Stokes preconditioning for the inverse Arnoldi method

L.S. Tuckerman[1], F. Bertagnolio[1], O. Daube[1], P. Le Quéré[1]
& D. Barkley[2]
[1]LIMSI-CNRS, B.P. 133, 91403 Orsay Cedex, France
[2] University of Warwick, Coventry CV4 7AL United Kingdom,

Summary

We have implemented a new method for computing leading eigenvalues for several two-dimensional hydrodynamic stability problems: spherical Couette flow and thermal convection in a rectangular container.

Our method for computing leading eigenvalues consists of the Arnoldi method carried out on the inverse matrix A^{-1} or, more generally, the shifted inverse matrix $(A-sI)^{-1}$, where $A \equiv N_U + L$ is the Jacobian of the discretized Navier-Stokes equations linearized about a previously computed steady state U. L represents the linear (viscous) part of the right-hand-side of the Navier-Stokes equations and N_U represents the nonlinear (advective) part linearized about U. To act with the inverse matrix, we solve a linear system via a conjugate gradient method, acclerated with the Stokes preconditioner.

1 Method

1.1 Arnoldi's method

We begin by reviewing the essential idea of the Arnoldi method [1, 2] for finding selected eigenvalues of an $M \times M$ matrix B. The Arnoldi method is a generalization of the power method. It consists of generating a sequence of K vectors $u_1, \ldots u_K$ by successive action of B. The set $\{u_1, \ldots u_K\}$ are orthogonalized to one another to form a set $\{w_1, \ldots w_K\}$, which is in turn normalized to form the basis set $\{v_1, \ldots v_K\}$. The linear space spanned by any of the sets $\{u_k\}$, $\{v_k\}$, $\{w_k\}$, $1 \leq k \leq K$ is called a Krylov space and the procedure which generates $u_{K+1}, v_{K+1}, w_{K+1}$ from the previous K-dimensional Krylov space is called an Arnoldi step. The large $M \times M$ matrix B is approximated by the much smaller $K \times K$ Hessenberg matrix H whose elements are $H_{jk} \equiv < v_j, Bv_k >$. Both the orthonormalization procedure and the main properties of the Arnoldi decomposition are summarized by the Arnoldi equation:

$$BV = VH + w_{K+1} e_K^T \tag{1.1}$$

where V is the $M \times K$ matrix whose elements are $V_{ij} \equiv v_j(i)$, the $M \times 1$ vector w_{K+1} is generated by an additional Arnoldi step, and e_K^T is the K^{th} unit row vector. H is diagonalized by calling a standard QR routine as in the packages EISPACK or LAPACK, leading to eigenpairs (λ, ϕ) satisfying:

$$H\phi = \lambda\phi \tag{1.2}$$

The eigenpairs of B are approximated by $(\lambda, V\phi)$. The accuracy of an eigenpair can be measured by the residual

$$r_{\text{eig}} \equiv \|BV\phi - \lambda V\phi\| = \|w_{K+1}\| \; |e_K^T\phi| \tag{1.3}$$

where we have used (1.1) and ϕ is assumed to be normalized.

Our implementation of the Arnoldi method is as follows. We choose a value of K, here $K = 4$, and an initial random vector u_1. We first take K Arnoldi steps by acting K times with the matrix B, thus generating the vectors $\{u_k\}$. We compute $\{v_k\}, \{w_k\}$ and the matrix H. Diagonalizing H gives K eigenpair estimates. To continue, we take one additional Arnoldi step, discard u_1, v_1, w_1 and rename $u_k \leftarrow u_{k+1}, v_k \leftarrow v_{k+1}, w_k \leftarrow w_{k+1}$. We then have a new Hessenberg matrices H which is again $K \times K$ and is diagonalized to give K updated eigenpair estimates.

1.2 Matrix transformations

In the form given above, the Arnoldi method is a generalization of the power method and will be most effective at computing dominant eigenvalues, i.e. those of largest magnitude. The eigenvalues of interest in hydrodynamic stability are not the dominant ones, which will be large and negative, reflecting primarily the spatial discretization, but the leading eigenvalues, i.e. those of largest real part which are responsible for bifurcations when the real part traverses zero. We therefore seek transformations of $A = N_U + L$ which give matrices B whose dominant eigenvalues are the leading eigenvalues of A. One choice is $B = \exp(At)$, which has precisely this property. Exact exponentiation of A would itself require its diagonalization and thus does not circumvent the necessity for computing eigenvalues (see, however, [3, 4]).

Time-integration of the linear evolution equation

$$\frac{du}{dt} = Au \tag{1.4}$$

is equivalent to computing $u(t) = \exp(At)u(0)$. Numerical time-integration schemes compute approximations to $u(t)$ by successively computing approximations to $u(t + \Delta t) = \exp(A\Delta t)u(t)$, each one accurate for small Δt. Thus any numerical time-integration scheme used on (1.4) is equivalent to an approximation to $\exp(A\Delta t)$. The drawback is the requirement that Δt be small. Let λ_1 be the leading eigenvalue of interest, and λ_2 the next leading eigenvalue. For the simple power method, the error at each step will decrease by a factor

$\exp(\lambda_2 \Delta t)/\exp(\lambda_1 \Delta t) = \exp(-(\lambda_1 - \lambda_2)\Delta t)$ which, for small Δt, will be very close to 1.

In order to converge more quickly, we wish to use the inverse transformation A^{-1}. For $B = A^{-1}$ the error in the power method decreases at each step by a factor λ_1/λ_2 which, for λ_1 near zero, is very small. Two drawbacks remain. One is that, while the exponential is available by merely linearizing a time-stepping code, the inverse is not. We will explain below how a time-stepping code may be modified to carry out action of the inverse. The other drawback is that the inverse power method or inverse Arnoldi method computes the least dominant eigenvalues, i.e. those of smallest magnitude, which are not necessarily identical with the leading ones. The least dominant eigenvalue does give essential important information: the threshold and eigenvector of the first steady bifurcation undergone by a flow. However, we may wish to study further steady bifurcations which may be imminent or interesting. More fundamentally, the system may undergo a Hopf bifurcation, in which the eigenvalues responsible for the bifurcation are a complex conjugate pair with zero real part but finite, and possibly large, imaginary part. These eigenvalues may not be those of smallest magnitude. The solution to this difficulty is well-known: to shift, applying $(A - sI)^{-1}$ rather than the inverse of A itself. Then, the eigenvalues computed are those closest to s, rather than those closest to zero. A real shift s is easy to carry out. For a complex shift, thought must be given to implementation of the method and interpretation of the results [5]. There is a trade-off in choosing s. The closer s is to an eigenvalue, the faster the Arnoldi method will converge to that eigenvalue. However, convergence to eigenvalues which are further will worsen. If s is very near an eigenvalue, it becomes impossible to compute any other eigenvalues.

1.3 Stokes preconditioning

In order to carry out the Arnoldi method on $(A - sI)^{-1}$, each Arnoldi step requires the solution of the linear system

$$(A - sI)\, u_{k+1} = u_k \tag{1.5}$$

Rather than compute the LU decomposition of $A - sI$ necessary for direct solution of (1.5), we solve (1.5) iteratively by one of the generalizations of the conjugate gradient method applicable to non-self-adjoint operators. Similar methods are discussed in [6, 7, 8, 9]. but with different iterative methods or preconditioners for solving (1.5). Diagonal or incomplete LU preconditioning is used with a conjugate-gradient-type method in [6, 7]. The solution of (1.5) via a class of methods called stationary solvers and containing, for example, Gauss-Seidel iteration, is discussed in [9].

Conjugate-gradient-type methods resemble the Arnoldi method for computing eigenvalues in that they generate an approximate solution to (1.5) by acting repeatedly with $A - sI$, starting with u_k as an initial vector. The rate at which such methods converge to an approximate solution of (1.5) depends strongly

on the condition number of the operator $A - sI$. The condition number can be greatly improved by use of a preconditioner, i.e. multiplying both sides of (1.5) by an operator approximating $(A - sI)^{-1}$.

We present the Stokes preconditioner [10, 11, 12, 13] for $A = N_U + L$ as follows:

$$\begin{aligned}
(N_U + L)u_{k+1} &= u_k \\
(I - \Delta tL)^{-1}\Delta t(N_U + L)u_{k+1} &= (I - \Delta tL)^{-1}\Delta t u_k \\
(I - \Delta tL)^{-1}\left[(I + \Delta tN_U) - (I - \Delta tL)\right]u_{k+1} &= (I - \Delta tL)^{-1}\Delta t u_k \\
\left[(I - \Delta tL)^{-1}(I + \Delta tN_U) - I\right]u_{k+1} &= (I - \Delta tL)^{-1}\Delta t u_k \quad (1.6)
\end{aligned}$$

The Stokes preconditioner $P \equiv (I - \Delta tL)^{-1}\Delta t$ has the following properties:

- In a typical time-integration code with implicit viscous timestepping, the action of $(I - \Delta tL)^{-1}$ is already available, and imposes the constraints of boundary conditions and, if necessary, incompressibility. The operator $(I - \Delta tL)^{-1}$ (or even L) is not constructed; considerable effort in computational fluid dynamics has been expended to develop algorithms which carry out this operation at low cost (closer to $O(M)$ than to $O(M^2)$) without constructing or inverting operators.

- For $\Delta t \gg 1$, $P = (I - \Delta tL)^{-1}\Delta t \approx -L^{-1}$. Since in (1.6), Δt plays the role of an algebraic parameter and not that of a time increment, taking Δt large poses no problem. Its value should be chosen to maximize the convergence rate of the conjugate-gradient-type iteration.

The operator appearing on the left-hand-side of (1.6) is merely the difference between two successive linearized timesteps of a typical time-integration code combining explicit advective timestepping with implicit viscous timestepping such as:

$$u(t + \Delta t) = (I - \Delta tL)^{-1}(I + \Delta tN_U)u(t) \quad (1.7)$$

In fact, the wide use of explicit-advective/implicit-viscous timestepping is related to the effectiveness of the preconditioner. The wide range of eigenvalues of L (rather than of N_U) leads to the problem of stiffness in integrating the time-dependent Navier-Stokes equations; this problem is alleviated by implicit timestepping of L. The same property of L leads to the poor condition number of A and consequent slow convergence of conjugate gradient iteration attempted directly on (1.5). This is why the Stokes preconditioner resembling L^{-1} is effective for $A = N_U + L$.

For the shifted operator $A - sI$ with s real, the calculation in (1.6) becomes

$$\begin{aligned}
(I - \Delta tL)^{-1}\Delta t(N_U + L - sI)u_{k+1} &= (I - \Delta tL)^{-1}\Delta t u_k \\
\left[(I - \Delta tL)^{-1}(I + \Delta t(N_U - sI)) - I\right]u_{k+1} &= (I - \Delta tL)^{-1}\Delta t u_k \quad (1.8)
\end{aligned}$$

Thus we additionally modify the time-stepping code (1.7) by including the shift $-sI$ in the nonlinear operator N_U. We may abbreviate (1.8) by

$$P(A - sI)\, u_{k+1} = Pu_k \tag{1.9}$$

and define the linear residual by

$$r_{\text{lin}} \equiv \frac{\|P(A - sI)u_{k+1} - Pu_k\|}{\|Pu_k\|} \tag{1.10}$$

Our method thus consists of both outer and inner iterations. The outer iterations are those of the Arnoldi method and consist primarily of carrying out $(A - sI)^{-1}$ as in (1.5), or rather its equivalent but better conditioned version $(P(A - sI))^{-1}P$ as in (1.8-1.9), along with the far less costly auxiliary operations explained in the summary of the Arnoldi method. This is accomplished by carrying out inner iterations of the conjugate-gradient-type method used in solving (1.8-1.9). These consist primarily of repeatedly acting with the operator $P(A - sI)$ on the left-hand-side of (1.8-1.9) on vectors chosen by the conjugate-gradient-type algorithm. We act with $P(A - sI)$, not by constructing explicitly any of the matrices in (1.8), but rather by adapting a time-stepping algorithm such as (1.7). By far the dominant expense of the combined method is acting with $P(A - sI)$ on a vector, close in cost and content to a single timestep of the original timestepping code.

There exists an inherent conflict between the outer Arnoldi and inner conjugate-gradient-type iterations. The Arnoldi method converges faster to the least dominant eigenvalue as the magnitude of this eigenvalue decreases (relative to the other eigenvalues). On the other hand, conjugate-gradient-type methods converge fastest when the condition number is closest to one, or when the matrix is farthest from singular, which is incompatible with an eigenvalue whose magnitude is much smaller than that of next larger eigenvalue. We will see in the numerical experiments on spherical Couette flow that this conflict need not in fact pose any practical difficulties.

1.4 Complex shifts

We discuss our implementation of a complex shift. With a complex shift $s = s_r + is_i$, (1.5) becomes:

$$(A - s_r I - is_i I)\, u_{k+1} = u_k \tag{1.11}$$

We decompose (1.11) into its real and imaginary parts:

$$\begin{pmatrix} A - s_r I & s_i I \\ -s_i I & A - s_r I \end{pmatrix} \begin{pmatrix} u_r \\ u_i \end{pmatrix}_{k+1} = \begin{pmatrix} u_r \\ u_i \end{pmatrix}_k \tag{1.12}$$

The matrix in (1.12) is $2M \times 2M$. Each eigenvalue λ of A corresponds to a pair of eigenvalues $\lambda - s_r \pm is_i$ of the matrix in (1.12). Thus, a real eigenvalue λ is transformed to a complex conjugate pair, and a complex conjugate pair to two complex conjugate pairs. To precondition (1.12), we multiply real and imaginary parts by the Stokes preconditioner, leading to:

$$\begin{pmatrix} P(A - s_r I) & Ps_i \\ -Ps_i & P(A - s_r I) \end{pmatrix} \begin{pmatrix} u_r \\ u_i \end{pmatrix}_{k+1} = \begin{pmatrix} Pu_r \\ Pu_i \end{pmatrix}_k \tag{1.13}$$

We then solve (1.13) by a conjugate-gradient-type method.

2 Spherical Couette flow

2.1 Description of spherical Couette flow

Spherical Couette flow is the flow between two concentric differentially rotating spheres. Spherical Couette flow with an outer stationary sphere is characterized by two dimensionless quantities, the Reynolds number $Re \equiv \Omega_1 r_1^2/\nu$ and the gap ratio $\sigma \equiv (r_2 - r_1)/r_1$ where r_1, r_2 are the inner and outer radii, Ω_1 is the angular velocity of the inner sphere, and ν is the kinematic viscosity. Like the better known cylindrical Couette flow, spherical Couette flow undergoes an instability as Re is increased, which leads to the formation of vortices. The physical mechanism responsible for the instability is the radial gradient in angular momentum, which is decreased by the radial mixing of fluid engendered by the vortices. One measure of this gradient is the torque required to rotate the inner sphere at angular velocity Ω_1 or, equivalently for a steady flow, to keep the outer sphere stationary.

Extensive studies [14, 15, 16, 17, 18, 19] of the case $\sigma = 0.18$ have led to the following conclusions: In the range $Re < 850$, all steady states are axisymmetric, i.e. independent of the angle about the axis of rotation, and reflection-symmetric about the equator. For brevity, we describe states or eigenvectors which are symmetric or antisymmetric with respect to reflection about the equator as symmetric and antisymmetric, respectively. Those which are neither are called asymmetric. There exist three types of steady states: the zero-vortex state, with no vortices, the one-vortex state, with one vortex in each hemisphere, and the two-vortex state with two vortices in each hemisphere.

The steady solutions obtained by gradually increasing Re from $Re = 0$ are located on what can be termed the basic branch. Along the basic branch, the zero-vortex state evolves continuously into the two-vortex state. Since the vortices are infinitesimal at onset, it is difficult to define a precise criterion for when this occurs, but $Re \approx 735$. The zero- and two- vortex states along the basic branch are unstable over the range $650 < Re < 775$. The eigenvalue responsible for this instability is real, and the corresponding eigenvector is antisymmetric. The endpoints of this interval correspond to subcritical pitchfork bifurcations, which means in this case that the asymmetric bifurcating branches originating at $Re = 650$ and at $Re = 775$ are unstable and not the final destinations of the transitions triggered by the instability. Instead, an asymmetric transient terminates at a steady stable symmetric one-vortex state.

The symmetry of the steady states has important mathematical and numerical consequences. The Jacobian $N_U + L$ about any symmetric state U is block

diagonal, with eigenvectors that are either symmetric or antisymmetric. All numerical calculations involving the Jacobian can in principle be decomposed into two decoupled problems, each of half the size of the original problem. Even if this economy is not explicitly taken advantage of in coding, the iterative Arnoldi and conjugate-gradient-type methods we use may converge more quickly because the Jacobian is block diagonal.

2.2 Numerical results on spherical Couette flow

We now describe the results of eigenvalue computations using the inverse Arnoldi method with Stokes preconditioning for this case of spherical Couette flow. Approximately 100-200 lines were added to an existing time-stepping program [18] to implement the method. This program uses a tensor-product basis set (Chebyshev polynomials in radius multiplied by trigonometric functions of meridional angle) to represent fields. This leads to a Laplacian which is highly structured (although not sparse) and thus to rapid action with $(I - \Delta t L)^{-1}$. Incompressibility is imposed to machine accuracy via the influence matrix technique. The program was previously adapted to calculate the leading eigenpair via the simple power method [19] or Arnoldi's method [11] on the approximate exponential (1.7), and also to calculate steady states by Newton's method with Stokes preconditioning [11]. The numerical resolution used is $M = 4096$. The value of Δt is 1000. (Recall that, while Δt plays the role of a time increment in the original time-stepping program (1.7), for the linear systems (1.6) or (1.8), Δt is merely an algebraic parameter chosen empirically to maximize convergence.) For the outer Arnoldi method, Krylov spaces of dimension $K = 4$ are constructed. For the inner conjugate-gradient-type solution, we use the NSPCG (NonSymmetric Preconditioned Conjugate Gradient) software package [20], allowing up to 100 iterations to achieve a linear residual of 10^{-5}. Most computations use the BCGS (BiConjugate Gradient Squared) method, but results using GMRES (Generalized Minimum Residual) are also reported.

Figure 1 shows up to four leading eigenvalues of the basic flow. One eigenvalue is positive over the range $650 < Re < 775$. This is the leading eigenvalue; the other positive quantity shown in figure 1 is the imaginary part of a complex conjugate pair to be discussed shortly. This is the eigenvalue responsible for the subcritical pitchfork bifurcations initiating transition to the one-vortex state described above; the corresponding eigenvector is antisymmetric. For $Re < 744$, the next leading eigenvalues are a complex conjugate pair. The corresponding eigenvectors are symmetric. The real and imaginary parts are both shown in figure 1. At $Re = 744$, these coalesce and become two real eigenvalues, the lower of which decreases so rapidly with Re that it is no longer visible on figure 1 for $Re > 755$. The last eigenvalue shown on figure 1 belongs to a second antisymmetric eigenvector.

The leading eigenvalue attains a maximum at $Re = 735$, precisely at the value where the torque is minimum and very near the value at which the basic

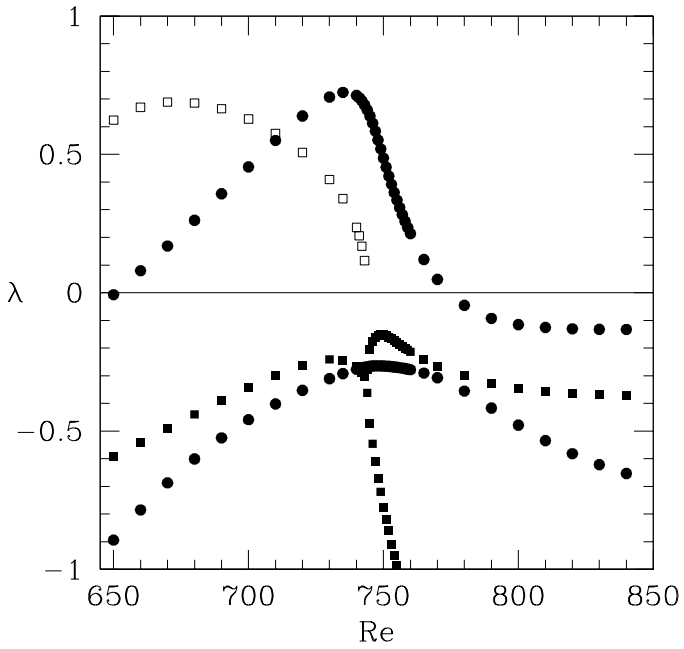


Figure 1 Four leading eigenvalues for spherical Couette flow. Corresponding to antisymmetric eigenvectors are two eigenvalues (circles), one of which is positive over $650 < Re < 775$. Corresponding to symmetric eigenvectors are two eigenvalues (squares) which form a complex conjugate pair for $Re < 744$; both real (filled squares) and imaginary (hollow squares) parts are shown.

flow evolves from a zero-vortex to a two-vortex flow. All of the other leading eigenvalues (or, in the case of the complex conjugate pair, their real part) also have maxima near, though not exactly at, this Reynolds number. This reflects the fact that the angular momentum gradient responsible for the instability has been alleviated by the radial fluid mixing of the vortices. In evolving from a zero-vortex to a two-vortex flow, the basic branch becomes less unstable and the eigenvalues decrease.

Figure 2 shows an example of the convergence of the inner conjugate-gradient-type iteration in solving the linear system (1.8) required by a single Arnoldi step. The Reynolds number is $Re = 750$. The upper part of figure 2 displays the residual r_{lin} defined in (1.10) of the linear system (1.8) as a function of BCGS iteration number. We investigate the influence of the singularity of the matrix by taking four values of the shift s approaching the eigenvalue of smallest magnitude $\lambda = -0.15181122$. In all cases, the convergence is highly irregular and non-monotonic (although it may be made monotonic with long piecewise-constant segments merely by retaining the solution with the lowest residual heretofore). For $s = 0$ and $s = -0.15$, the requested linear residual r_{lin} of 10^{-5} is attained at 47 and 63 BCGS iterations, respectively. For $s = -0.1518$, r_{lin}

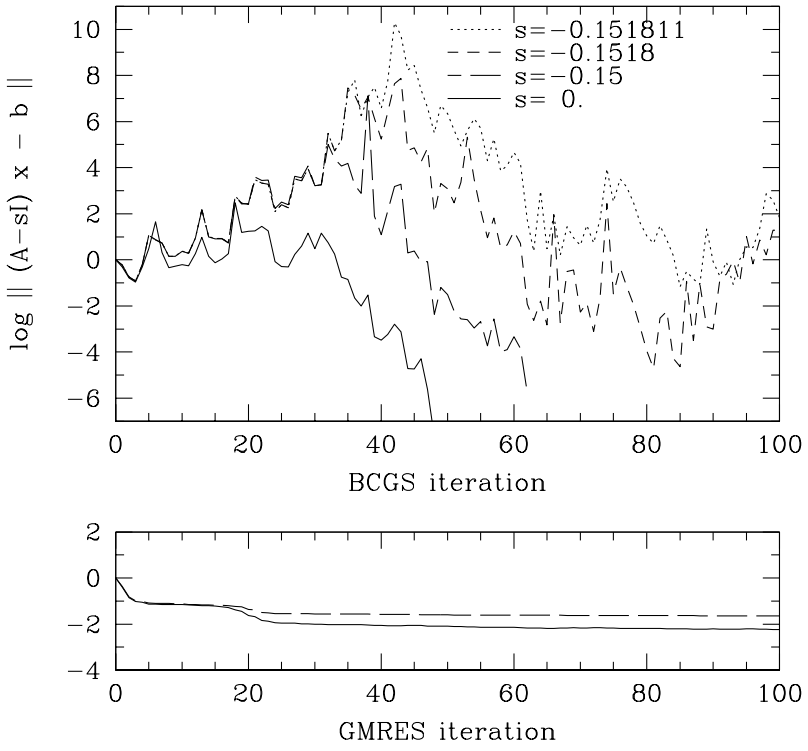


Figure 2 Convergence of conjugate-gradient-type method for shifts s approaching an eigenvalue $\lambda = -0.15181122$. Above: BCGS. Convergence is satisfactory for $|s - \lambda| \leq 10^{-5}$ Below: GMRES. Convergence is not satisfactory for any value of s.

never attains this value but reaches 2×10^{-5} at iteration 81. For $s = -0.151811$, the lowest value of r_{lin} achieved in 100 iterations is 7×10^{-2} at iteration 85. Full eigenvalue computations with this shift nevertheless succeed in computing λ to eight-digit accuracy, but not reliably: four (i.e. three spurious) copies of this eigenvalue are produced and the eigenvalue residual r_{eig} associated with the (accurately estimated) eigenvalue oscillates between 20 and 5000. We conclude that approaching singularity of the matrix does adversely affect the convergence of BCGS, but the shift must be taken very close to the eigenvalue before the effect is at all severe.

The lower part of figure 2 shows the convergence of the far more widespread method GMRES in solving the same linear system. We see that the linear residual decreases monotonically, a popular feature of GMRES, but that it never goes below 5×10^{-3} in 100 GMRES iterations, regardless of s. The reason for the far better performance of BCGS for this problem is unknown.

Figure 3 shows the convergence of the full method to the four leading eigen-

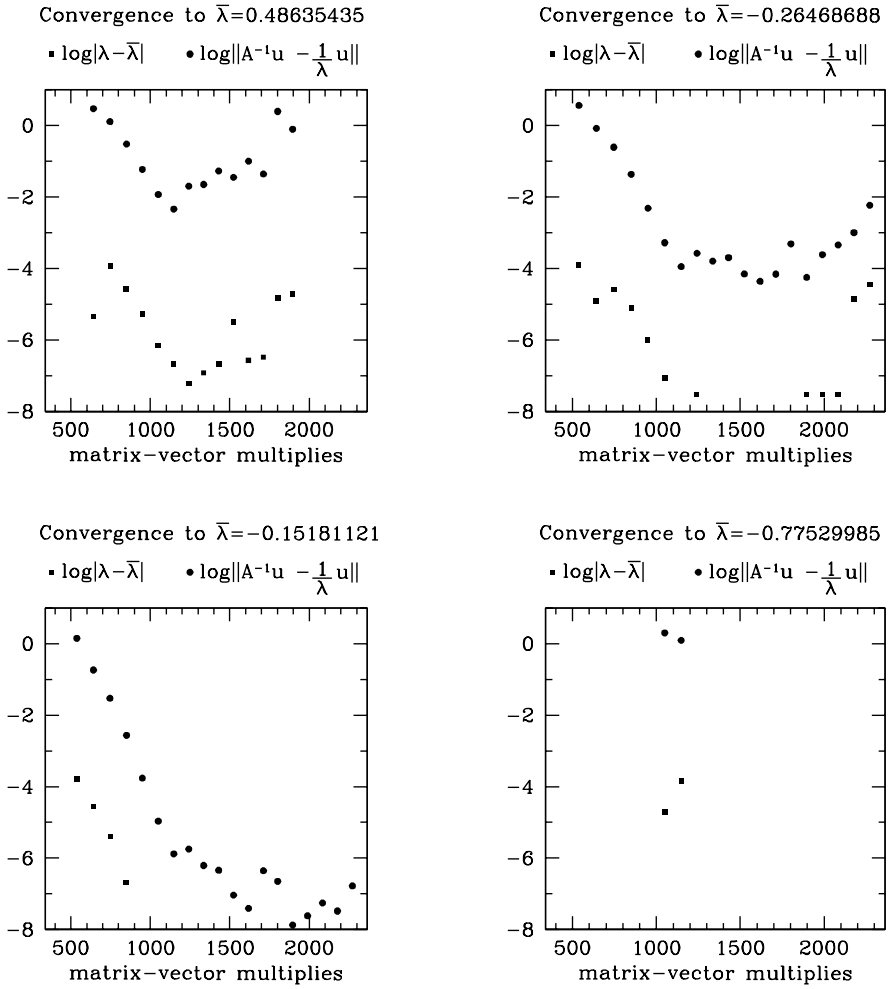


Figure 3 Convergence of inverse Arnoldi method with Stokes preconditioning and $s = 0$ to the four leading eigenvalues at $Re = 750$. The exact error (squares) is much smaller than the eigenvalue residual (circles): $|\lambda - \bar{\lambda}| \approx 10^{-4}\, r_{\text{eig}}$. Estimates of the eigenvalue $\bar{\lambda} = -0.15181122$ of smallest magnitude converge rapidly and monotonically. Estimates of the next two eigenvalues, -0.26468688 and 0.48635434, converge rapidly but worsen if too many steps are taken. The fourth eigenvalue, -0.77529992, is not computed satisfactorily with $s = 0$.

values at $Re = 750$ using the shift $s = 0$. The abcissa is the number of actions of $P(A - sI)$ of (1.9). Each BCGS iteration requires two actions and, as was seen on figure 1, typically 50 BCGS iterations are taken to solve the linear system to accuracy 10^{-5}. Hence each Arnoldi step requires about 100 actions. It is only after $K = 4$ Arnoldi steps that eigenvalue approximations are first generated. Afterwards, they are generated at each subsequent Arnoldi step. Two measures of accuracy are shown on figure 3. One measure is the exact *a posteriori* error $|\lambda - \bar{\lambda}|$ of each eigenvalue estimate λ determined after the eigenvalue $\bar{\lambda}$ has been determined to accuracy 10^{-8}, typically in another computation using s near $\bar{\lambda}$. The other measure is the *a priori* error, the eigenpair residual r_{eig} which (1.3) shows to be available at no additional cost after each diagonalization of H and without knowledge of the exact eigenvalue. We see from figure 3 that $|\lambda - \bar{\lambda}| \approx 10^{-4} r_{\text{eig}}$. This ratio is unexpectedly small, given that these two measures should theoretically be related by

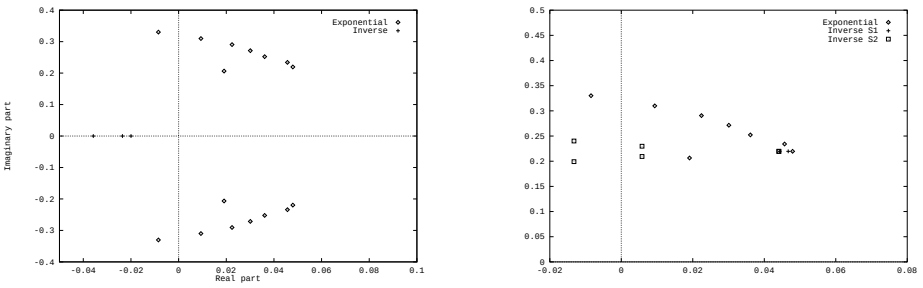
$$|\lambda - \bar{\lambda}| = |\lambda\bar{\lambda}| \, |\lambda^{-1} - \bar{\lambda}^{-1}| \leq |\lambda\bar{\lambda}| \, \kappa \, r_{\text{eig}} \tag{2.14}$$

where $\kappa \geq 1$ is the condition number of the eigenvector matrix and $|\lambda\bar{\lambda}| \approx \bar{\lambda}^2$ is between 0.02 and 0.6 for the eigenvalues shown on figure 3. Thus eigenpair residuals of order 1 might seem to be unacceptable, but in fact here correspond to four-digit accuracy in the eigenvalues.

The estimates of the eigenvalue $\lambda = -0.15181122$ of smallest magnitude converges most rapidly and monotonically, attaining eight-digit accuracy by 10 Arnoldi steps, i.e. 10 actions of $(A - sI)^{-1}$ accomplished via approximately 1000 actions of $P(A - sI)$. The eigenvalue of next largest magnitude $\lambda = -0.26468688$ achieves this soon after, by about 12 Arnoldi steps. However, as can be seen, the convergence is non-monotonic in that after 21 Arnoldi steps, the estimate begins to seriously degrade. This long-term degradation is to be expected for all but the eigenvalue closest to zero (or, more generally, to s). Thus, the best estimates for different eigenvalues are obtained at different Arnoldi iterations. The estimates for the eigenvalue of next large magnitude, $\lambda = 0.48635434$ behaves similarly. Finally, a reliable estimate for the eigenvalue of fourth largest magnitude, $\lambda = -0.77529992$, is not obtained for this shift of $s = 0$. Although it can be seen that three-digit accuracy is retained over five Arnoldi iterations, it is difficult to ascertain this without *a posteriori* knowledge of the exact eigenvalue and r_{eig} is never less than 1.

3 Thermal convection

Thermal convection is flow driven by buoyancy due to a temperature gradient whose intensity is quantified by the nondimensional Rayleigh number Ra. In the case we have studied, a horizontal temperature gradient is imposed across a rectangular container whose height is four times its width and $Ra = 3 \times 10^6$. The steady states and transitions undergone by this flow have been previously



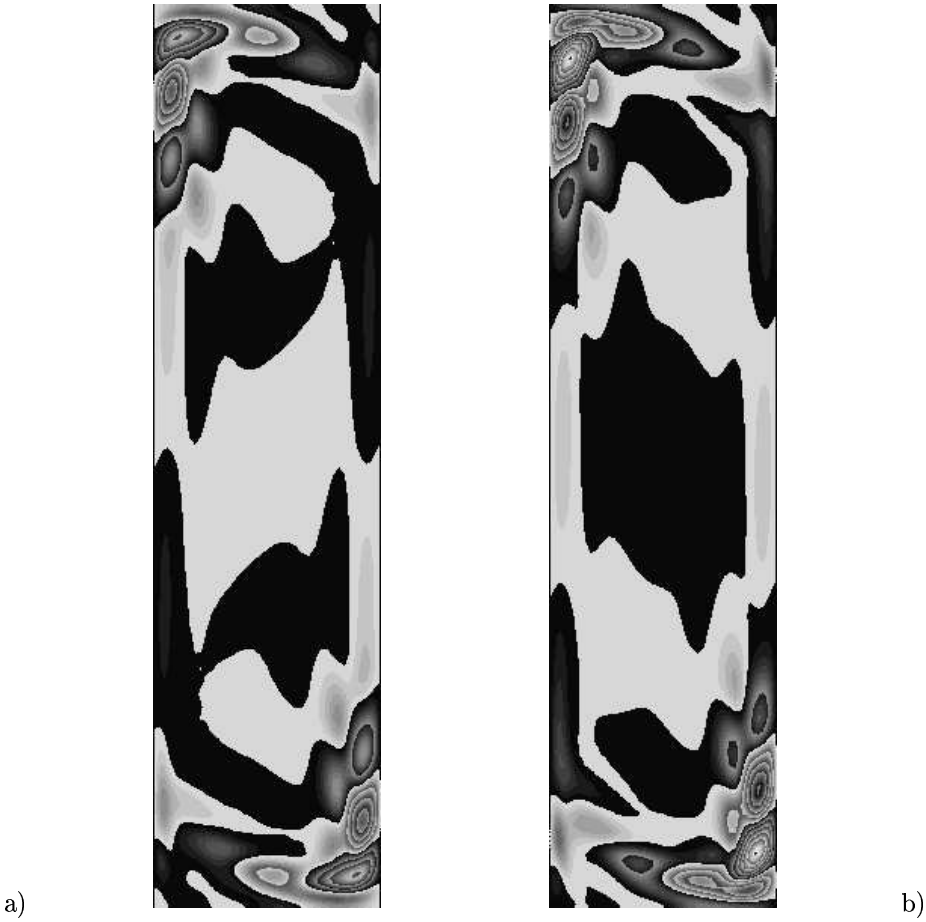
a) b)

Figure 4 a) Eigenvalues calculated by the Arnoldi method using the exponential (diamonds) and the inverse without shift (crosses). The inverse method does not calculate the leading eigenvalues sought, but instead those of smallest magnitude. b) Eigenvalues calculated by the Arnoldi method using the exponential (diamonds), the inverse with complex shift $s_1 = 0.1 + 0.219i$ (crosses), and with $s_2 = 0.06 + .0219i$ (squares). The leading eigenvalue $\lambda = 0.048 \pm 0.22i$ is calculated using s_1 but not using s_2.

studied via numerical time-integration [21]. Unstable steady states have been computed by adapting the time-stepping program to carry out Newton's method with Stokes preconditioning. Leading eigenvalues belonging to these unstable states have been computed by adapting the same time-stepping program to carry out Arnoldi's method on the approximate exponential. In order to see if these eigenvalues could be computed more rapidly, we implemented the inverse Arnoldi method with Stokes preconditioning. GMRES was used to solve the linear system.

Figure 4a shows the eigenvalues calculated by the exponential (diamonds) and by the inverse without shift (crosses). The leading eigenvalue pair is $\lambda = 0.0479 \pm 0.219i$ and the leading eigenvalues are not those of smallest magnitude. Hence the inverse Arnoldi method requires a shift. In addition to $s_0 = 0$, we have tried two other values for the shift: $s_1 = 0.1 + 0.219i$ and $s_2 = 0.06 + .0219i$. Figure 4b shows the eigenvalues calculated by the exponential (diamonds), inverse Arnoldi method with shift s_1 (crosses), and inverse Arnoldi method with shift s_2 (squares). With shift s_1, the leading eigenvalue pair is recovered. Although the error in the eigenvalue is fairly large, examination of the real part of the eigenvector (figure 5a) shows it to be the same as that calculated by the exponential. With shift s_2, the leading eigenvalue pair is not recovered. Shifts s_1 and s_2 both calculate an eigenvalue pair (the real part of whose eigenvector is shown on figure 5b) which is deemed to be spurious, since it is not one of those calculated by the exponential.

This case demonstrates that our method can indeed calculate the leading complex eigenpairs for this system. However, the accuracy achieved is poor and is greatly affected by the shift. We believe that the poor performance is due to the



a) b)

Figure 5 a) Real part of leading eigenvector calculated by the Arnoldi method on the exponential and on the inverse with shift s_1.
b) Real part of a spurious eigenvector calculated by the Arnoldi method on the inverse with shifts s_1 and s_2.

inadequate convergence of GMRES for the linear system, especially for the shift $s_2 \approx \lambda$. Tests are underway to determine if this is indeed the cause, and, if so, whether the accuracy can be better improved and controlled, either by use of BCGS or some other means.

Acknowledgments. This research was supported in part by a European COPERNICUS contract to the project STABLE.

Bibliography

[1] W.E. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Q. Appl. Math **9**, 17 (1951).

[2] Y. Saad, *Variations on Arnoldi's method for computing eigenelements of large unsymnmetric matrices*, Linear Algebra Appl. **34**, 269 (1980).

[3] E. Gallopoulos and Y. Saad, *Efficient solution of parabolic equations by Krylov approximation methods*, SIAM J. Sci. Statist. Comput. **13**, 1236 (1992).

[4] W.S. Edwards, L.S. Tuckerman, R.A. Friesner, and D. Sorensen, *Krylov methods for the incompressible Navier-Stokes equations*, J. Comput. Phys., **110**, 82 (1994).

[5] B.N. Parlett and Y. Saad, *Complex shift and invert strategies for real matrices*, Linear Algebra Appl. **88**, 575 (1987).

[6] H. Mittelmann, C.C. Law, D.F. Jankowski, and G.P. Neitzel, *A large, sparse, and indefinite generalized eigenvalue problem from fluid mechanics*, SIAM J. Sci. Stat. Comput. **13**, 411 (1992).

[7] H. Mittelmann, K.T. Chang, D.F. Jankowski, and G.P. Neitzel, *Iterative solution of the eigenvalue problem in Hopf bifurcation for teh Boussinesq equations*, SIAM J. Sci. Comput. **15**, 704 (1994).

[8] K. Meerbergen and D. Roose, *Matrix transformations for computing rightmost eigenvalues of large sparse non-symmetric eigenvalue problems*, IMA J. Numer. Anal. **16**, 297 (1996).

[9] K. Meerbergen and D. Roose, *The restarted Arnoldi method applied to iterative linear system solvers for the computation of rightmost eigenvalues*, SIAM J. Matrix Anal. Appl. **18**, 1 (1997).

[10] L.S. Tuckerman, *Steady-state solving via Stokes preconditioning; recursion relations for elliptic operators*, in Lecture Notes in Physics: Proc. of the 11th Int'l. Conf. on Numerical Methods in Fluid Dynamics, ed. by D.L. Dwoyer, M.Y. Hussaini, and R.G. Voigt, (Springer, New York, 1989), p. 573.

[11] C.K. Mamun and L.S. Tuckerman, *Asymmetry and Hopf bifurcation in spherical Couette flow*, Phys. Fluids **7**, 80 (1995).

[12] D. Barkley and L.S. Tuckerman, *Stokes preconditioning for the inverse power method*, in Lecture Notes in Physics: Proc. of the 15th Int'l. Conf. on Numerical Methods in Fluid Dynamics ed. by P. Kutler, J. Flores, and J.-J. Chattot (Springer, New York, 1997), p. 75.

[13] L.S. Tuckerman and D. Barkley, *Bifurcation analysis for time-steppers*, in Numerical Methods for Bifurcation Problems and Large-Scale Dynamical Systems ed. by E. Doedel, B. Fiedler, Y. Kevrekides, W.-J. Beyn, J. Lorenz, L.S. Tuckerman, E. Titi, H.B. Keller, and D. Aronson (Springer, New York, to appear).

[14] M. Wimmer, *Experiments on a viscous fluid flow between concentric rotating spheres*, J. Fluid Mech. **78**, 317 (1976).

[15] G. Schrauf, *Branching of Navier-Stokes equations in a spherical gap*, in Lecture Notes in Physics: Proc. of the 8th Int'l. Conf. on Numerical Methods in Fluid Dynamics, ed. by E. Krause (Springer, New York, 1982), p. 474.

[16] G. Schrauf and E. Krause, *Symmetric and asymmetric Taylor vortices in a spherical gap*, in Proc. of the 2nd IUTAM Symp. on laminar-turbulent transition, ed. by V.V. Kozlov (Springer, New York, 1985), p. 659.

[17] G. Schrauf, *The first instability in spherical Taylor-Couette flow*, J. Fluid Mech. **166**, 287 (1986).

[18] P.S. Marcus and L.S. Tuckerman, *Numerical simulation of spherical Couette flow. Part I: Numerical methods and steady states*, J. Fluid Mech. **185**, 1 (1987).

[19] P.S. Marcus and L.S. Tuckerman, *Numerical simulation of spherical Couette flow. Part II: Transitions*, J. Fluid Mech. **185**, 31 (1987).

[20] D.R. Kincaid, T.C. Oppe, and W.D. Joubert, *An overview of NSPCG: A nonsymmetric preconditioned conjugate gradient package*, Center for Numerical Analysis Technical Report CNA-228, Austin Texas, 1988.

[21] S. Xin and P. Le Quéré, *Direct simulation of chaotic natural convection in a differentially heated cavity of aspect ratio 4 with spectral methods*, J. Fluid Mech. **304**, 87 (1995).