Resolving thermomechanical coupling in two and three dimensions: spontaneous strain localization owing to shear heating
Thibault Duretz, L Räss, Y. Podladchikov, S Schmalholz

To cite this version:

HAL Id: insu-02060419
https://hal-insu.archives-ouvertes.fr/insu-02060419
Submitted on 7 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Resolving thermomechanical coupling in two and three dimensions: spontaneous strain localization owing to shear heating

T. Duretz,1,2 L. Räss,2,3 Y.Y. Podladchikov2,3 and S.M. Schmalholz 2

1Geosciences Rennes, Univ. Rennes 1, UMR CNRS 6118, 35042 Rennes, France. E-mail: thibault.duretz@univ-rennes1.fr
2Faculté des géosciences et de l’environnement, Institut des Sciences de la Terre, Université de Lausanne, 1015 Lausanne, Switzerland
3Swiss Geocomputing Centre, Université de Lausanne, 1015 Lausanne, Switzerland

Accepted 2018 October 18. Received 2018 September 19; in original form 2018 April 20

SUMMARY
Numerous geological processes are governed by thermal and mechanical interactions. In particular, tectonic processes such as ductile strain localization can be induced by the intrinsic coupling that exists between deformation, energy and rheology. To investigate this thermomechanical feedback, we have designed 2-D codes that are based on an implicit finite-difference discretization. The direct-iterative method relies on a classical Newton iteration cycle and requires assembly of sparse matrices, while the pseudo-transient method uses pseudo-time integration and is matrix-free. We show that both methods are able to capture thermomechanical instabilities when applied to model thermally activated shear localization; they exhibit similar temporal evolution and deliver coherent results both in terms of nonlinear accuracy and conservativeness. The pseudo-transient method is an attractive alternative, since it can deliver similar accuracy to a standard direct-iterative method but is based on a much simpler algorithm and enables high-resolution simulations in 3-D. We systematically investigate the dimensionless parameters controlling 2-D shear localization and model shear zone propagation in 3-D using the pseudo-transient method. Code examples based on the pseudo-transient and direct-iterative methods are part of the M2Di routines (Räss et al., 2017) and can be downloaded from Bitbucket and the Swiss Geocomputing Centre website.

Key words: Finite-difference methods; Continental deformation; Numerical techniques; Geodynamics.

1 INTRODUCTION
Thermomechanical feedback represents a first-order multiphysics coupling for geodynamic processes. For instance, thermomechanical coupling plays a major role in initiating and regulating convection currents at the scale of the Earth's mantle (Pekeris 1935; McKenzie et al. 1974; Parsons & McKenzie 1978). Thermal convection represents a type of Rayleigh–Bénard instability that is intrinsically linked to the temperature sensitivity of rock-forming mineral densities. Besides convection, thermally activated shear localization is another, yet far less explored, example of thermomechanical coupling in geodynamics. However, these processes have been used to explain the initiation of subduction (Regenauer-Lieb et al. 2001; Thielmann & Kaus 2012), the generation of deep earthquakes (Ogawa 1987; Hobbs & Ord 1988; Prieto et al. 2013; Ohuch & Schmalholz 2017), the formation of tectonic nappes (Jaquet & Schmalholz 2017) or the formation of thermally activated shear localization occurs when local temperature perturbations owing to shear heating (mechanical dissipation induced by irreversible deformation) are large enough not to be diffused away efficiently. Since ductile mineral strengths depend on temperature (e.g. Carter & Ave’Lallemant 1970), a local temperature increase results in thermal softening, which can further induce the focusing of strain into a localized shear zone (Yuen & Schubert 1979; Fleitout & Froidevaux 1980; Kaus & Podladchikov 2006). Since localization is driven by a self-regulating feedback process, it can either vanish, be maintained stably or evolve into a runaway instability given specific conditions (e.g. Rice & Fairbridge 1975; John et al. 2009; Braeck & Podladchikov 2007). Shear zones caused by shear heating have an inherent width (Duretz et al. 2014; Moore & Parsons 2015) in which the strain is focused although the thermal imprint may be diffused (Takeuchi & Fialko 2012; Schmalholz & Duretz 2015). Thermally activated shear localization is generally not an exclusive mechanism and can occur in conjunction with other physical processes such as microstructural evolution (Peters et al. 2015; Thielmann et al. 2015) and mineral reactions (Andersen et al. 2008) that can further promote strain localization.

In the following, we first proceed to a dimensional analysis and parameter reduction for thermomechanical equations, which are further used to model thermally activated shear localization. We
introduce two numerical modelling approaches, namely a classic direct-iterative (DI) method and a less conventional, pseudo-transient (PT) approach. The PT method’s viability is demonstrated by providing a quantitative comparison with numerical solutions achieved with the standard DI method. We assess both the accuracy and performance of PT solutions. We provide a systematic parametric analysis of 2-D thermally activated shear localization. Natural shear localization always occurs in 3-D, which is important for localization caused by thermal softening, since thermal diffusion of local heat sources is more efficient in 3-D than in 2-D. Yet, there are very few efficient 3-D algorithms that allow one to accurately simulate thermally activated strain localization in 3-D. Here, we show that the PT method is well suited for the efficient and accurate simulation of 3-D shear zone formation by thermal softening. Finally, we discuss how different numerical treatments of nonlinearities and time integration schemes can affect model predictions as well as performance. For reproducibility purposes, we provide the PT and DI numerical codes (MATLAB) used to solve thermo-mechanical problems. The MATLAB routines are part of M2Di (Räss et al., 2017) and are available for download from Bitbucket at https://bitbucket.org/lraess/m2di and from the Swiss Geocomputing Centre website http://wp.unil.ch/geocomputing/software/. The PT routines are located in the TM2Dpt folder and the DI routines (TM2Di) are located in the M2Di2 folder. The GPU (C-CUDA) routines are available upon request to the authors.

2 THE MATHEMATICAL MODEL

2.1 Thermomechanical coupling

The equations governing thermomechanics of slowly creeping incompressible power-law viscous fluids, in the absence of buoyancy forces, are

\[ \frac{\partial v_i}{\partial x_i} = 0, \]

\[ \frac{\partial \tau_{ij}}{\partial x_j} - \frac{\partial p}{\partial x_i} = 0, \]

\[ \tau_{ij} \dot{e}_{ij} + k \frac{\partial^2 T}{\partial x_i^2} - \rho C_p \frac{\partial T}{\partial t} = 0, \]

\[ \dot{e}_{ij} = \frac{1}{2} \left( \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) = \frac{1}{2} A \tau_{ij}^{n-1} \exp \left( \frac{-Q}{R T_0} \right) \tau_{ij}, \]

where \( v_i \) are components of the velocity vector in the \( x_i \) spatial direction, \( p \) is the pressure, \( T \) is the temperature deviation from the initial temperature \( T_0 \), \( \rho \) is the density, \( C_p \) is the specific heat, \( \dot{e}_{ij} \) is the strain rate tensor, \( \tau_{ij} \) and \( \tau_{ii} \) are the deviatoric stress tensor and the square root of its second invariant:

\[ \tau_{ii} = \sqrt{\frac{1}{2} \tau_{ij} \tau_{ij}}, \]

\( A \) is the pre-exponent, \( n \) is the stress exponent, \( Q \) is the activation energy and \( R \) is the universal gas constant. Since we mostly focus on small strains, we do not consider heat transport owing to advection.

Four independent scales

\[ \bar{T} = \frac{n R T_0^2}{Q}, \]

\[ \bar{\tau} = \rho C_p \bar{\dot{e}}, \]

\[ \bar{\dot{e}} = 2^{1-n} A^{-1} \tau^{-n} \exp \left( \frac{Q}{R T_0} \right) \text{ and} \]

\[ \bar{L} = \sqrt{\frac{k}{\rho C_p \bar{\tau}}} \]

for temperature, stress, time and length, respectively, and their dependent combinations, such as the velocity scale: \( \bar{V} = \bar{L} / \bar{t} \), are used to make all the variables dimensionless. Hereinafter, all variables are dimensionless, unless noted otherwise. Introducing dimensionless variables in eq. (1) results in the following dimensionless form of the governing equations (see Appendix Acc for details):

\[ \frac{\partial v_i}{\partial x_i} = 0, \]

\[ \frac{\partial \tau_{ij}}{\partial x_j} - \frac{\partial p}{\partial x_i} = 0, \]

\[ \tau_{ij} \dot{e}_{ij} + \frac{\partial^2 T}{\partial x_i^2} - \frac{\partial T}{\partial t} = 0, \]

\[ \dot{e}_{ij} = \frac{1}{2} \left( \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) = 2^{1-n} A^{-1} \exp \left( \frac{n T}{1 + \frac{T}{T_0}} \right) \tau_{ij}, \]

The model parameters are the dimensionless initial temperature \( T_0 \), the power-law exponent \( n \) as well as four parameters that arise from the initial and boundary conditions: the radius and the magnitude of the circular thermal perturbation, the length of the square computational domain, \( L \) and the boundary velocity, \( V_{BC} \) (Fig. 2).

3 THE NUMERICAL METHODS

The system of nonlinear equations (eq. 4) is discretized on a Cartesian staggered grid with regular grid-spacing. The time derivative of the heat equation is approximated by either a backward-Euler or a Crank–Nicolson scheme. The dimensionless viscosity is a nonlinear function of both the temperature and strain rate and is expressed, after eq. (4), as

\[ \eta = \bar{\epsilon}_{ii} \frac{1}{A \tau_{ii}} \exp \left( -\frac{T}{1 + \frac{T}{T_0}} \right), \]

where \( \bar{\epsilon}_{ii} \) is the square root of the second invariant of the deviatoric strain rate tensor:

\[ \bar{\epsilon}_{ii} = \sqrt{\frac{1}{2} \dot{e}_{ij} \dot{e}_{ij}}. \]

Obtaining a numerical solution that satisfies the nonlinear discrete thermomechanical equations is generally an iterative process. Arbitrary initial pressure, velocity and temperature fields will not satisfy the discrete equations as they would produce an imbalance. An implicit solution procedure will seek to iteratively reduce the imbalances until the thermomechanical equations are satisfied to a desired accuracy. To this end, the thermomechanical eq. (4) is
formulated as
\[ \frac{\partial \tau_{ij}}{\partial x_j} = f_r, \]
\[ \frac{\partial x_i}{\partial x_j} = f_p, \]
\[ \frac{\partial^2 T}{\partial x_i^2} + \tau_{ij} \delta_{ij} - \frac{\partial T}{\partial t} = f_t, \] (7)

where the \( \delta \) symbolises numerical approximation to partial derivatives. The right-hand side terms are nonlinear continuity, momentum and thermal residuals \( (f_p, f_r, f_t) \), which quantify the magnitude of the imbalance of the thermomechanical equations.

Two methods are employed to minimize the magnitude of these residuals and deliver accurate pressure, flow and temperature fields. The first method relies on Newton iterations and uses a DI scheme, thus requiring the assembly and factorization of sparse matrices. The second method uses a PT approach, which is fully iterative and matrix-free. Both numerical methods rely on a two-way coupling, since both the coupling term (shear heating) and rheology are treated implicitly.

3.1 The direct-iterative method

For the DI method, we employ a Newton scheme which allows us to obtain accurate nonlinear solutions within few iterations (see Appendix B). The numerical solution, \( x = [v, p, T]^T \), is iteratively corrected with the following update:
\[ x^{k+1} = x^k + \alpha \delta x^{k+1}, \] (8)
where the \( \delta \) operator stands for the correction of a quantity, \( \alpha \) is a scalar optimization parameter and \( k \) is the nonlinear iteration index. The Newton correction, \( \delta x = [\delta v, \delta p, \delta T]^T \), is obtained by applying the inverse of the Jacobian matrix, \( J_{TM} \), to the current nonlinear residual \( f = [f_v, f_p, f_T]^T \):
\[ \delta x^{k+1} = - J_{TM}^{-1} f^k. \] (9)
Prior to the solution update, we run a line search procedure to determine the optimization parameter \( \alpha \) (0 \( \leq \alpha \leq 1 \)) that yields:
\[ \min ||f (x^k + \alpha \delta x^{k+1})||_{L_2}. \] (10)

In the DI context, it is necessary to formulate and assemble the Jacobian matrix. The latter describes the gradient of the residuals with regard to the solutions; for example, the sensitivity of the momentum imbalance with regard to the velocity field. For the considered thermomechanically coupled flow, the Jacobian matrix takes the form of
\[ J_{TM} = \frac{\partial f}{\partial x} = \begin{bmatrix} J_{vv} & K_v & J_{vT} \\ K_{pv} & 0 & 0 \\ J_{Tv} & 0 & J_{TT} \end{bmatrix}. \] (11)

The \( J_{vv} \) corresponds to the mechanical Jacobian matrix, which arises from the strain rate dependence of viscosity. The \( J_{vT} \) block highlights the temperature dependence of viscosity and the \( J_{Tv} \) block arises from the linearization of the shear heating term. The \( K_v \) and \( K_{pv} \), respectively, represent the gradient operator and divergence discrete operators. The matrix \( J_{TT} \) is a modified Laplace operator that also includes contributions from the temperature dependence of the viscosity. The Newton corrections for the thermomechanical system can be formulated as
\[ \delta x = \begin{bmatrix} \delta v \\ \delta p \\ \delta T \end{bmatrix} = - \begin{bmatrix} J_{vv} & K_v & J_{vT} \\ K_{pv} & 0 & 0 \\ J_{Tv} & 0 & J_{TT} \end{bmatrix}^{-1} \begin{bmatrix} f_v \\ f_p \\ f_T \end{bmatrix}, \] (12)

and is obtained via DI procedure (see Appendix C).

The nonlinear iteration cycle is aborted once one of the following criteria \( ||f||_{L_2} / ||f||_{L_2}^{\text{initial}} < \theta_{\text{tol}} \) or \( ||f||_{L_2} < \theta_{\text{tol}}^{\text{nonlin}} \) is met; meaning that the thermomechanical balance equations are satisfied to the desired accuracy. The main steps of the DI approach are summarized in Fig. 1(a). This code is written in MATLAB language; it is based on and made available as part of the M2Di routines (Räss et al. 2017) under the name TM2Di (https://bitbucket.org/lraess/m2di).

3.2 The pseudo-transient method

The PT or relaxation method allows one to iteratively solve nonlinear problems in a single iteration loop in a matrix-free way. The relaxation method is a classical numerical technique to solve stationary (elliptic) problems (Frankel 1950). The method was extended in the 1960s to elastic problems (Otter et al. 1966) and more recently to elasto-plastic (Cundall 1982) and viscoelastic problems (Poliaikov et al. 1993). The PT method relies on introducing PT terms into steady-state equations. Given a set of initial and boundary conditions, solutions can be found by integrating the equations forward in pseudo-time (\( \tau \)) until steady state is attained; that is, the pseudo-time derivative vanishes. For example, the compressible Navier–Stokes equations incorporate right-hand side time derivatives for both the mass \( \left( \frac{\partial v_i}{\partial \tau} \right) \) and momentum \( \left( \frac{\partial \mathbf{v}}{\partial \tau} \right) \) balance equations. These latter represent the elastic bulk rheology (with \( \beta \) as compressibility) and the acceleration (with \( \rho \) as the density), respectively. A solution to the incompressible Stokes problem requires that both of these transient terms vanish.

The essence of the PT method is to integrate the balance equations in pseudo-time, \( \tau \), until the PT terms vanish. To this end, eq. (4) is expressed as
\[ \frac{\partial v_i}{\partial x_i} = \frac{\partial p}{\partial t} = \beta \frac{\partial v_i}{\partial \tau} - \beta \frac{\partial p}{\partial \tau}, \]
\[ \frac{\partial \tau_{ij}}{\partial x_j} = \frac{\partial p}{\partial x_i} = \beta \frac{\partial v_i}{\partial \tau}, \]
\[ \frac{\partial^2 T}{\partial x_i^2} + \tau_{ij} \delta_{ij} - \frac{\partial T}{\partial t} = \beta \frac{\partial T}{\partial \tau}. \] (13)

This approach is equivalent to iteratively reducing the magnitude of the residual of each equation (e.g. as in the DI method) since PT terms are equivalent to residuals.

The nonlinear viscosity \( \eta \) is evaluated at each PT iteration \( k \) using the current strain rate and temperature solution fields. The treatment of nonlinearities is greatly facilitated by using an effective viscosity \( (\eta_{\text{eff}}) \), which we formulate as
\[ \eta_{\text{eff}} k = \exp \left[ \theta_{\eta} \ln \left( \eta_{\text{eff}}^{k-1} \right) + \left( 1 - \theta_{\eta} \right) \ln \left( \eta^k \right) \right], \] (14)
where \( \theta_{\eta} \) (0 \( \leq \theta_{\eta} \leq 1 \)) corresponds to a relaxation factor. This approach is a continuation method, since the effective viscosity progressively relaxes towards the de facto physical viscosity \( (\eta_{\text{eff}} \rightarrow \eta) \) throughout the PT iterations.

The integration of the momentum, mass conservation and temperature evolution equations necessitates the definition of individual pseudo-time steps, \( \Delta \tau_v, \Delta \tau_p \) and \( \Delta \tau_T \). Hereinafter, we assume that \( \rho \) and \( \beta \) are equal to 1.0 and the pseudo-time steps are formulated
and \(\Delta_1\) the number of gridpoints in the use of the denominator of \(\Delta_1\tau\) production factors (step for viscous flow min(equations. It is constructed by multiplying the explicit CFL time which includes a numerical analogy to the bulk viscosity \(\eta\) as stands for the initial temperature, Figure 2.

Algorithmic flowchart for both methods used in the study: (a) the direct-iterative method (TM2Di code) and (b) the pseudo-transient method.

Figure 1. Algorithmic flowchart for both methods used in the study: (a) the direct-iterative method (TM2Di code) and (b) the pseudo-transient method.

\[
\begin{align*}
\Delta r_p &= \theta_p \frac{2.1 N_{\text{dim}} \eta_b^i (1 + \eta_b)}{\max(N_x, N_y)} , \\
\Delta r_{v_i} &= \theta_{v_i} \frac{\min(\Delta x, \Delta y)^2}{2.1 N_{\text{dim}} \eta_b^i (1 + \eta_b)} , \\
\Delta r_T &= \theta_T \frac{\min(\Delta x, \Delta y)^2}{2.1 N_{\text{dim}}} ,
\end{align*}
\]

where \(\eta_b\) is a numerical analogy of bulk viscosity, \(N_x\) and \(N_y\) are the number of gridpoints in the \(x\) and \(y\) direction, respectively, \(\Delta x\) and \(\Delta y\) are grid-spacings, \(\theta_{v_i}\), \(\theta_p\) and \(\theta_T\) are pseudo-time step reduction factors (\(\leq 1.0\)) and \(N_{\text{dim}}\) is the number of dimensions. \(\Delta r_T\) is the pseudo-time step used for explicit integration of diffusion equation satisfying the Courant–Friedrichs–Lewy condition (CFL). \(\Delta r_{v_i}\) is the pseudo-time step used for integrating the momentum equations. It is constructed by multiplying the explicit CFL time step for viscous flow \(\min(\Delta x, \Delta y)^2/(2.1 N_{\text{dim}} \eta_b^i)\) by \(1/(1 + \eta_b)\), which includes a numerical analogy to the bulk viscosity \(\eta_b\). The use of the denominator of \(\Delta r_{v_i}\) as nominator in the definition of \(\Delta r_T\) leads to an empirically derived pseudo-time step for the continuity equation. This criteria allows for an optimal convergence of the Stokes problem, making the iteration strategy less sensitive to the physical shear viscosity \(\eta\). A dimensional analysis confirms that the product of \(\beta^{-1} \Delta T_b\) [Pa s] by the divergence of velocities [s–1] produces dynamic pressure increments in [Pa]. Identical reasoning can be applied to the momentum balance equation, where \(\rho^{-1} \Delta \tau_v\) [m2 Pa–1 m–1] multiplies the force balance terms [Pa m–1] to produce increments of velocity [m s–1]. We further highlight that \(\eta_{\text{IT}}\) refers to entire fields (defined for every gridpoint); thus, pseudo-time step values are local to every gridpoint within the computational domain and analogous to the application of diagonal preconditioner in matrix-based solvers. At each PT iteration, the velocity, pressure and temperature fields are updated at each iteration using their current values of pseudo-time step and residual:

\[
\begin{align*}
p^k &= p^{k-1} + \Delta T_p f_p^k , \\
v^k &= v^{k-1} + \Delta r_{v_i} g_{v_i}^k , \\
T^k &= T^{k-1} + \Delta T_T f_T^k .
\end{align*}
\]

The use of damping greatly reduces the number of iterations needed for convergence of the PT iterations (Choi et al. 2013; Yang & Mittal 2014). To this end, the damped momentum residual \(g_{v_i}^k\) in eq. (16) is written as

\[
g_{v_i}^k = f_{v_i}^k + \left(1 - \frac{\nu}{N_i}\right) f_{v_i}^{k-1} 
\]

where optimal values of \(\nu\) reside within the range \((1 \leq \nu \leq 10)\) and \(N_i\) is the number of gridpoints in the direction \(i\). An analogue approach for elastic rheology is described by Cundall & Strack (1979) and is successfully used in the FLAC geotechnical software (Cundall et al. 1993). As for the DI method, the PT iterations are performed until one of the following criteria \(||f||_l^2 / ||f||_l^2\text{initial} < \text{tol_{nonlin}}\) or \(||f||_l^2 < \text{tol_{nonlin}}\) is verified.

3.3 Physical time integration

For both the DI and PT methods, the integration of the heat equation is done in physical time \(t\). An implicit (backward-Euler) or semi-implicit (Crank–Nicolson) solution is obtained by updating the heat fluxes and shear heating term at each nonlinear or PT iteration. In the following examples, we use a physical time step, \(\Delta t_p\), which is proportional to the CFL time step:

\[
\Delta t_T = \xi \frac{\min(\Delta x, \Delta y)^2}{2.1 N_{\text{dim}}} = \xi \Delta t_{\exp} ,
\]

where \(\xi\) corresponds to a time step ratio, \(\Delta t_T/\Delta t_{\exp}\). Despite the use of an implicit scheme, we did not obtain successful time integration of the nonlinear equations system for an arbitrary choice of time
4 A COMPARISON OF THE DIRECT-ITERATIVE AND THE PSEUDO-TRANSIENT METHODS

In this section, we demonstrate that both the DI method and the PT method can solve thermomechanical problems to the same accuracy level. We evaluate each method’s performance on current personal computers (single CPU and single GPU).

4.1 The reference model’s configuration

The numerical models were designed to study the propagation of shear zones owing to shear heating in a viscous medium subjected to far-field pure shear kinematics. For 2-D calculations, we consider a physical domain of dimensions $[0, L] \times [0, L]$. The normal boundary velocities are set to $v_x = -V_{BC}$ for $x = L$, $v_y = V_{BC}$ for $y = L$ and 0 elsewhere. Zero shear stress and zero heat flux boundary conditions are applied on all model sides. The initial temperature field, $T_0$, is set to 16.4423 and a perturbation of radius equals to 0.0857L and amplitude equals to 0.1T_0 is centred around the origin (Fig. 2).

The reference model is run for a boundary velocity of 66.4437, a model length of 0.86038 and a stress exponent of 3. For the 3-D models, the model domain extends up to $z = L$, where a zero normal boundary velocity, shear stress and heat flux are applied. Thus, the first model embeds an initial cylindrical (i.e. continuous in the $z$ direction) temperature perturbation and is similar to the 2-D model. In the second model, an initially spherical temperature perturbation is prescribed.

4.2 The temperatures and strain rates inside the shear zones

Thermally activated shear zones have inherent length scales that are proportional to material properties and loading conditions. Thus, it is possible to characterize dynamically evolving fields such as the temperature in such shear zones. If the numerical resolution is greater than the characteristic length scale, these modelling results are essentially independent of the numerical resolution (e.g. Duretz et al. 2014). A typical model evolution is depicted in Fig. 3, which shows the progression focusing of strain rate and temperature with time. Since the shear bands have a finite-length scale, it is possible to monitor the evolution of these quantities inside the shear band (Fig. 4). The strain rate overcomes the background strain rate by an order of magnitude within $1.5 \times 10^{-3}$ time units. The temperature increase follows a distinct evolution and tends towards a value of 3.5 for a model time of $3.0 \times 10^{-3}$. The simulations computed with the DI and PT methods clearly provide the same temperature and strain rate predictions (Fig. 4).

4.3 The nonlinear solvers’ accuracy

To show the PT method’s ability to handle nonlinear thermomechanical problems, we present a quantitative analysis of errors caused by the nonlinearity. A single time step numerical solution was computed with the DI method using the previously described setup and a resolution of $94^2$ numerical cells. The solution procedure was aborted once we attained machine precision for nonlinear residuals; the obtained effective strain rate, temperature and pressure serve as reference solution fields. We then computed a series of numerical solutions using larger nonlinear tolerances (iteration exit criteria) with both the PT and DI method.

The deviation of the numerical solutions with regard to the reference fields was calculated as

$$
\| E_{a} \|_{\ell^2} = \| \text{error}_{\text{numerical}} - \text{error}_{\text{reference}} \|_{\ell^2},
$$

where $a$ is either $\varepsilon$, $P$ or $T$. For the same nonlinear tolerance, the DI and PT provide the same deviation from the reference results (Fig. 5). The deviations obtained with either the DI or the PT decrease as the nonlinear tolerance is decreased. Thus, both methods converge towards the reference numerical solution with a linear trend. Nonlinear numerical solutions obtained with the PT method can reach the same accuracy level as those obtained with the DI method.

4.4 The conservation of energy

For thermomechanical problems, considering a purely viscous rheology, energy conservation postulates that mechanical work should be fully converted into heat. The numerical simulations’ accuracy depends on numerical schemes’ ability to conserve energy and therefore fulfill the energy conservation equation. The work per unit length is obtained by evaluating the following integrals (Green’s theorem):

$$
W = \int_V \int_{\Sigma} \tau_{ij} \partial_j \varepsilon_{ij} \, dV \, dt = \int_V \int_{S} \sigma_{ij} n_i v_j \, dS dt,
$$

where $n_i$ is the unit vector to the boundaries and $S$ is the surface of the domain. The thermal energy per unit length takes the form of

$$
E = \int_V \int_f \frac{\partial T}{\partial t} \, dV dt,
$$

where $V$ is the volume of the domain. The time evolution of $W$ and $E$ for 2-D numerical simulations using the configuration described in Section 4.1 is depicted in Fig. 6. The work and thermal energy computed from the PT and DI simulations all follow the same trend. For either methods, the work equals to thermal energy at any moment in time. Thus, the numerical solutions arising from the finite-difference discretization are conservative, independent of the solving procedure.

4.5 Performance

We evaluate the performance of the MATLAB CPU-based and C-CUDA GPU-based PT solvers using two different metrics. The first metric is an evaluation of the effective memory throughput (MTP\textsubscript{effective}). The second metric is a measurement of the wall-time needed to achieve convergence over a time step. Since the PT solvers perform stencil operations in a matrix-free approach, the memory transfers bound the algorithm and the number of floating-point operations per second are not affecting the performance. The used MTP\textsubscript{effective} metric (Omlin 2017) evaluates how efficiently data is transferred between the memory and the computation units, in Gigabytes per second (GB s\textsuperscript{-1})

$$
\text{MTP}_{\text{effective}} = \frac{(N_x \times N_y) \times N_t \times nO \times \text{precis}}{1 \times 9 \times \text{time}_{\text{lat}}},
$$

where $N_x$, $N_y$, and $N_t$ are the spatial and temporal size of the simulation, $nO$ is the number of operations per time step, and $\text{precis}$ is the precision of the data type.
Figure 3. Thermomechanical activated shear localization in 2-D. The model was run using the reference model parameter (see the main text) and initiated with a circular temperature perturbation (10% per cent increase of temperature). Pure shear was applied on boundaries normal to $x$ (inflow) and $y$ (outflow). All thermal boundary conditions were insulating (zero heat flow).

where $(N_x \times N_y)$ is the number of cells, $N_t$ is the number of time steps or iterations performed, nIO is the number of memory accesses performed, precis is the floating-point precision (either 4 or 8 bytes) and time$_{N_t}$ is the time (in seconds) needed to perform the $N_t$ steps. The number of memory accesses (nIO) defines the minimum number of read-and-write or read-only operations required to solve
Shear heating and localization

Figure 4. Strain rate amplification (upper panel) and temperature (lower panel) evolution in the shear zone (away from the model boundaries). Both PT and DI models were run with a relative tolerance of $10^{-9}$ and with backward-Euler time integration for the energy equation. For this example, PT models typically converge within about $10^5$ iterations.

Figure 5. Nonlinear accuracy PT and the DI method (TM2Di code). Errors in $P$, $T$ and $\dot{\varepsilon}_{II}$ were computed for different relative tolerance ($\epsilon_r$). The PT and the DI methods provide similar errors, and both converge linearly to the reference results with decreasing tolerance.

Figure 6. Evolution of mechanical work ($W$) and heat ($E$) in time. Both the PT and DI models were run with backward-Euler time integration for the energy equation.

the specific physics. For 2-D coupled thermomechanics, the read-and-write operations correspond to the updates of the degrees of freedom ($v_x$, $v_y$, $P$, $T$), and two additional read-only operations for converging the nonlinear viscosity; in our case, $nF = 10$. All the performance benchmark runs are performed using double-precision floating-point arithmetic (precis = 8 bytes) for fair comparison in particular between MATLAB and C-CUDA implementations. The MTP\textsubscript{effective} values reported in Fig. 7(a) represent the efficiency of memory transfers for the PT solvers, for both vectorized MATLAB CPU and C-CUDA single-GPU implementations. The obtained numbers should be compared to the peak memory throughput values (MTP\textsubscript{peak}) for the specific hardware, here an Intel i5 CPU and an Nvidia Titan X (Maxwell) GPU. MTP\textsubscript{peak} values are measured performing memory copy only without any computation. Values of MTP\textsubscript{peak} are in the order of 20 GB s$^{-1}$ for the Intel i5 CPU and in the order of 260 GB s$^{-1}$ for the Titan X (Maxwell) GPU. The MATLAB CPU implementation runs at about 2 per cent of the MTP\textsubscript{peak}, while the C-CUDA GPU codes run above 10 per cent of the MTP\textsubscript{peak}. The optimized memory bandwidth as well as the inherent parallelism on the GPU chip could explain such differences. The resulting overall performance gain of the parallel GPU implementation versus the serial CPU is more than two orders of magnitude. The GPU MTP\textsubscript{effective} values show that some optimization steps could still be performed in order to achieve closer to MTP\textsubscript{peak} values. Such considerations are beyond the scope of this study, but could include increased number of on-the-fly computation, kernel rearranging and register queues.

Although the MTP\textsubscript{effective} provides the efficiency of hardware utilization for a specific implementation of the thermomechanical solver, a more relevant metric should be used to compare the memory-bounded stencil PT iterative approach to the DI solver TM2Di. Here, the wall-time metric is chosen to assess the overall time to solution of a nonlinear step converged to $\text{tol rel} = 10^{-8}$ (Fig. 7b). The Newton-based DI solver TM2Di shows a close to linear increase of wall-time with increasing problem size (DoF). It is implemented in MATLAB and an Intel i5 (2016) CPU on a system equipped with 16 GB of RAM is used for computations. The maximal 2-D problem size fitting in 16 GB RAM represents a numerical domain of 960$^2$ gridpoints, solved in a wall-time close to 2.5 min. In comparison, 15 per cent less time was needed to converge the same problem using the C-CUDA GPU-based PT solver on an Nvidia Titan X (Maxwell). Nonetheless, the key benefit of this method is the maximal problem size that can be resolved using the available 12 GB of on-chip RAM of the GPU; 163 MDoF represents a numerical 2-D domain size of 6400$^2$ cells, which is a very high numerical resolution compared to the resolution currently used in geodynamic numerical simulations. In terms of wall-time, the PT GPU solver outperforms the Newton DI-based solver TM2Di for the investigated setup.

4.6 Explicit and implicit coupling strategies

The numerical solution of multiphysics problems can be achieved by means of various coupling strategies. For thermomechanical flow, a two-way coupling implies an implicit treatment of nonlinear coupling terms (Popov & Sobolev 2008; Kaus et al. 2016), namely the viscous-dissipation term and the strain rate and temperature dependence of the viscosity. In the two-way coupling, the viscosity is thus a function of the strain rate and temperature evaluated at new time index: $\eta(\dot{\varepsilon}_{II}^{i+\Delta t}, T^{i+\Delta t})$. One-way coupling represents an alternative coupling strategy commonly used in geodynamic modelling.
One-way coupling resides in (1) obtaining a nonlinear solution of the purely mechanical problem (thus only considering the strain rate dependence of viscosity) and (2) solving the energy equation using the mechanical dissipation obtained from the mechanical solution (Kaus & Podladchikov 2006). In other words, one-way coupling uses strain rates at the new time index but temperature of the old time index to evaluate the viscosity, \( \eta \left( \dot{\varepsilon}_m^{\text{II}}, T \right) \). A fully explicit coupling strategy can also be envisaged (Gerya & Yuen 2003, 2007). This approach relies on an explicit treatment of both coupling terms and rheological equations, which results in a linear mechanical problem. With this approach, the viscosity is evaluated using solutions from the previous time index: \( \eta \left( \dot{\varepsilon}_m^{\text{II}}, T^{\text{old}} \right) \). To evaluate the impacts of the different coupling strategies, we have run our reference simulation (Section 4.1) with the different strategies. Strain localization can be achieved with an explicit coupling strategy, but with the least intensity. We have monitored the gains in strain rate amplification and temperature using the solutions obtained using explicit coupling as reference (Fig. 8). For the same value of physical time step, \( \Delta t \), one-way coupling leads to an increase of 25 per cent in strain rate amplification and two-way coupling results in a 40 per cent increase in strain rate amplification. The impacts on maximum temperatures were less pronounced, since increments of 8 and 13 per cent were obtained for one-way and two-way coupling, respectively. Refining the time step can be used to improve the accuracy of the one-way coupling approach. For instance, twice smaller time step led to a 30 per cent gain in strain rate; however, with a smaller time step, the gain in strain rate rapidly saturates (here, to about 32 per cent) and does not catch up the values obtained with a two-way coupling. Alternative coupling strategies, such as time averaging whereby \( \eta \left( \dot{\varepsilon}_m^{\text{II}}, T^{\text{old}}+\Delta t \right) \), were not considered here but could also be envisaged.

4.7 Time integration

The choice of a time integration scheme is also critical when solving coupled nonlinear transient equations. Using our configuration (Section 4.1), successful time integrations were not possible for arbitrary large time step values. Time step values were often required to be on the order of the CFL criteria for diffusion despite the use of an implicit integration scheme (backward-Euler). This was especially true when using low grid resolutions (50\(^2\) cells, \( h = 1.7 \times 10^{-2} \)). We have reported the range of time step variation factors \( \xi \) that provided stable integration for various grid resolutions (Fig. 9). Restrictions on time step values usually decrease with increasing grid resolution. At high resolution (500\(^2\) cells, \( h = 1.7 \times 10^{-3} \)), a time step variation factor in the order of 50 was achievable using backward-Euler. The Crank–Nicolson scheme generally proved to have a larger stability domain, and time step variation factors could reach 80 at high resolution. In practice, the combination of a Crank–Nicolson scheme with an adaptive time stepping procedure (e.g. using time step bisection based on the magnitude of nonlinear residuals (Popov & Sobolev 2008; Duretz et al. 2015) can allow for stable and flexible time integration.

5 THERMALLY ACTIVATED SHEAR LOCALIZATION: APPLICATIONS

Here, we present applications computed with the PT approach that demonstrate the method’s flexibility as well as its power.

5.1 Thermomechanical strain localization in 2-D: a systematic study

We have studied the relative importance of the boundary velocity (\( V_{BC} \)) and bulk strain (\( \varepsilon \)) on shear zone development and evolution, performing 139 systematic 2-D simulations. Each simulation completed 1000 time steps with a resolution of 190\(^2\) cells. We have used a Crank–Nicolson time integration and relative nonlinear tolerance...
Shear heating and localization

Figure 8. The impacts of different multiphysics coupling strategies (one-way coupling, two-way coupling and explicit coupling) on numerical solution. Panels (a) and (b) depict the evolution of the temperature increase and strain rate amplification in the shear zone for the different coupling strategies using the same constant physical time step ($\Delta t$). Panels (c) and (d) show the gains in temperature and strain rate when using one-way coupling and two-way coupling in comparison to the explicit coupling (used as a reference here). We also report results obtained with the one-way strategy but with lower time step values (colour lines). Backward-Euler time integration was used to integrate the energy equation.

Figure 9. Solvability of the thermomechanical nonlinear system (two-way coupling). Models were run for different values of $\xi = \Delta t_1/\Delta t^{\text{max}}$ and various grid resolutions $N_x$. At larger resolutions, models with a larger physical time step become more solvable. These results were obtained using the DI method. Models were run for the reference parameters until a final time of $3 \times 10^{-3}$. Simulations for which the DI solver failed to converge for the requested value of $\xi$ are considered unsuccessful. In a general case, the time step could be adapted through the simulations in case the linear or nonlinear solves are unsuccessful.

5.2 The development of shear zones in 3-D

To demonstrate the flexibility of the presented PT algorithm, we have extended our GPU code to study 3-D thermomechanical deformation. Two models characterized by different initial thermal conditions were performed. The models were run with a resolution of $158^3$ and a relative tolerance of $10^{-5}$ was achieved at each time step. Both simulations ran for 5000 time steps, and each took...
about 1 day on a single Nvidia GTX Titan X (Maxwell) GPU card. For both models, we applied zero shear stress on each side and only boundaries normal to the x- and y-axis had a non-zero normal velocity component. Thus, in the first setup, we considered a cylindrical initial thermal perturbation; this configuration is equivalent to the 2-D model discussed above (Section 4.1) and leads to the rapid development of a cylindrical shear zone (Fig. 11a). The second model embedded spherical initial thermal perturbation. For such a configuration, more time is required to propagate the shear zone in 3-D (Fig. 11b). Since more mechanical work is dissipated prior to localization, the shear zone’s walls experienced a higher temperature than in the cylindrical case. The maximum temperature in the centre of the shear zones was about 3.5 for both models, which is similar to what was obtained in the 2-D models (Fig. 4).

6 DISCUSSION

6.1 Benefits of the pseudo-transient method and perspectives

The reported results confirm the PT method as performant alternative to a more classical DI solver type to address nonlinear coupled problems in geodynamics. In this study, we proved that both methods are capable of resolving the complex nonlinear physics and converge to an identical solution, even over a large number of time steps. The DI solvers are robust and weakly sensitive to large contrasts in material properties (e.g. viscosity). However, they may require a long and non-trivial development phase. A performant Cholesky factorization of the symmetrical and positive definite finite-difference matrix delivers acceptable time-to-solution for 2-D setups, but is inclined to hit the maximum system wide available RAM memory already at low spatial resolutions for 3-D problems on the considered personal computers: for example, about 60³ cells. We further report that implicit methods require physical time step values close to CFL if significant nonlinearities are involved.

In contrast, PT solvers result from a simple implementation of the coupled system of equations. The inherent parallelism in PT iterative-based solvers enables a straightforward vectorization, which shows promising implementation on multiple-core accelerators such as GPUs. GPU-based PT solvers outperformed CPU-based DI solvers regarding wall-time, even for 2-D setups. The reported solution of nonlinear thermomechanical problems are identical when computed with either the PT or the DI solver types, and validates the robustness and accuracy of the PT solver implementation. Further, the PT algorithms are succinct codes that enhance readability and make them less error prone. Besides faster times-to-solution, the benefits of PT stencil-based matrix-free solver approaches reside in lower memory footprint, optimal usage of actual hardware and straightforward parallelization, since only neighbours’ access is required. Since the investigation of large 3-D setups may require the use of more than a single-GPU accelerator, the GPU-based PT solver can readily be extended to run on a distributed-memory machine, via message-passing interface (MPI). Implementing an MPI point-to-point communication type for subdomain boundary exchange enables the PT solvers to scale on the largest supercomputers and show by construction a close to optimal parallel efficiency (Omlin et al. 2017a, b).

In this contribution, we focused on the small-strain limit and thus did not treat advection. However, the PT method is not restricted to this specific case and will be extended to convection-type problems in the future. To this end, the PT method will be coupled to either Eulerian (i.e. upwind type) or Eulerian–Lagrangian
Figure 11. 3-D numerical models of thermomechanical shear localization. Two different configurations were considered: a cylindrical initial thermal perturbation (a) and a spherical initial thermal perturbation. Both perturbations evolve into a single shear zone in response to the mechanical work exerted by the boundaries. Pure shear was applied on boundaries normal to y (inflow) and z (outflow); boundaries normal to x were free to slip. All thermal boundary conditions were insulating.

(i.e. characteristics-based) advection solvers. By analogy with the treatment of nonlinear couplings (see Section 4.6), the PT method will provide a simple framework for implementing either explicit or implicit advection solvers (Furuichi & May 2015). The latter discretization is a method of choice in order to avoid numerical instabilities in convection problems involving a free surface (Furuichi & May 2015).
7 CONCLUSIONS

In the perspective of quantifying and simulating fully coupled thermomechanical processes, such as ductile strain localization owing to shear heating, we have presented two numerical methods based on the finite-difference discretization. The first method is a thermomechanical extension to the DI M2Di solver (Räss et al. 2017) that relies on sparse matrix assembly and factorization, and the second method is based on a fully matrix-free PT method. Both methods can model thermomechanically activated shear localization in 2-D and provide consistent results. For 2-D models on a standard desktop computer, the PT method is as accurate and can be as efficient (in terms of wall-time) as the DI approach. We also investigated the impacts of different nonlinear coupling strategies and could show that no matter how much the time step is decreased, solutions obtained with one-way coupling never achieves the accuracy of the two-way coupling.

The significant advantage of the PT method is that it can be extended for high-resolution 3-D numerical simulations without significant modification of the 2-D algorithm and without a drastic increase in memory requirements as the latter scales linearly with the number of grid cells. We show that the PT method is suitable to perform high-resolution 3-D simulations of thermomechanically activated shear localization and other relevant coupled physics in geodynamics (Räss et al. 2018). Further, the efficiency of the thermomechanical codes makes it suitable for systematic analysis of the parameters that control the dynamics of shear zone development. Based on 139 2-D simulations, we show that a consistent data collapse of the shear zone temperature and strain rate can be established, which further demonstrate the PT method’s robustness.

ACKNOWLEDGEMENTS

The authors acknowledge two anonymous reviewers that provided thorough and constructive reviews of the original draft. The authors are grateful to Daniel Kiss for constructive discussions and Samuel Omlin for his long-term partnership in the high performance code development. We acknowledge the Swiss Geocomputing Centre at the University of Lausanne for support and computing resources. This work was supported by the University of Lausanne.

REFERENCES


APPENDIX A: NON-DIMENSIONALIZATION OF THE THERMOMECHANICAL EQUATIONS

Dimensionless thermomechanical equations (eq. 4) can be obtained by introducing the characteristic scales (eq. 3) in the dimensional thermomechanical eq. (1). The dimensionless equations are obtained by substituting all dimensional variables by the product of their characteristic values and dimensionless value. For example, the temperature can be expressed as \( T = \tilde{T} \tilde{T} \), where the tilde sign stands for the dimensionless temperature.

Here the temperature evolution equation is expressed as

\[
0 = \tau_{ij} \hat{\varepsilon}_{ij} + k \frac{\partial^2 \tilde{T}}{\partial x_i \partial x_j} - \rho C_v \frac{\partial \tilde{T}}{\partial t} = \frac{\tilde{\tau}}{\tilde{T}} \hat{\varepsilon}_{ij} + k \frac{\tilde{T}}{L^2} \frac{\partial^2 \tilde{T}}{\partial x_i \partial x_j} - \rho C_v \frac{\tilde{T}}{\tilde{T}} \frac{\partial \tilde{T}}{\partial t}.
\]

By dividing the above expression by \( \tilde{T} / \tilde{T} \) yields

\[
0 = \hat{\varepsilon}_{ij} + k \frac{\tilde{T}}{L^2} \frac{\partial^2 \tilde{T}}{\partial x_i \partial x_j} - \rho C_v \frac{\tilde{T}}{\tilde{T}} \frac{\partial \tilde{T}}{\partial t}.
\]

After introducing the characteristic length, \( L = \sqrt{\frac{k}{\rho C_v}} \) and stress, \( \tilde{\tau} = \rho C_v \tilde{T} \), one finally gets

\[
0 = \hat{\varepsilon}_{ij} + k \frac{\tilde{T}}{L^2} \frac{\partial^2 \tilde{T}}{\partial x_i \partial x_j} - \rho C_v \frac{\tilde{T}}{\tilde{T}} \frac{\partial \tilde{T}}{\partial t}.
\]

The dimensional constitutive relationship is spelled as

\[
\dot{\varepsilon}_{ij} = \frac{1}{2} A \tau_{ij}^{-1} \exp \left( -\frac{Q}{R (T_0 + T)} \right) \tau_{ij},
\]

One first introduces the characteristic time and stress to express the dimensionless strain rate:

\[
\dot{\varepsilon}_{ij} = \tilde{T} A \tau_{ij}^{-1} \tilde{T}^{-1} \tilde{\tau}_{ij}^{-1} \exp \left( -\frac{Q}{R (T_0 + \tilde{T})} \right) \tilde{\tau}_{ij},
\]

which can be recasted as

\[
\dot{\varepsilon}_{ij} = \tilde{T} A \tau_{ij}^{-1} \tilde{T}^{-1} \tilde{\tau}_{ij}^{-1} \exp \left( -\frac{Q}{R (T_0 + \tilde{T})} \right) \tilde{\tau}_{ij},
\]

Expressing the Arrhenius term with a common denominator and introducing the dimensionless temperature yields:

\[
\dot{\varepsilon}_{ij} = 2^{-n} \tau_{ij}^{-1} \exp \left( \frac{Q \tilde{T} \tilde{T}}{R T_0^2 + R T_0 \tilde{T} \tilde{T}} \right) \tilde{\tau}_{ij}.
\]

Finally one may introduce a dimensionless reference temperature, \( \tilde{T}_0 = \tilde{T} / \tilde{T} \), which simplifies the expression in the following way:

\[
\dot{\varepsilon}_{ij} = 2^{-n} \tau_{ij}^{-1} \exp \left( \frac{n \tilde{T}}{1 + \tilde{T} \tilde{T}} \right) \tilde{\tau}_{ij}.
\]

The dimensionalization of the momentum equations (in the absence of body forces) and of the continuity equation is straightforward, hence they will not be detailed here. Please note that in the main body of the text, we omit the ˜ superscript to enhance the readability.

APPENDIX B: SOLVING THE NONLINEAR THERMOMECHANICAL SYSTEM WITH THE DIRECT-ITERATIVE SCHEME

An implicit (or semi-implicit) discretization (e.g. backward-Euler, Crank–Nicolson) results in a linear system of equation of the form:

\[
K_{TM} \mathbf{x} = \mathbf{b},
\]

The systems of equation couples the different solution fields \( \mathbf{v}, \mathbf{p}, \mathbf{T} \) in the following way:

\[
\begin{bmatrix}
K_{vv} & K_{vp} & 0 \\
K_{pv} & 0 & 0 \\
K_{Tv} & 0 & K_{TT}
\end{bmatrix}
\begin{bmatrix}
\mathbf{v} \\
\mathbf{p} \\
\mathbf{T}
\end{bmatrix}
= \begin{bmatrix}
\mathbf{b}_v \\
\mathbf{b}_p \\
\mathbf{b}_T
\end{bmatrix}.
\]

\[
K_{TM}
\]
where the matrices $K_{TM}$ contains coefficients resulting from the discretization (here, centred finite differences). The blocks $K_{vv}$ and $K_{Tv}$ represent the deviatoric stress gradient and shear heating operators, respectively. The $K_{pp}$ block represents the discrete gradient operator, which equals minus the transpose of the divergence operator, that is, $K_{pp} = -K_{p}^T$. The $K_{TT}$ is the Laplace operator, which arises from the diffusion term in the temperature evolution equation. The $b$ contains contributions resulting from the time discretization of the transient terms (here in $b_1$) as well as contributions from the boundary conditions (e.g. Dirichlet and Newman boundary values).

In the case of a linear problem, the solutions $x$ can readily be obtained by applying the inverse of the matrix $K_{TM}$ to the right-hand side vector $b$:

$$x = K_{TM}^{-1} b.$$  \hfill \text{(B3)}

In the present case, the nonlinear dependence of the viscosity on the temperature and the velocity (i.e. via the strain rate dependence) causes the global system of equation to be nonlinear. The nonlinearity results in an imbalance of the global equations:

$$b - K_{TM} x = f \neq 0,$$  \hfill \text{(B4)}

where $f$ is the nonlinear residual vector. In order to minimize the imbalance, quantified by the magnitude $\| f \|$, it is convenient to solve this problem iteratively. To this end, the solution vector is updated within a cycle of $k$ nonlinear iterations:

$$x^{k+1} = x^k + \alpha \delta x^{k+1},$$  \hfill \text{(B5)}

where $\delta x^k$ is the nonlinear correction to the solution. The scalar parameter $\alpha$ is determined during a line search procedure in order to minimize the magnitude of the residuals:

$$\min \| f(x^k + \alpha \delta x^{k+1}) \|_{L_2}. $$  \hfill \text{(B6)}

In the context of Picard iterations, these corrections are obtained by evaluating the matrix $K_{TM}$ at each iteration, using the values of viscosities evaluated with the current temperature and velocity fields, and by computing the correction as

$$\delta x^{k+1} = (K_{TM}^{-1})^k f^k,$$  \hfill \text{(B7)}

where $k$ is the nonlinear iteration index. The vector $f$ can be obtained with the following operation:

$$\begin{bmatrix} f_v \\ f_p \\ f_T \end{bmatrix} = \begin{bmatrix} b_v \\ b_p \\ b_T \end{bmatrix} - \begin{bmatrix} K_{vv}(v, T) & K_{vp} & 0 \\ K_{pv} & 0 & 0 \\ K_{Tv}(v, T) & 0 & K_{TT} \end{bmatrix} \begin{bmatrix} v \\ p \\ T \end{bmatrix}.$$  \hfill \text{(B8)}

Here, we emphasize the nonlinearity held in the blocks $K_{vv}$ and $K_{Tv}$ due to the nonlinear dependence of viscosity on the velocity and temperature. Nonlinear iterations need to be performed until the magnitude of the nonlinear residual has decreased below a given level of tolerance, for example, $\| f^k \| < \text{tol}$. The Picard iterations only deliver a linear rate of convergence and therefore, a high number of iterations is generally required before reaching the desired nonlinear accuracy (i.e. several tens of iterations).

To overcome this severe restriction, we have used a Newton linearization. To this end, the iteration matrix in eq. (B11) is substituted by the Jacobian matrix $J_{TM}$. The Jacobian matrix takes the form of

$$J_{TM} = \frac{\partial f}{\partial x},$$  \hfill \text{(B9)}

and thus contains information about the gradient of the residual with regard to the solution. In practice, for the considered case, the Jacobian matrix can be written as

$$J_{TM} = \begin{bmatrix} J_{vv} & K_{vp} & J_{vT} \\ K_{pv} & 0 & 0 \\ J_{Tv} & 0 & J_{TT} \end{bmatrix},$$  \hfill \text{(B10)}

where the blocks $J_{vv}$ and $J_{Tv}$ differ from $K_{vv}$ and $K_{Tv}$, since they contain additional contributions from the gradients of viscosity with regard to the velocity. The new block $J_{vT}$ as well as the block $J_{TT}$ contain information from the gradients of viscosity with regard to the temperature. The structure of the matrix operator $J_{TM}$ shares similarities with that employed by Wilson et al. (2017), who applied similar linearization to study thermomechanical convection. The Newton correction is hence obtained with the following operation:

$$\delta x^{k+1} = (J_{TM}^{-1})^k f^k,$$  \hfill \text{(B11)}

which allows to reach the desired nonlinear accuracy with a quadratic rate of convergence (i.e. less than 10 iterations).

**APPENDIX C: THE DIRECT-ITERATIVE SOLVER**

We seek a solution of the following linear system:

$$\begin{bmatrix} J_{vv} & K_{vp} & J_{vT} \\ K_{pv} & 0 & 0 \\ J_{Tv} & 0 & J_{TT} \end{bmatrix} \begin{bmatrix} \delta v \\ \delta p \\ \delta T \end{bmatrix} = - \begin{bmatrix} f_v \\ f_p \\ f_T \end{bmatrix}.$$  \hfill \text{(C1)}

To facilitate and enhance the linear solve procedure, we introduce the pre-conditioner:

$$J_{TM}^{-1} = \begin{bmatrix} J_{vv} & K_{vp} & J_{vT} \\ K_{pv} & 0 & 0 \\ J_{Tv} & 0 & J_{TT} \end{bmatrix},$$  \hfill \text{(C2)}

where the block matrix $J_{pp} = \gamma^{-1} I$ corresponds to a weakly compressible contribution.

The linear residuals are defined as

$$\begin{align*}
    r_v &= f_v - J_{vv} \delta v - K_{vp} \delta p - J_{vT} \delta T \\
    r_p &= f_p - K_{pv} \delta v \\
    r_T &= f_T - J_{Tv} \delta v - J_{TT} \delta T
\end{align*}$$  \hfill \text{(C3)}

The solutions are found iteratively:

$$\begin{align*}
    \delta v^{i+1} &= \delta v^i + \delta \delta v \\
    \delta p^{i+1} &= \delta p^i + \delta \delta p \\
    \delta T^{i+1} &= \delta T^i + \delta \delta T,
\end{align*}$$  \hfill \text{(C4)}

where $\delta \delta v$, $\delta \delta p$ and $\delta \delta T$ are iterative corrections and $i$ is the iteration count.

The iterative velocity correction is obtained by solving:

$$\delta \delta v = J_{vv}^{-1} r_v,$$  \hfill \text{(C5)}

where $J_{vv} = J_{vv} - K_{vp} (J_{pp}^{-1} K_{pv})$ and $r_v = r_v - K_{pv} (J_{pp}^{-1} r_p) - K_{vT} \delta \delta T$. Applying the inverse of $J_{pp}$ is a trivial operation, since $J_{pp}$ is a diagonal block matrix. However, applying the inverse of $J_{vv}$ is a cumbersome task, since $J_{vv}$ may not be a symmetrical matrix. Here, we use an iterative Krylov subspace solver (generalized conjugate residuals, Eisenstat et al. 1983) where the Cholesky factors of the symmetrical operators resulting from a Picard linearization are used for preconditioning. This approach is described in detail by Rass et al. (2017). Subsequently, the iterative pressure correction is obtained in a trivial way by evaluating

$$\delta \delta p = J_{pp}^{-1} (r_p - K_{pv} \delta \delta v).$$  \hfill \text{(C6)}
Finally, the iterative temperature correction is calculated as follows:

$$\delta\delta T = J_{TT}^{-1} (r_T - J_{Tv} \delta v). \quad (C7)$$

Here, $J_{TT}$ is a symmetric positive definite matrix and its inverse can be efficiently applied using pre-computed Cholesky factors. The iteration is repeated until the $L_2$ norm of all linear residual vectors has decreased below a given threshold value.
Key words

Authors are requested to choose key words from the list below to describe their work. The key words will be printed underneath the summary and are useful for readers and researchers. Key words should be separated by a semi-colon and listed in the order that they appear in this list. An article should contain no more than six key words.

COMPOSITION and PHYSICAL PROPERTIES
Composition and structure of the continental crust
Composition and structure of the core
Composition and structure of the mantle
Composition and structure of the oceanic crust
Composition of the planets
Creep and deformation
Defects
Elasticity and anelasticity
Electrical properties
Equations of state
Fault zone rheology
Fracture and flow
Friction
High-pressure behaviour
Magnetic properties
Microstructure
Permeability and porosity
Phase transitions
Plasticity, diffusion, and creep

GENERAL SUBJECTS
Core
Gas and hydrate systems
Geomechanics
Geomorphology
Glaciology
Heat flow
Hydrogeophysics
Hydrology
Hydrothermal systems
Infrasound
Instrumental noise
Ionosphere/atmosphere interactions
Ionosphere/magnetosphere interactions
Mantle processes
Ocean drilling
Structure of the Earth
Thermochronology
Tsunamis
Ultra-high pressure metamorphism
Ultra-high temperature metamorphism

GEODESY and GRAVITY
Acoustic-gravity waves
Earth rotation variations
Geodetic instrumentation
Geopotential theory
Global change from geodesy
Gravity anomalies and Earth structure
Loading of the Earth
Lunar and planetary geodesy and gravity
Plate motions
Radar interferometry
Reference systems
Satellite geodesy
Satellite gravity
Sea level change
Seismic cycle
Space geodetic surveys
Tides and planetary waves
Time variable gravity
Transient deformation

GEOPHYSICAL METHODS
Downhole methods
Fourier analysis
Fractals and multifractals
Image processing

GEOMAGNETISM and ELECTROMAGNETISM
Archaeomagnetism
Biogenic magnetic minerals
Controlled source electromagnetics (CSEM)
Dynamo: theories and simulations
Electrical anisotropy
Electrical resistivity tomography (ERT)
Electromagnetic theory
Environmental magnetism
Geomagnetic excursions
Geomagnetic induction
Ground penetrating radar
Magnetic anomalies: modelling and interpretation
Magnetic fabrics and anisotropy
Magnetic field variations through time
Magnetic mineralogy and petrology
Magnetostriatigraphy
Magnetotellurics
Marine electromagnetics
Marine magnetics and palaeomagnetics
Non-linear electromagnetics
Paleointensity
Palaeomagnetic secular variation
Palaeomagnetism
Rapid time variations
Remagnetization
Reversals: process, time scale, magnetostriatigraphy
Rock and mineral magnetism
Satellite magnetics

GEOPHYSICAL METHODS
Downhole methods
Fourier analysis
Fractals and multifractals
Image processing

Instability analysis
Interferometry
Inverse theory
Joint inversion
Neural networks, fuzzy logic
Non-linear differential equations
Numerical approximations and analysis
Numerical modelling
Numerical solutions
Persistence, memory, correlations, clustering
Probabilistic forecasting
Probability distributions
Self-organization
Statistical methods
Thermobarometry
Time-series analysis
Tomography
Waveform inversion
Wavelet transform

PLANETS
Planetary interiors
Planetary volcanism

SEISMOLOGY
Acoustic properties
Body waves
Coda waves
Computational seismology
Controlled source seismology
Crustal imaging
Earthquake dynamics
Earthquake early warning
Earthquake ground motions
Earthquake hazards
Earthquake interaction, forecasting, and prediction
Earthquake monitoring and test-ban treaty verification
Earthquake source observations
Guided waves
Induced seismicity
Interface waves
Palaeseismology
Rheology and friction of fault zones
Rotational seismology
Seismic anisotropy
Seismic attenuation
Seismic instruments
Seismic interferometry
Seismicity and tectonics
Seismic noise
Seismic tomography
Site effects
Statistical seismology
Surface waves and free oscillations
Theoretical seismology
Tsunami warning
Volcano seismology
Wave propagation
Wave scattering and diffraction

TECTONOPHYSICS
Backarc basin processes
Continental margins: convergent
Continental margins: divergent
Continental margins: transform
Continental neotectonics
Continental tectonics: compressional
Continental tectonics: extensional
Continental tectonics: strike-slip and transform
Cratons
Crustal structure
Diapirism
Dynamics: convection currents, and mantle plumes
Dynamics: gravity and tectonics
Dynamics: seismotectonics
Dynamics and mechanics of faulting
Dynamics of lithosphere and mantle
Folds and folding
Fractures, faults, and high strain deformation zones
Heat generation and transport
Hotspots
Impact phenomena
Intra-plate processes
Kinematics of crustal and mantle deformation
Large igneous provinces
Lithospheric flexure
Mechanics, theory, and modelling
Microstructures
Mid-ocean ridge processes
Neotectonics
Obduction tectonics
Oceanic hotspots and intraplate volcanism
Oceanic plateaus and microcontinents
Oceanic transform and fracture zone processes
Paleoseismology
Planetary tectonics
Rheology: crust and lithosphere
Rheology: mantle
Rheology and friction of fault zones
Sedimentary basin processes
Subduction zone processes
Submarine landslides
Submarine tectonics and volcanism
Tectonics and climatic interactions
Tectonics and landscape evolution
Transform faults
Volcanic arc processes

VOLCANOLOGY
Atmospheric effects (volcano)
Calderas
Effusive volcanism
Eruption mechanisms and flow emplacement
Experimental volcanism
Explosive volcanism
Lava rheology and morphology
Magma chamber processes
Magma genesis and partial melting
Magma migration and fragmentation
Mud volcanism
Physics and chemistry of magma bodies
Physics of magma and magma bodies
Planetary volcanism
Pluton emplacement
Remote sensing of volcanoes
Subaqueous volcanism
Tephrochronology
Volcanic gases
Volcanic hazards and risks
Volcaniclastic deposits
Volcano/climate interactions
Volcano monitoring
Volcano seismology

Key words