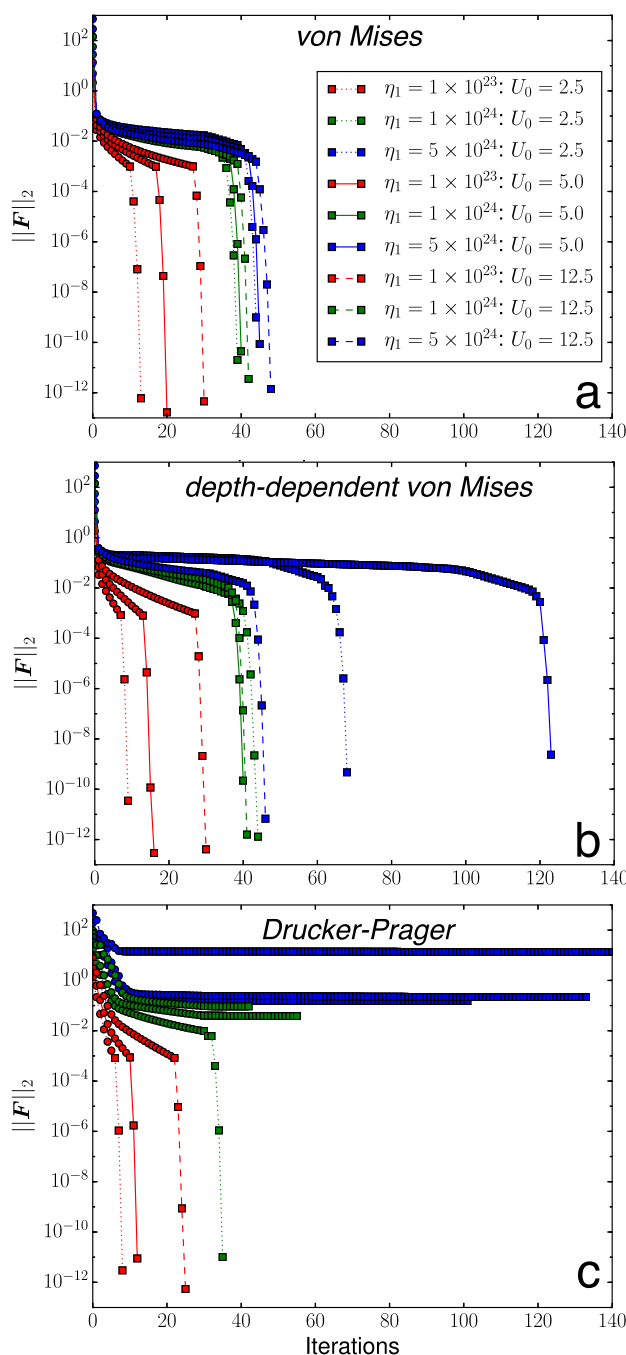


Figure 5. Comparison of convergence behavior of the combined Picard-Newton scheme for the compressive notch test with the three plasticity models (a) von Mises, (b) depth-dependent von Mises, and (c) Drucker-Prager. Models are run for the range of upper layer viscosities and compression velocities given in the legend. Approximate Newton steps are shown as circles, Newton steps are indicated via squares. For these parameters, both von Mises and depth-dependent von Mises models converge for all runs, however, Drucker-Prager fails to converge for stronger layers or larger compression velocities.

clearly that the regularized Drucker-Prager model becomes ill-posed as the problem approaches pure plasticity. For the more complicated compression problem, the supporting information provides PDF flipbooks that visualize the velocity field \mathbf{v} , the velocity residual \mathbf{F}_v , and the velocity correction $\delta\mathbf{v}$ during each iteration of the combined Picard-Newton algorithm for two runs. The first is the converged Drucker-Prager

much harder to solve). An open question remains as to whether these predicted pressures are geologically reasonable; however, they are a natural outcome of this plasticity model. It is worth pointing out that since Stokes solves for both velocity and pressure it is critical to evaluate the pressure fields as well the strain rate fields for physical relevance, particularly if the constitutive relationships depend on pressure.

Finally we can also gain some insight into how convergence breaks down (and what poorly converged Drucker-Prager solutions can look like) by considering a solver strategy of continuation in α from 0 (depth-dependent von Mises) to 1 (full Drucker-Prager). Figure 7 shows the nonlinear residuals for a sequence of Newton solves starting from a fully converged depth-dependent von Mises model ($\alpha = 0$) with layer parameters $\eta_1 = 2 \times 10^{24}$ Pa s, $U_0 = 2.5$ mm/yr (i.e., twice the viscosity of the solution in Figure 6). α is then incremented by 0.1 with each iteration. Figure 7 shows good convergence up to $\alpha = 0.5$, after which the solver no longer converges.

Figure 8 shows a sequence of images showing the strain rate field ($\dot{\epsilon}_{II}$) at the end of each α continuation step. Note, only the first six figures are actually converged repeatable solutions. The remainder show very narrow shear bands and multiple spurious shear bands, which while visually attractive, are not actually solutions of the underlying PDE's.

More careful examination of the convergence behavior of Drucker-Prager viscoplasticity under both Picard and Newton solvers suggests that as the problem approaches the purely plastic limit, the Jacobian become increasingly ill-conditioned and likely singular. Appendix C provides analysis for a simpler but related problem of a viscoplastic layer in *simple shear* and shows

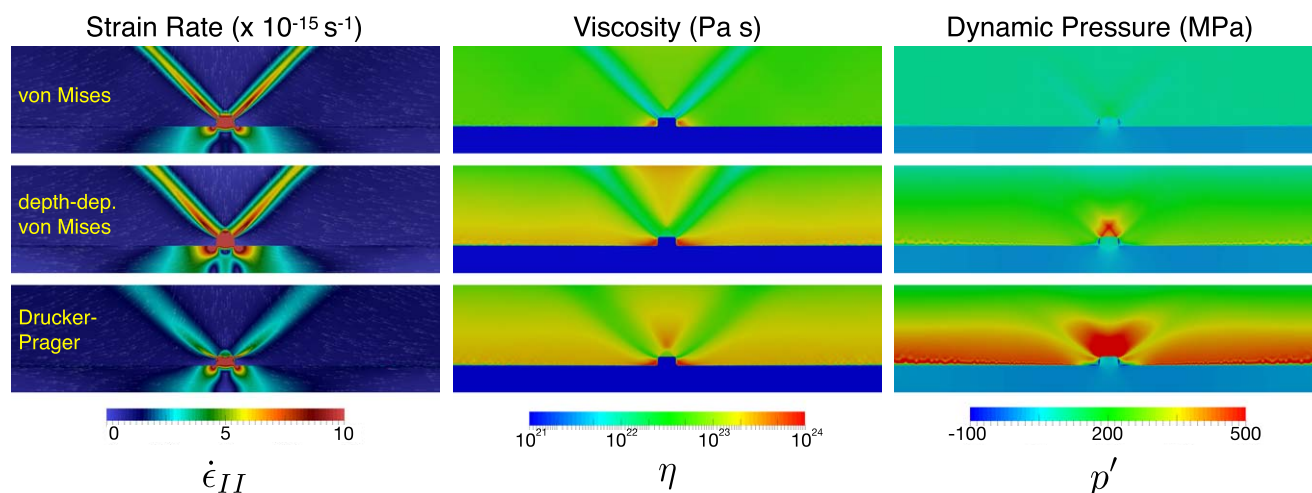


Figure 6. Comparison of strain-rate invariant $\dot{\epsilon}_{II}$, viscosity η , and dynamic pressure p' fields for the three plasticity models: von Mises (top row), depth-dependent von Mises (middle row), and Drucker-Prager (bottom row). All solutions use an upper layer reference viscosity $\eta_1 = 10^{24}$ Pa s, lower layer viscosity $\eta_2 = 10^{21}$ Pa s, and compressive velocity $U_0 = 2.5$ mm/yr. These are the most extreme parameters explored where all three models give converged solutions with nonlinear residuals $\|\mathbf{F}\|_2 \leq 10^{-10}$. Plasticity parameters for each model are von Mises: $[A=C, B=0]$, depth-dependent von Mises: $[A=C\cos(\varphi), B=\sin(\varphi), \varphi=30^\circ, \alpha=0]$, Drucker-Prager: $[A=C\cos(\varphi), B=\sin(\varphi), \varphi=30^\circ, \alpha=1]$. For dynamic pressure (third column), the color bar is clipped at 500 MPa for better comparison of the models. In the Drucker-Prager model, p' can exceed 1 GPa locally near the notch.

solution shown in the bottom row of Figure 6 ($\eta_1 = 10^{24}$ Pa s, $U_0 = 2.5$ mm/yr), the second has the same viscosity but twice the compressive velocity ($U_0 = 5.0$ mm/yr) and does not converge. Examination of the first case shows clearly how Picard can stall by producing a small correction $\delta \mathbf{v}$ but retaining a large residual. When Newton is applied to the Picard solution, both the residual and correction converge toward zero in a few iterations. For the problem that does not converge, Picard proceeds as before, however, the first iteration of Newton produces a large correction, including a suite of shear bands conjugate to the primary bands. Because of the line-search controls in PETSc SNES, only a small fraction of this correction is added to the solution. However, small perturbations in the solution produce very large deviations in the correction $\delta \mathbf{x}$ and many of these corrections take the form of spurious shear bands. This is indicative of a singular Jacobian with a very feature rich null-space which is being amplified by the direct solvers used here. Initial experiments using iterative solvers do not converge.

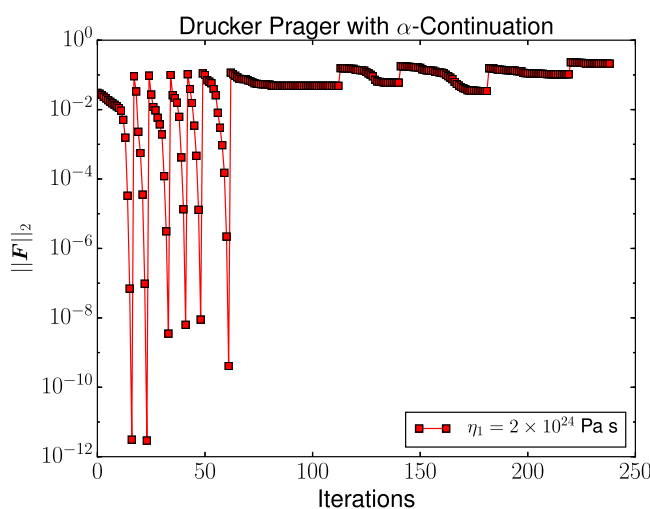


Figure 7. Convergence behavior for a continuation scheme with α -Drucker-Prager viscosity where α is incremented from 0 to 1 in increments of 0.1. A Newton solve is attempted for each value of α . Layer parameters are $\eta_1 = 2 \times 10^{24}$ Pa s, $U_0 = 2.5$ mm/yr. Plasticity parameters are $A=C\cos(\varphi)$, $B=\sin(\varphi)$, $\varphi=30^\circ$. Newton converges within ~ 10 iterations up to $\alpha=0.5$. Newton fails to converge for $\alpha > 0.5$.

At any rate, these results suggest, that any discussion of shear bands and solutions of incompressible Stokes with viscoplastic rheologies, requires that a solution to the nonlinear problem has been found, e.g., convergence needs to be demonstrated. Moreover, the results also highlight that care needs to be taken when using a viscosity which depends on the dynamic pressure in incompressible media.

5.4. Extension

While all of the above results presented so far have been for compressive systems, changing the sign of U_0 to make the problem extensional does not change the overall results. The convergence behavior using a combined Picard-Newton scheme for rheologies that do not depend on the

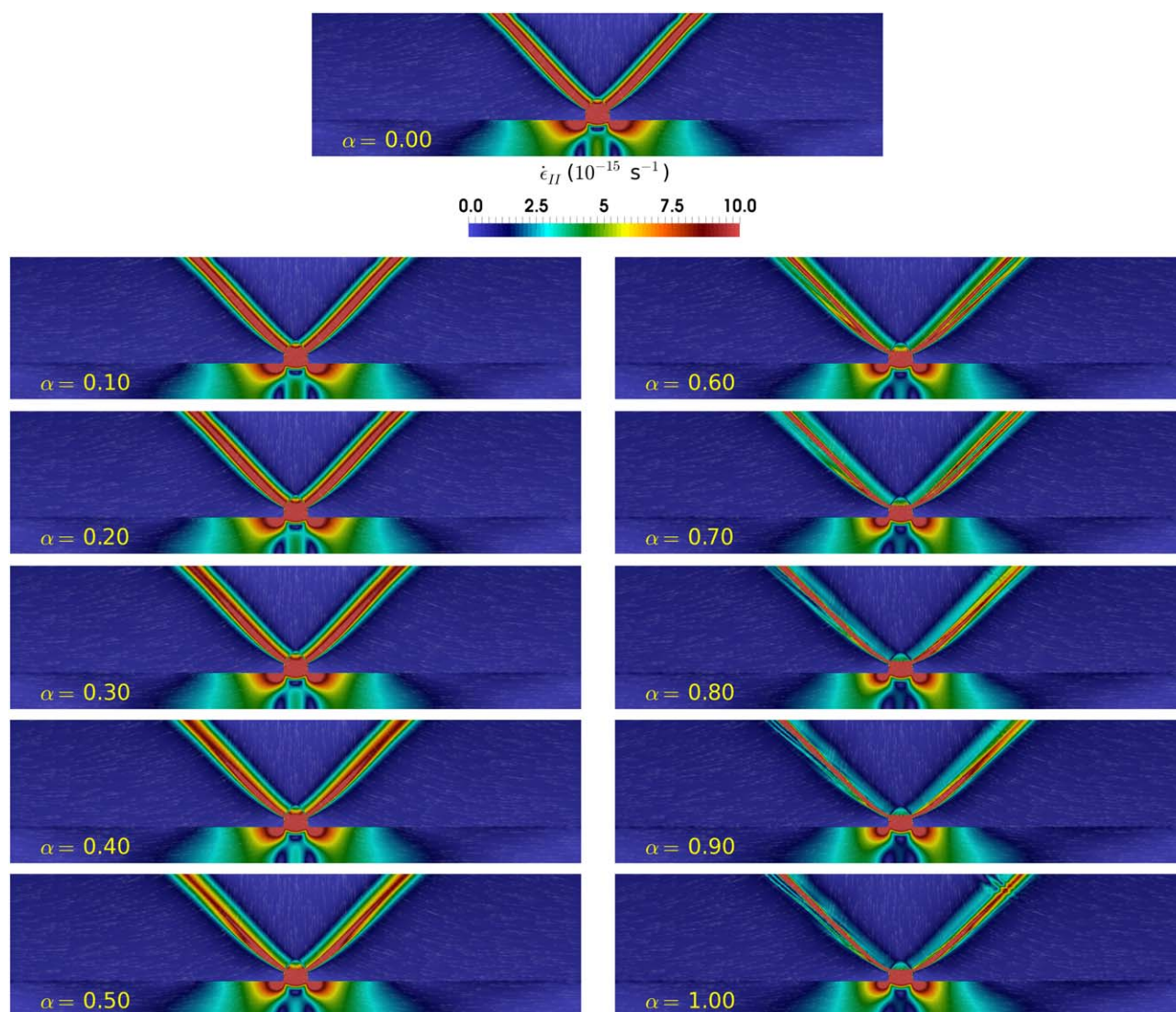


Figure 8. Second invariant of strain-rate $\dot{\epsilon}_{II}$ as a function of the continuation parameter α . Note, only $0 < \alpha \leq 0.5$ are actually converged solutions. Plots for $\alpha > 0.5$ are not converged and show spurious shear bands and asymmetry.

dynamic pressure (von Mises and depth-dependent von Mises) show behavior very similar to Figures 5a and 5b. However, Drucker-Prager proved even more difficult to solve and failed to converge for all values of η_1 and U_0 used here when started from an initial isoviscous solution.

This result is not surprising as in extension, the dynamic pressure can be strongly negative, bringing the stress closer to the yield surface (equation (20)). In the extreme case

$$p'_{\text{crit}} \leq -(p_{\text{lith}} + C \cot(\varphi)) \quad (37)$$

the effective viscosity is ≤ 0 , the problem loses ellipticity and thus has no solution. For a strong *isoviscous* layer under extension, the dynamic pressure around the heterogeneity is of order $\eta_1 U_0 / H$ which is ~ -2.6 GPa for $\eta_1 = 10^{24}$ Pa s and $U_0 = -2.5$ mm/yr, whereas equation (37) gives a value of p'_{crit} at ~ 1 GPa. Thus the initial isoviscous guess generates nonphysical viscosities for any nonlinear solver (including Picard) and thus the solver fails to converge. Starting from a “softer” initial guess, such as the solution from von Mises plasticity, for all plasticity models show similar behavior to Figure 5. That is, von Mises and depth-dependent von Mises converge for all parameters but Drucker-Prager only converges for relatively soft layers $\eta_1 = 10^{23}$ Pa s.

For sufficiently strong layers, theoretical analysis [Lanzendörfer, 2011] suggests that the problem has no solution (see also Appendices B and C). In the following discussion section, we will apply this analysis to the α -DP formulation and use it to estimate some bounds on degree of pressure dependence which permits nonlinear solutions to be found.

6. Discussion

A number of enigmatic issues associated with the usage of Drucker-Prager rheologies within crustal scale and lithospheric-scale simulations have been raised in the geodynamics literature. Based on our examination of the “solvability” of Drucker-Prager, and other viscoplastic, rheologies we wish to first emphasize that any comparison of simulation results of a viscoplastic (or any nonlinear problem) should be conducted on the basis of fully converged nonlinear solutions. As illustrated in Figure 4, the difference in strain rate invariant (which can be used to define an effective shear band) can be dramatically different between converged and nonconverged solutions. This result was also highlighted in Popov and Sobolev [2008].

Shear band orientation is commonly thought to be connected to the friction angle φ used in the Drucker-Prager yield criteria. Our numerical experiments show that the orientation of the shear bands close to the heterogeneity is largely dictated by the geometry of the very strong key stone structure which develops immediately above the heterogeneity. The shear bands obtained within our simulations accommodate the rigid key stone by “bending” around it. The strength of the key stone stems from our a priori assumption that the material is incompressible. In an incompressible regime, the dynamic pressure around weak (or strong) inclusions at a shallow depth can be comparable to the lithostatic pressure and have arbitrary sign. For compressional problems, the total pressure is much greater than lithostatic, causing the Drucker-Prager yield criterion to predict (locally) a very strong material. Consequently, we conclude that the shear band orientation has little to do with an intrinsic property of the material, but rather is defined by the geometry of the heterogeneity used to initiate localization and the pressure field associated with the heterogeneity.

Our experiments are instantaneous and thus independent of time. Consequently our models do not introduce any weakening mechanisms via strain softening applied to either the cohesion C , or friction angle φ . Softening is commonly used to mimic the behavior that rocks which have undergone failure become weaker than they were in their original (pristine) form [Lavie et al., 1999]. Strain weakening is a function of the strain rate invariant $\dot{\epsilon}_{II}$. Hence, given a symmetric heterogeneity and symmetric boundary conditions (about some plane), the nonlinear \mathbf{v}, p solution and thus $\dot{\epsilon}_{II}$ must also be symmetric about the same plane. Consequently, strain weakening cannot introduce asymmetric shear bands within a converged nonlinear solution. In our calculations, the symmetry plane is $x=L/2$, and we observe that all converged solutions exhibit symmetry of p and $\dot{\epsilon}_{II}$ about this plane. While this argument does not rule out eventual physical instabilities that break symmetry in time-dependent calculations, we note that *nonconverged* solutions (e.g., α -Drucker-Prager, $\alpha \geq 0.5$ – Figure 8, or supporting information FlipBook S2) display asymmetry even when the instantaneous solution should be symmetric. If this solution was evolved forward in time and strain weakening was employed, one of the two active shear bands would become weaker and localize more than its counterpart. This may be interpreted as a form of mode selection behavior, due to noise or some other physical process, however the “noise” in this instance is purely numerical and stems from integrating a nonconverged nonlinear solution forward in time. Again, to have a meaningful discussion of shear band evolution requires demonstrating that discrete solutions are converged at all times. Last we note that introducing strain weakening within the friction angle serves to reduce the overall pressure dependence within η_p , thus rendering the nonlinear Drucker-Prager rheologies potentially more solvable.

The desire to compare converged nonlinear solutions involving incompressible flow and a Drucker-Prager rheology is problematic. From our experiments, we find that Newton fails to converge with full Drucker-Prager viscoplasticity when the plastic layer has a high initial strength (and is thus closer to the purely plastic limit). We have explored various means to assist the convergence of Newton using continuation strategies similar to that adopted in Figure 8 where we incrementally increased the α parameter. We have also tried incrementally increasing the background velocity (U_0), or reference viscosity (η_1), but all approaches are found to lead to failure of the Newton solver as the incremental value (U'_0, η'_1) approaches the desired value (U_0, η_1). We associate the failure of Newton with the fact that the Jacobian defined via Drucker-Prager (Appendix A) becomes increasingly ill-conditioned in the terminal phase of the continuation

Table 2. Maximum Values of $|\partial\tau/\partial p'|$ Obtained in the Drucker-Prager Simulations Shown in Figure 5c

η_1	U_0	$\ F\ _2$	$ \partial\tau/\partial p' _{\max}$	$1 - \partial\tau/\partial p' _{\max}/\Theta_{\max}$	Θ^*	Converged
1×10^{23}	2.5	2.994490e-12	0.585155	1.724654e-01	1.299030e-02	Yes
1×10^{23}	5.0	8.908157e-12	0.668862	5.408613e-02	6.526998e-03	Yes
1×10^{23}	12.5	5.420185e-13	0.700951	8.705113e-03	2.618496e-03	Yes
1×10^{24}	2.5	1.011092e-11	0.706848	3.660920e-04	1.310535e-03	Yes
1×10^{24}	5.0	3.825603e-02	0.707075	4.526265e-05	6.555899e-04	No
1×10^{24}	12.5	9.339700e-02	0.707104	3.801896e-06	2.623133e-04	No
5×10^{24}	2.5	1.520565e-01	0.707106	1.275644e-06	2.623133e-04	No
5×10^{24}	5.0	2.167817e-01	0.707107	2.509049e-07	1.311696e-04	No
5×10^{24}	12.5	1.365168e+01	0.707107	3.486708e-10	5.247093e-05	No

process. While failure of Newton due to a singular Jacobian is not equivalent to the nonexistence of solutions, analysis of Stokes systems with pressure-dependent viscosities [Lanzendörfer, 2011; Hirn et al., 2012] suggests that there will be a limited range of pressure-dependent rheologies that admit well-posed solutions to incompressible Stokes solutions. The key bound appears to be that $|\partial\tau/\partial p'| \leq \Theta$ for some constant $\Theta < 1$, i.e., the change in deviatoric stress with pressure must be bounded. Appendix B provides some analysis based on the rheologies used here and shows that

$$\left| \frac{\partial\tau}{\partial p'} \right| = \sqrt{2\alpha B} \left[\frac{1}{1 + \eta_p/\eta_1} \right]^2 \quad (38)$$

which obtains a maximum value of $\Theta_{\max} = \sqrt{2\alpha B}$ in the perfectly plastic limit ($\eta_1 \gg \eta_p$). Table 2 tabulates $|\partial\tau/\partial p'|_{\max}$ and its relative difference from Θ_{\max} for the nine full Drucker-Prager runs shown in Figure 5c. While there is no sharp cut-off, for the problems shown here, those with a maximum value of equation (38) within a factor of $\sim 10^{-5}$ of Θ_{\max} do not converge using a combined Picard-Newton iteration scheme. The exact bound is likely to be problem-dependent, however, given the almost one-to-one correspondence between the theoretical ill-posedness bound provided by equation (38) and failure of our nonlinear solver, we boldly suggest that for these models, lack of convergence reflects ill-posedness. While in general, failure of Newton does not mean that a problem is ill-posed, the available analysis (Appendices B and C), together with our experimental results (see supporting information) that illustrate how Newton fails for this problem support this idea.

Moreover we suggest a very rough heuristic for predicting which problems may be ill-posed (and thus non-solvable) for full Drucker-Prager plasticity is to calculate

$$\Theta^* = 1 - \left[\frac{1}{1 + A\ell/(U_0\eta_1)} \right]^2, \quad (39)$$

where $A\ell/U_0$ is an estimate of the plastic viscosity at strain rate U_0/ℓ and ℓ is a measure of the width of the shear bands. Table 2 shows values of Θ^* calculated using $\ell = r = 0.02H$ where r is the radius of the smoothed corners in the basal heterogeneity. For problems with sharp discontinuities, we suggest using the grid spacing as a measure of ℓ . While not a particularly sharp estimator, Θ^* is easy to calculate and may provide some insight into whether a given problem may present convergence or solvability issues. We also note that the above analysis is based on the inverse viscosity mixing model which regularizes this problem. Initial numerical experiments suggests that convergence is even more difficult using the minimum viscosity model (equation (24)) as it is likely that the purely plastic limit is singular (Appendix C provides a concrete example). This raises additional issues as the inverse viscosity mixing model, where it does converge, is likely to produce weaker shear-bands as it allows some plastic deformation below the yield surface.

For von Mises ($B = 0$) or depth-dependent von Mises ($\alpha = 0$) plasticity models, however, there is no dependence of the effective viscosity on the dynamic pressure p' and $|\partial\tau/\partial p'|_{\max} = 0$ leading to overall better convergence behavior. Moreover, since the dynamic pressure arising from incompressible Stokes can lead to strongly nonlithostatic pressures near heterogeneities, we suggest that depth-dependent von Mises viscoplasticity is a more conservative plasticity model if the primary design is to limit plastic behavior to shallow

depths. However, it is only a reasonable approximation if the total pressure does not deviate significantly from lithostatic.

6.1. Viscoelastic-Plastic Models

Another approach to model brittle failure includes adding quasi-static elasticity into the description of the rheology [e.g., *Moresi et al.*, 2003; *Gerya and Yuen*, 2007; *Gerya*, 2010]. The inclusion of elasticity has been suggested to improve the convergence behavior of Newton solvers [*Kaus*, 2010]. While this may be the case, we wish to emphasize that viscoelasto-plasticity is a fundamentally different problem than viscoplasticity as it introduces time dependence in the form of stress evolution. Thus demonstrating solvability requires showing convergence in *both* the nonlinear residual and in time. A full discussion is beyond the scope of this paper, however, we feel it is important to at least outline the issues here.

As described in *Moresi et al.* [2003], the basic approach is to solve an approximation of the Maxwell viscoelastic rheology

$$\dot{\epsilon} = \frac{\tau}{2\eta_{\text{eff}}} + \frac{1}{2G} \frac{D\tau}{Dt} \quad (40)$$

along with the basic quasi-static force balance given by equation (1). Here G is the elastic shear modulus and η_{eff} is the effective viscoplastic rheology as previously defined. Equation (40) is inherently time-dependent and requires tracking the material derivative of the stress tensor with time. However, this equation is usually solved by applying finite differences in time to approximate the last term giving

$$\dot{\epsilon} \approx \frac{\tau}{2\eta_{\text{eff}}} + \frac{1}{2G\Delta t} (\tau - \tau^0), \quad (41)$$

where Δt is the discrete time step and τ^0 is the deviatoric stress tensor at the previous time step along a characteristic of the velocity field. Equation (41) can be rearranged as

$$\tau = 2\eta^* \dot{\epsilon} + \frac{\eta^*}{G\Delta t} \tau^0, \quad (42)$$

where

$$\eta^* = \frac{G\Delta t \eta_{\text{eff}}}{G\Delta t + \eta_{\text{eff}}} = \frac{\eta_{\text{eff}}}{1 + t_m/\Delta t} \quad (43)$$

is a new effective viscosity that incorporates the discrete time step (Δt) and the Maxwell relaxation time, $t_m = \eta_{\text{eff}}/G$. When the stress from equation (42) is substituted into the incompressible Stokes equations (equation (1)), one obtains a Stokes problem with an effective viscosity η^* and a forcing term due to the stress state from time $t - \Delta t$. The key effect here is that for time steps smaller than the Maxwell relaxation time, the effective viscosity is reduced by a term of size $1/(1 + t_m/\Delta t)$ which in general will make the instantaneous Stokes system easier to solve and should improve the convergence behavior of Newton solvers. However, this is a slightly misleading result. In the limit $\Delta t \rightarrow 0$, the solution is not controlled by viscous strain at all but by the elastic response and prior stress. Because this problem is fundamentally time-dependent, proper numerical convergence requires that not only the individual nonlinear Stokes solves converge, but that the whole problem is time-convergent in that the solution is insensitive to time-step and yields a consistent, stable and well-posed solution for some fixed time interval $T \gg \Delta t$. It may well be that including Maxwell viscoelasticity is more realistic and resolves some of the solvability issues with incompressible viscoplasticity, however, this needs to be demonstrated thoroughly and is beyond the scope of this paper.

7. Conclusions

Including some form of localized weakening is essential for geodynamic models of lithospheric deformation and global convection with plate tectonics. However, developing physically reasonable, tractable, and *solvable* continuum descriptions of this process remains an open challenge. The purpose of this paper is not to introduce new plasticity models, but rather to elucidate some of the critical mathematical issues arising from the most common viscoplasticity models used in geodynamics. We first want to emphasize that

viscoplasticity plus incompressible Stokes is a highly nonlinear problem and any discussion of results requires, at the very minimum, demonstration that the discrete solution is converged. This requires both a reasonable scaling for the problem and a small nonlinear residual e.g., $\|\mathbf{F}\|_2 < 10^{-7}$. For the problems discussed here, Picard iterations alone can lead to slow convergence and/or stalling at large nonlinear residuals that have significant errors relative to well-converged solutions. However, a combined Picard-Newton solver can improve convergence significantly for viscoplasticity models that do not depend on the dynamic pressure (e.g., von Mises or depth-dependent von Mises).

Additional computational problems exist for viscoplastic models that couple the dynamic pressure to the yield stress (e.g., Drucker-Prager). These models increase the degree of nonlinearity to the point where Newton methods often fail due to singular Jacobians. A characteristic feature of these nonconverged solutions is spurious shear bands and asymmetry (for numerically symmetric problems) which can pollute further subsequent solutions in a time-dependent problem. Moreover, theoretical analysis of incompressible Stokes problems with a pressure-dependent viscosity indicate that as these problems approach the purely plastic limit, they are fundamentally ill-posed and thus have no unique solutions [e.g., Lanzendörfer, 2011; Hirn *et al.*, 2012, see also Appendix C]. For the simple representative lithospheric dynamics model used in this paper, layers with an intrinsically large reference viscosity ($> 10^{24}$ Pa s) resulted in nonlinear problems which could not be solved using any combination of Picard, or Newton's method. Based on the analysis of Lanzendörfer [2011], we proposed a heuristic, which, for the test problem considered, provides a measure of whether the Drucker-Prager rheology will result in an ill-posed problem. Moreover, we suggest that for this problem, the failure of Newton is indicative of ill-posedness (although this remains to be proven rigorously).

It should be stressed that Drucker-Prager viscoplasticity can admit well-posed problems and thus fully converged solutions are possible to obtain in specific parameter ranges. However, these solutions have additional consequences for the pressure field that need to be assessed to determine whether the full solution for both velocity and pressure is geologically meaningful. For the compression tests shown here, Drucker-Prager viscoplasticity actually acts as a strengthening mechanism and can produce dynamic pressures that are comparable to the lithostatic pressure. If these pressure fields have observational consequences (e.g., for metamorphic phase assemblages), then they would significantly change our inferences about the depths of burial if the pressure was assumed to be lithostatic. In general, incompressible Stokes equations solve for *both* velocity and pressure and it is always important to visualize and evaluate p' , particularly if it is coupled to other constitutive or thermodynamic relationships.

Of the viscoplastic models explored here, depth-dependent von Mises provides the simplest, tractable model that attempts to restrict plastic failure to shallow depths. However, this approximation will only be consistent for problems where the dynamic pressure is small compared to lithostatic pressure. Moreover, fully converged solutions of this model suggest that the angle of the shear bands is not necessarily controlled by the "internal friction angle" ϕ , but by the pressure-induced viscosity field.

For clarity, we have just explored the behavior of viscoplastic models coupled to incompressible Stokes which describe only some of the existing geodynamic models. Viscoelastic-plastic models, are also commonly used and may resolve some of the computational issues involved with viscoplasticity. These problems, however, are fundamentally time-dependent and a similar study needs to demonstrate that solutions converge both nonlinearly and in time. In addition, future work should carefully evaluate the relative merits of using alternative plasticity models developed in the engineering community together with the inclusion of elasticity and potentially relaxing the notion of incompressibility. The search for a rheology which is simultaneously geologically meaningful and solvable remains an open challenge for the lithospheric dynamics community.

Appendix A: Discretization Using Finite Elements

In finite elements, we solve the weak form of the governing equations (equations (23) and (2)) assuming a discrete function space for the system $\mathbf{x} = (\mathbf{v}, p')$. For this paper, we use the stable mixed element $\mathbf{x} \in (P_2 \times P_2, P_1)$ with quadratic velocities and linear continuous pressures (i.e., stable Taylor Hood elements).

Since our general problem is nonlinear, we write the weak form in terms of a nonlinear residual for the momentum equation (\mathbf{F}_v) and continuity equation ($F_{p'}$):

$$\begin{aligned} \mathbf{F}_v = & \int_{\Omega} \left[2\eta_{\text{eff}} \dot{\epsilon}(\mathbf{v}_t) : \dot{\epsilon}(\mathbf{v}_t) - p'_t \nabla \cdot \mathbf{v}_t - \mathbf{v}_t \cdot \mathbf{f}_m \right] dV \\ & - \int_{\Gamma_N} \mathbf{v}_t \cdot \tilde{\boldsymbol{\sigma}} dS, \\ F_{p'} = & - \int_{\Omega} p'_t \nabla \cdot \mathbf{v}_t dV, \\ \mathbf{F} = & \mathbf{F}_v + F_{p'}, \end{aligned} \quad (\text{A1})$$

where $\mathbf{x}_i = (\mathbf{v}_i, p'_i)$ is the velocity and dynamic pressure at iteration i , and $\mathbf{x}_t = (\mathbf{v}_t, p'_t)$ are test functions for the mixed function space. The surface traction ($\tilde{\boldsymbol{\sigma}}$) associated with the pressure split form introduced in the above momentum residual is given by:

$$\tilde{\boldsymbol{\sigma}} = (2\eta_{\text{eff}} \dot{\epsilon} - p'_t \mathbf{I}) \cdot \mathbf{n}.$$

To impose a free surface boundary condition, $\boldsymbol{\sigma} \cdot \mathbf{n} = \mathbf{0}$ on Γ_N with the pressure split formulation [e.g., *Kramer et al.*, 2012], we require that

$$\tilde{\boldsymbol{\sigma}} = p_{\text{lith}} \mathbf{I} \cdot \mathbf{n}.$$

For the instantaneous models performed here, we used a model domain with a topographic profile which was horizontal, and we defined the datum associated with the lithostatic pressure such that $p_{\text{lith}} = 0$ on the upper surface (Γ_N). Under these conditions, $\tilde{\boldsymbol{\sigma}}$ will equal zero. However, we wish to emphasize that this is not a general result. For instance, if a different datum was used, or if the free surface was not horizontal, for example if temporal evolution was considered, then $\tilde{\boldsymbol{\sigma}} \neq 0$.

In equation (A1), \mathbf{F} is the weak form of the entire nonlinear residual, and if $\mathbf{x}_i, \mathbf{x}_t$ are chosen to be in a discrete function space, then \mathbf{F} assembles into a real vector of length $N_v + N_{p'}$ where $N_v, N_{p'}$ are the number of degrees of freedom of the velocity and dynamic pressure, respectively. Given \mathbf{F} , the weak form of the Jacobian can be calculated analytically (or with automatic differentiation) using equation (35). Again, if we consider discrete function spaces, the assembled Jacobian and nonlinear residual within the Newton update

$$\begin{bmatrix} \mathbf{F}_{v,v} & \mathbf{F}_{v,p'} \\ \mathbf{F}_{p',v} & \mathbf{F}_{p',p'} \end{bmatrix} \begin{bmatrix} \delta \mathbf{v} \\ \delta p' \end{bmatrix} = - \begin{bmatrix} \mathbf{F}_v \\ \mathbf{F}_{p'} \end{bmatrix}, \quad (\text{A2})$$

where, for example $\mathbf{F}_{v,v}$ is the functional derivative of \mathbf{F}_v with respect to the velocity. Due to the absence of any nonlinear coupling in the conservation of mass, the term $\mathbf{F}_{p',v}$ reduces to the discrete divergence operator (\mathbf{G}^T) and $\mathbf{F}_{p',p'} = \mathbf{0}$. For a linear problem where the viscosity does not depend on the current solution (i.e., is independent of both \mathbf{v} and p'), the Jacobian reduces to the standard linear Stokes operator

$$\mathbf{J} = \begin{bmatrix} \mathbf{K} & \mathbf{G} \\ \mathbf{G}^T & \mathbf{0} \end{bmatrix}, \quad (\text{A3})$$

where \mathbf{K} is the discrete divergence of the deviatoric stress tensor and \mathbf{G} is the discrete gradient operator.

When the viscosity depends on velocity (through the strain-rate invariant) and/or pressure, then additional blocks appear in the Jacobian. For our viscoplastic problem with an α -Drucker-Prager plasticity model, the Jacobian becomes

$$\mathbf{J} = \begin{bmatrix} \mathbf{K}^* & \mathbf{G} + \mathbf{G}^* \\ \mathbf{G}^T & \mathbf{0} \end{bmatrix}, \quad (\text{A4})$$

where \mathbf{K}^* is the matrix that arises from assembly of the bilinear form

$$(\dot{\epsilon}_t, \delta \mathbf{v}) = \int_{\Omega} \dot{\epsilon}_t : \frac{\partial \boldsymbol{\tau}}{\partial \mathbf{v}} \delta \mathbf{v} dV \quad (\text{A5})$$

which includes \mathbf{K} along with entries due to the strain rate dependence of the viscosity. The matrix \mathbf{G}^* arises from assembly of the bilinear form

$$(\dot{\epsilon}_t, \delta p') = \int_{\Omega} \dot{\epsilon}_t : \frac{\partial \tau}{\partial p'} \delta p' dV \quad (\text{A6})$$

which depends on derivatives of the deviatoric stress tensor with dynamic pressure (but not lithostatic pressure as that is not a variable in the linear solve). Analysis of Stokes systems with pressure-dependent viscosities [Lanzendörfer, 2011] shows that solutions to the continuous problem only exist when $|\partial \tau / \partial p'| \leq \Theta$ for some constant $\Theta < 1$. This is supported by our numerical results that indicate that as $|\partial \tau / \partial p'|$ approaches its maximum value for any given problem (see Appendix B) then the Jacobian becomes extremely ill-conditioned, if not singular, and Newton fails to find a solution.

Appendix B: Solvability analysis

In our experiments, we defined the deviatoric stress as

$$\tau_{ij} = 2\eta_{\text{eff}} \dot{\epsilon}_{ij}, \quad (\text{B1})$$

where the viscoplastic effective viscosity in the upper layer was given by

$$\eta_{\text{eff}} = \left[\frac{1}{\eta_1} + \frac{2\dot{\epsilon}_{II}}{A + B(p_{\text{lith}} + \alpha p')} \right]^{-1} \quad (\text{B2})$$

For stability (e.g., to maintain ellipticity), analysis suggests that

$$\left| \frac{\partial \tau}{\partial p'} \right| \leq \Theta \quad (\text{B3})$$

where $\Theta < 1$ is some constant [Lanzendörfer, 2011]. This quantity is directly related to the magnitude of the off-diagonal block \mathbf{G}^* in the Jacobian (equation (A4)) which affects the convergence behavior of Newton's method. For the composite rheology of equation (B2), we can calculate $|\partial \tau / \partial p'|$ exactly. From equations (B1) and (B2), we have that

$$\frac{\partial \tau_{ij}}{\partial p'} = \left[\frac{2\eta_1}{(A + B(p_{\text{lith}} + \alpha p')) + 2\eta_1 \dot{\epsilon}_{II}} \right]^2 \dot{\epsilon}_{II} \alpha B \dot{\epsilon}_{ij} \quad (\text{B4})$$

and using the tensor norm

$$|\mathbf{A}| = \sqrt{A_{ij} A_{ij}},$$

we have that

$$\left| \frac{\partial \tau}{\partial p'} \right| = \sqrt{2} \alpha B \left[\frac{2\eta_1 \dot{\epsilon}_{II}}{(A + B(p_{\text{lith}} + \alpha p')) + 2\eta_1 \dot{\epsilon}_{II}} \right]^2. \quad (\text{B5})$$

Thus, equation (B3) can be nicely expressed as the following inequality

$$\sqrt{2} \alpha B \left[\frac{\eta_1}{\eta_p + \eta_1} \right]^2 \leq \Theta \quad (\text{B6})$$

Equation (B6) has an upper bound $\Theta_{\text{max}} = \sqrt{2} \alpha B$ in the limit of strong layers with $\eta_1 \gg \eta_p$. A lower bound of 0 is attained for $\alpha B = 0$ i.e., either von Mises or depth-dependent von Mises plasticity, or in the limit $\eta_p / \eta_1 \rightarrow \infty$ which is essentially isoviscous. Rheologies with $|\partial \tau / \partial p'| = 0$ do not have the extra \mathbf{G}^* block in the Jacobian and converge under Newton iteration. For pressure-dependent rheologies, we find that problems where $|\partial \tau / \partial p'| \rightarrow \Theta_{\text{max}}$ do not converge. For a fixed value of $\alpha B \neq 0$, a measure of how close the problem is to its upper bound is

$$\Theta^* = 1 - \frac{\sqrt{2} \alpha B \left[\frac{\eta_1}{\eta_p + \eta_1} \right]_{\text{max}}^2}{\Theta_{\text{max}}} = 1 - \left[\frac{1}{1 + \eta_p / \eta_1} \right]_{\text{max}}^2 \quad (\text{B7})$$

which is calculated in Table 2.

Appendix C: Analysis of a Viscoplastic Layer in Simple Shear

This appendix provides analysis of a related problem of a viscoplastic layer in simple shear. We show that the problem is singular for a purely plastic rheology where von Mises plasticity admits an infinite number of solutions while Drucker-Prager (or any depth-dependent yield criterion) has no solutions. We then show that the inverse viscosity mixing model, equation (28) regularizes the problem such that the von Mises model has unique solutions for all values of parameters while Drucker-Prager is well-posed only for a range of parameters. Moreover, this analysis suggests that for the parameters where the regularized DP *compression* model fails, the simple shear model is ill-posed.

Consider simple shear flow in a layer of thickness h , with $\mathbf{u}=(u, v)$, $\mathbf{x}=(x, y)$. Let the shear flow be in the x -direction, and layer height in y . Moreover, let $v = 0$ and all the remaining variables (u, p, η) are only functions of y . With these approximations, the 2-D governing equations for incompressible Stokes

$$\frac{\partial}{\partial x} \left(2\eta \frac{\partial u}{\partial x} \right) + \frac{\partial}{\partial y} \left(\eta \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \right) - \frac{\partial p}{\partial x} = 0, \quad (C1)$$

$$\frac{\partial}{\partial x} \left(\eta \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \right) + \frac{\partial}{\partial y} \left(2\eta \frac{\partial v}{\partial y} \right) - \frac{\partial p}{\partial y} = \rho g, \quad (C2)$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0, \quad (C3)$$

reduce to

$$\frac{d}{dy} \left(\eta \frac{du}{dy} \right) = 0, \quad (C4)$$

$$\frac{dp}{dy} = -\rho g, \quad (C5)$$

and the continuity equation is automatically satisfied. Integrating the y -momentum equation over the layer thickness, we obtain

$$p = \rho g(h - y),$$

such that $p = 0$ at the top of our 1-D model domain ($y = h$).

C1. Perfectly Plastic Media

Now, consider a perfectly plastic material in which

$$\eta = \frac{Y}{2\dot{\epsilon}_{II}}, \quad (C6)$$

where $\dot{\epsilon}_{II}$ is given by equation (9) and $Y(y)$ is any yield criterion. For the simple shear flow problem, we have that

$$\dot{\epsilon}_{II} = \sqrt{\dot{\epsilon}_{xy}^2} = \frac{1}{2} \left| \frac{du}{dy} \right|.$$

Substituting the viscosity into equation (C4) yields

$$\frac{d}{dy} \left(Y \left| \frac{du}{dy} \right|^{-1} \frac{du}{dy} \right) = 0. \quad (C7)$$

This resembles the p -Laplacian equation

$$-\nabla \cdot (|\nabla u|^{p-2} \nabla u) = 0, \quad (C8)$$

with $p = 1$ which is known to be degenerate for this value of p [Evans, 2007]. To see the degeneracy, we integrate equation (C7) once to yield

$$Y \left| \frac{du}{dy} \right|^{-1} \frac{du}{dy} = \bar{\tau}, \quad (\text{C9})$$

where $\bar{\tau}$ is the *constant* shear stress in the layer, independent of height y . However, the term

$$\left| \frac{du}{dy} \right|^{-1} \frac{du}{dy} = \begin{cases} +1 \\ -1 \end{cases} = \text{sgn} \left(\frac{du}{dy} \right) \quad (\text{C10})$$

depending on whether the function u is strictly monotonically increasing (+1), or decreasing (−1).

For a constant yield, criteria such as von Mises ($Y = A$), equation (C9) states that the layer will be everywhere at the yield stress $\bar{\tau} = Y$ for *any* strictly monotonically increasing or decreasing function u . I.e., a perfectly plastic layer with von Mises yield criterion has an infinite number of solutions such that $\bar{\tau} = A$. However, if Y is a function of y , such as a Drucker-Prager model with $Y^{DP} = A + Bp(y)$, then there are clearly no solutions u that satisfy equation (C9).

C2. Viscoplastic Media Using the Mixing Model

If we change our viscosity from purely plastic to the inverse viscosity mixed model (see equation (26)) with a constant background viscosity η_0 , we obtain

$$\eta = \left[\frac{1}{\eta_0} + \frac{1}{Y} \right]^{-1} = \left[\frac{1}{\eta_0} + \frac{1}{Y} \left| \frac{du}{dy} \right| \right]^{-1}. \quad (\text{C11})$$

Inspection of equation (C11) indicates that the mixing model effectively regularizes the problem as η is defined for any value of du/dy .

It is straightforward to show that the regularized model can have unique solutions for the simple shear problem with boundary conditions $u(0) = 0$, $u(h) = u_0 > 0$. We begin by integrating equation (C4) and substituting equation (C11) to yield

$$\frac{du}{dy} = \bar{\tau} \left(\frac{1}{\eta_0} + \frac{1}{Y} \left| \frac{du}{dy} \right| \right). \quad (\text{C12})$$

If we just seek strictly monotonically increasing solutions ($du/dy > 0$) consistent with the boundary conditions, we can rearrange equation (C12) as

$$\left(1 - \frac{\bar{\tau}}{Y} \right) \frac{du}{dy} = \frac{\bar{\tau}}{\eta_0}. \quad (\text{C13})$$

In the limit $\eta_0 \rightarrow \infty$, the problem returns to the perfectly plastic layer which has an infinite number of solutions if $\bar{\tau} = Y$ (equation (C9)). However, for finite background viscosity, we can rewrite equation (C13) as

$$\frac{du}{dy} = \frac{1}{\eta_0} \frac{\bar{\tau} Y}{Y - \bar{\tau}}, \quad (\text{C14})$$

which admits strictly monotonic increasing solutions for all $0 < \bar{\tau} < \min(Y)$. To determine the constant shear stress $\bar{\tau}$ requires integrating both sides with respect to y and applying Dirichlet boundary conditions $u(0) = 0$, $u(h) = u_0$, i.e.,

$$u_0 = \frac{1}{\eta_0} \int_0^h \frac{\bar{\tau} Y}{Y - \bar{\tau}} dy \quad (\text{C15})$$

and solving the resulting implicit problem for $\bar{\tau}$.

C2.1. Regularized von Mises

For von Mises viscoplasticity $Y = A$, evaluating the integral in equation (C15) yields

$$\bar{\tau} = \frac{A \tau_0}{A + \tau_0}, \quad (\text{C16})$$

where $\tau_0 = \eta_0 u_0 / h$ is the reference shear stress for an isoviscous layer with viscosity η_0 in simple shear. Thus the constant shear stress in the layer goes smoothly from $\bar{\tau} = \tau_0$ if $\tau_0 \ll A$ (i.e., well below yield) to $\bar{\tau} = A$ for

which is the plastic limit. Given $\bar{\tau}$, one can integrate equation (C14) to obtain an expression for the velocity profile

$$u(y) = \frac{1}{\eta_0} \int_0^y \frac{\bar{\tau} Y(\zeta)}{Y(\zeta) - \bar{\tau}} d\zeta \quad (C17)$$

which for the von Mises models is simply a linear profile

$$u(y) = u_0 \frac{y}{h}. \quad (C18)$$

Thus, the regularized problem looks exactly like the simple shear profile for an isoviscous layer with effective viscosity

$$\eta = \frac{\eta_0 \eta_p}{\eta_0 + \eta_p}, \quad (C19)$$

where in this case the plastic viscosity is $\eta_p = Ah/u_0$. While the unregularized (perfectly plastic) model admits an infinite number of solutions, the inverse viscosity mixing model makes the problem unique.

C2.2. Regularized Drucker-Prager

Analysis for the regularized Drucker-Prager model ($Y^{DP} = A + Bp$, $p = \rho g(h - y)$), is similar, but shows that the problem becomes increasingly ill-posed as the layer stress approaches the minimum yield stress. Equation (C14) shows that at the top of the layer where $p = 0$, $Y = A$, and the velocity gradients $du/dy \rightarrow \infty$ as the layer stress approaches yield ($\bar{\tau} \rightarrow A$). In this regime, the upper boundary condition is effectively decoupled from the interior causing the solution to be ill-posed as $u(y)$ will be highly sensitive to small changes in parameters or boundary conditions.

We can quantify the parameter regimes where the regularized DP model is ill-posed. Integration of equation (C15) yields an implicit relation for $\bar{\tau}$

$$1 = \frac{\bar{\tau}}{\tau_0} \left[1 + \frac{\bar{\tau}}{Bp_{\max}} \log \left(\frac{A + Bp_{\max} - \bar{\tau}}{A - \bar{\tau}} \right) \right], \quad (C20)$$

where τ_0 is defined above and $p_{\max} = \rho gh$ is the pressure at the base of the layer. Alternatively if we scale all the stresses by τ_0 and define $\bar{\tau}' = \bar{\tau}/\tau_0$, $A' = A/\tau_0$ and $B' = Bp_{\max}/\tau_0$, equation (C20) becomes

$$1 = \bar{\tau}' \left[1 + \frac{\bar{\tau}'}{B'} \log \left(1 + \frac{B'}{A' - \bar{\tau}'} \right) \right]. \quad (C21)$$

Supporting information Figure S1 shows $\bar{\tau}'$ as a function of A' and B' calculated using a numerical root finder. Examination of equation (C21) (or (C20)) together with supporting information Figure S1 shows the following,

1. In the limit $B \rightarrow 0$, the solution reduces to equation (C16).
2. If we assume the layer stress is just below yield, i.e., $\bar{\tau}' = A' - \epsilon$ where ϵ is a very small parameter, then the largest value of B' that admits solutions for a given value of A' is given implicitly by

$$B' = \frac{(A' - \epsilon)^2}{1 - A' + \epsilon} \log \left[1 + \frac{B'}{\epsilon} \right] \quad (C22)$$

$$\approx \frac{A'^2}{1 - A'} \log \left[1 + \frac{B'}{\epsilon} \right] \quad \text{for } \epsilon \ll A' \quad (C23)$$

3. The problem is well-posed for $A' \geq 1$. (i.e., there are no bounds on B' in this regime which corresponds to low stress layers with $\tau_0 < A$).
4. For $A' < 1$ the problem becomes increasingly ill-posed for even small values of B' . Supporting information Figure S1 shows contours for several values of ϵ and shows that the numerical root finder fails for $\epsilon \lesssim 10^{-6}$. I.e., any value of A' , B' that falls in the white region of this figure is less than $10^{-6}\tau_0$ of yield and is ill-posed.
5. The blue dots in Figure S1, show values of A' and B' calculated for the compressive regularized Drucker-Prager models calculated in this paper. While this is a different problem, the figure suggests that parameters regimes where Newton fails for the compressive problem, are exceptionally ill-conditioned in the simple shear problem.

Acknowledgments

We thank Boris Kaus, Jed Brown and Anton Popov for insightful and useful reviews. We also thank Paul Tackley and the GFD group at ETH Zürich for providing a visiting professorship and generous hosting at ETH Zürich. U.S. National Science Foundation grants OCE-1358091, EAR-1141976, and EAR-1520732 provide partial funding for this work. This is LDEO contribution 8015. All supporting information required to reproduce the models and figures shown in this paper are available in the git repository <http://bitbucket.org/mspieg/plasticitymodels>.

References

- Anderson, E. M. (1905), The dynamics of faulting, *Trans. Edinburgh Geol. Soc.*, 8(3), 387–402.
- Balay, S., J. Brown, K. Buschelman, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang (2012a), PETSc Web page. [Available at <http://www.mcs.anl.gov/petsc>.]
- Balay, S., J. Brown, K. Buschelman, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang (2012b), *PETSc Users Manual*, Tech. Rep. ANL-95/11, Revision 3.3, Argonne Natl. Lab, Chicago.
- Buiter, S. J. H. (2012), A review of brittle compressional wedge models, *Tectonophysics*, 530, 1–17.
- Buiter, S. J. H., A. Y. Babeyko, S. Ellis, T. V. Gerya, B. J. P. Kaus, A. Kellner, G. Schreurs, and Y. Yamada (2006), The numerical sandbox: Comparison of model results for a shortening and an extension experiment, *Geol. Soc. Spec. Publ.*, 253(1), 29–64.
- Byerlee, J. (1978), Friction of rocks, *Pure Appl. Geophys.* 116(4–5), 615–626.
- Davis, T. A. (2004), Algorithm 832: UMFPACK V4.3, An unsymmetric-pattern multifrontal method, *ACM Trans. Math. Software*, 30(2), 196–199.
- Drucker, D. C., and W. Prager (1952), Soil mechanics and plastic analysis or limit design, *Q. Appl. Math.*, 10, 157–165.
- Evans, B., and C. Goetze (1979), The temperature variation of hardness of olivine and its implication for polycrystalline yield stress, *J. Geophys. Res.*, 84(B10), 5505–5524.
- Evans, L. (2007), The 1-Laplacian, the ∞ -Laplacian and differential games, in edited by H. Berestycki, and H. Brézis, *Perspectives in Nonlinear Partial Differential Equations: In Honor of Haim Brezis*, Contemporary mathematics. American Mathematical Society, Providence, R. I.
- Gerya, T. (2010), *Introduction to Numerical Geodynamic Modelling*, Cambridge Univ. Press, Cambridge, U. K.
- Gerya, T. V., and D. A. Yuen (2007), Robust characteristics method for modelling multiphase visco-elasto-plastic thermo-mechanical problems, *Phys. Earth Planet. Int.*, 163(1), 83–105.
- Ham, D. A., P. E. Farrell, G. J. Gorman, J. R. Maddison, C. R. Wilson, S. C. Kramer, J. Shipton, G. S. Collins, C. J. Cotter, and M. D. Piggott (2009), Spud 1.0: Generalising and automating the user interfaces of scientific computer models, *Geosci. Model Dev.*, 2(1), 33–42, doi:10.5194/gmd-2-33-2009.
- Hirn, A., M. Lanzendorfer, and J. Stebel (2012), Finite element approximation of flow of fluids with shear-rate- and pressure-dependent viscosity, *IMA J. Numer. Anal.*, 32(4), 1604–1634, doi:10.1093/imanum/drr033.
- Huismans, R. S., and C. Beaumont (2003), Symmetric and asymmetric lithospheric extension: Relative effects of frictional-plastic and viscous strain softening, *J. Geophys. Res.*, 108(B10), 2496, doi:10.1029/2002JB002026.
- Huismans, R. S., and C. Beaumont (2008), Complex rifted continental margins explained by dynamical models of depth-dependent lithospheric extension, *Geology*, 36(2), 163–166.
- Huismans, R. S., S. J. H. Buiter, and C. Beaumont (2005), Effect of plastic-viscous layering and strain softening on mode selection during lithospheric extension, *J. Geophys. Res.*, 110, B02406, doi:10.1029/2004JB003114.
- Kameyama, M., D. A. Yuen, and S.-I. Karato (1999), Thermal-mechanical effects of low-temperature plasticity (the Peierls mechanism) on the deformation of a viscoelastic shear zone, *Earth Planet. Sci. Lett.*, 168(1), 159–172.
- Karato, S. (2011), *Deformation of earth materials: An introduction to the rheology of solid earth*, Cambridge Univ. Press, Cambridge.
- Karato, S.-I., and P. Wu (1993), Rheology of the upper mantle: A synthesis, *Science*, 260(5109), 771–778.
- Kaus, B. J. P. (2010), Factors that control the angle of shear bands in geodynamic numerical models of brittle deformation, *Tectonophysics*, 484(1), 36–47.
- Kramer, S. C., C. R. Wilson, and D. R. Davies (2012), An implicit free surface algorithm for geodynamical simulations, *Phys. Earth Planet. Int.*, 194195, 25–37, doi:10.1016/j.pepi.2012.01.001.
- Lanzendorfer, M. (2011), Flows of incompressible fluids with pressure-dependent viscosity (and their application to modelling the flow in journal bearing), PhD thesis, Charles Univ. in Prague, Prague, Czech Republic.
- Lavier, L. L., R. W. Buck, and A. N. B. Poliakov (1999), Self-consistent rolling-hinge model for the evolution of large-offset low-angle normal faults, *Geology*, 27(12), 1127–1130.
- Lemiale, V., H.-B. Mühlhaus, L. Moresi, and J. Stafford (2008), Shear banding analysis of plastic models formulated for incompressible viscous flows, *Phys. Earth Planet. Int.*, 171(1), 177–186.
- Logg, A. (Ed.) (2012), *Automated Solution of Differential Equations by the Finite Element Method The FEniCS Book*, Springer, Berlin.
- Moresi, L., and V. Solomatov (1998), Mantle convection with a brittle lithosphere: Thoughts on the global tectonic styles of the earth and venus, *Geophys. J. Int.*, 133(3), 669–682.
- Moresi, L., F. Dufour, and H.-B. Mühlhaus (2003), A Lagrangian integration point finite element method for large deformation modeling of viscoelastic geomaterials, *J. Comput. Phys.*, 184(2), 476–497.
- Moresi, L., H.-B. Mühlhaus, V. Lemiale, and D. May (2007), Incompressible viscous formulations for deformation and yielding of the lithosphere, *Geol. Soc. London Spec. Publ.*, 282(1), 457–472.
- Moresi, L.-N., and V. S. Solomatov (1995), Numerical investigation of 2d convection with extremely large viscosity variations, *Phys. Fluids*, 7(9), 2154–2162.
- Popov, A. A., and S. V. Sobolev (2008), SLIM3D: A tool for three-dimensional thermomechanical modeling of lithospheric deformation with elasto-visco-plastic rheology, *Phys. Earth Planet. Int.*, 171(1), 55–75.
- Ranalli, G. (1995), *Rheology of the Earth*, 2nd ed., Chapman & Hall, London, U. K.
- Scholz, C. H. (2002), *The Mechanics of Earthquakes and Faulting*, 2nd ed., Cambridge Univ. Press, Cambridge, U. K.
- Solomatov, V. S., and L.-N. Moresi (1996), Stagnant lid convection on Venus, *J. Geophys. Res.*, 101(E2), 4737–4753.
- Stein, C., J. Schmalzl, and U. Hansen (2004), The effect of rheological parameters on plate behaviour in a self-consistent model of mantle convection, *Phys. Earth Planet. Int.*, 142(3–4), 225–255, doi:10.1016/j.pepi.2004.01.006.
- Tackley, P. J. (2000), Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations, *Geochim., Geophys. Geosyst.*, 1(8), 1021.
- Tosi, N., et al. (2015), A community benchmark for viscoplastic thermal convection in a 2-D square box, *Geochim. Geophys. Geosyst.*, 16, 2175–2196, doi:10.1002/2015GC005807.
- Trompert, R., and U. Hansen (1998), Mantle convection simulations with rheologies that generate plate-like behaviour, *Nature*, 395(6703), 686–689.
- van Heck, H. J., and P. J. Tackley (2008), Planforms of self-consistently generated plates in 3D spherical geometry, *Geophys. Res. Lett.*, 35, L19312, doi:10.1029/2008GL035190.

- Vermeer, P. A., and R. de Borst (1984), Non-associated plasticity for soils, concrete and rock, in *HERON*, vol. 29, no. 3. University of Delft, Delft, Netherlands.
- von Mises, R. (1913), Mechanik der festen körper im plastisch-deformablen zustand, in *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, pp. 582–592.
- Wilson, C. R., and M. Spiegelman (2015), Software: TerraFERMA v1.0: The Transparent Finite Element Rapid Model Assembler. [Available at open-source repository: terraferma.github.io.]