NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2015; **00**:1–10
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nla

Aspects of Multigrid for Mantle Convection

S. Baumann[1]*, M. Mohr[1]

[1]*Dept. of Earth and Environmental Sciences, Ludwig-Maximilians-Universität München*

SUMMARY

Convection in the Earth's mantle can be described by a Stokes-type equation with strongly varying viscosity values coupled to an energy equation. The dynamics of the mantle depend essentially on the underlying viscosity structure. Computing the quasi-stationary flow field in each time step dominates the computational cost of Earth mantle simulators. Therefore constructing efficient solvers is crucial for simulating Earth mantle dynamics. In this article we consider a geometric multigrid method for the viscous operator of the Stokes-type system and study the convergence behaviour for different smoothers and transfer operators. We focus on two aspects, the influence of operator complexity, starting with a simplified form and ending with its most general formulation, and the effects of highly discontinuous viscosity parameters. Systematic numerical tests help to identify the most efficient multigrid components. Copyright © 2015 John Wiley & Sons, Ltd.

Received ...

KEY WORDS: Multigrid, Mantle Convection, Jumping Coefficients, Stokes problem, Anelastic Model

1. INTRODUCTION

While to us as human beings the ground on which we walk may appear 'rock-solid' the surface of our planet is actually in constant albeit very slow motion. The continental plates move at a rate of centimetres per year. The reason for this movement are enormous forces acting deep below our feet. Convective processes in the Earth's mantle help the planet to rid itself of excess energy that is either left from the time of its formation or generated by continued radioactive decay. The mantle is a layer of Earth starting from below the crust at roughly 60 km and extending down to the core-mantle-boundary at a depth of about 3.000 km. On geologic time-scales the rocks inside the mantle behave like a highly viscous fluid. A single overturn of the material in the mantle takes about 100 mio. years.

A detailed understanding of these processes is of fundamental interest to geophysics, as they are the driving force behind phenomena such as mountain building, back-arc volcanism and finally earthquakes. While the basic principles are those of fluid dynamics and, thus, well understood, there are essential details and parameters of mantle convection that are only poorly understood. Key among these is the (dynamic) viscosity which describes the rheology of the mantle.

The interior of the Earth is not directly accessible to us. What we know about the inner structure of our planet is derived from indirect observations and inverse reconstructions, such as measurements of the Earth's Geoid or Seismic Tomography. Consequently many research questions can only be studied through numerical simulation. Therefore computational techniques have for a long time played an important role in Geophysics.

A general mathematical description of mantle convection can be derived from the principles of conservation of momentum, mass and energy. It is common practice to neglect inertial terms and

*Correspondence to: simon.bauman@lmu.de

Copyright © 2015 John Wiley & Sons, Ltd.

the coriolis force due to their relative insignificance. For a detailed explanation and justification see e.g. [1]. Doing so one arrives at

$$\operatorname{div} \boldsymbol{\sigma} + \rho \mathbf{g} = 0 \tag{1a}$$

$$\partial_t \rho + \operatorname{div}(\rho \mathbf{u}) = 0 \tag{1b}$$

$$\partial_t (\rho e) + \operatorname{div}(\rho e \mathbf{u}) + \operatorname{div} \mathbf{q} - H - \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} = 0 \ . \tag{1c}$$

Here $\mathbf{u}$ represents velocity, ρ density, $\boldsymbol{g}$ gravitational acceleration, e internal energy density, H volumetric radiogenic heat production rate and $\boldsymbol{q}$ heat flux per unit area. The terms $\boldsymbol{\sigma}$ and $\dot{\boldsymbol{\varepsilon}}$ are the stress and rate of strain tensor, respectively. These are coupled to each other, to velocity and to pressure p via

$$\boldsymbol{\sigma} = 2\mu \left(\dot{\boldsymbol{\varepsilon}} - \frac{\operatorname{tr} \dot{\boldsymbol{\varepsilon}}}{3} \mathbf{I} \right) - p\mathbf{I} \ , \quad \dot{\boldsymbol{\varepsilon}} = \frac{1}{2} \left(\nabla \mathbf{u} + (\nabla \mathbf{u})^T \right) \ , \tag{2}$$

where $\mathbf{I}$ represents identity and μ dynamic viscosity. The energy equation (1c) is re-cast in terms of temperature T and the system is closed by adding an equation of state relating density to pressure and temperature $\rho = \rho(p, T)$. The precise details of the latter vary between models as the composition and mineralogical behaviour of the mantle is also an open research problem.

Density can be split into two contributions $\rho = \rho_{\text{ref}} + \hat{\rho}$. Here ρ_{ref} is a so called *background density* that is derived from the time-constant hydrostatic pressure. The term $\hat{\rho}$ then represents the density variations resulting from thermal expansion that drive the convection. The latter, however, are very small compared to ρ_{ref}. This allows to consider their effects only in the buoyancy term of the momentum equation (1a) and to neglect them in the continuity equation (1b). This results in the so called *anelastic approximation*

$$\operatorname{div}(\rho_{\text{ref}} \mathbf{u}) = 0 \ .$$

This is similar to the *Boussinesq approximation*, but takes into account that the background density increases by a factor of about two from the top to the bottom of the mantle, see e.g. [2]. Setting $\partial_t \rho = 0$ and employing (2) results in the following *generalised Stokes problem*

$$\mathrm{L}(\mathbf{u}) - \nabla p = \boldsymbol{F}(T) \ , \quad \operatorname{div}(\mathbf{v}) = 0 \ , \tag{3}$$

with an elliptic *viscous operator* L. In the case of an anelastic model we have $\mathbf{v} = \rho_{\text{ref}} \mathbf{u}$ and L is given by

$$\mathrm{L}(\mathbf{u}) = \operatorname{div} \left(\mu \left(\nabla \mathbf{u} + (\nabla \mathbf{u})^T \right) \right) - \frac{2}{3} \nabla \left(\mu \operatorname{div} \mathbf{u} \right) \ . \tag{4}$$

We denote (4) as the *full operator*. If, instead of the anelastic the Boussinesq approximation is employed, we have $\mathbf{v} = \mathbf{u}$ and the operator reduces to the *epsilon operator*

$$\mathrm{L}(\mathbf{u}) = \operatorname{div} \left[\mu \left(\nabla \mathbf{u} + (\nabla \mathbf{u})^T \right) \right] \ . \tag{5}$$

In the case of constant viscosity $\mu = \mu_0$ the epsilon operator further simplifies to

$$\mathrm{L}(\mathbf{u}) = \mu_0 \Delta \mathbf{u} \tag{6}$$

and (3) becomes the standard Stokes problem.

Note that for both, the anelastic and the Boussinesq approximation, the explicit time-dependence of the model is contained solely in the energy equation (1c) and enters the generalised Stokes part only via changes in the buoyancy term. While the temporal discretisation of the energy equation requires some care, see e.g. [3] and the references therein, the computationally most intensive part is the solution of the generalised Stokes problem (3) one or more times per time-step. This is especially challenging in global geodynamical simulations as they require meshes resolving the whole mantle at resolutions fine enough to allow using realistic earth-like parameters, see e.g. [3, 4, 5].

A standard discretisation of the generalised Stokes problem, e.g. with Finite Elements as in [3, 6, 7], leads to an algebraic system of the form

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \tilde{\mathbf{B}} & -\mathbf{C} \end{bmatrix} \begin{bmatrix} u \\ q \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix} \ . \tag{7}$$

Copyright © 2015 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

Here $\mathbf{A}$ represents the discrete viscous operator $\mathrm{L}(\mathbf{u})$, $\mathbf{B}^T$ a discrete gradient, while $\tilde{\mathbf{B}}$ stems from the discretisation of the continuity equation $\mathrm{div}(\mathbf{v}) = 0$. Global convection models prescribe no-outflow boundary conditions, which together with standard ansatz functions result in $\tilde{\mathbf{B}} = \mathbf{B}$ for a formulation employing $\mathbf{v} = \mathbf{u}$. The $\mathbf{C}$ block might represent a stabilisation term, that is needed only, if the ansatz functions for velocity and pressure are not inf-sup stable, see [7] for an example. Vectors u and q contain the coefficients of the discrete velocity and pressure, while f and g represent the discrete right-hand sides, potentially including contributions from the boundary conditions.

Due to the ellipticity of the viscous operator, most mantle convection codes today employ multigrid as part of the solution strategy for (7) in one way or the other. One variant is to solve the system using a pressure-correction scheme, which can also be seen as an outer iteration applied to the system's Schur complement S. It can be shown that S has bounded condition number independent of the mesh size and, thus, the Schur complement system can be solved efficiently with a (diagonally preconditioned) conjugate gradient method [6, 8].

The latter step involves an application of $\mathbf{A}^{-1}$. Here multigrid can be used in an inner iteration to solve $\mathbf{A}x = b$ and, thus, to evaluate the effect of multiplying S with a vector. This is the classic method described in [8] and employed e.g. in [7, 9, 10, 11]. Other models solve (7) using directly a suitable Krylov subspace method, such as e.g. MINRES, if (7) is symmetric. Multigrid then commonly plays a role in the preconditioning of the A block. In [3] e.g. a flexible GMRES method is employed for solving (7) which employs a right-preconditioner involving an approximate inverse of $\mathbf{A}$ that is computed using a multigrid preconditioned CG method. A third alternative is to attempt to solve the full saddle point problem (3) with multigrid directly. The difficulty here then lies with finding a smoother that is efficient w.r.t. both smoothing property and algorithmic cost. In [12, 13] a specially developed *ACuTE* smoother, an extension of the artificial compressibility method, is employed for this purpose. As this third alternative is still less commonly employed we concentrate here on multigrid methods for solving systems $\mathbf{A}x = b$.

In geodynamic models the mantle is typically represented as a thick spherical shell. The latter is commonly discretised either using an icosahedral mesh [10, 14], a Yin-Yang grid [13] or a (modified) cubed sphere [4, 15]. In all three approaches the computational mesh is generated by structured refinement of a coarse base grid. This natural availability of a grid hierarchy makes it possible to use geometric multigrid methods. This avoids the setup cost and the difficulties of algebraic multigrid for systems of equations. It further permits a highly efficient scalable matrix-free implementation [16]. Notable exceptions are models with fully adaptive unstructured meshes, see e.g. [17], and models constructed using general purpose libraries such as [3].

While multigrid for the (vector) Laplacian has been extensively studied, including the pure Stokes problem, the situation is different for the full operator (4). In this paper we focus on a detailed investigation of the influence of growing complexity of the viscous operator $\mathrm{L}(\mathbf{u})$ on the performance of geometric multigrid. We will do so first for constant viscosity and then also consider the effects of large viscosity contrasts on the convergence properties.

The latter is of great relevance for convection models, since the radial viscosity profile of Earth is known to exhibit strong discontinuities of several orders of magnitude. While the average viscosity of the upper and lower mantle is believed to be on the order of 10^{21} Pa s and 10^{22} Pa s respectively, the upper mantle contains a mechanically weak zone, the asthenosphere, with a significantly lower viscosity. The extent of that zone and of the associated viscosity drop, which are coupled together, are presently unknown, but the viscosity contrast may be up to four orders in magnitude large. This radial variability is overlaid by lateral variations, as viscosity is influenced by temperature, pressure, flow velocity and chemical composition. For further information see e.g. [18] and the references therein.

2. STUDY SETUP

We now specify the setup of our study. The goal of the article is to present a detailed and comprehensive investigation of the multigrid convergence properties in a prototype 2D setting. We

Copyright © 2015 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

consider the problem of solving

$$\begin{aligned} \mathrm{L}(\mathbf{u}) &= \boldsymbol{f} \ , \quad \forall (x,y) \in \Omega \\ \mathbf{u} &= 0 \ , \quad \forall (x,y) \in \partial\Omega \end{aligned}$$

where L is the viscous operator of the *generalised Stokes problem* (3) and $\Omega = (0,1)^2 \subset \mathbb{R}^2$. The unknown function $\mathbf{u}$ is vector-valued with components $\mathbf{u} = (u_1, u_2)$. Note that only the cases where L is not the vector Laplacian (6) will exhibit a coupling between the two velocity components.

We discretise the spatial domain $\bar{\Omega}$ using a regular grid Ω_h composed of rectangular cells. A nested hierarchy of grids for the multigrid method is constructed by full coarsening, i.e. we set

$$\Omega_h^l := \left\{ (ih_l, jh_l) \, : 0 \leqslant i,j \leqslant 2^l \, , h_l = 2^{-l} \right\} \quad , \quad l \in \{1, ..., L\} \quad .$$

For our numerical experiments we will use $L = 5$ which results in a finest grid of 33×33 nodes. Following the approaches in [9], [10] and [7] we employ nodal ansatz functions of 1^{st} order, i.e. we discretise with bi-linear finite elements. This results for every interior node of our mesh in a 9-point stencil whose values are given by (2×2) tensors that express the coupling between the two components of $\mathbf{u}$ at the two coupled nodes. Denoting by $\mathcal{S}_{\text{lap}}$ the stencil resulting for the vector Laplacian, by $\mathcal{S}_{\text{eps}}$ and $\mathcal{S}_{\text{full}}$ the stencils resulting for operators (5) and (4) and setting viscosity to be $\mu \equiv 1$ we obtain

$$\mathcal{S}_{\text{lap}} = \frac{1}{3} \left\{ \begin{array}{ccc} \begin{bmatrix} 1 & \cdot \\ \cdot & 1 \end{bmatrix} & \begin{bmatrix} 1 & \cdot \\ \cdot & 1 \end{bmatrix} & \begin{bmatrix} 1 & \cdot \\ \cdot & 1 \end{bmatrix} \\ \begin{bmatrix} 1 & \cdot \\ \cdot & 1 \end{bmatrix} & \begin{bmatrix} -8 & \cdot \\ \cdot & -8 \end{bmatrix} & \begin{bmatrix} 1 & \cdot \\ \cdot & 1 \end{bmatrix} \\ \begin{bmatrix} 1 & \cdot \\ \cdot & 1 \end{bmatrix} & \begin{bmatrix} 1 & \cdot \\ \cdot & 1 \end{bmatrix} & \begin{bmatrix} 1 & \cdot \\ \cdot & 1 \end{bmatrix} \end{array} \right\} \ , \tag{8}$$

$$\mathcal{S}_{\text{eps}} = \frac{1}{4} \left\{ \begin{array}{ccc} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} & \begin{bmatrix} \cdot & \cdot \\ \cdot & 4 \end{bmatrix} & \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \\ \begin{bmatrix} 4 & \cdot \\ \cdot & \cdot \end{bmatrix} & \begin{bmatrix} -16 & \cdot \\ \cdot & -16 \end{bmatrix} & \begin{bmatrix} 4 & \cdot \\ \cdot & \cdot \end{bmatrix} \\ \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} & \begin{bmatrix} \cdot & \cdot \\ \cdot & 4 \end{bmatrix} & \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \end{array} \right\} \ , \tag{9}$$

$$\mathcal{S}_{\text{full}} = \frac{1}{36} \left\{ \begin{array}{ccc} \begin{bmatrix} 14 & -3 \\ -3 & 14 \end{bmatrix} & \begin{bmatrix} 8 & \cdot \\ \cdot & 20 \end{bmatrix} & \begin{bmatrix} 14 & 3 \\ 3 & 14 \end{bmatrix} \\ \begin{bmatrix} 20 & \cdot \\ \cdot & 8 \end{bmatrix} & \begin{bmatrix} -112 & \cdot \\ \cdot & -112 \end{bmatrix} & \begin{bmatrix} 20 & \cdot \\ \cdot & 8 \end{bmatrix} \\ \begin{bmatrix} 14 & 3 \\ 3 & 14 \end{bmatrix} & \begin{bmatrix} 8 & \cdot \\ \cdot & 20 \end{bmatrix} & \begin{bmatrix} 14 & -3 \\ -3 & 14 \end{bmatrix} \end{array} \right\} \ . \tag{10}$$

Note that, since our discrete ansatz functions are not divergence-free, S_{full} does not naturally reduce to S_{eps} and though we use constant μ, S_{eps} is different from S_{lap}, as $\operatorname{div}\left(\nabla \mathbf{u} + (\nabla \mathbf{u})^T\right) = \Delta \mathbf{u} + \nabla(\operatorname{div} \mathbf{u})$.

In the following we will study solvers for the linear system $\mathbf{A}_L u_L = f_L$ resulting from the discretisation on the finest grid level. An efficient multigrid method requires an optimal interplay of different components. Besides the mesh coarsening strategy itself, one can adjust the type of pre- and post-smoother, the number of smoothing steps, the cycling strategy, the intergrid transfer operators used for prolongation ($\mathbf{P}$) and restriction ($\mathbf{R}$) and finally the approach for representing the operator $\mathbf{A}_L$ on coarser levels.

Throughout the paper we will keep the following components fixed. Our cycling strategy will be the standard V-cycle, since this is expected to be advantageous compared to more complex cycle

Copyright © 2015 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

types in a parallel setting. We also select the transpose of the prolongation operator for restriction, $\mathbf{R} := \mathbf{P}^T$. As prolongation operator itself we use bi-linear interpolation for the cases with constant viscosity. This results in a natural embedding of the coarse grid finite element space into the finer one. The Galerkin coarse grid approximation (GCA), i.e. the construction of the coarse grid operator as $\mathbf{A}_{l-1} := \mathbf{R}\mathbf{A}_l\mathbf{P}$ then coincides with the direct coarse grid approximation (DCA) which constructs $\mathbf{A}_{l-1}$ by employing the same discretisation approach on the coarse mesh Ω_h^{l-1}. This will be different for the case of operator-dependent transfers in the variable viscosity case, as described below.

In this article, we will test five different smoothers, namely Jacobi (JAC), Gauss-Seidel / Successive over-relaxation in red-black (RB) and four-colour (4C) ordering as well as horizontal (X) and vertical (Y) line-smoothing in zebra ordering. As we deal with a vector-valued function $\mathbf{u}$ the smoothers are applied in a point-block fashion, i.e. both unknowns at a grid point are always relaxed together. Since our discretisation leads to a 9-point stencil, a red-black ordering is not sufficient to decouple the grid nodes into independent sets and, thus, a four-colour ordering was chosen. Additionally we will vary the number $\nu = (\nu_1, \nu_2)$ of pre- and post-smoothing steps and introduce a relaxation parameter ω. For this we limit our tests to the combinations $\nu \in \{(0,1), (1,1), (2,2)\}$ and $\omega \in \{0.8, 1.0, 1.5\}$.

Our main criterion for the analysis of the multigrid performance will be the computation of the spectrum of the iteration matrix M. For our problem and choice of smoothers multigrid is a linear iterative method. Thus, it can be expressed in matrix notation as $\Phi^{\mathrm{MG}}(u, \mathbf{A}, f) = Mu + Nf$, where M is the iteration matrix and $N = (I - M)\mathbf{A}^{-1}$. The iteration then reads $u^{(k+1)} := Mu^{(k)} + Nf$ for $k \in \mathbb{N}_0$. For a detailed account of linear iteration schemes and their representation see e.g. [19]. The spectrum of M provides essential information on the convergence behaviour. The spectral radius represents the asymptotic convergence rate. Furthermore the distribution of eigenvalues yields additional insights. If there are isolated outliers of the spectrum that prevent better convergence rates, then e.g. Krylov subspace acceleration may lead to much better performance.

Technically computation of the matrix M is straightforward. We select a basis of the $2\left(2^L - 1\right)^2$ dimensional space of degrees of freedom on the finest grid Ω_h^L given by

$$e_{ij}^m(x_{pq}, k) := \delta_{ip}\delta_{jq}\delta_{mk} \quad , \quad i, j \in \{1, \ldots, 2^L - 1\}, m \in \{1, 2\}$$

with x_{pq} being an interior node of Ω_h^L and $k \in \{1, 2\}$ selecting one of the two velocity components at the node. The columns of M are then obtained by applying a single V-cycle to e_{ij}^m as initial guess using a zero right-hand side. The spectrum can then be computed using standard tools.

Next, we investigate the influence of discontinuous coefficients on the multigrid performance. We define viscosity μ as a cell-wise quantity and consider three different scenarios denoted by μ_1, μ_2 and μ_3. The first scenario represents the 2D analogue of a radial viscosity jump across an interface, where we associate y with depth.

$$\mu_1(x, y) = \begin{cases} 1, & 0 < y < 0.5 \\ \eta, & 0.5 < y < 1 \end{cases} . \tag{11}$$

The strength of the contrast can be adapted by choosing η. The second scenario is a chequerboard pattern as representative for a combination of radial and lateral variations

$$\mu_2(x, y) = \begin{cases} 1, & (0 < x < 0.5 \wedge 0.5 < y < 1) \vee (0.5 < x < 1 \wedge 0 < y < 0.5) \\ \eta, & \text{otherwise.} \end{cases} \tag{12}$$

As viscosity is a cell-wise quantity, the two scenarios above are such that the jumps are resolved on all levels of the mesh hierarchy. In realistic models, where viscosity depends on other quantities such as temperature and pressure, and is, thus, unknown in advance, this cannot be achieved with a static mesh. Thus, in the third scenario we position the interface such that it is resolved only on the finest mesh and does not coincide with coarser meshes.

$$\mu_3(x, y) = \begin{cases} 1, & (0 < x < 0.5 + h_L \wedge 0.5 + h_L < y < 1) \vee \\ & \quad (0.5 + h_L < x < 1 \wedge 0 < y < 0.5 + h_L) \\ \eta, & \text{otherwise.} \end{cases} \tag{13}$$

Copyright © 2015 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

It is well known that the convergence of multigrid may significantly deteriorate in the presence of strongly varying or discontinuous coefficients. This can be improved by using operator-dependent transfers as detailed e.g. in [20]. We are going to compare standard bi-linear interpolation to an approach originally presented in [9]. In a nutshell the idea there is as follows. As viscosity is isotropic and, thus, represented by a scalar quantity, prolongation for both velocity components, resp. their corrections, is constructed similarly. The first step is to reduce the stencils $\mathcal{S}_{\text{eps}}$ or $\mathcal{S}_{\text{full}}$ to scalar 9-point stencils by replacing the tensor-valued entries by their traces.

In the second step interpolation weights are computed for the fine grid nodes lying on coarse grid edges. This construction follows [20], i.e. the stencil for the fine grid node is compacted to 1D by operator lumping and the two weights for the two coarse grid values are chosen such that

$$\mathbf{A}_{1\text{D}}\mathbf{P}v = 0 \ . \tag{14}$$

Here $\mathbf{A}_{1\text{D}}$ is the local 1D stencil at the fine grid node and v is a coarse grid function. As a last step, four weights are computed to interpolate from the values at the four corners of a coarse grid cell to the fine grid node in its centre. This can be done using the condition $\mathbf{AP}v = 0$ at the central fine grid node using its 2D stencil. The difference between [20] and [9] is that in the former the fine grid values that are already interpolated are used along the cell's edges. In the latter these are re-interpolated using (14) again, but with a 1D stencil that is constructed not by lumping weights, but by discarding those 2D weights not directly involved. This simplifies computation of the interpolation weights as it reduces the number of couplings that enter the computation.

For computation of coarse grid operators we use GCA in the case of operator-dependent transfers and DCA for bi-linear interpolation. In the latter case we transfer viscosity to the coarser levels by computing the arithmetic mean over the four fine grid cells that are combined into a coarse grid cell.

3. NUMERICAL RESULTS

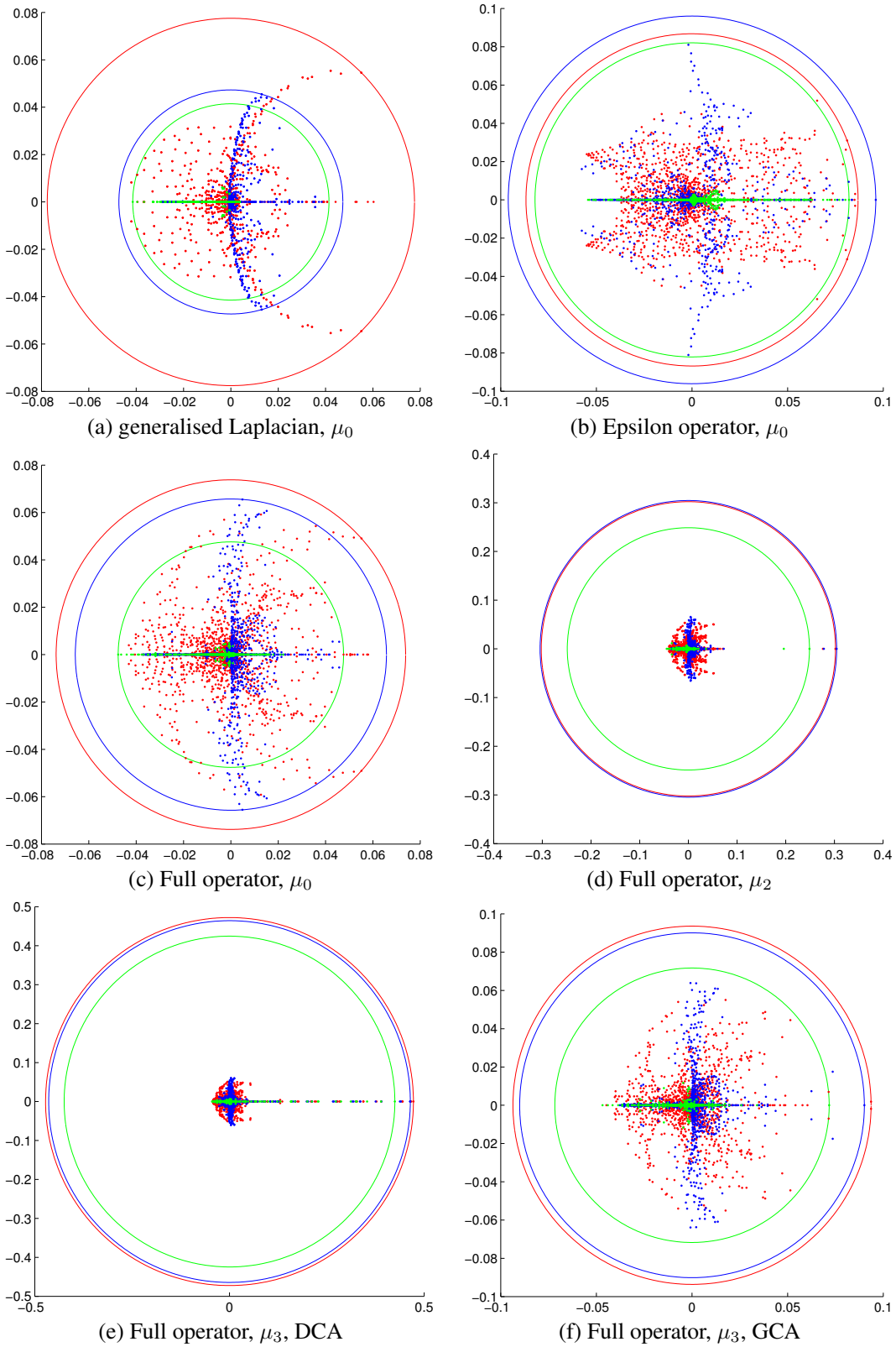
3.1. Constant Viscosity

We start our analysis with a constant viscosity profile setting $\mu \equiv 1$ and refer to this scenario as μ_0. This will provide us with reference values for the variable viscosity case below. As discussed in Sec. 2 both GCA and DCA here lead to the same coarse grid operator $\mathbf{A}_{l-1}$. Hence for constant viscosity we only consider DCA. In order to evaluate the effect of the increased complexity of operators in the *generalised Stokes* formulation we compare the discrete versions of the viscous operators (5) and (4) to the *generalised Laplacian* $\mathrm{L}(\mathbf{u}) = \operatorname{div}(\mu\nabla\mathbf{u})$. The latter for μ_0, of course, reduces to the vector Laplacian (6).

Table I lists the asymptotic convergence rates, given by the spectral radius $\rho(M)$ of the iteration matrix, for the different operators and smoothing strategies. We only show results for $\omega = 1$, as in our experiments this yielded the best results for the range tested. The exception is the Jacobi smoother where it is well known that under-relaxation performs better and results are given for $\omega = 0.8$. Considering the entries of the stencils, see (8) - (10), it is obvious that both vertical and horizontal line-smoothing must give identical results, as from a block-point of view the coupling in both directions is the same. Note that we can observe excellent convergence rates for all three operators. Actually, results for $\mathcal{S}_{\text{full}}$ are even better than for $\mathcal{S}_{\text{eps}}$. The latter operator performs slightly worse for all variations. This is due to a loss of coupling in the discretisation. As can be seen from (9) the u_1 component has no coupling to nearest neighbours in vertical direction, while the u_2 component has none in horizontal direction. A standard mode analysis and investigation of the eigenvectors of M for the JAC smoother indicate that this leads to certain error modes, being highly oscillatory in one component and smooth in the other one, for which our smoothers are not very efficient. Since these modes cannot be removed by the coarse grid correction either, a poorer multigrid convergence is obtained. The full operator (10) does not suffer from such a loss in coupling. Also note that the (RB) and (4C) point-smoothers are on a par with the, in practice more costly, line-smoothers.

Copyright © 2015 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla



(a) generalised Laplacian, μ_0

(b) Epsilon operator, μ_0

(c) Full operator, μ_0

(d) Full operator, μ_2

(e) Full operator, μ_3, DCA

(f) Full operator, μ_3, GCA

Figure 1. Spectrum of iteration matrix for 4-colour (blue), red-black (red) and (Y) line-smoother (green) for $\nu = (1, 1)$ and $\omega = 1$. Plots (a)-(c) show the spectra for the three different operators for a constant viscosity setting. In (d)-(f) the spectrum of the full operator is shown for the chequerboard scenarios μ_2 and μ_3. While for μ_2 (d), DCA and GCA have an identical spectrum, GCA (f) clearly beats DCA (e) for μ_3.

Copyright © 2015 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

Table I. Spectral radii for the three different operator forms in scenarios μ_0 and μ_1. With $\omega = 1$ for all smoothers, except JAC with $\omega = 0.8$. Rows correspond to $\nu = (0,1), (1,1)$ and $(2,2)$.

gen. Laplacian				Epsilon				Full			
JAC	RB	4C	X/Y	JAC	RB	4C	X/Y	JAC	RB	4C	X/Y
0.399	0.190	0.203	**0.152**	0.598	**0.325**	0.329	0.331	0.484	0.212	0.215	**0.203**
0.191	0.077	0.047	**0.041**	0.363	0.086	0.096	**0.082**	0.249	0.073	0.065	**0.047**
0.085	0.023	**0.008**	0.018	0.160	0.043	0.039	**0.024**	0.105	0.027	**0.018**	0.020

In Fig. 1 (a)-(c) the spectra of the multigrid iteration matrix are plotted for different smoothers. We observe that the eigenvalues exhibit characteristic patterns in which they are rather densely distributed and $\rho(M)$ is not dominated by separated outliers. This indicates that using an outer Krylov subspace method will generally yield only a moderate acceleration effect.

3.2. Variable Viscosity

We now turn to the influence of viscosity jumps on the convergence behaviour, by first investigating the two viscosity profiles μ_1 and μ_2 given in (11) and (12). We successively increase the parameter η from 10^1 up to 10^9 in steps of one order of magnitude.

The size of the jump in μ_1 has no influence on the speed of convergence. Here the grid hierarchy (2) resolves that discontinuity down to the coarsest grid level. As a consequence one can show that, as for the constant case μ_0, the operator-dependent transfer reduces to bi-linear interpolation. Also the viscosity averaging for DCA does not smear out the jump and, thus, DCA and GCA coincide. While the point smoothers "see" the change in coupling strength along the interface, this does not affect their performance. For increasing η the interface is simply seen as a Dirichlet resp. a Neumann boundary condition from the two halves of the domain. While the restricted residuals are larger in the η-half, the correction will have the correct scaling, since the coarse grid stencils on that part are also weighted by η in their setup.

The situation changes for the chequerboard scenario μ_2. We observe a monotonous increase in $\rho(M)$ up to $\eta = 10^3$. From that point on the rates remain almost constant for all larger values. The deterioration is mainly due to the junction point, see [20] for further details. However, the results for $\eta = 10^3$ in Tab. II show that even in this scenario we can obtain satisfactory convergence rates with standard components, thanks to the fact that the jumps are resolved on all levels. This does again hold for all three operators. Note also that the differences between the point- and line-smoothers are somewhat larger in this setting.

As expected results change significantly for scenario μ_3 from (13), see Tab. III. We observe a considerable deterioration of the convergence rates for DCA, as in this case the jumps are resolved on the finest mesh only. In this scenario the operator-dependent transfers can play out their full potential. They allow to obtain convergence rates close to the constant case μ_0. This holds not only for the line- but also for the point-smoothers.

In our experiments for μ_2 and μ_3^{DCA} the bulk of the eigenvalues is located within a circle of small radius. In case of the spectra in Fig. 1 (d)-(f) for the full operator with $\nu = (1,1)$, $\omega = 1$ and $\eta = 10^3$ e.g., we observe that most eigenvalues for μ_2 and μ_3^{DCA} are located inside a disk of radius 0.1. This is about the largest spectral radius we obtained for μ_3^{GCA} and also the distributions within that circle exhibit similar patterns than for GCA or the constant case μ_0. This is an indicator that an

Table II. Spectral radii for scenario μ_2 with $\eta = 10^3$ and $\omega = 1$ (JAC $\omega = 0.8$). Rows correspond to $\nu = (0,1), (1,1)$ and $(2,2)$. DCA and GCA are identical in this case.

gen. Laplacian				Epsilon				Full			
JAC	RB	4C	X/Y	JAC	RB	4C	X/Y	JAC	RB	4C	X/Y
0.688	0.502	0.521	**0.469**	0.762	**0.583**	0.609	0.590	0.716	0.529	0.552	**0.514**
0.495	0.274	0.276	**0.211**	0.604	0.366	0.369	**0.332**	0.532	0.302	0.305	**0.249**
0.318	0.166	0.158	**0.125**	0.417	0.221	0.211	**0.197**	0.348	0.180	0.172	**0.146**

Copyright © 2015 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

Table III. Spectral radii for scenario μ_3 with $\eta = 10^3$. With $\omega = 1$ (JAC $\omega = 0.8$).

		DCA				GCA			
ν	operator	JAC	RB	4C	X (Y)	JAC	RB	4C	X (Y)
(0,1)	gen.	0.815	0.652	0.650	**0.602**	0.486	0.213	**0.200**	0.216
(1,1)	Laplacian	0.678	0.455	0.449	**0.375**	0.271	0.071	0.056	**0.042**
(2,2)		0.500	0.279	0.273	**0.239**	0.119	0.021	**0.013**	0.017
(0,1)		0.853	0.708	**0.705**	0.707	0.655	0.413	**0.400**	0.413
(1,1)	Epsilon	0.734	0.514	0.506	**0.501**	0.454	0.181	0.174	**0.165**
(2,2)		0.561	0.318	0.315	**0.308**	0.252	0.107	0.101	**0.079**
(0,1)		0.827	0.669	0.666	**0.645**	0.554	0.286	**0.272**	0.291
(1,1)	Full	0.695	0.472	0.464	**0.425**	0.331	0.094	0.090	**0.072**
(2,2)		0.517	0.289	0.284	**0.261**	0.153	0.041	0.031	**0.019**

outer iteration may lead to a significant acceleration. As an initial test we use a V(1,1)-cycle with $\omega = 1$ as a preconditioner for the conjugate gradient method (PCG). In Fig. 2 we show approximate asymptotic convergence rates, computed from the residual norm reduction as

$$\rho_{\text{arf}} = \left(\frac{\|r^{(15)}\|}{\|r^{(20)}\|} \right)^{\frac{1}{5}}$$

for pure MG and PCG. Here we use a zero right-hand side and a random initial guess. Note that the rates for MG are close to the spectral radii shown in the tables above. Since the MG methods used so far lack the symmetry usually required for preconditioners in the CG method, we also show symmetric variants (MG symm) obtained by inverting the node resp. line order in the post-smoothing step. For MG symm alone this leads, effectively, to less smoothing when a pre-smoothing step follows on a post-smoothing one, and reduced convergence. This observation is in line with the theory in [21]. The effect is significantly more pronounced for the zebra-ordering than for the 4-colour smoother, as in the latter only one colour, or a forth of the nodes, is affected.

Overall we observe that PCG with an unsymmetric (4C) smoother is not a good choice, while with the line-smoother we get convergence improvements for μ_2 and μ_3^{DCA} even in this case. Note that the spectra for (Y) line-smoother are closer to that of a symmetric matrix. Using symmetric MG we always obtain a significant reduction in convergence rates by PCG. However, especially for the most interesting μ_3^{DCA} case rates are close to what we would obtain for two applications of MG. Thus, it depends strongly on the computational cost of the outer CG compared to the inner MG iteration whether there will be an actual run-time gain. Also we find that the PCG for the DCA case still cannot beat the GCA convergence rates for the μ_3 case.

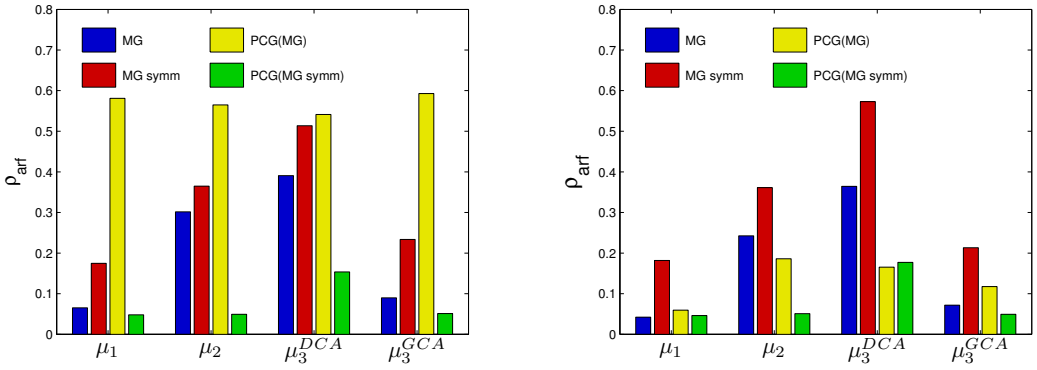


Figure 2. Approximate asymptotic convergence rates for MG and PCG variants with 4-colour smoother (left) and (Y) line-smoother (right) applied to different viscosity scenarios and the full operator.

Copyright © 2015 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

We finally remark that contrary to the previous cases, a slight improvement in asymptotic convergence rates was obtained for μ_2 and μ_3^{DCA} when using over-relaxation ($\omega = 1.5$) for the non-Jacobi smoothers.

4. CONCLUSION AND FUTURE WORK

Our study demonstrates that the two operators of the *generalised Stokes* problem as they appear in sophisticated models of mantle convection do not pose a challenge for the convergence of multigrid methods employing standard components. Results are comparable to those of the standard and generalised Laplacian in both constant and discontinuous viscosity settings. Furthermore we have demonstrated that, in a 2D setting, viscosity jumps that are resolved by the mesh hierarchy can satisfactorily be treated with standard components alone, while in other cases methods based on operator-dependent transfers are preferable. In future work we plan to extend our investigation to 3D and other finite-element ansatz functions, and we will study the behaviour of the methods when used as building blocks for generalised Stokes solvers.

REFERENCES

1. Ricard Y. Physics of Mantle Convection. *Treatise of Geophysics*, vol. 7, Schubert G (ed.). Elsevier, 2007.
2. Dziewonski AM, Anderson D. Preliminary reference Earth model. *Phys. Earth Plan. Int.* 1981; **25**:297–356.
3. Kronbichler M, Heister T, Bangerth W. High Accuracy Mantle Convection Simulation through Modern Numerical Methods. *Geophysical Journal International* 2012; **191**(1):12–29.
4. Burstedde C, Stadler G, Alisic L, Wilcox LC, Tan E, Gurnis M, Ghattas O. Large-scale adaptive mantle convection simulation. *Geophysical Journal International* 2013; **192**(3):889–906.
5. Oeser J, Bunge HP, Mohr M, Igel H. Frontiers in Computational Geophysics: simulations of mantle circulation, plate tectonics and seismic wave propagation. *100 Volumes NNFM and 40 Years Numerical Fluid Mechanics*, *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*, vol. 100, Hirschel E, Krause E (eds.). Springer, 2009.
6. Elman H, Silvester D, Wathen A. *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*. Oxford University Press, 2014.
7. Gmeiner B, Rüde U, Stengel H, Waluga C, Wohlmuth B. Performance and Scalability of Hierarchical Hybrid Multigrid Solvers for Stokes Systems. *SIAM Journal on Scientific Computing* 2015; Accepted for publication.
8. Verfürth R. A Combined Conjugate Gradient-Multigrid Algorithm for the Numerical Solution of the Stokes Problem. *IMA Journal of Numerical Analysis* 1984; **4**:441–455.
9. Yang WS, Baumgardner JR. A matrix-dependent transfer multigrid method for strongly variable viscosity infinite Prandtl number thermal convection. *Geophysical and Astrophysical Fluid Dynamics* 2000; **92**(3–4):151–195.
10. Davies DR, Davies JH, Bollada PC, Hassan O, Morgan K, Nithiarasu P. A hierarchical mesh refinement technique for global 3-D spherical mantle convection modelling. *Geoscientific Model Development* 2013; **6**(4):1095–1107.
11. Oeser J, Bunge HP, Mohr M. Cluster Design in the Earth Sciences: TETHYS. *High Performance Computing and Communications – Second International Conference, HPCC 2006, Munich, Germany, September 13-15, 2006, Proceedings*, *Lecture Notes in Comput. Sci*, vol. 4208, Gerndt M, Kranzlmüller D (eds.), Springer, 2006; 31–40.
12. Kameyama M. ACuTEMan: A multigrid-based mantle convection simulation code and its optimization to the Earth Simulator. *Journal of the Earth Simulator* 2005; **4**:2–10.
13. Kameyama M, Kageyama A, Sato T. Multigrid-based simulation code for mantle convection in spherical shell using Yin–Yang grid. *Phys. Earth Plan. Int.* 2008; **171**:19–32.
14. Baumgardner JR, Frederickson PO. Icosahedral discretization of the two-sphere. *SIAM Journal on Numerical Analysis* 1985; **22**(6):1107–1115.
15. Stemmer K, Harder H, Hansen U. A new method to simulate convection with strongly temperature and pressure-dependent viscosity in a spherical shell: Applications to the Earth's mantle. *Phys. Earth Plan. Int.* 2006; **157**:223–249.
16. Gmeiner B, Rüde U, Stengel H, Waluga C, Wohlmuth B. Towards Textbook Efficiency for Parallel Multigrid. *Numerical Mathematics Theory, Methods and Applications* 2015; Accepted for publication.
17. Davies DR, Wilson CR, Kramer SC. Fluidity: A fully unstructured anisotropic adaptive mesh computational modeling framework for geodynamics. *Geochem. Geophys. Geosyst.* 2011; **12**(6):20 pp.
18. Weismüller J, Gmeiner B, Rüde U, Wohlmuth B, Bunge HP. Channeled flow in the Asthenosphere in a synthetic high-resolution mantle convection model. *Geophysical Research Letters* ; In preparation.
19. Hackbusch W. *Iterative Solution of Large Sparse Systems of Equations*, *Applied mathematical sciences*, vol. 95. Springer, 1994.
20. Alcouffe RE, Brandt A, Dendy Jr JE, Painter JW. The Multi-Grid Method for the Diffusion Equation with Strongly Discontinuous Coefficients. *SIAM Journal on Scientific and Statistical Computing* 1981; **2**(4):430–454.
21. Holst M, Vandewalle S. Schwarz Methods: To Symmetrize or Not to Symmetrize. *SIAM Journal on Numerical Analysis* 1997; **34**(2):699–722.

Copyright © 2015 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla