Commun. Comput. Phys.
doi: 10.4208/cicp.281114.290615a

Vol. **18**, No. x, pp. 1-16
x 2015

Laplacian Preconditioning for the Inverse Arnoldi Method

Laurette S. Tuckerman[1],*

[1] *PMMH (UMR 7636 CNRS – ESPCI – UPMC Paris 6 – UPD Paris 7 – ParisTech – PSL), 10 rue Vauquelin, 75005 Paris, France.*

Received 28 November 2014; Accepted (in revised version) 29 June 2015

Abstract. Many physical processes are described by elliptic or parabolic partial differential equations. For linear stability problems associated with such equations, the inverse Laplacian provides a very effective preconditioner. In addition, it is also readily available in most scientific calculations in the form of a Poisson solver or an implicit diffusive timestep. We incorporate Laplacian preconditioning into the inverse Arnoldi method, using BiCGSTAB to solve the large linear systems. Two successful implementations are described: spherical Couette flow described by the Navier-Stokes equations and Bose-Einstein condensation described by the nonlinear Schrödinger equation.

AMS subject classifications: 37M20, 65F08, 65F18, 65P10, 65P30, 76E07

Key words: To be provided by author

1 Introduction

Many physical systems are governed by parabolic evolution equations of the general form

$$\partial_t U = LU + N(U), \tag{1.1}$$

where L is the Laplacian operator and N represents some combination of nonlinear terms or a multiplicative potential. Two examples which we will consider are the Navier-Stokes equations

$$\partial_t U = -(U\cdot\nabla)U - \nabla P + \nu\nabla^2 U, \tag{1.2a}$$

$$\nabla\cdot U = 0, \tag{1.2b}$$

and the nonlinear Schrödinger equation

$$-i\partial_t \Psi = \left[\frac{1}{2}\nabla^2 + \mu - V(\mathbf{x}) - a|\Psi|^2\right]\Psi. \tag{1.3}$$

*Corresponding author. *Email address:* `laurette@pmmh.espci.fr` (L. S. Tuckerman)

Steady solutions of (1.1) satisfy

$$0=LU+N(U) \tag{1.4}$$

and the Jacobian operator evaluated at U is defined by

$$A\equiv L+N_U, \tag{1.5}$$

where N_U is the linearization of N at U. Steady bifurcations from U occur when an eigenvalue of A crosses zero. For this reason, we are interested in the eigenvalues of (1.5) which are closest to zero. These eigenvalues can be calculated by the classic inverse power method, generalized to the inverse Arnoldi method [1]. The sequence $\{u_k\equiv A^{-(k-1)}u_1;\, k=1,\cdots K\}$ is generated by solving

$$Au_{k+1}=u_k. \tag{1.6}$$

This sequence is orthonormalized by the usual Arnoldi process to yield the basis $\{v_k\}$ for the Krylov space and the upper Hessenberg matrix

$$H_{jk}\equiv\langle v_j,A^{-1}v_k\rangle. \tag{1.7}$$

H is directly diagonalized, yielding

$$H\phi_k=\lambda_k\phi_k, \tag{1.8}$$

with estimated eigenpairs $(\lambda_k^{-1},V\phi_k)$ for A, where V is the rectangular matrix whose j^{th} column is v_j. A shift s can, as usual, be used to accelerate convergence of the Arnoldi method to a desired eigenvalue. In this case, we solve

$$(A-sI)u_{k+1}=u_k. \tag{1.9}$$

Solving the linear systems (1.6) or (1.9) is by far the most time-consuming part of the algorithm. This means that it is far more difficult to find the smallest eigenvalues of A than the largest ones, since acting with A is usually far easier than acting with its inverse. The purpose of this paper is to present a method for quickly formulating and solving the linear systems (1.6) or (1.9), assuming that we have a time-stepping code for integrating the time-dependent equation (1.1).

A related scheme has been used to compute steady states via Newton's method [2–6]. This scheme has been proposed and used as a method for calculating eigenvalues in [7–10]. Here we provide a study of its convergence.

2 Method

2.1 Laplacian preconditioning

Our method for solving (1.6) is based on the BiCGSTAB variant of the conjugate gradient method [11]. Since A results from the spatial discretization of a partial differential equation, its size may be quite large. Denoting by M the number of points or modes necessary to represent the variation in each spatial dimension D, we assume that $10 \le M \le 1000$. The size of A is then $10^D \le M^D \le 10^{3D}$, i.e. $10^2 \le M^2 \le 10^6$ for problems with two spatial dimensions and $10^3 \le M^3 \le 10^9$ for three-dimensional problems. In addition, A is poorly conditioned, primarily because of the wide range of eigenvalues of the Laplacian L. The smallest and largest eigenvalues of L can be estimated roughly as -1 and $-DM^2$, yielding a condition number for L of 3×10^4 for a three-dimensional case with $M=100$. Thus, preconditioning by the inverse L^{-1} of the Laplacian will be very effective. In addition, multiplication by the inverse Laplacian, i.e. solution of the Poisson equation, is a ubiquitous problem for which a great deal of computational technology has been developed, for all sorts of spatial discretizations.

The problem we solve instead of (1.6) uses as a preconditioner either the Poisson operator L^{-1}:

$$L^{-1}(L+N_U)u_{k+1} = L^{-1}u_k \tag{2.1}$$

or the Helmholtz operator $(I-\Delta tL)^{-1}\Delta t$:

$$(I-\Delta tL)^{-1}\Delta t(L+N_U)u_{k+1} = (I-\Delta tL)^{-1}\Delta t u_k. \tag{2.2}$$

Appropriate boundary conditions must be imposed on either equation; this is assumed to be encompassed in the notation L^{-1} or $(I-\Delta tL)^{-1}$. Use of (2.2) instead of (2.1) is motivated by the utilization of timestepping schemes in which the evolution of the diffusive terms is calculated implicitly. Implicit timestepping is required because the wide range of eigenvalues of L leading to poor conditioning in the context of the linear system (1.6) also leads to stiffness of the evolution equation (1.1). The simplest such timestepping scheme is the backward-Euler/forward Euler first-order algorithm. For the linearized operator $L+N_U$, this algorithm reads:

$$u(t+\Delta t) = (I-\Delta tL)^{-1}(I+\Delta tN_U)u(t). \tag{2.3}$$

For small Δt, (2.3) necessarily approximates the action on $u(t)$ of the exponential of $\Delta t(L+N_U)$. The difference between two consecutive linearized timesteps can be written:

$$\begin{aligned} u(t+\Delta t)-u(t) &= (I-\Delta tL)^{-1}(I+\Delta tN_U)u(t)-u(t) \\ &= (I-\Delta tL)^{-1}\left[(I+\Delta tN_U)-(I-\Delta tL)\right]u(t) \\ &= (I-\Delta tL)^{-1}\Delta t(N_U+L)u(t), \end{aligned} \tag{2.4}$$

which is seen to be the action of the operator on the left hand side of (2.2). In order for the Helmholtz operator $(I-\Delta tL)^{-1}\Delta t$ to be an effective preconditioner, Δt must be set to a large value, in contrast to the small value required for timestepping. Varying Δt can also provide a way of testing the preconditioning, since $\Delta t \to 0$ is equivalent to no preconditioning, while $\Delta t \to \infty$ is equivalent to preconditioning by L^{-1}.

Thus (2.2) is carried out by the equivalent procedure:

$$\left[(I-\Delta tL)^{-1}(I+\Delta tN_U)-I\right]u_{k+1} = (I-\Delta tL)^{-1}\Delta t u_k, \tag{2.5}$$

where the action of the operator on the left-hand-side is carried out by taking the difference between two widely spaced linearized timesteps as in (2.4), and the action on the right-hand-side is the implicit part of the timestepping scheme (2.3).

In the two examples we have implemented, we have used a *pseudospectral* spatial discretization [12, 13]. Functions are represented both as series of basis functions, such as Fourier series or Chebyshev polynomials (spectral representation), and also by their values on a spatial grid. Actions and inversions of the Laplacian L are carried out in the spectral representation, while the actions of the multiplicative operator N or N_U are carried out on the grid representations; all of these operations scale approximately linearly in M^D, the number of gridpoints or basis functions. Fourier or Chebyshev transforms are used to pass between the spectral and grid representations in a time proportional to $DM^D \log M$.

2.2 Implementation of the Arnoldi method

Our implementation of the Arnoldi method is as follows. Choosing a small value of K, typically $2 \le K \le 6$, and an initial random vector u_1, we take K Arnoldi steps, generating the K Krylov vectors $\{u_k\}$, the $K \times K$ matrix H, and the K eigenpair estimates. To continue, we take one additional Arnoldi step, discard u_1, redefine $u_k \leftarrow u_{k+1}$, and generate an updated H and eigenpair estimates. The procedure is halted when the residual $||(A^{-1} - \lambda_k^{-1} I) V \phi_k||$ or $||(A - \lambda_k I) V \phi_k||$ of the eigenpair sought is sufficiently small.

Inaccuracy in the computed eigenvalues can arise from several sources, any of which may be so large as to prevent the use of the method. One source, characteristic of all Arnoldi methods, is the projection of a high-dimensional operator onto a low-dimensional H. Another source of error is the iterative solution of the linear system (2.5), the preconditioned version of (1.6). The errors incurred correspond to the two roles played by the procedure (1.6)-(1.7): to generate a Krylov space $\{u_k\}$ via (1.6), and to generate a low-dimensional projection H to A^{-1} via (1.7). The usual Arnoldi process combines the two functions, but they can be decoupled, reducing or eliminating the error incurred in (1.7). If direct action by A is feasible, inexpensive, and more accurate than action by A^{-1} via iterative solution, we can replace (1.7) by

$$H_{jk} \equiv \langle v_j, A v_k \rangle \tag{2.6}$$

to construct a Krylov space representation of A rather than of A^{-1}; the eigenvalues of H are then estimates of those of A. Action by A is not desirable in generating the Krylov space (which would then be focused on eigenvectors corresponding to eigenvalues of largest magnitude). Thus, we continue to generate the Krylov vectors by acting with A^{-1} via (1.6), targeting the method to eigenvalues closest to 0 or to s. We then carry out the additional multiplications by A in (2.6), separately from the Arnoldi procedure, to generate H. This adds little to the cost, while eliminating an additional source of inaccuracy.

If we are following eigenvalues along a branch of steady states which depends on a parameter such as a Reynolds number, then a very accurate estimate s of the desired eigenvalue is that obtained for a neighboring parameter value. A shift by s can be incorporated into the action of N_U, so that Eqs. (2.1) and (2.2) are replaced by

$$L^{-1}(L+N_U-sI)u_{k+1}=L^{-1}u_k, \tag{2.7}$$

$$(I-\Delta tL)^{-1}\Delta t(L+N_U-sI)u_{k+1}=(I-\Delta tL)^{-1}\Delta tu_k. \tag{2.8}$$

The overall method consists of a sequence of outer Arnoldi iterations, each of which requires a sequence of inner BiCGSTAB iterations. There exists an inherent conflict between the outer Arnoldi and inner BiCGSTAB iterations: the Arnoldi method should converge fastest when the eigenvalues differ most, while BiCGSTAB should converge fastest when the matrix is well conditioned. We will see in the applications that this conflict posed no practical difficulty in the case of our spherical Couette flow problem (Navier-Stokes equations), but may be responsible for problems encountered in the case of our Bose-Einstein condensation problem (the nonlinear Schrödinger Equation).

3 Application to spherical Couette flow

3.1 Physical description of flow and instabilities

Spherical Couette flow is the flow between two concentric differentially rotating spheres. When the outer sphere is held fixed, spherical Couette flow 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 [14–19] 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 as symmetric or antisymmetric the steady states and eigenvectors which are reflection-symmetric or antisymmetric about the equator, as well as the corresponding eigenvalues. 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 it is approximately $Re=735$. 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 775 are unstable and not the final destinations of the transitions triggered by the instability. Instead, evolution via a sequence of asymmetric transient states terminates at a steady stable symmetric one-vortex state.

Fig. 1 shows up to four leading eigenvalues of the basic flow, calculated using the inverse Arnoldi method. The eigenvalue which is positive over the range $650<Re<775$ is that responsible for the subcritical pitchfork bifurcations initiating transition to the one-vortex state described above. This is the leading eigenvalue and it is antisymmetric. For $Re<744$, the next leading eigenvalues are a symmetric complex conjugate pair whose real and imaginary parts are both shown in Fig. 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 Fig. 1 for $Re>755$. The last eigenvalue shown on Fig. 1 belongs to a second antisymmetric eigenvector.

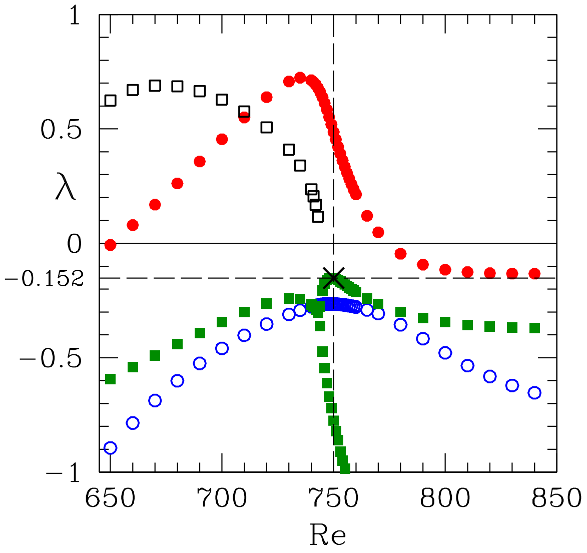

Figure 1: Four leading eigenvalues for spherical Couette flow. Two eigenvectors are antisymmetric and their corresponding eigenvalues are shown as solid red and hollow blue circles. One of these eigenvalues (solid red circles) is positive over the range $650<Re<775$. Two eigenvectors are symmetric. Their eigenvalues form a complex conjugate pair for $Re<744$; their real part is shown as solid green squares and the imaginary part as hollow black squares. For $Re>744$, the corresponding eigenvalues are both real and shown as solid green squares. The cross at $(Re=750,\lambda=-0.152)$ shows the eigenvalue which we target for our test case.

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 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.

3.2 Numerical results

We now describe the results of eigenvalue computations for this case of spherical Couette flow. Approximately 50 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 $(I-\Delta t L)^{-1}$. The azimuthal velocity and the meridional streamfunction are used to represent the axisymmetric fields and 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 [3] on the approximate exponential (2.3), and also to calculate steady states by Newton's method with Laplacian preconditioning [3] (called Stokes preconditioning in this context). The numerical resolution usually used is 16×128 which, with two fields, leads to matrices of size 4096×4096. Our non-dimensionalization is such that the rotation period of the inner sphere is 70 and the timestep used for time-integration in this regime is $\Delta t = 1$.

We focus on the calculation of the eigenvalue of smallest magnitude at Reynolds number 750, whose value is $\bar{\lambda} = -0.15181122$. Krylov spaces of dimension $K=2$ are used for the Arnoldi iterations. To solve the linear systems, BiCGSTAB is given a stopping criterion of

$$||Au_{k+1}-u_k||/||u_k|| \le 10^{-7}, \tag{3.1}$$

with a maximum number of iterations of 2000. We measure CPU time by the number of matrix-vector multiplications, each approximately equivalent to a timestep. We recall that for this pseudospectral code, the cost of such a multiplication increases only slightly faster than linearly in the number of gridpoints of basis functions. Fig. 2 shows the convergence of the error $|\lambda-\bar{\lambda}|$ as a function of the number n of matrix-vector multiplications. Each point corresponds to a single Arnoldi iteration, i.e. a single solution of (2.2), and thus to a sequence of BiCGSTAB iterations. The timestep Δt, which determines the efficiency of the Helmholtz preconditioning $(I-\Delta t L)^{-1}$, is varied from 0.01 to 100. Very fast convergence is seen for $\Delta t = 10$. Six-digit accuracy is obtained after 7 Arnoldi steps, requiring $n=754$ matrix-vector multiplies. Further increase of Δt has little effect, as shown by the similar values corresponding to $\Delta t = 100$; this shows that the preconditioning operator is essentially the inverse Laplacian L^{-1}. As Δt is decreased, the preconditioning also decreases drastically. For $\Delta t = 1$, the number of matrix-vector

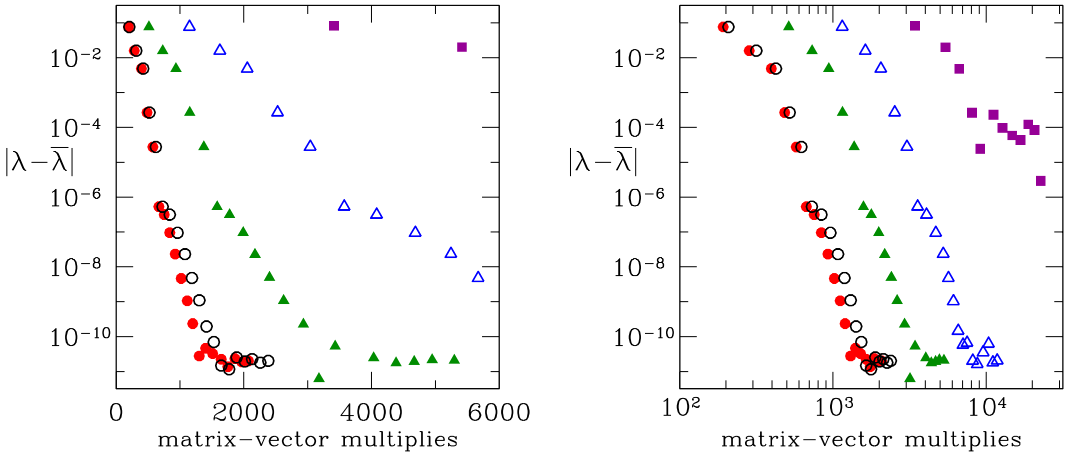

Figure 2: Left: Convergence of error $|\lambda-\bar{\lambda}|$ as a function of the number of matrix-vector multiplications for different values of Δt. For $\Delta t=100$ (solid red circles) or $\Delta t=10$ (hollow black circles), inverse Laplacian preconditioning is achieved and convergence is rapid. As Δt is decreased through 1 (solid green triangles) and 0.1 (hollow blue triangles), convergence decreases dramatically. For $\Delta t=0.01$ (solid purple squares), the preconditioner has become so ineffective that BiCGSTAB does not converge to the requested precision in the maximum number of iterations allowed. Right: Same data in log-log representation indicates that $|\lambda-\bar{\lambda}|(n,\Delta t)\sim(n\Delta t^{\alpha})^{-\beta}$ with $\alpha\approx 1/4$ and $\beta\approx 14$.

multiplications required by BiCGSTAB to converge to the same accuracy increases from 100 to 300, while for $\Delta t=0.1$, this number is on the order of 600. For $\Delta t=0.01$, convergence would require more than the maximum number of iterations we have allowed for BiCGSTAB.

We quantify the dependence of convergence on Δt further by plotting the same data logarithmically in the number of matrix-vector multiplications. Each sequence $|\lambda-\bar{\lambda}|(n)$ has approximately the same slope. The sequences are displaced rightwards by the same interval as Δt is decreased by factors of 10 from 10 to 1 to 0.1. This implies that

$$|\lambda-\bar{\lambda}|(n,\Delta t)\sim(n\Delta t^{\alpha})^{-\beta} \tag{3.2}$$

and Fig. 2 yields the estimated dependence

$$|\lambda-\bar{\lambda}|(n,\Delta t)\sim(n\Delta t^{1/4})^{-14}. \tag{3.3}$$

Fig. 3 shows the convergence of $|\lambda-\bar{\lambda}|$ as shifts progressively closer to $\bar{\lambda}$ are employed, more specifically $s=0$, $s=-0.1$, $s=-0.15$, and $s=-0.152$. Two unexpected conclusions can be drawn from Fig. 3. The first is that as s approaches $\bar{\lambda}$, convergence continues to improve, despite the fact that the matrix must become less well conditioned as $s\to\bar{\lambda}$. The second is that convergence does not improve monotonically as s approaches $\bar{\lambda}$; convergence for $s=-0.1$ is slower than that for $s=0$. In seeking to understand this, we notice that BiCGSTAB requires fewer matrix-vector multiplications for each Arnoldi iteration for the case $s=-0.1$ than for the case $s=0$. We therefore forced BiCGSTAB to

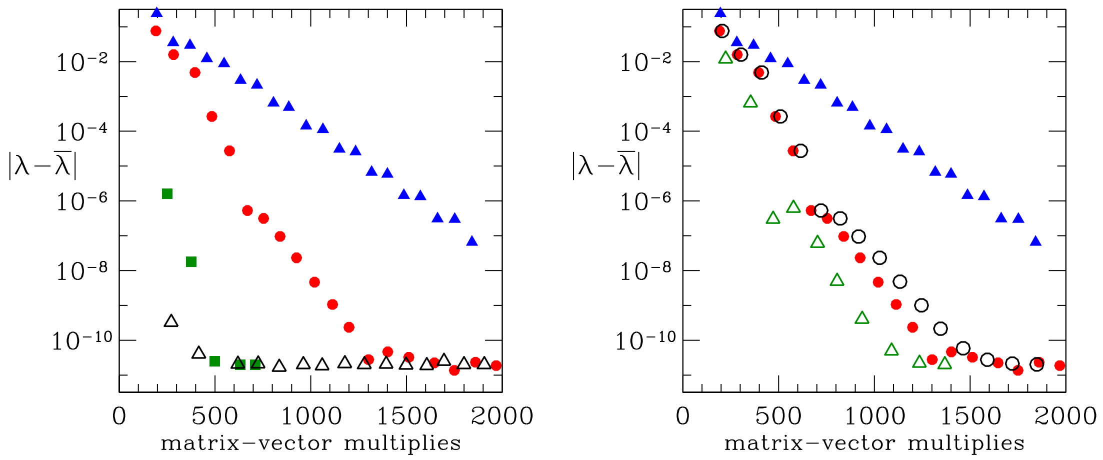

Figure 3: Left: Convergence of error $|\lambda-\bar{\lambda}|$ as a function of the number of matrix-vector multiplications for different values of shift s, with Δt fixed at 100 and BiCGSTAB stopping criterion fixed at 10^{-7}. As the shift approaches $\bar{\lambda}=-0.15181122$ from $s=0$ (solid red circles), through $s=-0.1$ (solid blue triangles) $s=-0.15$ (solid green squares), to $s=-0.152$ (hollow black triangles), convergence greatly accelerates except for $s=-0.1$. Right: Convergence of error $|\lambda-\bar{\lambda}|$ for $s=0$ (circles) and $s=-0.1$ (triangles) and for BiCGSTAB stopping criterion 10^{-7} (solid) and 10^{-9} (hollow). Reducing the stopping criterion improves convergence for $s=-0.1$ but has little effect for $s=0$.

increase the number of matrix-vector multiplications by reducing its stopping criterion from 10^{-7} to the more stringent value of 10^{-9}. Fig. 3 shows that this change greatly improves the convergence of the $s=-0.1$ case but has little effect on the $s=0$ case. This suggests that the outer Arnoldi and inner BiCGSTAB iterations are intermeshed in a way which is more complicated to capture than by the simple value of the stopping criterion. We have not explored the effect of varying s from one Arnoldi iteration to the next.

Fig. 4 shows that convergence is almost unaffected by an increase in the size of the Krylov space from $K=2$ to $K=4$. We also increased the spatial resolution from $2\times16\times128$ to $2\times32\times256$. This increases the size of the matrix by a factor of 4 and the cost of each matrix-vector multiplication by slightly more than a factor of 4, and changes the value of the eigenvalue of the spatially discretized problem to $\bar{\lambda}=-0.15197992$. We see, however, that the dependence of $|\lambda-\bar{\lambda}|$ on the number of matrix-vector multiplications is almost unaffected. This demonstrates our claim that the cost of this method is approximately linear in the number of gridpoints or modes, i.e. in the size of the matrix.

4 Bose-Einstein condensation

The nonlinear Schrödinger equation (1.3), also called the Gross-Pitaevskii equation [20, 21], has been used to described the behavior of a Bose-Einstein condensate [22–26], in which atoms are cooled so drastically that they populate the same quantum-mechanical state.

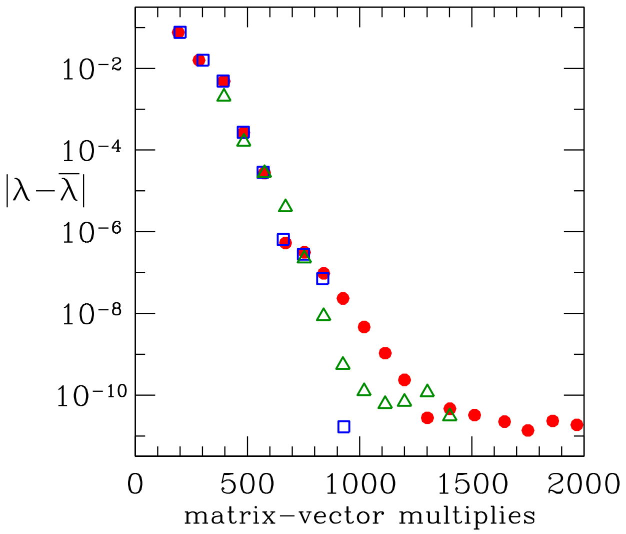

Figure 4: Convergence of error $|\lambda-\bar{\lambda}|$ as a function of Krylov space dimension and spatial resolution. Solid red circles: $K=2$ and spatial resolution 16×128. Hollow green triangles: $K=4$ and spatial resolution 16×128. Hollow blue squares: $K=2$ and 32×256.

The steady states are a stable (elliptic) and unstable (hyperbolic/elliptic) pair which meet at a Hamiltonian saddle-node bifurcation [5,27,28] described as follows. The eigenvalues occur in pairs $\pm\lambda$ or $\pm i\lambda$. Along the elliptic branch, all are imaginary. As this branch is followed towards the Hamiltonian saddle-node bifurcation, one imaginary eigenvalue pair $\pm i\lambda$ approaches zero, becoming zero at the saddle-node bifurcation. As we leave the bifurcation along the unstable hyperbolic branch, the eigenvalue pair $\pm\lambda$ is real, with absolute value increasing along the branch. The rate at which the critical eigenvalue $|\lambda|$ approaches and recedes from zero determines the rate at which the Bose-Einstein condensate decays [5,28].

The low temperature needed for Bose-Einstein condensation is modeled by a confining harmonic potential $\frac{1}{2}|\boldsymbol{\omega}\cdot\mathbf{x}|^2$. Two types of calculations of the steady states and eigenvalues have previously been carried out. First, a variational technique which approximates steady states as Gaussians yields analytic estimates of the critical eigenvalues [5, 10, 27, 28]. Second, if the potential is isotropic (spherically symmetric), and this assumption is made throughout, then the problem has effectively only one spatial dimension and the stability matrix is small enough to be directly diagonalized [5]. We describe an implementation of the inverse Arnoldi method with Laplacian preconditioning which calculates the eigenvalues in a general geometry and present results for the two cylindrically symmetric potentials, termed a cigar and a pancake [10,29,30].

We write the nonlinear Schrödinger equation in the abbreviated form:

$$-i\partial_t\Psi = L\Psi + W(\Psi), \tag{4.1}$$

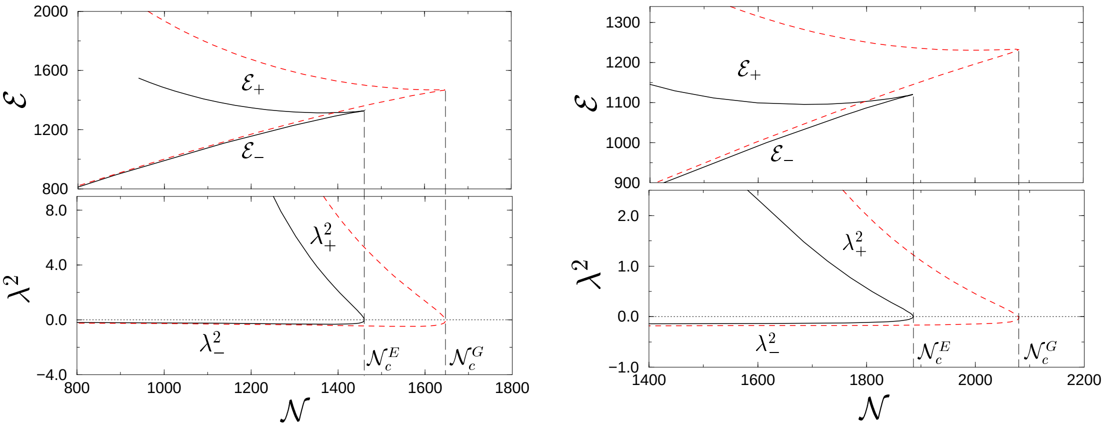

Figure 5: Stationary solutions of the GP equation as a function of particle number $\mathcal{N}$ for two non-isotropic potentials with $\omega_z = \omega_r/5$ (cigar, left) and $\omega_r = \omega_z/5$ (pancake, right). Top: value of the energy functional $\mathcal{E}_+$ on the unstable (hyperbolic) branch and $\mathcal{E}_-$ on the stable (elliptic) branch. Bottom: square of the bifurcating eigenvalue ($\lambda^2_\pm$); $|\lambda_-|$ is the energy of small excitations around the stable branch. Solid lines: exact solution of the GP equation. Dashed lines: Gaussian approximation.

where

$$L\Psi \equiv \frac{1}{2}\nabla^2\Psi, \tag{4.2}$$

$$W(\Psi) \equiv \left[\mu - \frac{1}{2}|\boldsymbol{\omega}\cdot\mathbf{x}|^2 - a|\Psi|^2\right]\Psi, \tag{4.3}$$

$$|\boldsymbol{\omega}\cdot\mathbf{x}|^2 = \omega_r r^2 + \omega_z z^2. \tag{4.4}$$

We begin by presenting our results for two geometries, a cigar ($\omega_z = \omega_r/5$), and a pancake ($\omega_r = \omega_z/5$) [10,29,30] in Fig. 5. The particle number $\mathcal{N}$ and energy $\mathcal{E}$ are defined by:

$$\mathcal{N} = \int d^3x |\Psi|^2, \tag{4.5}$$

$$\mathcal{E} = \int d^3x \left[\frac{1}{2}|\nabla\Psi|^2 + \frac{1}{2}|\boldsymbol{\omega}\cdot\mathbf{x}|^2|\Psi|^2 + \frac{a}{2}|\Psi|^4\right]. \tag{4.6}$$

The control parameter is the particle number, which is conserved by the nonlinear Schrödinger equation. (The solutions can also be indexed by μ.) When the particle number is below $\mathcal{N}_c$, there exist two solutions, one stable (elliptic) and the other unstable (hyperbolic). The Hamiltonian saddle-node bifurcation takes place at $\mathcal{N}_c$, and for a particle number exceeding $\mathcal{N}_c$, no Bose-Einstein condensate exists.

The operators L and W defined in (4.2) and (4.3) are spatially discretized using the pseudospectral method. We assume a three-dimensional periodic Cartesian domain, on which Ψ and $|\boldsymbol{\omega}\cdot\mathbf{x}|^2$ are expanded as three-dimensional trigonometric (Fourier) series. In this representation, solution to the Poisson equation is trivial, since each Fourier component is merely divided by the square of its wavenumber:

$$\begin{aligned}&\nabla^2 \sum_{l_x,l_y,l_z} f_{l_x,l_y,l_z} \exp(il_x x+il_y y+il_z z)\\ =&-\sum_{l_x,l_y,l_z} |l|^2 f_{l_x,l_y,l_z} \exp(il_x x+il_y y+il_z z).\end{aligned} \tag{4.7}$$

(The $|l|=0$ component is determined by the boundary conditions in solving the Poisson equation and can be treated arbitrarily when L^{-1} is used merely as a preconditioner.) The resolution M in each direction is 50 or 100, so the total number of gridpoints or trigonometric functions is as high as 10^6. The time required for action by L or L^{-1} is again proportional to the number of gridpoints or modes, while action by W includes a Fourier transform and so takes a time proportional to $M^3 \log M$. Stable and unstable steady states were previously obtained [5] by adapting a time-stepping code to carry out Newton's method, using Laplacian preconditioning and BiCGSTAB to solve the resulting linear systems.

In order to correctly formulate the linear stability problem, it is necessary to decompose the eigenvector $\psi=\psi^R+i\psi^I$. We have

$$W_\Psi \psi = W_\Psi^R \psi^R + i W_\Psi^I \psi^I, \tag{4.8}$$

where

$$W_\Psi^R \equiv \mu - \frac{1}{2}|\boldsymbol{\omega}\cdot\mathbf{x}|^2 - 3a\Psi^2, \tag{4.9a}$$

$$W_\Psi^I \equiv \mu - \frac{1}{2}|\boldsymbol{\omega}\cdot\mathbf{x}|^2 - a\Psi^2. \tag{4.9b}$$

The equation governing the eigenmodes of (4.1) is

$$\lambda \begin{pmatrix} \psi^R \\ \psi^I \end{pmatrix} = \begin{bmatrix} 0 & -(L+W_\Psi^I) \\ L+W_\Psi^R & 0 \end{bmatrix} \begin{pmatrix} \psi^R \\ \psi^I \end{pmatrix}, \tag{4.10}$$

but it is more convenient to work with the square of the matrix in (4.10):

$$\lambda^2 \begin{pmatrix} \psi^R \\ \psi^I \end{pmatrix} = \begin{bmatrix} -(L+W_\Psi^I)(L+W_\Psi^R) & 0 \\ 0 & -(L+W_\Psi^R)(L+W_\Psi^I) \end{bmatrix} \begin{pmatrix} \psi^R \\ \psi^I \end{pmatrix} \tag{4.11}$$

and to calculate λ^2. Because (4.11) is block diagonal, it can be separated into the two problems:

$$\lambda^2 \psi^R = -(L+W_\Psi^I)(L+W_\Psi^R)\psi^R, \tag{4.12a}$$

$$\lambda^2 \psi^I = -(L+W_\Psi^R)(L+W_\Psi^I)\psi^I. \tag{4.12b}$$

Problems (4.12a) and (4.12b) are closely related, since if ψ^R is an eigenvector of (4.12a) with eigenvalue λ, then $(L+W_\Psi^R)\psi^R$ is an eigenvector of (4.12b) with the same eigenvalue.

Thus, we solve only (4.12a). As the critical eigenvalue pair $\pm\lambda$ makes the transition from imaginary to real, λ^2 passes from negative to positive.

To generate the Krylov space, we shift the operator in (4.12a) and precondition with the square of the inverse Laplacian:

$$L^{-2}[-(L+W^I_\Psi)(L+W^R_\Psi)-sI]\psi_{k+1}=L^{-2}\psi_k. \tag{4.13}$$

For this problem, we find that it is better to construct H as an approximation to A via (2.6) rather than as an approximation to A^{-1} via (1.7). The inverse Arnoldi method usually converges in 3 to 10 iterations, each of which requires several hundred BiCGSTAB iterations (matrix-vector multiplications) in order to solve its associated linear system. We adjust s empirically, both to target the critical eigenvalue and also to improve BiCGSTAB convergence. For some cases, we find that these goals are incompatible: with the shift required for the desired eigenvalue, BiCGSTAB becomes unable to converge as the Arnoldi iteration progresses, even when allowed a very large number of matrix-vector multiplies. This is the reason that the branch of eigenvalues calculated for the cigar case terminates prematurely in Fig. 5.

5 Towards a complex shift

The method described in Section 2 can calculate complex conjugate pairs of eigenvalues, since these may appear in (1.8) when the matrix H is diagonalized. However, if complex eigenvalues have a substantial imaginary part, then their inverses are not close to the origin, and the inverse power method will locate other eigenvalues instead. Complex eigenvalues with zero real part are important, regardless of the size of their imaginary part, since they are associated with Hopf bifurcations.

Complex eigenvalues cannot be shifted to the origin with a real shift as in (1.9); instead, an imaginary or complex shift must be used. The goal of the techniques described here is to use an existing timestepping code and its data structures to calculate leading eigenvalues, including those with large imaginary parts, so it is desirable to avoid complex arithmetic. We now describe a technique for carrying out what is effectively an imaginary shift without resorting to complex arithmetic. If A has a pair of imaginary eigenvalues $\pm i\lambda_I$ and eigenvectors $u_R\pm iu_I$, then

$$\left.\begin{array}{r} Au_R=-\lambda_I u_I \\ Au_I=\lambda_I u_R \end{array}\right\} \leftrightarrow\ A^2u_R=-\lambda_I^2 u_R. \tag{5.1}$$

(We recall that the real and imaginary parts of a complex eigenvector have no intrinsic significance, since these parts will be transformed by multiplying $u_R\pm iu_I$ by any complex number.) The application of a negative real shift $-s_I^2$ to A^2 leads to

$$(A^2+s_I^2)u_R=(-\lambda_I^2+s_I^2)u_R. \tag{5.2}$$

If s_I is chosen to be near λ_I, then u_R is an eigenvector of $(A^2+s_I^2)$ with eigenvalue near zero, and should be obtained by acting repeatedly with $(A^2+s_I^2)^{-1}$. This inverse action can be carried out by solving the pair of equations

$$A\tilde{u}+s_I^2u^{k+1}=u^k, \tag{5.3a}$$

$$-\tilde{u}+Au^{k+1}=0, \tag{5.3b}$$

for the doubled solution vector $(u^{k+1},\tilde{u})$, since then

$$(A^2+s_I^2)u^{k+1}=A\tilde{u}+s_I^2u^{k+1}=u^k. \tag{5.4}$$

This calculation can be generalized to eigenvalues and shifts which contain real parts as well as imaginary ones. The possibility of solving system (5.3) by methods similar to those in Section 2 is presently under investigation.

6 Conclusion

Methods based on inverse iteration are recognized as the fastest way to extract eigenvectors close to zero or any desired value. In problems arising from stability analyses of partial differential equations, the corresponding matrices are too large to be inverted. Instead, action by the inverse must be carried out by matrix-free iterative solution of the linear system, such as BICGSTAB, whose rate of convergence is governed by the condition number of the matrix.

Time-stepping codes necessarily approximate the exponential for a small timestep. We have shown that the action of an implicit-explicit time-stepping code can be modified to carry out a preconditioned version of the Jacobian if the timestep is taken to be very large. Based on this, we have proposed a method for converting a time-stepping code to rapidly compute small eigenvalues of large systems via the inverse power method. We have tested this method for spherical Couette flow and for the nonlinear Schrödinger equation. By varying the timestep, the preconditioner can be tuned between the identity and the inverse Laplacian, which are the least and most effective preconditioners, respectively, as demonstrated in Fig. 2. The crucial point, demonstrated by Fig. 4, is that the number of matrix-vector multiplications required by BICGSTAB is independent of the spatial resolution, which demonstrates the optimality of the preconditioning.

References

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

[2] L.S. TUCKERMAN, *Steady-state solving via Stokes preconditioning; recursion relations for elliptic operators*, in Lecture Notes in Physics (No. 323): Proc. of the 11th Int'l. Conf. on Numerical Methods in Fluid Dynamics, ed. by D.L. Dwoyer, M.Y. Hussaini & R.G. Voigt, (Springer, New York, 1989), p. 573–577.

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

[4] A. Bergeon, D. Henry, H. BenHadid and L.S. Tuckerman, *Marangoni convection in binary mixtures with Soret effect*, J. Fluid Mech. **375**, 143 (1998).

[5] C. Huepe, S. Métens, G. Dewel, P. Borckmans and M.E. Brachet, *Decay rates in attractive Bose-Einstein condensates*, Phys. Rev. Lett. **82**, 1616 (1999).

[6] O. Batiste, E. Knobloch, A. Alonso and I. Mercader, J. Fluid Mech. **560**, 149 (2006).

[7] 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 and L.S. Tuckerman (Springer, New York, 2000), p. 452–466.

[8] 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. 75–76.

[9] L.S. Tuckerman, C. Huepe and M.-E. Brachet, *Numerical methods for bifurcation problems*, in Instabilities and non-equilibrium structures IX, ed. by O. Descalzi, J. Martinez and S. Rica (Kluwer, Dordecht, 2004).

[10] C. Huepe, L.S. Tuckerman, S. Métens, and M.E. Brachet, *Stability and decay rates of non-isotropic attractive Bose-Einstein condensates*, Phys. Rev. A. **68**, 023609 (2003).

[11] H.A. van der Vorst, *Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput. **13**, 631 (1992).

[12] D. Gottlieb and S. A. Orszag, *Numerical Analysis of Spectral Methods* (SIAM, Philadelphia, 1977).

[13] J.P. Boyd, *Chebyshev and Fourier Spectral Methods* (Dover, New York, 2001).

[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] L.S. Tuckerman & P.S. Marcus, *Formation of Taylor vortices in spherical Couette flow*, in Lecture Notes in Physics (No. 218): Proc. of the 9th Int'l. Conf. on Numerical Methods in Fluid Dynamics, ed. by Soubbarameyer & J.P. Boujot, (Springer, New York, 1985), p. 552–556.

[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] E.P. Gross, Nuovo Cimento **20** 454 (1961).

[21] L.P. Pitaevskii, *Vortex lines in an imperfect Bose gas*, Zh. Eksp. Teor. Fiz. **40**, 646 (1961) [Sov. Phys. JETP **13**, 451 (1961)].

[22] S. Bose, *Plancks Gesetz und Lichtquantenhypothese*, Z. Phys. **26**, 178 (1924).

[23] A. Einstein, *Quantentheorie des einatomigen idealen gases: Zweite Abhandlung*, Sitzungber. Preuss. Akad. Wiss. **1925**, 3 (1925).

[24] M.H. Anderson, J.R. Ensher, M.R. Matthews, C.E. Wieman and E.A. Cornell, *Observation of Bose-Einstein condensation in a dilute atomic vapor*, Science **269**, 198 (1995).

[25] K.B. Davis, M.-O. Mewes, M.R. Andrews, N.J. van Druten, D.S. Durfee, D.M. Kurn and W. Ketterle, *Bose-Einstein condensation in a gas of sodium atoms*, Phys. Rev. Lett. **75**, 3969 (1995).

[26] C.C. Bradley, C.A. Sackett, J.J. Tollett and R.G. Hulet, *Evidence of Bose-Einstein condensation in an atomic gas with attractive interactions*, Phys. Rev. Lett. **75**, 1687 (1995).

[27] P.A. Ruprecht, M.J. Holland, K. Burnett, and M. Edwards, *Time-dependent solution of the nonlinear Schrödinger equation for Bose-condensed trapped neutral atoms*, Phys. Rev. A **51**, 4704 (1995).

[28] M. Ueda and A.J. Leggett, *Macroscopic quantum tunneling of a Bose-Einstein condensate with attractive interaction*, Phys. Rev. Lett. **80**, 1576 (1998).

[29] J.L. Roberts, N.R. Claussen, S.L. Cornish, E.A. Donley, E.A. Cornell and C.E. Wieman, *Controlled collapse of a Bose-Einstein condensate*, Phys. Rev. Lett. **86**, 4211 (2001).

[30] A. Gammal, T. Frederico and L. Tomio, *Critical number of atoms for attractive Bose-Einstein condensates with cylindrically symmetrical traps*, Phys. Rev. A **64** 055602 (2001).