JOURNAL OF COMPUTATIONAL PHYSICS 121, 246–260 (1995)

A Direct Method for Computation of Simple Bifurcations

MAXIM POLIASHENKO AND CYRUS K. AIDUN

Institute of Paper Science and Technology and The George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

Received June 4, 1993; revised March 28, 1995

In the present study, we focus on the computational analysis of partial differential equations with emphasis on the stability of the equilibrium states and on their bifurcations. In practical applications, it is not sufficient to obtain an equilibrium solution at a point in the parameter space. The equilibrium solution branches, their stability characteristics, and particularly the critical points of transition from one state to another (e.g., bifurcation points), are required for understanding the physics of the problem. In principle, the linear stability of an equilibrium state can be investigated by solving an eigenvalue problem, and consequently, the points of bifurcations can be detected. We review alternative techniques for detecting bifurcation points which are direct and numerically efficient and, therefore, more practical. Starting with a large dimension dynamical system, which represents a projection of a set of coupled partial differential equations onto a basis function, we discuss the relative effectiveness of the time evolution approach, the test function approach, and the direct method. We will then extend the direct method for a more practical and efficient implementation. With this technique, we compute the sequence of transitions from steady state to chaotic flow in a two-dimensional lid-driven cavity of aspect ratio 0.8, 1.0, and 1.5. We demonstrate the effectiveness of this technique by computing interesting new dynamics in this relatively simple hydrodynamic system. In particular, we show that depending on the aspect ratio, the first transition from steady state could be through a supercritical or a subcritical Hopf bifurcation leading the system to a time periodic state. We construct the destabilizing disturbance structure and conclude that the first bifurcation of the primary steady state is due to the centrifugal instability of the primary eddy. The mechanism of transition to chaos is low-dimensional. The transition to chaos occurs after a secondary Hopf bifurcation.

1. INTRODUCTION

Questions of stability of nonlinear dynamical systems are extremely important, both from the theoretical and practical points of view. For the systems with small numbers of degrees of freedom, a significant progress has been made in development of computational methods that allow effective detection and analysis of typical bifurcations that affect stability [1]. However, for systems with large or infinite numbers of degrees of freedom, for instance, those that arise in hydrodynamics, most of the conventional methods, such as Routh–Hurwitz criteria or eigen-analysis, are not readily applicable because of their excessive computational demands. Alternative approaches for detection and pinpointing of even simple bifurcations need to be developed. In this paper, we review some of these alternative techniques for analysis of large dynamical systems. We then outline a practical technique that can be effectively used for direct computation of bifurcations of steady state solutions of hydrodynamic flow equations.

As one of the parameters that control the dynamics of a nonlinear system varies, the stability of its steady state may typically change via bifurcations of codimension one, such as saddle-node, Hopf, or pitchfork bifurcations. Let us consider a general system of ordinary differential equations that may also describe the dynamics of spatially discretized partial differential equations.

$$M\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, \mu), \quad \mathbf{F} : D \subset \mathbf{X} \times \Re \to \mathbf{X}, \tag{1}$$

where $\mathbf{x} \in \mathbf{X}$ and μ is a system parameter. M is usually called a mass matrix and F is a system matrix. Then steady state solutions $\mathbf{x}_0$ of system (1) are determined by the equation

$$\mathbf{F}(\mathbf{x}_0, \mu) = 0 \tag{2}$$

and their stability is governed by a linear differential equation for disturbances $\mathbf{x}_1 = \mathbf{x} - \mathbf{x}_0$ about a steady state $\mathbf{x}_0$

$$M\frac{d\mathbf{x}_1}{dt} = F_x(\mathbf{x}_0, \mu)\,\mathbf{x}_1 \equiv J(\mathbf{x}_0, \mu)\,\mathbf{x}_1. \tag{3}$$

Thus, the local stability of an equilibrium state is fully determined by the following generalized eigenvalue problem involving the mass matrix, M, and the Jacobian matrix, $J(\mathbf{x}_0, \mu)$, at that state,

$$M\lambda_k\boldsymbol{\xi}_k = J(\mathbf{x}_0, \mu)\boldsymbol{\xi}_k, \tag{4}$$

where $\lambda_k \in \mathbb{C}$ is an eigenvalue and $\boldsymbol{\xi}_k \in \mathbb{C}^N$ is the corresponding eigenvector. So, if all the eigenvalues have negative real parts, a small disturbance,

0021-9991/95 $12.00

$$\mathbf{x}_1(t) = \sum_{k=1}^{N} e^{\lambda_k t} a_k \boldsymbol{\xi}_k, \tag{5}$$

about $\mathbf{x}_0$ will decay. If one of the eigenvalues λ_c has a positive real part, then all the disturbances along the eigenvectors $\boldsymbol{\xi}_k$ with $k \neq c$ will decay, whereas they will grow along the critical eigenvector $\boldsymbol{\xi}_c$ approaching it asymptotically. Thus, critical values of μ, where $\mathrm{Re}(\lambda_c) = 0$, mark thresholds of bifurcations in system (1). It should also be noted that since the eigenvalue problem (4) involves real matrices M and J, the eigenvalues of that problem are real or pairs of complex-conjugate. We will call a bifurcation simple if only one real or one complex-conjugate pair has zero real part at the point of bifurcation.

A saddle-node, or turning (limit) point bifurcation occurs when two branches of steady states meet with the continuous tangential direction in the absence of algebraic constraints. At least one of these branches represents a hyperbolic state and is unstable, since one of the real eigenvalues of Eq. (4) also changes sign from one branch to another. So, a turning point can be characterized by the following set of conditions:

$$\begin{aligned} &\lambda_c = 0 \\ &\boldsymbol{\xi}_c \neq \mathbf{0} \\ &\frac{d\lambda_c}{d\mu}(\mu_c) \to \infty \quad \text{or} \quad \frac{d\mu}{d\lambda_c}(\lambda_c = 0) = 0. \end{aligned} \tag{6}$$

Approaching the turning point while tracing the stable branch of equilibria will result in the abrupt loss of stability of the steady state. Consequently, there will be transition to another steady state, periodic, quasiperiodic, or chaotic motion.

A Hopf bifurcation occurs when a pair of the complex-conjugate eigenvalues λ_c, λ_c^* of Eq. (4) crosses the imaginary axis in the complex plane. The steady state then loses its stability and gives rise to a limit cycle with frequency equal to the imaginary part of the critical eigenvalues. The amplitude of the limit cycle increases from zero as square root of the bifurcation parameter $(\mu - \mu_c)$. This cycle may be locally stable or unstable, depending on whether the bifurcation is super-critical or sub-critical. The structure of the time periodic solution just after the bifurcation point $\mu = \mu_c$ can be constructed according to the formula:

$$\mathbf{x}(t) = \mathbf{x}_0 + a\,\boldsymbol{\xi}_R \cos \omega t - a\,\boldsymbol{\xi}_I \sin \omega t; \tag{7}$$

here, ω is the imaginary part of the critical eigenvalue pair, $\boldsymbol{\xi}_R$ and $\boldsymbol{\xi}_I$ are real and imaginary parts of the complex eigenvector $\boldsymbol{\xi}_c$ corresponding to the critical eigenvalue, $a(\mu)$ is the radius of the limit cycle with characteristic square-root dependence on the bifurcation parameter $(\mu - \mu_c)$, i.e.,

$$a(\mu) = \sqrt{(d/b)\,(\mu - \mu_c)}. \tag{8}$$

Constant $b \neq 0$ can be found empirically and determines criticality and stability of the bifurcating limit cycle. If $b > 0$, it is locally stable; otherwise, it is unstable and the Hopf bifurcation is sub-critical.

For a normal Hopf bifurcation, an additional condition for the real part of the critical eigenvalues is given by

$$\frac{d}{d\mu}(\mathrm{Re}\{\lambda(\mu_c)\}) = d \neq 0. \tag{9}$$

A pitchfork bifurcation can generically occur only if certain inherent symmetries of the system become broken [1]; this gives rise to the two asymmetric solution branches that can be sub-critical or supercritical and stable or unstable, respectively. At a bifurcation point, a symmetric solution branch loses stability as a real eigenvalue crosses to a positive region in the spectrum of its linearized matrix. So, this bifurcation can be characterized by the following set of conditions:

$$\begin{aligned} &\lambda_c = 0 \\ &\boldsymbol{\xi}_c \neq \mathbf{0} \\ &\frac{d\lambda_c}{d\mu}(\mu = \mu_c) \neq 0 \\ &\mathbf{F}_\mu(\mathbf{x}, \mu_c) \notin \text{range}\,\{F_x(\mathbf{x}, \mu_c)\} \end{aligned} \tag{10}$$

Considering these properties of simple codimension-one bifurcations, there are three strategies that are most commonly used to detect them in dynamical systems of large dimensions:

- Time evolution approach, which treats the system as an initial value problem. With this method, the time evolution of the system is observed and certain conclusions are made from the change in the system behavior [2].
- Test function approach, which analyzes stability of the equilibrium state by evaluating a certain test function. It is normally used in conjunction with continuation of steady state solution branches as some control parameter is varied and is applied at each parameter step. This allows one to detect change in the stability of the equilibrium branch [3, 4].
- Direct approach, where the bifurcation problem is reduced to an appropriate algebraic system with the bifurcation point being an isolated solution of this system [3, 5, 6].

There are a number of studies advocating particular methods, based on the above strategies. However, in our opinion, the three strategies are complementary and can be used most effectively depending on the specific problems and objectives. Below, we outline the advantages and drawbacks of each approach as applied to the Hopf, saddle-node, and pitchfork bifurcation problems. For each technique, we will review how one can detect simple bifurcations with particular attention to the Hopf

bifurcation, since it is more complicated and computationally demanding to locate.

2. TIME EVOLUTION APPROACH

Since a steady state continues to exist after a Hopf bifurcation and just loses its stability, the transition can be detected in principle by integrating the system of equations in time, using steady state as the initial condition for different parameter values. However, a small disturbance should be applied to the steady state, since if the initial conditions are chosen precisely enough, the system may very slowly evolve away from the unstable steady state. This could result in tremendously long integration time needed to reach a stable attractor. On the other hand, the disturbance should not be too large, because in the case of multi-stability, it may result in an attraction to a coexisting equilibrium or to another stable attractor, even if the steady state being tested is still stable.

This approach allows relatively reliable detection of the Hopf bifurcations and is quite simple, however, it has serious drawbacks. First, it does not work close to the bifurcation point, where the time of attraction, both to the steady state and to the emerging limit cycle is very long and is virtually infinite at the point of the bifurcation. Therefore, the Hopf bifurcation point cannot be pinpointed directly by the time integration technique. Second, this method cannot locate a sub-critical Hopf bifurcation. Third, numerical time integration replaces continuous ODE/PDE with a discrete map whose time evolution is computed in reality and with the time step becoming an artificial bifurcation parameter of the map. It turns out that as the time step increases, solutions and their stability properties of the discrete map, may become quite different from the ones of the underlying continuous system. This makes the *time evolution approach* a particularly vulnerable tool. For a more comprehensive discussion of this subject, the reader is addressed to [38, 39].

Yet, with a sufficiently small time step, it is still possible to make a rough prediction about the point of super-critical Hopf bifurcation by using the fact that close to that point the amplitude grows quadratically from zero with the bifurcation parameter. By precisely measuring the stationary oscillation amplitudes at any representative point in the domain for two or more different parameter values close to the onset of oscillations, one can estimate the point of bifurcation by extrapolation.

One of the important advantages of the time-evolution technique is that it is probably the most efficient way to distinguish between sub- and supercritical bifurcation, if the Hopf bifurcation point is already known. Indeed, if the characteristic square-root dependence of the amplitude of the limit cycle at a representative point on the bifurcation parameter is not observed as we approach the Hopf point, and if the periodic oscillations are non-decaying above and below the threshold of the bifurcation, this indicates that the bifurcation is sub-critical. At a certain parameter value below the sub-critical Hopf bifurcation point, a saddle-node bifurcation of the periodic orbits usually occurs with stable and unstable limit cycles merging and disappearing. This results in an abrupt loss of finite amplitude oscillations below the critical parameter value. If one performs time integration further with changing the control parameter backwards, so as to approach the Hopf point, the system should remain in the stable steady state, displaying, thus, a hysteresis between that state and a time-periodic motion, which is characteristic of sub-critical Hopf bifurcation.

Applying time-integration continuation to the emerged limit cycle past the Hopf point, allows further studies of its nonlinear evolution, including possible period-doubling or secondary Hopf bifurcations including a transition to chaos.

A pitchfork bifurcation can also be detected in a similar fashion by doing time integration of Eq. (1) in conjunction with a continuation of its equilibrium branch, starting from a slightly disturbed steady state. Past the bifurcation point, the solution will be attracted to one of the asymmetric steady states. This can be detected either by deviation from the initial equilibrium state or by the asymmetric structure of the final state.

As the threshold of the bifurcation is approached, the time-evolution of the system will show increasing relaxation times. This, and the fact that right near the bifurcation point, asymmetric steady states are very close to the symmetric states, will cause uncertainties in the determination of the transition point.

Time integration of the system (1) in case of saddle-node bifurcation is of limited use since, beyond the turning point, the equilibrium branch no longer exists, which can be easily detected by continuation of that branch.

3. TEST FUNCTION APPROACH

Among the test functions used to detect occurrence of simple codimension-one bifurcations, the most comprehensive is the test for eigenvalues. However, for more specific problems, determinant-based criteria or spectrum-slicing are applicable and are more efficient [4].

Unlike the time-evolution technique, the efficiency of the linear eigenvalue analysis is not dependent on the closeness to the bifurcation point. This method provides all the necessary information, both about the critical parameter value and about the structure of the critical disturbance and its natural frequency. The point of the bifurcation can be easily pinpointed using assumption (9), along with a secant iteration,

$$\mu_{n+1} = \frac{\mathrm{Re}(\lambda(\mu_{n-1}))\cdot\mu_n - \mathrm{Re}(\lambda(\mu_n))\cdot\mu_{n-1}}{\mathrm{Re}(\lambda(\mu_{n-1})) - \mathrm{Re}(\lambda(\mu_n))}. \qquad (11)$$

A serious drawback of the eigen-analysis is that the most robust methods that can treat non-symmetric generalized eigenvalue problems are based on the QZ algorithm and require computation of all the eigenvalues. This becomes computationally demanding and impractical as the size of the system in-

creases. Computing the full spectrum of the eigenvalues is redundant since only those with smallest real parts are of interest. Some iterative techniques are used in order to avoid computing all the eigenvalues, concentrating only on the leading ones. Among those, Arnoldi's method [7–9], inverse iteration [10, 11], and orthogonal iteration [4] methods are noted.

Generally, much more efficient than the full spectrum approach, the inverse iteration technique is based on an iterative formula

$$(J(\mathbf{x}_0, \mu) - sM)\,\Delta\mathbf{x}^k = (\lambda^k M - J(\mathbf{x}_0, \mu))\,\mathbf{x}^k$$

$$\lambda^{k+1} = \frac{(x^k + \Delta x^k)_i}{\lambda^k\,[x^k]_i} \tag{12}$$

$$\mathbf{x}^{k+1} = \frac{\mathbf{x}^k + \Delta\mathbf{x}^k}{[x^k + \Delta x^k]_i},$$

where λ^k and $\mathbf{x}^k$ are the kth approximations of the eigenvalue and of the corresponding eigenvector, respectively, and $[x]_i$ denotes the component of a vector $\mathbf{x}$ that has the largest modulus [10]. This iterative formula provides convergence to an eigenvector and an eigenvalue that is closest to the complex shift s with the rate of convergence r determined from the quotient

$$r = \left|\frac{s - \lambda^*}{s - \lambda^{**}}\right|, \tag{13}$$

where λ^* and λ^{**} are the nearest and the next nearest eigenvalue to s.

As it follows from Eq. (13), this method converges very poorly when there is another eigenvalue close to the critical one. Since shift s is complex, the size of the matrix $(J - sM)$ to be factorized at each iteration step is twice as large as matrix J. Another serious shortcoming of the inverse iteration method is that it requires a prior knowledge about the structure of the eigenvalue spectrum. If this information is not available, then several values of s need to be considered to assure that the computed eigenvalue is the leading one.

Among other methods of solving large eigenvalue problem (4) is a group of subspace iteration methods, based on the orthogonal (Galerkin) projection on a subspace of a smaller dimension or the oblique projection on a Krylov subspace. These methods are extensions of the symmetric Lanczos algorithm to nonsymmetric matrices. An orthogonal projection technique onto a Krylov subspace, known as Arnoldi's method [7, 8], is one of the most popular methods among this class. This method favors the outer part of the spectrum, so it must be combined with shift-and-invert techniques to converge to an eigenvalue closest to the complex shift. Unfortunately, this eigenvalue is not necessarily the leading one. Arnoldi's method with Schur–Wieland deflation can be used to overcome this difficulty by calculating several eigenvalues at a time [7]. However, for this method to converge to the leading part of the eigenvalue spectrum, one must use rather sophisticated combinations of accelerations and preconditioning, as done in [9]. Another problem encountered by authors when using Arnoldi's method with deflation was insufficient accuracy of the computed eigenvalues in comparison with results provided by QZ algorithm when the size of the Krylov subspace was small in comparison with the size of the matrix.

A serious difficulty in applying eigenvalue-based methods to hydrodynamic problems is the singularity of mass matrix M. In incompressible confined flows, this singularity is due to the mass conservation equation which contains no time derivatives. Those algebraic equations contribute essentially infinite eigenvalues to the generalized eigenvalue problem. When applying most of the iterative algorithms, favoring the outer part of the spectrum, infinite eigenvalues may cause all sorts of problems with convergence, as well as generate spurious modes and eigenvalues that can be mistaken for leading modes [9]. By using the penalty function approach, we have eliminated this problem for confined flows. This approach also decreased the size of the problem to be solved, as will be shown in Section 6.

With penalty approach, a test for the sign of the determinant of the Jacobian matrix can be used as a test for saddle-node and pitchfork bifurcations. It is based on the fact that in the absence of singularity of the mass matrix M, the latter can be, in principle, inverted and then a generalized eigenvalue problem (4) can be reformulated as a regular eigenvalue problem for matrix $J * M^{-1}$. If one of the generalized eigenvalues is zero, then $\det(J * M^{-1}) = \det(J) * \det(M^{-1}) = 0$. Since M is nonsingular, then it implies $\det(J) = 0$. The test for the sign of $\det(J)$ is computationally inexpensive, since it requires only one LU-decomposition of J and often it comes free as a by-product of a Newton-based procedure to solve the base problem.

As another example of a test function approach, a spectrum-slicing technique can be mentioned, which may be efficient for a certain class of spectral methods which deal with a regular eigenvalue problem using an orthogonal set of basis functions [4]. By applying just one LU- decomposition to the symmetric part of the Jacobian matrix $\frac{1}{2}(J + J^{\mathrm{T}})$, one can determine the inertia of this matrix, i.e., number of eigenvalues with positive, negative, and zero real parts. This enables detection and pinpointing of Hopf and pitchfork bifurcations, but it does not provide the structure of the emerging solutions.

4. DIRECT METHODS

Both time evolution and test-function strategies are used in conjunction with the continuation of equilibrium branches techniques. This requires marching along such a branch as one parameter changes, applying test function or performing time integration, in order to determine the parameter value at which a bifurcation occurs. However, there is a group of direct methods that allows finding both the structure of leading disturbances as well as the critical parameter value as an isolated solution

of a suitably constructed algebraic system. The idea of this approach is to augment problem (4) for a specific eigenvalue with certain constraints in order to treat the critical parameter μ_c in Eq. (4) as a variable. Then, μ_c, along with the leading eigenvector, can be found as an isolated solution of the enlarged system. In principle, this can be done for any given eigenvalue, as follows. Let us rewrite Eq. (4) in the equivalent form:

$$(J(\mathbf{x}_0, \mu) - \lambda_k M)\xi_k = \mathbf{0}. \tag{14}$$

If the eigenvalue λ_k and the eigenvector ξ_k are complex, i.e.,

$$\begin{aligned} \lambda_k &= \alpha + i\omega \\ \xi_k &= \xi_R + i\xi_I \end{aligned} \tag{15}$$

then Eq. (14) can be further decomposed to a pair of real equations:

$$\begin{aligned} (J(\mathbf{x}_0, \mu) - \alpha M)\xi_R + \omega M\xi_I &= \mathbf{0} \\ (J(\mathbf{x}_0, \mu) - \alpha M)\xi_I - \omega M\xi_R &= \mathbf{0}, \end{aligned} \tag{16}$$

where $\mathbf{x}_0$ can be determined from Eq. (2). Since a well-posed algebraic system is sought, the freedom in scaling the eigenvectors should be removed by applying two normalization conditions for ξ_R and ξ_I, given by

$$\begin{aligned} l_1[\xi_R, \xi_I] &= 0 \\ l_2[\xi_R, \xi_I] &= 0, \end{aligned} \tag{17}$$

where l_1 and l_2 are additional conditions, such as linear functionals. So, for given α and ω, Eqs. (16) and (17) along with Eq. (2) have the following unknowns: ξ_R, ξ_I, $\mathbf{x}_0$, and μ, which add up to $3N + 1$ totally, and we have $3N + 2$ equations, so the problem is overdetermined. In this case, we can free α or ω as an unknown.

Even if we chose l_1 and l_2 as linear functionals, the full system is nonlinear due to the nonlinearity in Eq. (2). So, generally, the full system might not have a solution. Similarly, one can define an augmented system for a given real eigenvalue.

In the context of simple bifurcations, we are interested in the critical value of the parameter μ, where eigenvalue λ is zero (turning point and pitchfork bifurcation) or pure complex (Hopf bifurcation). According to this, the following two systems can be written:

$$\begin{aligned} \mathbf{F}(\mathbf{x}_0, \mu) &= \mathbf{0} \\ J(\mathbf{x}_0, \mu)\xi &= \mathbf{0} \\ l(\xi) &= 0 \end{aligned} \tag{18}$$

for turning point and pitchfork bifurcation, and

$$\begin{aligned} \mathbf{F}(\mathbf{x}_0, \mu) &= \mathbf{0} \\ J(\mathbf{x}_0, \mu)\xi_R + \omega M\xi_I &= \mathbf{0} \\ J(\mathbf{x}_0, \mu)\xi_I - \omega M\xi_R &= \mathbf{0} \\ l_1[\xi_R, \xi_I] = l_2[\xi_R, \xi_I] &= 0 \end{aligned} \tag{19}$$

for Hopf bifurcation.

Equation (18) was first proposed by Moore and Spence [5], who proved that a quadratic turning point, at which

$$\psi J(\mathbf{x}_0, \mu)\xi\xi \neq 0, \tag{20}$$

where ψ is a left nullvector of $J(\mathbf{x}_0, \mu)$, is a regular solution of that system. This system, with variations in normalization conditions, was later used in a number of studies [12–14]. Equation (18) was also used in [3] for continuation of steady-state branches through the turning point by using pseudo-arclength parametrization.

If condition (20) is not satisfied, then higher order singularities may arise [15]. In particular, under the conditions

$$\begin{aligned} \psi \mathbf{F}_\mu &= 0 \\ \psi(J_\mu \xi + J_x \xi\zeta) &\neq 0, \end{aligned} \tag{21}$$

where ζ is a null-vector of the extended system

$$J\zeta + F_\mu = 0. \tag{22}$$

Werner and Spence [16] proved that a pitchfork bifurcation point is an isolated solution of Eq. (18).

Equation (19) was first proposed by Jepson [6] and later studied by Griewank and Reddien [17]. Some variations of that system were used by Roose and Hlavaček [18], Jackson [19], and others. Methods based on equivalent criteria are reviewed and compared in [20].

Let us also note that conditions of types (14) and (16) are also suitable to detect other types of bifurcations, where an eigenvalue with a nonzero real part is sought. Examples of such situations are maps that can, for instance, be defined as

$$\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n, \mu), \quad \mathbf{f}: D \subset \mathbf{X} \times \mathcal{R} \to \mathbf{X}, \tag{23}$$

where x_n and x_{n+1} are subsequent iterations of this map and can represent, for instance, the velocity field of Eq. (1) through the period of its oscillations. The stability of a fixed point,

$$\mathbf{x}_0 = \mathbf{f}(\mathbf{x}_0, \mu), \tag{24}$$

of this map is determined by the eigenvalues of its Jacobian $j(\mathbf{x}_0, \mu) \equiv f_x(\mathbf{x}_0, \mu)$ which are called Floquet multipliers of the map (23). A fixed point $\mathbf{x}_0$ is stable if and only if all the Floquet multipliers are within the unit circle in the complex plane [1].

A saddle-node bifurcation of a pair of stable and unstable fixed points occurs when one of the multipliers becomes $+1$. This corresponds to a turning point of a periodic branch of the original flow Eq. (1). The analog of a pitchfork bifurcation, period-doubling bifurcation occurs, when one of the multipliers turns into -1. A (secondary) Hopf bifurcation occurs when

$$\tilde{\alpha}^2 + \tilde{\beta}^2 = 1, \tag{25}$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ are real and imaginary parts of a leading Floquet multiplier. Then condition (14), together with the proper normalization condition and with Eq. (24), is easily applicable to the saddle-node or period-doubling bifurcation

$$(j(\mathbf{x}_0, \mu) - \lambda I)\boldsymbol{\xi} = \mathbf{0}, \tag{26}$$

where λ is either $+1$ or -1, respectively. For the Hopf bifurcation, we have to consider both real and imaginary parts of the critical multiplier being nonzero, and either augment Eq. (16) with the constraint (25) or to replace $\tilde{\alpha} + i\tilde{\beta}$ by $\cos(\theta) + i \sin(\theta)$, as it has been implemented in the software package, AUTO, by Doedel [3].

Generally, the direct methods for computation of simple bifurcations require solution of an extended nonlinear algebraic system,

$$\mathbf{K}(\mathbf{x}_0, \boldsymbol{\xi}, \lambda, \mu_c) \equiv \mathbf{K}(\boldsymbol{\chi}) = \mathbf{0}, \tag{27}$$

which, in principle, can be done by using one of the Newton-type iterations

$$\begin{aligned} K_\chi(\boldsymbol{\chi}_n)\,\Delta\boldsymbol{\chi}_n &= -\mathbf{K}(\boldsymbol{\chi}_n) \\ \boldsymbol{\chi}_{n+1} &= \boldsymbol{\chi}_n + \Delta\boldsymbol{\chi}_n, \end{aligned} \tag{28}$$

where K_χ denotes the Jacobian of vector $\mathbf{K}$. The solution, $\boldsymbol{\chi}$, includes a value of the critical parameter, μ_c, basis state, $\mathbf{x}_0$, leading eigenvalue, $\lambda \in \mathbb{C}$, and corresponding eigenvector, $\boldsymbol{\xi} \in \mathbb{C}^n$. Such an approach can be superior to other methods, as it requires only a few matrix factorizations to complete the task.

Unlike time integration and test function methods, which are based on the continuation in one parameter, direct methods are suitable to be used for continuations of bifurcation curves in two or more parameters [3, 21, 22].

However, there are several things that limit the efficiency of the direct methods. First, as can be seen from Eqs. (18) and (19), the size of the problem grows at least twice for a saddle-node and pitchfork bifurcation, and at least three times for a Hopf bifurcation. For a large system of ODEs that arise, for example, in hydrodynamics after spatial discretization of the governing equations, this can be a serious limitation. The second issue is the requirement of a good initial guess so that the Newton-type methods would converge. Below, we formulate a technique within the class of direct methods which, in many cases, is more efficient and practical than others.

5. A NEW DIRECT TECHNIQUE

A Newton method, Eq. (28), applied to Eq. (19), will provide a quadratic convergence for a good initial guess, which implies solving the following linear system at each iteration step

$$\begin{aligned} J\,\Delta\mathbf{x} + \mathbf{F}_\mu\,\Delta\mu &= -\mathbf{F} \\ J_x\boldsymbol{\xi}_R\,\Delta\mathbf{x} + J\,\Delta\boldsymbol{\xi}_R + \omega M\,\Delta\boldsymbol{\xi}_I & \\ + J_\mu\boldsymbol{\xi}_R\,\Delta\mu + M\boldsymbol{\xi}_I\,\Delta\omega &= -J\boldsymbol{\xi}_R - \omega M\boldsymbol{\xi}_I \\ J_x\boldsymbol{\xi}_I\,\Delta\mathbf{x} + J\,\Delta\boldsymbol{\xi}_I - \omega M\,\Delta\boldsymbol{\xi}_R & \\ + J_\mu\boldsymbol{\xi}_I\,\Delta\mu - M\boldsymbol{\xi}_R\,\Delta\omega &= -J\boldsymbol{\xi}_I + \omega M\boldsymbol{\xi}_R \\ \frac{\partial l_1}{\partial \boldsymbol{\xi}_R}\Delta\boldsymbol{\xi}_R + \frac{\partial l_1}{\partial \boldsymbol{\xi}_I}\Delta\boldsymbol{\xi}_I &= -l_1[\boldsymbol{\xi}_R, \boldsymbol{\xi}_I] \\ \frac{\partial l_2}{\partial \boldsymbol{\xi}_R}\Delta\boldsymbol{\xi}_R + \frac{\partial l_2}{\partial \boldsymbol{\xi}_I}\Delta\boldsymbol{\xi}_I &= -l_2[\boldsymbol{\xi}_R, \boldsymbol{\xi}_I]. \end{aligned} \tag{29}$$

Here Eq. (29) is solved for corrections, $\Delta\mathbf{x}$, $\Delta\boldsymbol{\xi}_R$, $\Delta\boldsymbol{\xi}_I$, $\Delta\mu$, and $\Delta\omega$. Then the values of the unknowns at the next step are updated as follows: $\mathbf{x}' = \mathbf{x} + \Delta\mathbf{x}$, etc.

Here, we assumed mass matrix M is constant. This is justified for confined flows with parameter μ that does not affect the geometry of the mesh or basis projection functions. Otherwise, additional terms involving derivatives of M with respect to $\mathbf{x}$ and parameter μ must be included. The Jacobian $J \equiv F_x$ in many cases can be computed directly; as with most discretization techniques, expressions for J can be derived analytically. In some cases, $\mathbf{F}_\mu$ can also be evaluated directly. However, derivatives of J (and if necessary, of M) in most cases have to be evaluated numerically at each iteration step [23].

It has been recognized from the beginning that inflating the size of the original problem by more than three-fold can be disadvantageous as the problem's size increases. This is particularly true for hydrodynamic systems where many expansion terms or grid points are required for accuracy. First Griewank and Reddien [17], and later Roose and Hlavaček [18] and Jackson [19] proposed two techniques to replace solving linear system (29) of the size $3N + 2$ by a two-stage solution. The last one is based on the following idea.

From the first equation, $\Delta\mathbf{x}$ can be expressed in terms of $\Delta\mu$ as

$$\Delta\mathbf{x} = \mathbf{a} + \Delta\mu\,\mathbf{b} \tag{30}$$

with

$$\mathbf{a} = -J^{-1} * \mathbf{F}, \quad \mathbf{b} = -J^{-1} * \mathbf{F}_\mu. \tag{31}$$

This decouples the rest of the equations from the first, in a

sense that now the following problem can be solved autonomously,

$$\begin{aligned} &J\,\Delta\xi_R + \omega M\,\Delta\xi_I + (J_\mu\xi_R + J_x\xi_R\,\mathbf{b})\,\Delta\mu \\ &\qquad + M\xi_I\,\Delta\omega = -J\xi_R - \omega M\xi_I - J_x\xi_R\mathbf{a} \\ &J\,\Delta\xi_R - \omega M\,\Delta\xi_I + (J_\mu\xi_I + J_x\xi_I\,\mathbf{b})\,\Delta\mu - M\xi_I\,\Delta\omega \\ &\qquad = -J\xi_I + \omega M\xi_R - J_x\xi_I\mathbf{a} \\ &\frac{\partial l_1}{\partial\xi_R}\Delta\xi_R + \frac{\partial l_1}{\partial\xi_I}\Delta\xi_I = -l_1[\xi_R, \xi_I] \\ &\frac{\partial l_2}{\partial\xi_R}\Delta\xi_R + \frac{\partial l_2}{\partial\xi_I}\Delta\xi_I = -l_2[\xi_R, \xi_I], \end{aligned} \tag{32}$$

and $\Delta\mathbf{x}$ is then evaluated using Eq. (30). This would require additional factorization of matrix J and two back-substitutions at each step to solve Eq. (31) for $\mathbf{a}$ and $\mathbf{b}$, plus evaluating F_μ and J_x. In most cases these derivatives are not available analytically, so they have to be taken numerically, along each vector, separately. For instance,

$$J_x(\mathbf{x}, \mu)\,\mathbf{b}\xi = \lim_{\varepsilon\to 0}[J(\mathbf{x} + \varepsilon\mathbf{b}, \mu)\xi - J(\mathbf{x}, \mu)\xi]. \tag{33}$$

A similar procedure has been proposed by Griewank and Reddien [17]. Instead of eliminating $\Delta\mathbf{x}$ by Eq. (30), they replace $\Delta\mathbf{x}$ and $\Delta\mu$ with another variable, τ, according to the formula

$$\mathbf{z} = \begin{pmatrix}\mathbf{x}\\ \mu\end{pmatrix}, \quad \Delta\mathbf{z} = \mathbf{r} + \tau\mathbf{b}, \tag{34}$$

where $\mathbf{r}$ and $\mathbf{b}$ can be found from the following algebraic sytems:

$$\begin{pmatrix}J|\mathbf{F}_\mu\\ \mathbf{b}^T\end{pmatrix}\mathbf{r} = -\begin{pmatrix}\mathbf{F}\\ 0\end{pmatrix}, \quad (J|\mathbf{F}_\mu)\mathbf{b} = \mathbf{0}. \tag{35}$$

Although more complex, Griewank and Reddien's technique is applicable in the vicinity of a singular point (det $J = 0$), where omega is close to zero and system (31) becomes ill-conditioned. This is the case, for instance, in the vicinity of the cusp point on a Hopf bifurcation branch as it merges with a saddle-node fold. Slight generalization of the latter technique is described by Moore and Spence [5] for the case of a turning point and can be easily extrapolated for the Hopf bifurcation. However, it would require the same number of factorizations as the approach of Griewank and Reddien, with the increased number of back-substitutions and algebraic operations, and hence, more programming effort.

We propose a technique based on decoupling of Eq. (19) instead of the Newton–Raphson equations (29). Let us treat the steady state equation, $\mathbf{F}(\mathbf{x}_0, \mu) = 0$, as a constraint relationship that makes $\mathbf{x}_0$ a function of the parameter μ. Then the Jacobian in Eq. (19) can be considered some nonlinear function of the parameter only, and then, Newton–Raphson equations for such a system can be written as

$$\begin{aligned} &J(\mu)\,\Delta\xi_R + \omega M\,\Delta\xi_I + \tilde{J}_\mu\xi_R\,\Delta\mu + M\xi_I\,\Delta\omega = -J\xi_R - \omega M\xi_I \\ &J(\mu)\,\Delta\xi_I - \omega M\,\Delta\xi_R + \tilde{J}_\mu\xi_I\,\Delta\mu - M\xi_R\,\Delta\omega = -J\xi_I + \omega M\xi_R \\ &\frac{\partial l_1}{\partial\xi_R}\Delta\xi_R + \frac{\partial l_1}{\partial\xi_I}\Delta\xi_I = -l_1[\xi_R, \xi_I] \\ &\frac{\partial l_2}{\partial\xi_R}\Delta\xi_R + \frac{\partial l_2}{\partial\xi_I}\Delta\xi_I = -l_2[\xi_R, \xi_I]. \end{aligned} \tag{36}$$

It is necessary to note that the derivative $\tilde{J}_\mu$ in Eq. (36) is different from J_μ in Eq. (32), since in our case it is

$$\tilde{J}_\mu = \frac{\partial J(\mathbf{x}_0, \mu)}{\partial\mu} + \frac{\partial J(\mathbf{x}_0, \mu)}{\partial\mathbf{x}}\cdot\frac{d\mathbf{x}}{d\mu}. \tag{37}$$

If J_μ is used instead of $\tilde{J}_\mu$, then only linear convergence can be achieved. However, we never use formula (37) for computation of $\tilde{J}_\mu$. Instead, we compute it numerically as

$$\tilde{J}_\mu = \frac{1}{\varepsilon}[J(\mu + \varepsilon) - J(\mu)], \tag{38}$$

where ε is small, compared to μ.

So, in comparison with two-stage methods discussed above, at each step of the Newton–Raphson iteration (36), we avoid computing F_μ and derivatives (33) of the Jacobian J_x (and of the mass matrix M, if necessary) along the various directions and matrix–vector products involving those matrices, as well as solving Eq. (35). Instead, we have to provide the Jacobian J as a function of the parameter μ, which we obtain as a by-product of solving steady-state problem (2) for a given parameter value.

Our method has another advantage over the methods proposed in [18, 19]. In the vicinity of a singular point, where $\det(J) \to 0$ (this can happen, for instance, at the origin of a Hopf bifurcation curve [22, 24]), Eq. (31) and, thus, Eq. (30) would not apply and the method will diverge. With our method, we can still solve a steady-state problem and construct a Jacobian. The latter assures applicability of this technique to decouple a steady-state equation from the rest of the augmented system close to the origin of Hopf bifurcation. The same is also true in the case of pitchfork and saddle-node bifurcations.

Thus, at each step of the Newton–Raphson iteration, our method requires:

1. solution of the steady state problem and construction of

the Jacobian J and its derivative J_μ; this step consumes about 3% of the total computational time;

2. assembly of the Jacobian of the extended system and the vector of the residuals in Eq. (36), which requires approximately 12% of the computational time; and

3. solution of the linear system (36) for the corrections which we did using ESSL solver for sparse matrices, and this step took the remainder of the computation time.

One of the important issues for any of Newton–Raphson type methods is finding the initial approximation. Roose and Hlavaček [18] obtained starting values for basis state $\mathbf{x}_0$, μ, and ω by tracing the branch of the steady states and using eigenvalues of the Jacobian matrix as a test function. Once satisfactory approximations for those quantities are found, they propose to compute an initial guess for the eigenvector using inverse iteration procedure. A similar technique is also used for that solution and

$$l_1(0, 0) + l_2(0, 0) \neq 0.$$

In [17, 19], normalization conditions were used of the form

$$\langle \mathbf{c}, \xi_R \rangle = 0, \quad \langle \mathbf{c}, \xi_I \rangle = 1, \tag{41}$$

where $\mathbf{c}$ is some constant vector and $\langle \cdot, \cdot \rangle$ denotes scalar product. Roose and Hlavaček [18] used the following normalization:

$$\langle \mathbf{c}, \xi_R \rangle = 0, \quad \langle \xi_R, \xi_R \rangle = 1. \tag{42}$$

Griewank and Reddien [17] proved that also the turning point can be a solution of system (19) with $\omega = 0$ and $\xi_R = 0$ if normalization (41) is used whereas normalization (42) does

Here, ξ is some vector that belongs to span(ξ_R, ξ_I). If our technique, or other techniques discussed above are applied, Eq. (44) can be solved in two stages involving only matrices of the size N and $N + 2$. This would greatly accelerate calculations since the cost of factorization grows exponentially with the matrix size.

Because system (44) does not involve both eigenvectors corresponding to the degenerate eigenvalue ω^2, the structure of the oscillating flow is undefined. Therefore, this system can be used only if the information about the critical parameter and eigenfrequency is required, like in the case of two-parameter continuation of a bifurcation curve. To overcome this shortcoming, the approach of [17] can be recommended to further partition Eq. (36) with $M = I$ into two sub-systems with the cost of treating the squared Jacobian matrix with the increased condition number and performing additional algebraic operations. Unfortunately, both techniques are not applicable for $M \neq I$ since J is non-symmetric and does not commute with M.

It should also be noted that our technique of decoupling the steady-state equation would also apply to pitchfork and saddle-node bifurcations with the Jacobian J close to singular, since we do not invert J. However, it would require a technique other than the Newton method to solve for the basis state. An alternative approach can be used, first to pinpoint the critical parameter value by using the det(J) test in combination with some simple iterative method such as secant iterations. Once the critical parameter value is determined, then the structure of the null-vector ξ can be easily obtained by solving the linear system

$$\hat{J}(\mathbf{x}_0 \mu_c)\, \xi = \mathbf{e}, \tag{45}$$

where $\hat{J}$ is the Jacobian J with kth row replaced by the normalization condition for ξ, and $\mathbf{e}$ is parallel to the kth ort. Nullvector ξ can be used, for instance, for branch switching at the bifurcation point [21].

In the next section, we apply our technique to a two-dimensional lid-driven cavity and show that the first transition point in that system is a Hopf bifurcation.

6. STABILITY OF FLOW IN A TWO-DIMENSIONAL LID-DRIVEN CAVITY

As a test problem, we considered a viscous flow in a two-dimensional lid-driven cavity (LDC) of height h and width d shown in Fig. 1 with the boundary conditions for the velocity field $\mathbf{u}$ at the walls given in Cartesian coordinates (x, y) by

$$\mathbf{u}(0, y) = \mathbf{u}(d, y) = \mathbf{u}(x, 0) = (0, 0), \quad \mathbf{u}(x, h) = (V, 0). \tag{46}$$

This problem is important because of fundamental and practical reasons since LDC is a simple system that can be used to study stability properties of confined flows with closed streamlines. It also serves as a model for many manufacturing devices,

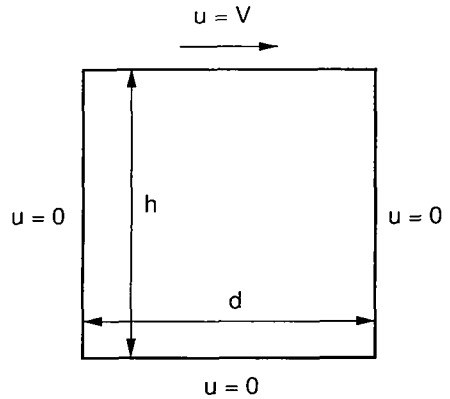

FIG. 1. Two-dimensional lid-driven cavity of aspect ratio $A = h/d$; direction of the lid velocity is indicated by an arrow.

such as short-dwell coaters and flexible blade coaters, used, for example, in the production of high-grade paper and photographic films [25, 26]. Flow instabilities in such systems may have adverse effect on the quality of the manufactured products. Therefore, understanding the stability properties of the flow in an LDC not only increases our knowledge of fundamental issues but also provides the basis for analysis and improvement of the performance of a broad class of engineering and manufacturing systems that share common features with the simple LDC.

The steady state flow of two-dimensional LDC was first studied in detail by Burggraf [27], who solved Navier–Stokes equation in a square cavity ($d = h$) for Reynolds numbers $R = Vd/\nu$ from 0 to 400 (here, V, d, and ν are the velocity of the lid, width of the cavity, and kinematic viscosity, respectively). Pan and Acrivos [28] studied the problem experimentally to examine the theorem of Prandl and Batchelor [29]. They found that for finite aspect ratios and at high Reynolds numbers (up to $R = 3000$), the LDC flow forms inviscid cores of uniform vorticity surrounded by thin shear layers at the boundaries, which is consistent with Prandl and Batchelor's theorem. However, they established that for the infinite aspect ratio, the flow does not become inviscid even if $R \to \infty$. Relevant experiments on stability of a three-dimensional LDC were performed by Koseff and Street [30] and, more recently, by Aidun *et al.* [25]. Their experiments concluded the presence of multistability of steady states as the Reynolds number increases as well as the recent studies by Benson and Aidun [31] reported the transition to a low-dimensional chaos in 3D LDC.

Accurate numerical solutions of the steady basis flow for the square geometry of LDC and for Reynolds numbers up to 10,000 have been obtained by several investigators, including Ghia *et al.* [32] (using a multigrid method), Schreiber and Keller [33] (utilizing a continuation technique to trace the branch of the steady states), and more recently, by Thomson and Ferziger (up to $R = 5000$) [34]. However, the stability of the two-dimensional steady state flow has remained unexplored. For example, the critical Reynolds number, where the primary

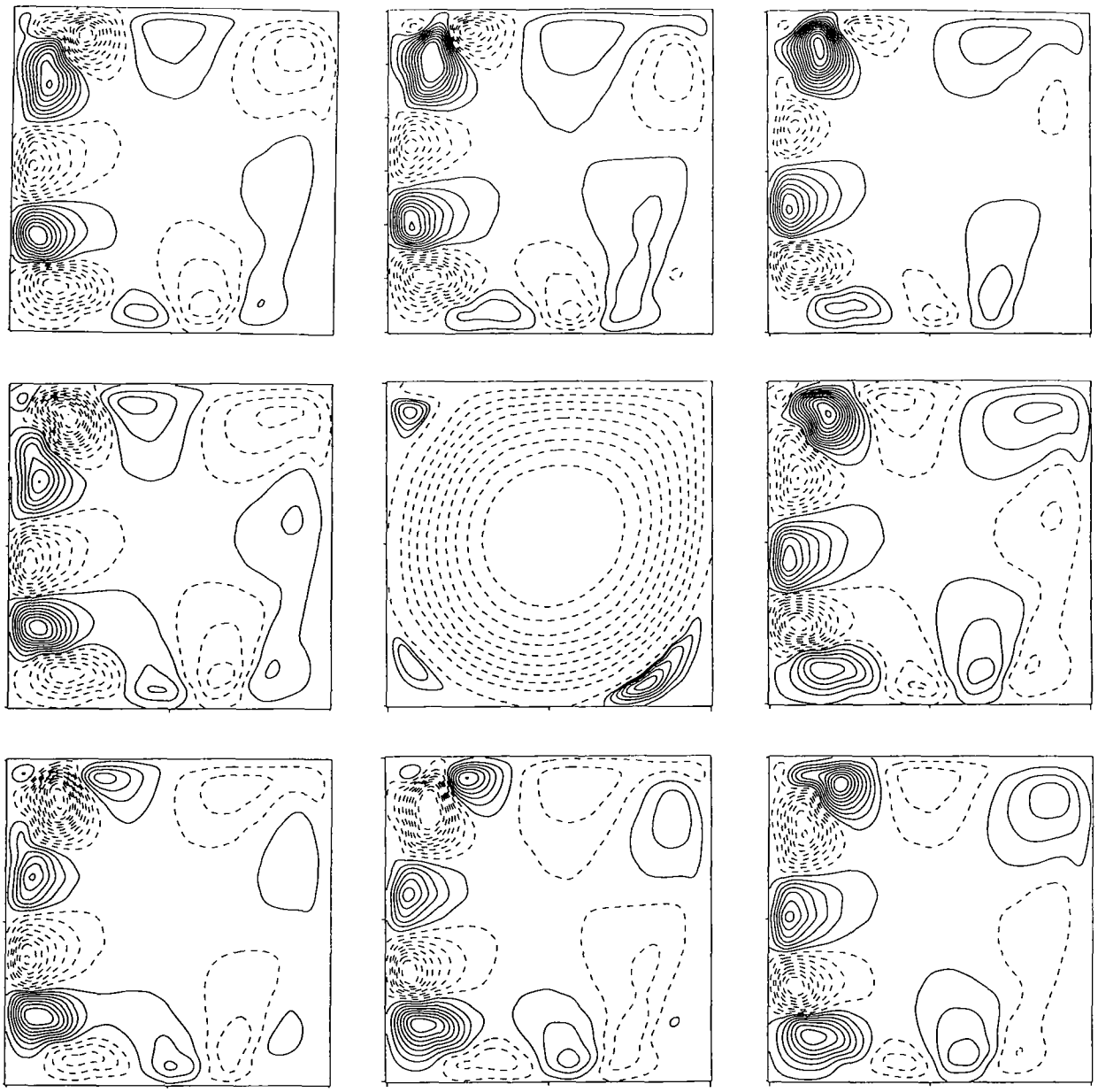

FIG. 2. The disturbance streamlines plotted through eight equal phase increments during the oscillation period with time progressing clockwise. The base-state streamlines are shown in the center.

steady state becomes unstable, has not been determined. Furthermore, it is not clear if the primary steady state destabilizes to a secondary steady state or if it is replaced by a time dependent flow. In this section, we examine the stability of the two-dimensional LDC flow to two-dimensional disturbances. We use our computational approach, discussed in the previous section, to show that the primary branch of steady states becomes unstable via the Hopf bifurcation as R increases, which can be either supercritical or subcritical, depending on the aspect ratio of the cavity. This happens, for instance, in the case of a square cavity at $R = 7763$ and gives rise to a time-periodic flow that may turn chaotic after a secondary Hopf bifurcation occurs.

The dynamics of the flow inside LDC can be described by Navier–Stokes and continuity equations. We consider isothermal incompressible Newtonian flow, so only momentum and continuity equations given, respectively, as

$$\nabla \cdot \mathbf{u} = 0 \tag{47}$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla p + R^{-1} \nabla^2 \mathbf{u} \tag{48}$$

are considered. Here, the velocity vector $\mathbf{u}$, pressure p, and time t are scaled with the lid velocity, V, pressure scale, $\rho \nu V/l$

Square LDC. Results of Newton Iterations to Detect Hopf Point

Mesh size	Number of iterations	Reynolds number R	Frequency ω	ΔR	$\|\Delta\xi_R\|$	$\|\Delta\xi_I\|$
	Initial	8250.0	2.9200			
37	1	8033.1	2.8430	−216.9	0.89	2.04
×	2	8211.8	2.9268	178.7	$3\ 10^{-1}$	$5\ 10^{-2}$
37	3	8195.7	2.9247	−16.1	$6\ 10^{-3}$	$6\ 10^{-3}$
	4	8195.5	2.9247	−0.2	$7\ 10^{-5}$	$7\ 10^{-5}$
	Initial	7650.0	2.9200			
47	1	8118.7	2.9077	468.7	1.36	1.24
×	2	7574.6	2.8930	−544.1	$1.8\ 10^{-1}$	$1.7\ 10^{-1}$
47	3	7614.3	2.8956	39.7	$2\ 10^{-2}$	$2\ 10^{-2}$
	4	7615.3	2.8957	1.0	$4\ 10^{-4}$	$5\ 10^{-4}$
57	Initial	7950.0	2.8700			
×	1	7763.7	2.8635	−186.3	$5\ 10^{-3}$	$6\ 10^{-3}$
57	2	7763.4	2.8634	−0.3	$6\ 10^{-4}$	$6\ 10^{-4}$

d, and time scale d^2/ν, respectively. The cavity width, d, is used as a length scale. So the Reynolds number, R, and the aspect ratio, $A \equiv h/d$, are the only parameters in Eqs. (47) and (48). The approximate solution of these equations for vanishing R at the corners of the cavity are available in closed form [28].

With a finite elements approach, the domain of the flow is divided into small elements, each defined by a fixed number of nodal points. Inside each element, the two components of the velocity field and pressure are approximated by

$$u_{1,2}(x, y, t) = \langle \boldsymbol{\phi}(x, y), \mathbf{U}_{1,2}(t) \rangle$$
$$p(x, y, t) = \langle \boldsymbol{\psi}(x, y), \mathbf{P}(t) \rangle, \tag{49}$$

where $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ are vectors of interpolating functions of the dimension equal to the number of nodes in the elements; $\mathbf{U}_{1,2}$ and $\mathbf{P}$ are vectors of nodal unknowns. After substituting Eq. (49) into the continuum equations (47), (48) and applying the Galerkin form of the method of weighted residuals, the following matrix equations are obtained:

$$G\mathbf{U} = \mathbf{0}$$
$$M\frac{d\mathbf{U}}{dt} + A(\mathbf{U})\,\mathbf{U} - G^{\mathrm{T}}\mathbf{P} = \mathbf{0}. \tag{50}$$

Here, $\mathbf{U} \equiv (\mathbf{U}_1^{\mathrm{T}}, \mathbf{U}_2^{\mathrm{T}})^{\mathrm{T}}$, M is the mass matrix, and $A(\mathbf{U})$ is the system matrix.

For the reasons discussed in the section Test-Function Approach, we prefer using penalty formulation of the Navier–Stokes equations. With this approach, the continuity equation (47) is substituted by an equation of a slightly compressible form

Results of Eigenvalue Problem Solution

Mesh size	Reynolds number R	$\mathrm{Re}(\lambda_c)$	$\mathrm{Im}(\lambda_c)$	Interpolated critical R	Interpolated critical ω
37 ×	8200	$1.9\ 10^{-4}$	2.9268		
37	8180	$-7.5\ 10^{-4}$	2.9218	8196	2.9258

$$p = -\frac{1}{\varepsilon}\nabla \cdot \mathbf{u}, \tag{51}$$

where ε is a penalty parameter. Then pressure in the momentum equation (48) can be substituted for $\nabla \cdot \mathbf{u}$, eliminating the continuity equation (47). This approach is in a way equivalent to the divergence-free basis functions used in spectral methods. The divergence equation and a dependent variable, namely the pressure, are removed from the problem. The validity of the penalty approach relies upon the result [35, 36]

$$\|\mathbf{u} - \mathbf{u}_\varepsilon\| + \|p - p_\varepsilon\| \le C\varepsilon, \tag{52}$$

where $\mathbf{u}_\varepsilon$ and p_ε are solutions of the penalty problem, and the constant C is dependent only on the dimension of the Galerkin subspace. Thus, the solutions of the penalty problem can be made arbitrarily close to the solution of the original system (47), (48), with a proper choice of the penalty parameter, ε.

In practice, the penalty parameter should be chosen sufficiently small to assure reasonable accuracy of the solution (Eq. (52), yet large enough, in comparison with machine accuracy, to avoid round-off errors which corrupt matrices making them ill-conditioned. So, usually the penalty parameter is chosen according to the formula [36]

$$\varepsilon = c/D,$$

TABLE III

Direct Method Iterations for Other Aspect Ratios

A	Mesh size	Number of iterations	Reynolds number R	Frequency ω	ΔR	$\|\Delta\xi_R\|$	$\|\Delta\xi_I\|$
		Initial	5425.0	2.2000			
	47	1	5213.0	2.2167	−212.0	$9\ 10^{-2}$	$9\ 10^{-2}$
0.8	×	2	5223.7	2.2170	10.7	$7\ 10^{-3}$	$9\ 10^{-3}$
	47	3	5225.0	2.2170	1.3	$3\ 10^{-3}$	$3\ 10^{-3}$
		Initial	6500.0	2.7600			
	65	1	6988.8	2.7780	488.8	1.00	1.00
1.5	×	2	7223.0	2.7730	234.2	$2\ 10^{-1}$	$2\ 10^{-1}$
	47	3	7216.8	2.7727	−6.2	$6\ 10^{-3}$	$7\ 10^{-3}$
		4	7216.9	2.7727	0.1	$5\ 10^{-5}$	$5\ 10^{-5}$

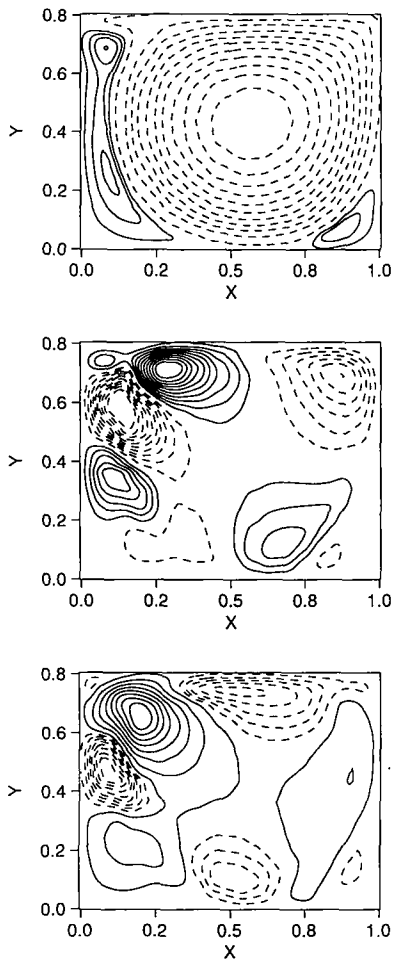

FIG. 3. (a) Base-state, (b) real, and (c) imaginary parts of the critical eigenvector for LDC with $A = 0.8$.

where c depends on the machine accuracy and can be assumed 10^{-6} for word lengths of 60–64 bits. D is a dominant contribution in the system stiffness matrix A and can be chosen as $\max\{\nu, \nu R\}$.

With finite element technique, a penalty formulation is usually applied to the discretized system (50) rather than to the original Eqs. (47) and (48) which requires assembly of an additional penalty matrix as discussed in [37]. So, the penalty formulation of the discrete problem can be, in general, presented by the following matrix equation:

$$M\frac{d\mathbf{U}}{dt} + B(\mathbf{U})\,\mathbf{U} = \mathbf{0}. \tag{53}$$

Thus, the finite element spacial discretization with penalty approach reduces the original problem to a system of ODEs of the order $2N$, where N is the total number of nodes not including the boundary nodes. We treat the resulting ODE system as a dynamical problem with R and A as parameters.

The mesh is graded according to the distribution of the velocity gradients in the cavity. All the calculations use nine node quadrilateral elements with biquadratic interpolation functions.

We used the continuation technique to trace the branch of equilibria originating from a Stokes solution at $R = 0$. By choosing a combined strategy for solving Eq. (2), applying

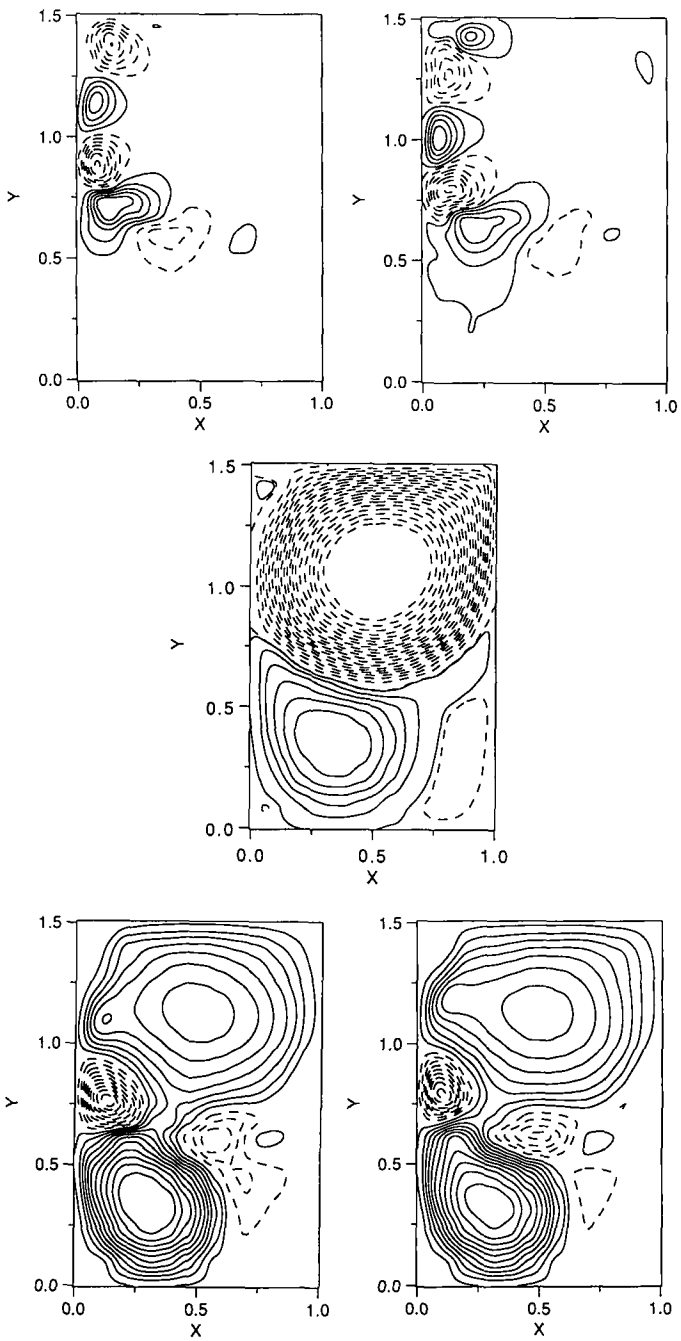

FIG. 4. (a) Base-state, (b) real, and (c) imaginary parts of the critical eigenvector at the point of sub-critical Hopf bifurcation ($R = 7217$) for LDC of $A = 1.5$; (d) and (e) show the structure of the time-periodic part of the oscillatory solution (with the steady state subtracted) existing below the Hopf bifurcation, shifted by $\frac{1}{4}$ of the period.

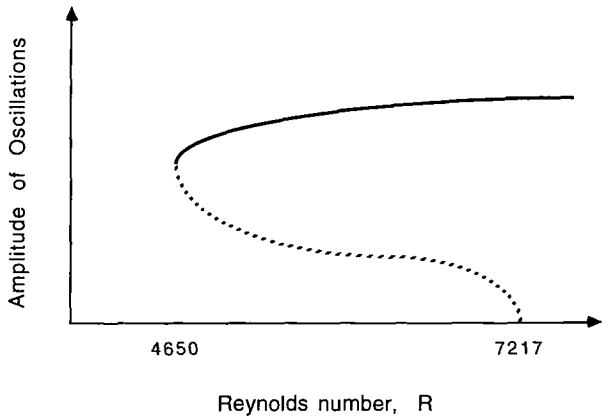

FIG. 5. Schematic bifurcation diagram for LDC of $A = 1.5$ reflecting subcritical character of Hopf bifurcation. The stable and unstable time-periodic branches are depicted by the solid and broken lines, respectively.

first the successive substitution method with a larger radius of convergence and, then, the Newton–Raphson method, we can make steps in parameter R values as large as 1000 and 500 for $R > 4000$. At each parameter step, we also do time-integration of Eq. (1), starting from the steady state solution. Once the onset of the oscillations is detected, we construct initial approximation for the direct method to pinpoint the Hopf bifurcation and the structure of the critical disturbance.

When applying our direct technique to Eq. (19), at each step of the Newton–Raphson iteration we also have to solve a steady state problem for an updated parameter value. This can be done efficiently by using Newton iterations again. We use a well-established standard finite element package (FIDAP, [36]) to construct the Jacobian and the mass matrix necessary to solve Eq. (36). It is worth noting that with finite-element discretization, both the Jacobian and the mass matrix have a banded structure that can be effectively utilized for reducing the costs of storage and computation; however, the matrix of system (36) does not have this advantage. To restore the banded structure of this matrix, the following renumbering of the equations in (36) should be performed: $1 \rightarrow 1$, $N + 1 \rightarrow 2$, $2 \rightarrow 3$, $N + 2 \rightarrow 4$, ..., $N \rightarrow 2N - 1$, $2N \rightarrow 2N$.

In the center of Fig. 2, the steady state for a square LDC is shown at the Hopf bifurcation point. We found that application of our method to system (19) results in almost quadratic convergence to the solution sought. Figure 2 also displays the structure of the critical disturbance for eight consecutive phases of the period of oscillation, obtained as a solution of Eq. (19) for a mesh of 57×57 nodal points chosen to represent the square domain of the LDC. The critical disturbance that gives rise to a time-periodic flow, consists of five primary pairs of counterrotating vortices, convected around the main core with maximum intensity at the left wall. This structure of the critical disturbance, computed directly, is identical to the corresponding structure obtained from the time-integration of the transient equations. This proves the correctness and validity of our technique. From the structure of the destabilizing disturbance, we conclude that the centrifugal instability of the primary vortex in the cavity is responsible for transition to time-periodic state.

Table I provides data about convergence of the method at each step for three different sizes of the mesh: 37×37, 47×47, and 57×57 nodal points. As one can see, our method shows almost quadratic convergence, where three to four Newton iterations are usually sufficient to satisfy the convergence criterion

$$\frac{\Delta_n R}{\Delta_1 R} + \frac{\Delta_n \omega}{\Delta_1 \omega} + \frac{\|\Delta \xi_R\|_n}{\|\Delta \xi_R\|_1} + \frac{\|\Delta \xi_I\|_n}{\|\Delta \xi_I\|_1} < \gamma, \tag{54}$$

where $\|\Delta\|_n$ denotes the norm of the Newton–Raphson correction at nth step and γ is normally chosen to be less than 10^{-2}.

As one may expect, the values of the parameters computed at the Hopf point differ for different sizes of the discretization mesh. The results, presented in Table I, allow predicting the exact point of Hopf bifurcation in a square cavity at Reynolds number $R = 7763 \pm 2\%$ with dimensionless frequency $\omega = 2.86 \pm 1\%$.

The dependence of the size of the limit cycle on the Reynolds number for each of the above grids shows good agreement with the square root law of the periodic branch amplitude near a Hopf point given by Eq. (8). The critical values of Reynolds number, R_c, extrapolated from the values of the amplitudes of the limit cycles at two points, close to the bifurcation point, by using formula (8), are $R_c = 8171$ for 37×37 mesh, $R_c = 7619$ for 47×47 mesh, and $R_c = 7740$ for 57×57 mesh which constitutes 0.3%, 0.05% and 0.3% errors with the correspondent results of the direct method, presented respectively in Table I. The latter comparison shows that the Hopf bifurcation in the square LDC is supercritical.

The accuracy of the method can also be proved by direct solution of the eigenvalue problem for the coarse grid. Table II presents results of eigenvalue calculations for 37×37 grid and the results of interpolation obtained by secant formula using real parts of the leading eigenvalue as a test function. As can be seen from Table II, the predicted bifurcation threshold is within 0.01% of the one obtained by the direct method. The structure of the leading eigenvector, computed by eigenvalue solver, is also in agreement with Fig. 2.

The solution of the eigenvalue problem also indicates that at the threshold of Hopf bifurcation, there are four other complex-conjugate pairs of eigenvalues with relatively small real parts and one small real eigenvalue. This may imply higher order bifurcations and interchange of leading modes as the second parameter of the system is varied.

We have also studied the variation of the Hopf bifurcation point as the aspect ratio $A = h/d$ changes. The convergence results for our method with geometries of the cavity of aspect ratio 0.8 and 1.5 are presented in Table III. The results suggest that the square geometry of the cavity is most stable with

24. D. Roose and B. de Dier, *SIAM J. Sci. Stat. Comput.* **10,** 671 (1989).
25. C. K. Aidun, N. G. Triantafillopoulos, and J. D. Benson, *Phys. Fluids A* **3,** 2081 (1991).
26. G. L. Booth and N. Millman, *TAPPI Ser.,* No. 28 (1965).
27. O. R. Burggraf, *J. Fluid Mech.* **24,** 113 (1966).
28. F. Pan and A. Acrivos, *J. Fluid Mech.* **28,** 643 (1967).
29. L. Prandtl, NACA Tech. Memo. No. 452, 1904; G. K. Batchelor, *J. Fluid Mech.* **1,** 177 (1956).
30. J. R. Koseff and R. L. Street, *J. Fluid Eng.* **106,** 21,385,390 (1984).
31. J. D. Benson and C. K. Aidun, *Phys. Fluids A* **4,** 2316 (1992).
32. U. Ghia, K. N. Ghia, and C. T. Shin, *J. Comput. Phys.* **48,** 387 (1982).
33. R. Schreiber and H. B. Keller, *J. Comput. Phys.* **49,** 310 (1983).
34. M. C. Thompson and J. H. Ferziger, *J. Comput. Phys.* **82,** 94 (1989).
35. M. Berkovier, These de Doctorat d'Etat, Rouen, 1976.
36. Fluids Dynamics International, Inc., *FIDAP* 7.0 *Theory Manual,* 1993.
37. M. S. Engelman, R. L. Sani, P. M. Gresho, and M. Bercovier, *Int. J. Numer. Methods Fluids* **2,** 25 (1984).
38. M. Poliashenko and C. K. Aidun, *Int. J. Bifurcation Chaos* **5,** 1 (1995).
39. H. C. Yee, P. K. Sweby, and D. F. Griffiths, *J. Comput. Phys.* **97,** 249 (1991).