*Japan J. Indust. Appl. Math., 26 (2009), 443–463* Area ⟨2⟩

# Numerical Verification Method of Solutions for Elliptic Equations and Its Application to the Rayleigh–Bénard Problem

Yoshitaka WATANABE* and Mitsuhiro T. NAKAO†

** Research Institute for Information Technology, Kyushu University*
*6–10–1 Higashi-Ku Hakozaki, Fukuoka 812–8581, Japan*
*Tel: +8192–642–2295*
*Fax: +8192–642–2294*
*E-mail: watanabe@cc.kyushu-u.ac.jp*

*† Faculty of Mathematics, Kyushu University*
*E-mail: mtnakao@math.kyushu-u.ac.jp*

Received March 26, 2008
Revised November 15, 2008

We first summarize the general concept of our verification method of solutions for elliptic equations. Next, as an application of our method, a survey and future works on the numerical verification method of solutions for heat convection problems known as Rayleigh–Bénard problem are described. We will give a method to verify the existence of bifurcating solutions of the two-dimensional problem and the bifurcation point itself. Finally, an extension to the three-dimensional case and future works will be described.

*Key words*: numerical verification method, elliptic equations, Rayleigh–Bénard problem, bifurcation point

## 1. Summary of the verification method

In this section, we will describe the basic concept of our verification method. The verification principle of this method was first originated in 1988 by one of the authors [15], and then several improvements have been done up to now. This method consists of the projection and the error estimates by the effective use of the compactness property of concerning operator, and it can be represented as rather generalized form in the below.

Let $X$, $\hat{X}$ and $Y$ denote the Hilbert spaces with compact imbedding $\hat{X} \hookrightarrow X$, and let a nonlinear operator $f\colon X \to Y$ be continuous, Fréchet differentiable and also maps a bounded set in $X$ into a bounded set in $Y$. Further, let $L$ be a linear elliptic differential operator from $\hat{X}$ into $Y$.

Then, we consider the following nonlinear problem:

$$Lu = f(u) \tag{1}$$

under some suitable boundary conditions.

Typical examples of the problem (1) are

$$\begin{cases} -\Delta u = f(x, u, \nabla u), & x \in \Omega, \\ u = 0, & x \in \partial\Omega, \end{cases} \tag{2}$$

or fourth order problem such as

$$\begin{cases} \Delta^2 u = f(x, u, \nabla u, \Delta u), & x \in \Omega, \\ u = \dfrac{\partial u}{\partial n} = 0, & x \in \partial\Omega, \end{cases} \tag{3}$$

where $\Omega$ is a bounded convex domain in $\mathbb{R}^n$ $(n = 1, 2, 3)$ with piecewise smooth boundary $\partial\Omega$.

Now, we also assume that the linear problem of eq. (1):

$$Lv = g \tag{4}$$

has a unique solution $v \in \hat{X}$ for each $g \in Y$. Therefore, when we denote the solution of (4) by $v \equiv Ag$, the map $A\colon Y \to X$ is compact, and the nonlinear map

$$Fu := Af(u)$$

is also a compact map on $X$, and we get the following fixed-point equation of the compact operator $F$ corresponding with eq. (1):

$$u = Fu. \tag{5}$$

In the case of the problem (2), we can choose, for example, $X = H_0^1(\Omega)$, $\hat{X} = H^2(\Omega) \cap H_0^1(\Omega)$ and $Y = L^2(\Omega)$. Thus, if we find a nonempty, bounded, convex and closed subset $U$ in $X$ satisfying

$$FU = \{Fu \mid u \in U\} \subset U,$$

then Schauder's fixed-point theorem asserts that there exists an element $u \in FU$ such that $u = Fu$. We call such a set $U$ expected to be $FU \subset U$ as a *candidate set*.

Next we will describe how to construct the candidate set in computer. In our verification method, the projection into a finite dimensional subspace and the constructive error estimates of the projection play an important and essential role. Let $X_h$ be an appropriate finite dimensional subspace of $X$ dependent on the parameter $h$. For example, $X_h$ is taken to be a finite element subspace with mesh size $h$. Now, let $P_h\colon X \to X_h$ denote an orthogonal projection and we suppose the following approximation property of $P_h$:

$$\|v - P_h v\|_X \le C(h)\|Lv\|_Y, \quad \forall v \in \hat{X}, \tag{6}$$

where $C(h) > 0$ is a positive constant which is *numerically determined* with the property that $C(h) \to 0$ as $h \to 0$. Note that the orthogonality of $P_h$ is not necessary, but the estimates (6) is indispensable in our argument. And also notice that the compactness of the imbedding $\hat{X} \hookrightarrow X$ is essential in getting the estimates (6). This assumption (6) holds for many finite element subspaces [17] or subspaces constituted by finite Fourier series [27] with adequate order for $h$. For example, in the case of the problem (2), when we define $P_h$ as the usual $H_0^1$-projection, it can be taken as $C(h) = h/\pi$ and $h/(2\pi)$ for bilinear and biquadratic element, respectively, for the rectangular mesh on the square domain [16]. And $C(h) = 0.493h$ for the linear and uniform triangular mesh of the convex polygonal domain [9].

Since $X_h$ is the closed subspace of $X$, each element of $X$ can be uniquely represented as the direct sum of the element of $X_h$ and $X_*$. Here $X_*$ stands for the orthogonal complement subspace of $X_h$ in $X$. Then, the fixed-point equation $u = Fu$ in $X$ can also be uniquely decomposed as the finite dimensional (projection) part and the infinite dimensional (error) part such that

$$\begin{cases} P_h u = P_h F u, \\ (I - P_h)u = (I - P_h)Fu. \end{cases} \tag{7}$$

Next in order to obtain a solution satisfying eqs. (7), we fix an approximate solution $u_h \in X_h$ of eq. (1) and introduce the Newton-like method. We define the nonlinear operator $N_h \colon X \to X_h$ by

$$N_h u := P_h u - [I - P_h F'[u_h]]_h^{-1} P_h (u - Fu),$$

where $F'[u_h]$ means the Fréchet derivative of $F$ at $u_h$ and $[I - P_h F'[u_h]]_h^{-1}$ denotes the inverse on $X_h$ of the restriction operator $P_h(I - P_h F'[u_h])|_{X_h}$. Note that the existence of the inverse $[I - P_h F'[u_h]]_h^{-1}$ is equivalent to the invertibility of a corresponding matrix which is able to be numerically checked in the actual verification process. According to the fact that $P_h u = P_h F u$ and $P_h u = N_h u$ are equivalent, defining the operator $T$ on $X$ by

$$Tu := N_h u + (I - P_h)Fu, \tag{8}$$

the fixed-point problem $u = Tu$ is also equivalent to $u = Fu$. Therefore since the operator $T$ on $X$ is compact, Schauder's fixed-point theorem asserts that if for a nonempty, bounded, convex and closed set $U \subset X$,

$$TU = \{Tu \mid u \in U\} \subset U,$$

holds, then there exists a fixed-point of $T$ in $U$. When the approximate solution $u_h \in X_h$ is sufficiently good, the finite dimensional part of (8), $N_h u$, will be possibly contractive near $u_h$ by the property of the Newton-like operator. On the other hand, the infinite dimensional part of $T$, i.e., $(I - P_h)Fu$, is expected to be a contraction

map when the parameter $h$ of $X_h$ are taken to be sufficiently small because of the approximation property (6).

We now consider the way to determine a candidate set $U$. If it is chosen such as

$$U := u_h + U_h + U_*, \quad U_h \subset X_h,\ U_* \subset X_*, \tag{9}$$

then the verification condition $TU \subset U$ can be written by

$$\begin{cases} N_h U - u_h \subset U_h, \\ (I - P_h)FU \subset U_*. \end{cases} \tag{10}$$

Depending on the problem, there are several choices on $U_h$ and $U_*$. Usually, the set $U_h$ is taken to be a set of linear combinations of base functions $\{\phi\}_{j=1}^N$ of $X_h$ with interval coefficients such that

$$U_h = \sum_{j=1}^{N} [\underline{A}_j, \overline{A}_j]\phi_j,$$

where $N := \dim X_h$ and $U_*$ a ball with radius $\alpha > 0$ of the form

$$U_* = \{v_* \in X_* \mid \|v_*\|_X \le \alpha\}.$$

Then, for a candidate set $U$, the former condition in (10) is validated by algebraic procedures, for example, solving a linear system of equation, interval Cholesky decomposition, and computation of the largest (or smallest) singular value of a matrix. Particularly, the interval enclosure of solutions for the linear system of equations with interval right-hand side plays an essential role in the verification for the former half of the condition (10). There are some useful algorithms with result verification to estimate rigorous bound for these problems [25]. Of course, prior to getting algebraic formulation, we need some additional computations on the function value evaluation or norm estimations and so on. On the other hand, $(I-P_h)FU$ cannot be estimated by any finite procedures but the latter part of (10) can be confirmed by comparing two nonnegative real numbers which correspond to the radii of balls such that

$$C(h) \sup_{u \in U} \|f(u)\|_Y \le \alpha$$

by using the property of the error estimates (6). We also need some interval-like approaches to evaluate the supremum in the left-hand side of the above.

In the actual computation, we use the simple iterative acceleration method which is called "$\varepsilon$-inflation" [26].

In addition to the condition $TU \subset U$, if there exists $k < 1$ such that

$$\|Tu_1 - Tu_2\| \le k\|u_1 - u_2\|, \quad \forall u_1, u_2 \in U \tag{11}$$

for a norm $\|\cdot\|$ of $X$, then the fixed-point is unique within the set $U$ by the Banach fixed-point theorem. Based on our verification principle and the mean value theorem, we can extend a procedure to derive sufficient condition for verifying not only the existence but also the local uniqueness of solutions to fixed-point equation $u = Fu$ [29, 28].

In the following sections, as an application of our method, we will give a survey on the numerical verification of solutions for heat convection problems known as Rayleigh–Bénard problem with some numerical examples.

## 2. Rayleigh–Bénard problem

In the end of nineteenth century, Bénard observed the establishment of a regular, steady pattern of flow cells in a thin horizontal layer of molten spermaceti with a free upper surface heated from below [1]. The principal facts are: a certain critical adverse temperature gradient must be exceeded before instability can set in; and the motions which ensure on surpassing the critical temperature gradient have a stationary cellular character [2]. His experiments are regarded as the starting point for the formation of contemporary knowledge on convection as the manifestation of an important class of hydrodynamic instabilities [4].

The theoretical foundation were laid by Load Rayleigh [24]. He considered a plane horizontal layer ($0 \le z \le h$) of an incompressible viscous fluid heated from below as shown in Fig. 1. At the lower boundary: $z = 0$ the layer of fluid is maintained at temperature $T + \delta T$ and the temperature of the upper boundary ($z = h$) is $T$.

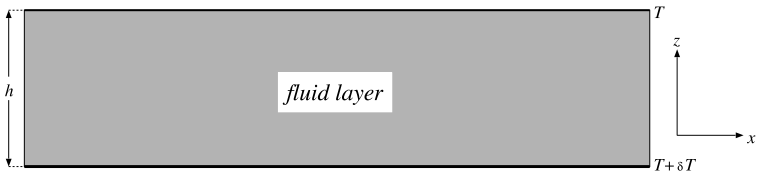


Fig. 1. Model of fluid layer.

According to the Oberbeck–Boussinesq approximations [2, 4], the equations governing convection in a layer are described as follows:

$$\begin{cases} \dfrac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = -\dfrac{1}{\rho_0}(\nabla p + g\rho \boldsymbol{e}_z) + \nu \Delta \boldsymbol{u}, \\ \operatorname{div} \boldsymbol{u} = 0, \\ \dfrac{\partial \theta}{\partial t} + (\boldsymbol{u}\cdot\nabla)\theta = \kappa \Delta \theta. \end{cases} \tag{12}$$

In the above system (12), $\boldsymbol{u} = (u, v, w)$ is the velocity vector field in the respective direction $(x, y, z)$; $p$ the pressure field; $\theta$ the temperature; $\rho$ the fluid density; $\rho_0$ the density at temperature $T + \delta T$; $\nu$ the kinematic viscosity; $g$ the gravitational acceleration; $\boldsymbol{e}_z = (0, 0, 1)$ and $\kappa$ is the coefficient of thermal diffusivity.

The Oberbeck–Boussinesq approximation also requires that the fluid density is to be independent of pressure and depends linearly on the temperature $\theta$, namely, $\rho$ can be represented by

$$\rho - \rho_0 = -\rho_0 \alpha(\theta - T - \delta T),$$

where $\alpha$ is the coefficient of thermal expansion. Therefore, the Oberbeck–Boussinesq equations (12) have a stationary solution

$$\boldsymbol{u}^* = \boldsymbol{0}, \quad \theta^* = T + \delta T - \frac{\delta T}{h} z, \quad p^* = p_0 - g\rho_0 \left( z + \frac{\alpha \delta T}{2h} z^2 \right) \tag{13}$$

representing the purely heat conducting state, where $p_0$ means a constant reference pressure.

Let us consider the perturbation from this conduction solution (13). By setting

$$\hat{\boldsymbol{u}} = (\hat{u}, \hat{v}, \hat{w}) := \boldsymbol{u}, \quad \hat{\theta} := \theta^* - \theta, \quad \hat{p} := p^* - p,$$

the perturbed equations

$$\begin{cases} \dfrac{\partial \hat{\boldsymbol{u}}}{\partial t} + (\hat{\boldsymbol{u}} \cdot \nabla)\hat{\boldsymbol{u}} = \dfrac{1}{\rho_0} \nabla \hat{p} - g\alpha \hat{\theta} \boldsymbol{e}_z + \nu \Delta \hat{\boldsymbol{u}}, \\ \operatorname{div} \hat{\boldsymbol{u}} = 0, \\ \dfrac{\partial \hat{\theta}}{\partial t} + \dfrac{\delta T \hat{w}}{h} + (\hat{\boldsymbol{u}} \cdot \nabla)\hat{\theta} = \kappa \Delta \hat{\theta} \end{cases} \tag{14}$$

are obtained. Moreover, transforming to dimensionless variables

$$t \to t/\kappa, \quad \boldsymbol{u} \to \hat{\boldsymbol{u}}/\kappa, \quad \theta \to \hat{\theta} h/\delta T, \quad p \to \hat{p}/(\rho_0 \kappa^2)$$

of eqs. (14), the dimensionless equations

$$\begin{cases} \dfrac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = \nabla p - \mathcal{P}\mathcal{R}\theta \boldsymbol{e}_z + \mathcal{P} \Delta \boldsymbol{u}, \\ \operatorname{div} \boldsymbol{u} = 0, \\ \dfrac{\partial \theta}{\partial t} + w + (\boldsymbol{u} \cdot \nabla)\theta = \Delta \theta \end{cases} \tag{15}$$

are led, where

$$\mathcal{R} := \frac{\delta T \alpha g}{\kappa \nu h}$$

is the Rayleigh number and

$$\mathcal{P} := \frac{\nu}{\kappa}$$

is the Prandtl number. Note that the Rayleigh number is sometimes defined by $\mathcal{R} = (\delta T \alpha g h^3)/(\kappa \nu)$ when the dimensionless equations are reduced to the domain of

$0 \leq z \leq 1$. In the perturbed Oberbeck–Boussinesq equations (15), $p$ and $\theta$ represent the perturbation from the equilibrium state.

Rayleigh considered the linearized stability of eqs. (15) for both the upper and lower boundaries are taken to be stress-free, and showed that for small $\mathcal{R}$ the fluid conducts heat diffusively, and instability must set in when $\mathcal{R}$ exceeds a certain critical value $\mathcal{R}_{\mathcal{C}}$; and that when $\mathcal{R}$ just exceeds $\mathcal{R}_{\mathcal{C}}$, a a stationary pattern of motions (*Rayleigh–Bénard convection*) must come to prevail [24].

Although a large number of studies have been made on the Rayleigh–Bénard convection [4, 12], theoretical results about the Rayleigh–Bénard convection are very few. It has been shown by Joseph [6] that eqs. (15) has a unique trivial solution for $\mathcal{R} < \mathcal{R}_{\mathcal{C}}$. Iudovich [5] and Rabinowitz [23] proved that, for each $\mathcal{R}$ slightly exceeding the critical Rayleigh number $\mathcal{R}_{\mathcal{C}}$, the equations (15) has at least two nontrivial steady-state solutions. The stability analysis of the bifurcated solution in a small neighbourhood of the bifurcation points is considered by Kagei and Wahl [7].

In this paper, we give an overview on the results of our computer assisted proofs for the Rayleigh–Bénard convection problem. First, in the following section, based on the basic principle mentioned in the Section 1, we will describe the formulation of the numerical verification method for the existence of the two-dimensional steady-state solutions and give some verified numerical examples. Next, in the Section 4, we will give a method to verify the bifurcation point itself, which should be an important information to clarify the global bifurcation structure, and show a numerical example. Further, in the Section 5, we will extend our verification method to the three-dimensional case for which we could no longer apply the formulation in two-dimensional cases. Finally, we will give some concluding remarks in the Section 6.

All numerical results discussed take into account of the effects of rounding errors in the floating point computations. Therefore, our results are considered as the mathematically rigorous proof.

## 3. Verification of two-dimensional non-trivial solutions

We now consider the two-dimensional ($x$-$z$) Rayleigh–Bénard heat convection model and describe on, based upon the numerical verification method in the Section 1, the fundamental formulation to prove the existence of the steady-state solutions. Then, since all variations with respect to $y$-direction are assumed to vanish, the stationary equation of the Oberbeck–Boussinesq approximations (15) are newly written as, for $\boldsymbol{u} = (u, w)$,

$$\begin{cases} -\mathcal{P}\Delta\boldsymbol{u} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = \nabla p - \mathcal{P}\mathcal{R}\theta\boldsymbol{e}_z, \\ \operatorname{div}\boldsymbol{u} = 0, \\ -\Delta\theta + (\boldsymbol{u}\cdot\nabla)\theta + w = 0. \end{cases} \tag{16}$$

We also assume that all fluid motion is confined to the rectangular region

$$\Omega := \{0 < x < 2\pi/a,\ 0 < z < \pi\},$$

where the given parameter $a > 0$ is called as a wave number which determines the aspect ratio of the rectangular domain. Let us impose periodic boundary condition (period $2\pi/a$) in the horizontal direction, stress-free boundary conditions ($u_z = w = 0$) for the velocity field and Dirichlet boundary conditions ($\theta = 0$) for the temperature field on the surfaces $z = 0, \pi$, respectively. Furthermore, we assume the following evenness and oddness conditions [3]:

$$u(x,z) = -u(-x,z), \quad w(x,z) = w(-x,z), \quad \theta(x,z) = \theta(-x,z).$$

We now use the stream function $\phi$ satisfying

$$\boldsymbol{u} = \mathbf{curl}\,\phi,$$

namely,

$$u = -\phi_z, \quad w = \phi_x$$

so that $u_x + w_z = 0$. Here, $*_\tau := \partial/\partial\tau$ ($\tau = x, z$). By some simple calculations with setting $\xi := \sqrt{\mathcal{PR}}\theta$, (16) is rewritten as

$$\begin{cases} \mathcal{P}\Delta^2\phi = \sqrt{\mathcal{PR}}\xi_x - \phi_z\Delta\phi_x + \phi_x\Delta\phi_z, \\ -\Delta\xi = -\sqrt{\mathcal{PR}}\phi_x + \phi_z\xi_x - \phi_x\xi_z. \end{cases} \tag{17}$$

In order to apply the argument in the Section 1, we formulate the problem concerned as a fixed-point equation of a compact map on the appropriate function space. From the boundary conditions, the functions $\phi$ and $\xi$ are supposed to be represented as the following double Fourier series:

$$\phi = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} A_{mn}\sin(amx)\sin(nz), \quad \xi = \sum_{m=0}^{\infty}\sum_{n=1}^{\infty} B_{mn}\cos(amx)\sin(nz). \tag{18}$$

We now define the following function spaces for integers $k \geq 0$:

$$X^k := \left\{ \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} A_{mn}\sin(amx)\sin(nz) \,\middle|\, A_{mn} \in \mathbb{R},\ \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}((am)^{2k}+n^{2k})A_{mn}^2 < \infty \right\},$$

$$Y^k := \left\{ \sum_{m=0}^{\infty}\sum_{n=1}^{\infty} B_{mn}\cos(amx)\sin(nz) \,\middle|\, B_{mn} \in \mathbb{R},\ \sum_{m=0}^{\infty}\sum_{n=1}^{\infty}((am)^{2k}+n^{2k})B_{mn}^2 < \infty \right\},$$

which are considered as closed subspaces of usual $k$-th order Sobolev space $H^k(\Omega)$.

In order to get the enclosure of the exact solutions for the problem (17), we need some appropriate finite dimensional subspaces. For $M_1, N_1, N_2 \geq 1$ and $M_2 \geq 0$, we set $N := (M_1, N_1, M_2, N_2)$ and define the finite dimensional approximate subspaces by

$$\begin{aligned}
S_N^{(1)} &:= \left\{ \phi_N = \sum_{m=1}^{M_1} \sum_{n=1}^{N_1} \hat{A}_{mn} \sin(amx) \sin(nz) \;\middle|\; \hat{A}_{mn} \in \mathbb{R} \right\}, \\
S_N^{(2)} &:= \left\{ \xi_N = \sum_{m=0}^{M_2} \sum_{n=1}^{N_2} \hat{B}_{mn} \cos(amx) \sin(nz) \;\middle|\; \hat{B}_{mn} \in \mathbb{R} \right\}, \\
S_N &:= S_N^{(1)} \times S_N^{(2)}.
\end{aligned}$$

Let $(\hat{\phi}_N, \hat{\xi}_N) \in S_N$ be an appropriate approximate solution of eqs. (17), and set

$$\begin{aligned}
f_1(\phi, \xi) &:= \sqrt{\mathcal{PR}}\xi_x - \phi_z \Delta \phi_x + \phi_x \Delta \phi_z, \\
f_2(\phi, \xi) &:= -\sqrt{\mathcal{PR}}\phi_x + \phi_z \xi_x - \phi_x \xi_z,
\end{aligned}$$

where

$$\phi = \hat{\phi}_N + u^{(1)}, \quad \xi = \hat{\xi}_N + u^{(2)}.$$

Then the problem (17) is rewritten as a system of equations with respect to $(u^{(1)}, u^{(2)}) \in X^4 \times Y^2$ satisfying

$$\begin{cases}
\mathcal{P}\Delta^2 u^{(1)} = f_1(\hat{\phi}_N + u^{(1)}, \hat{\xi}_N + u^{(2)}) - \mathcal{P}\Delta^2 \hat{\phi}_N, \\
-\Delta u^{(2)} = f_2(\hat{\phi}_N + u^{(1)}, \hat{\xi}_N + u^{(2)}) + \Delta \hat{\xi}_N,
\end{cases} \tag{19}$$

which is so-called a residual form of the original equation. Now, under the following correspondences with the function spaces in the Section 1,

$$X^3 \times Y^1 \sim X, \quad X^4 \times Y^2 \sim \hat{X}, \quad X^0 \times Y^0 \sim Y$$

we define

$$\begin{aligned}
u &:= (u^{(1)}, u^{(2)}), \\
f(u) &:= (f_1(\hat{\phi}_N + u^{(1)}, \hat{\xi}_N + u^{(2)}) - \mathcal{P}\Delta^2 \hat{\phi}_N, f_2(\hat{\phi}_N + u^{(1)}, \hat{\xi}_N + u^{(2)}) + \Delta \hat{\xi}_N), \\
L &:= (\mathcal{P}\Delta^2, -\Delta).
\end{aligned}$$

Then the compact nonlinear operator $F$ on $X^3 \times Y^1$ is defined as in the Section 1. Thus it is readily seen that the general arguments for problem (1) can also be applied to the present case (17). That is, for example, the projection

$P_h \equiv \left(P_N^{(1)}, P_N^{(2)}\right)$ from $X^3 \times Y^1$ to $S_N$ is defined by

$$\begin{cases} \left(\Delta\left(P_N^{(1)}\phi - \phi\right), \Delta v_N^{(1)}\right)_{L^2} = 0 & \forall v_N^{(1)} \in S_N^{(1)}, \\ \left(\nabla\left(P_N^{(2)}\xi - \xi\right), \nabla v_N^{(2)}\right)_{L^2} = 0 & \forall v_N^{(2)} \in S_N^{(2)}, \end{cases} \tag{20}$$

and the constructive error estimates of the projections $P_N^{(1)}$ and $P_N^{(2)}$ are determined [27] by

$$\left\|\psi - P_N^{(1)}\psi\right\|_{H^3} \le C_1 \|\Delta^2 \psi\|_{L^2},$$
$$\left\|\theta - P_N^{(2)}\theta\right\|_{H^1} \le C_2 \|\Delta\theta\|_{L^2}$$

for any $(\psi, \theta) \in X^4 \times Y^2$, where

$$C_1 = \max\left\{ \left(\sum_{\nu=1}^{4} \frac{1}{(a^2 + (N_1+1)^2)^\nu}\right)^{1/2}, \left(\sum_{\nu=1}^{4} \frac{1}{(a^2(M_1+1)^2 + 1)^\nu}\right)^{1/2} \right\},$$
$$C_2 = \max\left\{ \frac{(1 + (N_2+1)^2)^{1/2}}{(N_2+1)^2}, \frac{(2 + a^2(M_2+1)^2)^{1/2}}{a^2(M_2+1)^2 + 1} \right\}.$$

Another kind of error estimates can also be obtained and more detailed verification algorithm and computational procedures to construct the candidate set in $X^3 \times Y^1$ are shown in references [27, 19].

We have verified various kinds of bifurcating solutions corresponding to different Rayleigh numbers as shown in Fig. 2, where we fixed the wave number as $a = 1/\sqrt{2}$ and the Prandtl number $\mathcal{P} = 10$. The vertical axis stands for the absolute value of the coefficient of the approximate solution for $\xi$. And each dot in Fig. 2 means that the existence of an exact solution corresponding to the point

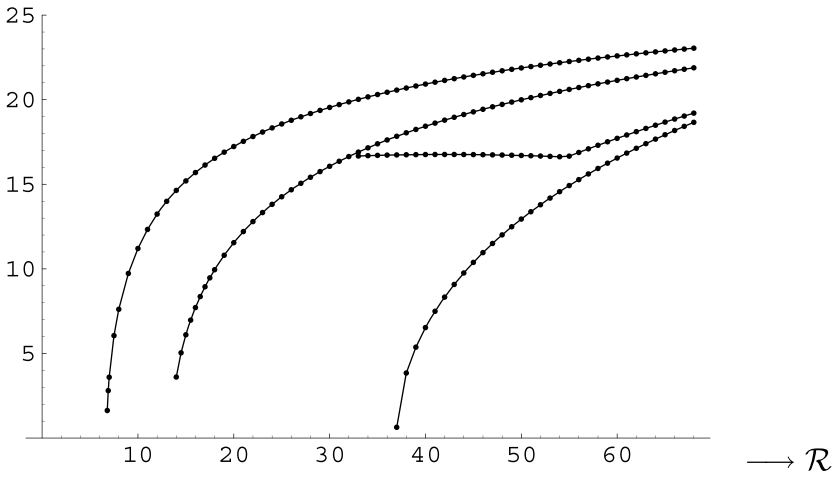


Fig. 2. Verified points on the bifurcation diagram.

was numerically verified. This result proves that, for example, at least four kinds of different bifurcating curves actually exist for certain long period, which could not be proved by any theoretical consideration up to the present. Therefore, it is a significant result in the sense of finding a new and mathematically rigorous result in the concerning research fields.

## 4. Existence proof of a bifurcation point

By considering Fig. 2, it seems that there exists a secondary bifurcation point around $\mathcal{R} \approx 32$. Thus we studied the detailed structure of the approximate solutions around those points. Then it was clarified that, for approximate solutions of the form

$$\phi_N = \sum_{m=1}^{M_1} \sum_{n=1}^{N_1} A_{mn} \sin(amx) \sin(nz), \quad \xi_N = \sum_{m=0}^{M_2} \sum_{n=1}^{N_2} B_{mn} \cos(amx) \sin(nz),$$

we have following two different kinds of approximate solutions satisfying

$$A_{mn} = B_{mn} = 0, \quad m = 1, 3, 5, 7, \dots \text{ with } \mathcal{R} = 32$$

and

$$A_{mn} \neq 0, \quad B_{mn} \neq 0, \quad m = 1, 3, 5, 7, \dots \text{ with } \mathcal{R} = 33.$$

Therefore, we expected that a secondary bifurcation should occur on the corresponding branch between $32 \leq \mathcal{R} \leq 33$, and that it must be a symmetry-breaking bifurcation point.

In the below, by using the similar arguments in [8], we consider the formulation of a numerical method to verify a bifurcation point itself and present an actually verified example. In order to obtain the enclosure of the bifurcation point, we define the "symmetric" operator $\mathcal{T} \colon X^0 \times Y^0 \to X^0 \times Y^0$ by

$$\begin{aligned} \mathcal{T}(\phi, \xi) &= (\mathcal{T}_1 \phi, \mathcal{T}_2 \xi) \\ &= (\phi(x + \pi/a, z), \xi(x + \pi/a, z)), \end{aligned}$$

then $X^k$ and $Y^k$ can be decomposed as the symmetric part and anti-symmetric part

$$X^k = X_s^k \oplus X_a^k, \quad Y^k = Y_s^k \oplus Y_a^k,$$

where

$$\begin{aligned} X_s^k &= \{\phi \in X^k \mid \mathcal{T}_1 \phi = \phi\}, \quad X_a^k = \{\phi \in X^k \mid \mathcal{T}_1 \phi = -\phi\}, \\ Y_s^k &= \{\xi \in Y^k \mid \mathcal{T}_2 \xi = \xi\}, \quad Y_a^k = \{\xi \in Y^k \mid \mathcal{T}_2 \xi = -\xi\}. \end{aligned}$$

Also, we set as $Z := X^3 \times Y^1$, and define the operator $G := I - F$, where $I$ is the identity map on $Z$ and $F$ is the compact nonlinear operator defined in the previous

section. Then, it is easily seen that $\mathcal{T}Gw = G\mathcal{T}w$ holds and $Z$ is also decomposed as two part

$$Z = Z_s \oplus Z_a,$$

where $Z_s = \{w \in Z \mid \mathcal{T}w = w\}$ and $Z_a = \{w \in Z \mid \mathcal{T}w = -w\}$.

Next, we define the extended operator $\mathcal{G}$ on $Z_s \times Z_a \times \mathbb{R}$ which includes $\mathcal{R}$ as a variable parameter, by

$$\mathcal{G}(w, v, \mathcal{R}) := \begin{pmatrix} G(w, \mathcal{R}) \\ D_w G[w, \mathcal{R}]v \\ \mathcal{H}(v) - 1 \end{pmatrix}. \tag{21}$$

Here $\mathcal{H}$ is a linear functional on $Z_a$ defined by for $v = (\psi, \eta)$

$$\mathcal{H}(v) := (\psi, \psi_0)_{L^2} + (\eta, \eta_0)_{L^2},$$

where we set $\psi_0 := 2a/\pi^2 \sin(ax)\sin(z)$ and $\eta_0 := 2a/\pi^2 \cos(ax)\sin(z)$. The basic principle of the verification for a symmetry-breaking bifurcation point is given by Kawanago [8] as follows:

Lemma 4.1. *$(w_0, \mathcal{R}_0) \in Z_s \times \mathbb{R}$ is a symmetry-breaking bifurcation point of $G(w, \mathcal{R}) = 0$ if*

1. *Extended system $\mathcal{G}(w, v, \mathcal{R}) = 0$ has an isolated solution $(w_0, v_0, \mathcal{R}_0) \in Z_s \times Z_a \times \mathbb{R}$.*
2. *$D_w G[w_0, \mathcal{R}_0]|_{Z_s} : Z_s \to X_s^0 \times Y_s^0$ is bijective.*

First we have to prove that the extended system $\mathcal{G}(w, v, \mathcal{R}) = 0$ has an isolated solution $(w_0, v_0, \mathcal{R}_0) \in Z_s \times Z_a \times \mathbb{R}$ by applying our verification method presented in the Section 1. We observe that a solution $[\phi, \xi, \psi, \eta, \mathcal{R}] \in Z_s \times Z_a \times \mathbb{R}$ of the equation $\mathcal{G}(w, v, \mathcal{R}) = 0$ is equivalently characterized as

$$\begin{cases} \mathcal{P}\Delta^2\phi - \sqrt{\mathcal{P}\mathcal{R}}\xi_x - J(\phi, \Delta\phi) = 0, \\ -\Delta\xi + \sqrt{\mathcal{P}\mathcal{R}}\phi_x + J(\phi, \xi) = 0, \\ \mathcal{P}\Delta^2\psi - \sqrt{\mathcal{P}\mathcal{R}}\eta_x - J(\phi, \Delta\psi) - J(\psi, \Delta\phi) = 0, \\ -\Delta\eta + \sqrt{\mathcal{P}\mathcal{R}}\psi_x + J(\phi, \eta) + J(\psi, \xi) = 0, \\ \mathcal{H}(v) - 1 = 0, \end{cases} \tag{22}$$

where $J(u, v) := u_x v_z - v_x u_z$. We compute an approximate solution $[\phi_N, \xi_N, \psi_N, \eta_N, R_N]$ of eq. (22) by using a usual Newton-like method in finite dimension. And, as in the previous section, setting the residual variables by

$$\phi := \phi_N + u^{(1)}, \quad \xi := \xi_N + u^{(2)}, \quad \psi := \psi_N + u^{(3)}, \quad \eta := \eta_N + u^{(4)}, \quad \mathcal{R} := \mathcal{R}_N + u^{(5)},$$

we rewrite the equation (22) as follows:

$$\begin{cases} \mathcal{P}\Delta^2 u^{(1)} = \sqrt{\mathcal{P}(R_N+u^{(5)})}(\xi_N+u^{(2)})_x + J(\phi_N+u^{(1)}, \Delta(\phi_N+u^{(1)})) - \mathcal{P}\Delta^2\phi_N, \\ -\Delta u^{(2)} = -\sqrt{\mathcal{P}(R_N+u^{(5)})}(\phi_N+u^{(1)})_x - J(\phi_N+u^{(1)}, \xi_N+u^{(2)}) + \Delta\xi_N, \\ \mathcal{P}\Delta^2 u^{(3)} = \sqrt{\mathcal{P}(R_N+u^{(5)})}(\eta_N+u^{(4)})_x + J(\phi_N+u^{(1)}, \Delta(\psi_N+u^{(3)})) \\ \qquad\qquad + J(\psi_N+u^{(3)}, \Delta(\phi_N+u^{(1)})) - \mathcal{P}\Delta^2\psi_N, \\ -\Delta u^{(4)} = -\sqrt{\mathcal{P}(R_N+u^{(5)})}(\psi_N+u^{(3)})_x - J(\phi_N+u^{(1)}, \eta_N+u^{(4)}) \\ \qquad\qquad - J(\psi_N+u^{(3)}, \xi_N+u^{(2)}) + \Delta\eta_N, \\ u^{(5)} = -(\psi_N+u^{(3)}, \psi_0)_{L^2} - (\eta_N+u^{(4)}, \eta_0)_{L^2} + 1 + u^{(5)}. \end{cases} \tag{23}$$

Note that the correspondences of function spaces between the present case and the Section 1 are

$$Z_s \times Z_a \times \mathbb{R} \sim X, \quad ((Z_s \times Z_a) \cap (X^4 \times Y^2)^2) \times \mathbb{R} \sim \hat{X}, \quad (X^0 \times Y^0)^2 \times \mathbb{R} \sim Y.$$

Therefore, when we define the nonlinear function $f(u)$ of $u{:=}(u^{(1)}, u^{(2)}, u^{(3)}, u^{(4)}, u^{(5)})$ by the right-hand side of the equation (23), and define the operator

$$L := (\mathcal{P}\Delta^2, -\Delta, \mathcal{P}\Delta^2, -\Delta, I),$$

the fixed-point formulation $u = Fu$ of the compact map on $Z_s \times Z_a \times \mathbb{R}$ can be obtained. We also easily define the finite dimensional subspaces satisfying

$$X_s^N \subset X_s^3, \quad Y_s^N \subset Y_s^1, \quad X_a^N \subset X_a^3, \quad Y_a^N \subset Y_a^1,$$

and the projections

$$P_N^{(1)} \colon X_s^3 \to X_s^N, \quad P_N^{(2)} \colon Y_s^1 \to Y_s^N, \quad P_N^{(3)} \colon X_a^3 \to X_a^N, \quad P_N^{(4)} \colon Y_a^1 \to Y_a^N,$$

as in the previous section. And we obtain the almost same constructive error estimates for these projections. Thus we can formulate the similar verification conditions as before.

We actually implemented numerical verification program incorporated with the some guaranteed computations by using interval arithmetic and proved that there exists an isolated solution of $\mathcal{G}(w_0, v_0, \mathcal{R}_0) = 0$. Then $\mathcal{R}_0$ is enclosed the following very small interval

$$\mathcal{R}_0 \in 32.04265510708193 + [-2.910, 2.910] \times 10^{-10}. \tag{24}$$

Next, it is necessary to prove the latter condition in Lemma 4.1 by numerically verifying the invertibility for linearized operator $D_w G[w_0, \mathcal{R}_0]$ on $Z_s$. Taking notice that the operator $D_w G[w_0, \mathcal{R}_0]$ is a Fredholm operator with index 0, it suffices to prove that

$$\begin{cases} \mathcal{P}\Delta^2\psi - \sqrt{\mathcal{P}\mathcal{R}_0}\eta_x - J(\phi_0, \Delta\psi) - J(\psi, \Delta\phi_0) = 0, \\ -\Delta\eta + \sqrt{\mathcal{P}\mathcal{R}_0}\psi_x + J(\phi_0, \eta) + J(\psi, \xi_0) = 0 \end{cases} \tag{25}$$

has a unique trivial solution $[\psi, \eta] = [0, 0]$ in $Z_s$, where $[\phi_0, \xi_0] := w_0$. Here, note that the system (25) has a set of equations because $w_0 = [\phi_0, \xi_0]$ and $\mathcal{R}_0$ are only given by sets of function and parameter as the solutions for the extended system (22). This difficulty, however, can be effectively overcome by appropriate use of interval estimates in the actual verification procedures.

We now for $u = [\psi, \eta]$ in $Z_s$, define

$$f(u) := \big(\sqrt{\mathcal{P}\mathcal{R}_0}\eta_x + J(\phi_0, \Delta\psi) + J(\psi, \Delta\phi_0), -\sqrt{\mathcal{P}\mathcal{R}_0}\psi_x - J(\phi_0, \eta) - J(\psi, \xi_0)\big),$$

and $L = (\mathcal{P}\Delta^2, -\Delta)$. Thus, the equation (25) can also be represented as the fixed-point form $u = Fu$ on $Z_s$, where, in this case, $F$ is a linear compact map on $Z_s$. The Newton-like operator $T$ on $Z_s$ is also well defined and we take a candidate set $U \subset Z_s$ including $[0, 0]$. If $TU \subset U^\circ$ is satisfied, then, by the linearity of the operator $T$, it is assured that $[0, 0] \in Z_s$ is the only solution of problem (25) (cf. [13]). Here, $U^\circ$ means the interior of $U$.

Based upon this principle, under the following computing environment which is same as above enclosing of $\mathcal{R}_0$, we actually succeeded to verify the invertibility of $D_w G[w_0, \mathcal{R}_0]$: Sun ONE Studio 7 Compiler Collection Fortran 95 on Fujitsu PRIMEPOWER 850 (CPU: SPARC64 V 1.35 GHz, OS: Solaris 8).

Therefore, it was proved that there exists a symmetry-breaking bifurcation point $\mathcal{R}_0$ satisfying (24).

## 5. Three-dimensional problems

In this section, we consider the verification of solutions for three-dimensional heat convection problems in which more realistic and interesting bifurcation phenomena are observed numerically in fluid mechanics [21]. Concerning Oberbeck–Boussinesq equations are the same as (16), but we could no longer apply the formulation in two-dimensional case by the use of stream functions. Our verification technique can also be extended to this case applying the verification method directly to the original three-dimensional Navier–Stokes equation (16). Modifying slightly in (16) as $p \to -p/\mathcal{P}$ and $\theta \to -\theta$, we consider the following steady-state problem:

$$\begin{cases} \dfrac{1}{\mathcal{P}}\boldsymbol{u} \cdot \nabla \boldsymbol{u} + \nabla p = \Delta \boldsymbol{u} + \mathcal{R}\theta\nabla z, \\ \operatorname{div} \boldsymbol{u} = 0, \\ (\boldsymbol{u} \cdot \nabla)\theta = \Delta\theta + w. \end{cases} \tag{26}$$

Here, $\boldsymbol{u} = (u, v, w)$. Given positive wave numbers $a, b \le 1$, we assume that all fluid motions are essentially confined to

$$\Omega = \left\{(x, y, z) \in \mathbb{R}^3 \;\middle|\; 0 \le x \le \frac{2\pi}{a},\ 0 \le y \le \frac{2\pi}{b},\ 0 \le z \le \pi\right\}, \quad |\Omega| = \frac{4\pi^3}{ab},$$

and impose parity conditions on new boundaries as in [21] which lead periodic boundary conditions in horizontal directions as in the Section 3. From these boundary conditions, the velocity field, the perturbations of temperature and pressure can be represented by the following Fourier series:

$$\boldsymbol{u} = \sum_{\alpha\neq\mathbf{0}} [u_\alpha \phi_1^\alpha, v_\alpha \phi_2^\alpha, w_\alpha \phi_3^\alpha], \quad \theta = \sum_{\alpha_3\neq 0} \theta_\alpha \phi_3^\alpha, \quad p = \sum_{\alpha\neq\mathbf{0}} p_\alpha \phi_4^\alpha, \tag{27}$$

where $\alpha$ is the multi-index of non-negative integers in $\mathbb{Z}_0^3$, and $(u_\alpha, v_\alpha, w_\alpha, \theta_\alpha, p_\alpha)$ are coefficients of $(\boldsymbol{u}, \theta, p)$ with respect to the base functions $\phi_i^\alpha$ defined by

$$\begin{aligned}
\phi_1^\alpha(x,y,z) &= K_\alpha \sin(a\alpha_1 x)\cos(b\alpha_2 y)\cos(\alpha_3 z),\\
\phi_2^\alpha(x,y,z) &= K_\alpha \cos(a\alpha_1 x)\sin(b\alpha_2 y)\cos(\alpha_3 z),\\
\phi_3^\alpha(x,y,z) &= K_\alpha \cos(a\alpha_1 x)\cos(b\alpha_2 y)\sin(\alpha_3 z),\\
\phi_4^\alpha(x,y,z) &= K_\alpha \cos(a\alpha_1 x)\cos(b\alpha_2 y)\cos(\alpha_3 z),
\end{aligned}$$

where $K_\alpha$ is the normalization factor with respect to the usual $L^2(\Omega)$ inner product $(\,\cdot\,,\,\cdot\,)_{L^2}$ defined by

$$K_\alpha := \sqrt{(2-\delta_{0\alpha_1})(2-\delta_{0\alpha_2})(2-\delta_{0\alpha_3})/|\Omega|}, \quad \delta_{ij} = \text{Kronecker delta on } i,\, j.$$

Setting $A_\alpha \equiv \sqrt{(a\alpha_1)^2 + (b\alpha_2)^2 + \alpha_3^2}$, $B_\alpha \equiv \sqrt{(a\alpha_1)^2 + (b\alpha_2)^2}$ and defining indices subsets by $I_1 \equiv [1,0,1] + \mathbb{Z}_0^3 \cup [0,1,1] + \mathbb{Z}_0^3$, $I_2 \equiv [1,1,0] + \mathbb{Z}_0^3$, we define the divergence free orthogonal base functions by

$$\begin{aligned}
\Phi^\alpha &= \left[-\frac{a\alpha_1\alpha_3}{A_\alpha B_\alpha}\phi_1^\alpha, -\frac{b\alpha_2\alpha_3}{A_\alpha B_\alpha}\phi_2^\alpha, \frac{B_\alpha}{A_\alpha}\phi_3^\alpha\right], \quad \alpha \in I_1,\\
\Psi^\alpha &= \left[\frac{b\alpha_2}{B_\alpha}\phi_1^\alpha, -\frac{a\alpha_1}{B_\alpha}\phi_2^\alpha, 0\right], \quad \alpha \in I_2.
\end{aligned}$$

Set $I_0 = I_1 \cup I_2$ and then define function spaces $V$ and $W$ with associated usual $H^1$ norm as follows:

$$\begin{aligned}
V &= \left\{ \boldsymbol{u} = \sum_{\alpha\in I_0} \{\xi_\alpha \Phi^\alpha + \eta_\alpha \Psi^\alpha\} \;\middle|\; \|\Delta \boldsymbol{u}\|_{L^2} < \infty \right\} \subset H^2(\Omega)^3,\\
W &= \left\{ \theta = \sum_{\alpha\in I_3} \theta_\alpha \phi_3^\alpha \;\middle|\; \|\Delta\theta\|_{L^2} < \infty \right\} \subset H^2(\Omega), \quad \text{where } I_3 \equiv [0,0,1] + \mathbb{Z}_0^3.
\end{aligned}$$

For a fixed number $N \geq 2$, we define the finite dimensional subspaces $V_N$ and $W_N$ of $V$ and $W$:

$$\begin{aligned}
V_N &\equiv \{\boldsymbol{u} \in V \mid \xi_\alpha = \eta_\alpha = 0, \text{ if } |\alpha| \equiv \alpha_1 + \alpha_2 + \alpha_3 > N\},\\
W_N &\equiv \{\theta \in W \mid \theta_\alpha = 0, \text{ if } |\alpha| > N\},
\end{aligned}$$

respectively. Set $X \equiv V \times W$ and $X_N \equiv V_N \times W_N$. Define projections $P_N : V \to V_N$ and $Q_N : W \to W_N$ as in previous sections:

$$\begin{aligned} &(\nabla(\boldsymbol{u} - P_N\boldsymbol{u}), \nabla\boldsymbol{v})_{L^2} = 0, \quad \forall \boldsymbol{v} \in V_N, \\ &(\nabla(\theta - Q_N\theta), \nabla\vartheta)_{L^2} = 0, \quad \forall \vartheta \in W_N, \end{aligned} \tag{28}$$

with respect to the inner product $(\,\cdot\,,\,\cdot\,)_{L^2}$ on $L^2(\Omega) \times L^2(\Omega)$ or properly extended spaces. Due to orthogonal relations of base functions in $X$, these projections $P_N$ and $Q_N$ coincide with truncation operators

$$P_N\boldsymbol{u} = \sum_{\alpha \in I_{0,N} \equiv I_0 \cap I_{,N}} \{\xi_\alpha \Phi^\alpha + \eta_\alpha \Psi^\alpha\}, \quad Q_N\theta = \sum_{\alpha \in I_{3,N} \equiv I_3 \cap I_{,N}} \theta_\alpha \phi_3^\alpha,$$

where $I_{,N} \equiv \{\alpha \in \mathbb{Z}_0^3 \mid |\alpha| \le N\}$.

Then, we have the following error estimates [10].

Theorem 5.1. *For any $(\boldsymbol{u}, \theta) \in X$ and $(P_N\boldsymbol{u}, Q_N\theta) \in X_N$ in* (28)*, the following holds:*

$$\|\boldsymbol{u} - P_N\boldsymbol{u}\|_{L^2} \le \frac{C_0^2}{(N+1)^2}\|\Delta\boldsymbol{u}\|_{L^2}, \quad \|\nabla(\boldsymbol{u} - P_N\boldsymbol{u})\|_{L^2} \le \frac{C_0}{N+1}\|\Delta\boldsymbol{u}\|_{L^2}, \tag{29}$$

$$\|\theta - Q_N\theta\|_{L^2} \le \frac{C_0^2}{(N+1)^2}\|\Delta\theta\|_{L^2}, \quad \|\nabla(\theta - Q_N\theta)\|_{L^2} \le \frac{C_0}{N+1}\|\Delta\theta\|_{L^2}, \tag{30}$$

*where $C_0 \equiv \sqrt{a^{-2} + b^{-2} + 1}$.*

We also obtain the following $L^\infty$ error estimates [10].

Corollary 5.1. *Under the same assumptions of* Theorem 5.1*, the following holds:*

$$\|\boldsymbol{u} - P_N\boldsymbol{u}\|_\infty \le 2C_1\sqrt{\frac{1}{N} - \frac{1}{3(N+1)^3}}\,\|\Delta\boldsymbol{u}\|_{L^2} < \frac{2C_1}{\sqrt{N}}\|\Delta\boldsymbol{u}\|_{L^2}, \tag{31}$$

$$\|\theta - Q_N\theta\|_\infty \le 2C_1\sqrt{\frac{1}{N} - \frac{1}{2(N+1)^2} + \frac{1}{6N^3}}\,\|\Delta\theta\|_{L^2} < \frac{2C_1}{\sqrt{N}}\|\Delta\theta\|_{L^2}, \tag{32}$$

*where $C_1 \equiv C_0^2|\Omega|^{-\frac{1}{2}}$.*

The steady state solution of (26) can be written as

$$\begin{cases} -\Delta\boldsymbol{u} + \nabla p = \boldsymbol{f}(\boldsymbol{u}, \theta), \\ \nabla \cdot \boldsymbol{u} = 0, \\ -\Delta\theta = g(\boldsymbol{u}, \theta), \end{cases} \tag{33}$$

where the right hand sides of (33) are defined by

$$\boldsymbol{f}(\boldsymbol{u}, \theta) = -\frac{1}{\mathcal{P}}(\boldsymbol{u} \cdot \nabla)\boldsymbol{u} + \mathcal{R}\theta\boldsymbol{e}_z, \quad g(\boldsymbol{u}, \theta) = -(\boldsymbol{u} \cdot \nabla)\theta + w.$$

Now, setting $F(\boldsymbol{u},\theta) \equiv (\boldsymbol{f}(\boldsymbol{u},\theta), g(\boldsymbol{u},\theta))$, the weak form of (33) is written as

$$(\nabla(\boldsymbol{u},\theta), \nabla(\boldsymbol{v},\vartheta))_{L^2} = (F(\boldsymbol{u},\theta), (\boldsymbol{v},\vartheta))_{L^2}, \quad \forall(\boldsymbol{v},\vartheta) \in X.$$

When the variables $\boldsymbol{u}$, $\theta$ in the right-hand side of (33) are replaced by some known functions, the first two and the last equations could be considered as the Stokes and Poisson equations, respectively. In such a case, we call the solution operator $\mathcal{S}$ for (33) with known right-hand side as Stokes operator. Thus $(\boldsymbol{u},\theta) = \mathcal{S}F(\boldsymbol{u},\theta)$ means

$$(\nabla\mathcal{S}F(\boldsymbol{u},\theta), \nabla(\boldsymbol{v},\vartheta))_{L^2} = (F(\boldsymbol{u},\theta), (\boldsymbol{v},\vartheta))_{L^2}, \quad \forall(\boldsymbol{v},\vartheta) \in X. \tag{34}$$

Note that we always have $\mathcal{S}^{-1}(\boldsymbol{u},\theta) = (-\Delta\boldsymbol{u} + \nabla p, -\Delta\theta)$ with an associated pressure $p = p(\boldsymbol{u},\theta)$. Now let $(\boldsymbol{u}_N,\theta_N) \in X_N$ be an approximate solution of (33), then we define the approximate pressure $p_N$ by

$$\nabla p_N \equiv \boldsymbol{f}_N(\boldsymbol{u}_N,\theta_N) + \Delta\boldsymbol{u}_N,$$

where $\boldsymbol{f}_N$ is the truncation up to $I_{,N}$ of the expansion of $\boldsymbol{f}$. For the solution $(\boldsymbol{u},\theta)$ of (33) with its associated pressure $p$, let $(\bar{\boldsymbol{u}},\bar{\theta}) \equiv (\boldsymbol{u}-\boldsymbol{u}_N, \theta-\theta_N)$ and let $\bar{p} \equiv p - p_N$. Then we have the following residual equations:

$$\begin{cases} -\Delta\bar{\boldsymbol{u}} + \nabla\bar{p} = \boldsymbol{f}(\boldsymbol{u}_N+\bar{\boldsymbol{u}}, \theta_N+\bar{\theta}) + \Delta\boldsymbol{u}_N - \nabla p_N, \\ \nabla\cdot\bar{\boldsymbol{u}} = 0, \\ -\Delta\bar{\theta} = g(\boldsymbol{u}_N+\bar{\boldsymbol{u}}, \theta_N+\bar{\theta}) + \Delta\theta_N. \end{cases} \tag{35}$$

Set

$$\begin{aligned} \bar{F}(\bar{\boldsymbol{u}},\bar{\theta}) &\equiv (\boldsymbol{f}(\boldsymbol{u}_N+\bar{\boldsymbol{u}}, \theta_N+\bar{\theta}) + \Delta\boldsymbol{u}_N - \nabla p_N, g(\boldsymbol{u}_N+\bar{\boldsymbol{u}}, \theta_N+\bar{\theta}) + \Delta\theta_N) \\ &\equiv (\bar{\boldsymbol{f}}(\bar{\boldsymbol{u}},\bar{\theta}), \bar{g}(\bar{\boldsymbol{u}},\bar{\theta})), \end{aligned}$$

then the Stokes operator $\mathcal{S}$ gives us a fixed-point problem from (35):

$$(\bar{\boldsymbol{u}},\bar{\theta}) = \mathcal{S}\bar{F}(\bar{\boldsymbol{u}},\bar{\theta}) \equiv \mathcal{K}(\bar{\boldsymbol{u}},\bar{\theta}). \tag{36}$$

Since $\mathcal{K}$ is a compact operator on $X$, we can formulate the verification procedure as in the previous sections.

In [10], we implemented the numerical verification program using the interval arithmetic by the PROFIL package [11] for gcc language on Linux Intel Pentium 4 (3.8 GHz) machine and we got some fundamental results of verification for several kinds of bifurcating solutions. Actually, we verified roll, rectangular and hexagonal type solutions for $a = 1/2\sqrt{2}$, $b = \sqrt{3}a$ and $\mathcal{P} = 10$. In this case, the critical Rayleigh number $\mathcal{R}_{\mathcal{C}} = 6.75$ can also be attained at some special mode as in two-dimensional case [21]. Figs. 3, 4 and 5 show the shape of roll type, rectangular type and hexagonal type solutions for $\mathcal{R}/\mathcal{R}_c = 1.1$, respectively. Here left is isothermal lines and right is contour lines of speed with streamlines. Fig. 6 shows the verified

points on bifurcation diagrams for each type of solutions. For roll type case, we can effectively use the unknown reduction due to the elimination of one space variable which comes from the fact that the solutions are independent of that variable. For other types, we used the property of the basic symmetry of solutions which make it possible to reduce the size of unknowns.

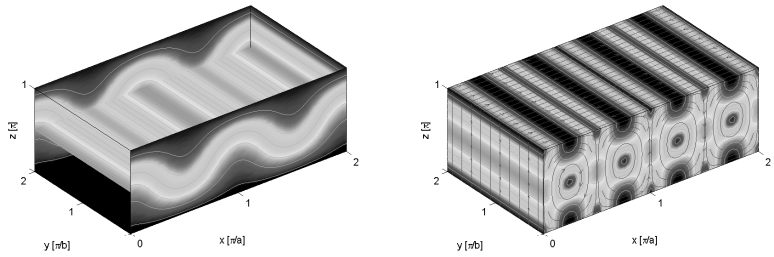


Fig. 3. Roll type solution for $\mathcal{R}/\mathcal{R}_c = 1.1$.

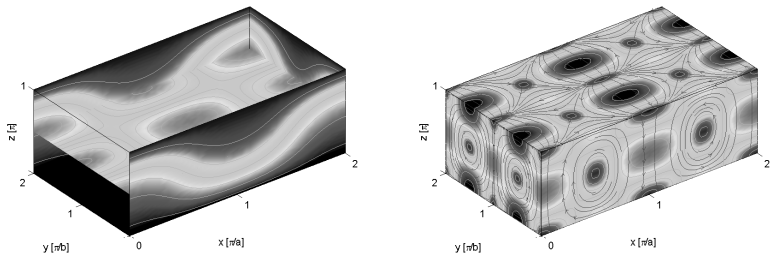


Fig. 4. Rectangular type solution for $\mathcal{R}/\mathcal{R}_c = 1.1$.

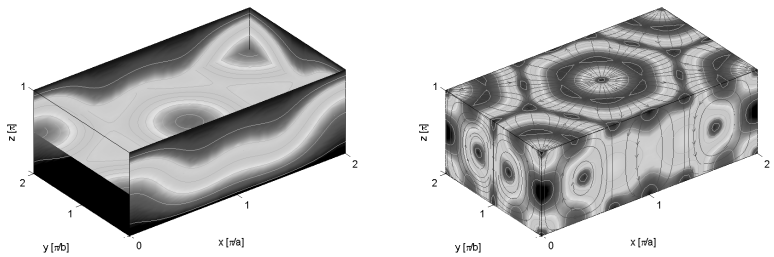


Fig. 5. Hexagonal type solution for $\mathcal{R}/\mathcal{R}_c = 1.1$.

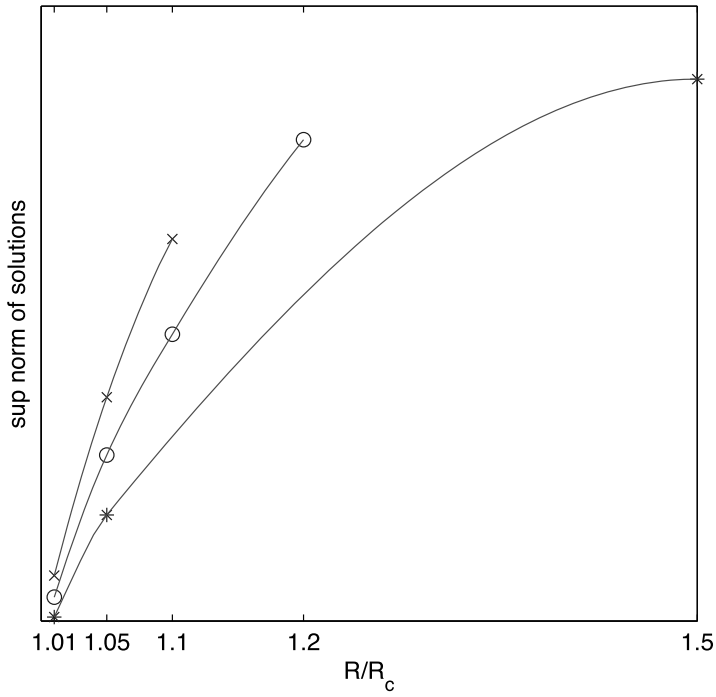


Fig. 6. Verified points: $\times$ for hexagonal, $\circ$ for rectangular, and $*$ for roll type cases.

## 6. Conclusion

We presented some computer assisted proofs to enclose the non-trivial solutions for two and three-dimensional heat convection problems as well as to prove the existence of a bifurcation point itself. These numerical computations are executed in mathematically rigorous sense and we actually succeeded to get numerical proofs of desired properties of the bifurcation structure.

Let us describe our future works. For two-dimensional heat convection problem, we want to trace non-trivial solution curve continuously with respect to the Rayleigh number. For three-dimensional case, Nishida et al. [21] suggested that the rectangular type solution becomes unstable at $\mathcal{R}/\mathcal{R}_c \approx 1.43$ and the mixed type solution bifurcates from it and goes in the hexagonal type solution at $\mathcal{R}/\mathcal{R}_c \approx 1.845$. We wish this bifurcation diagram will be justified by a computer assisted proof in the near future.

Remark 6.1. Recently we sometimes use a little bit of modified version of the original principle of our method as in [20, 18, 14] and so on. But, for the heat convection problem, in the present case, we have to use very delicate techniques to get efficient computation in the verification procedures. Such technique seem to be applicable only for our original method. However, since we have no experience by the modified methods for verification of this problem, we we could not mention

about the applicability or effectivity by the infinite dimensional Newton method such as [18] or by other method, e.g., Plum's method [22]. It should be one of the future subjects for us.

*Acknowledgment.* We would like to thank the referees for many helpful insights and comments. This work was supported by a Grant-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan (No. 18540127, No. 20224001, No. 16104001, No. 21540134).

## References

[1] H. Bénard, Les tourbillons cellulaires dans une nappe liquide. Revue Gén. Sci. Pure Appl., **11** (1900), 1261–1271, 1309–1328.

[2] S. Chandrasekhar, Hydrodynamic and Hydromagnetic Stability. Oxford University Press, 1961.

[3] J.H. Curry, Bounded solutions of finite dimensional approximations to the Boussinesq equations. SIAM J. Math. Anal., **10** (1979), 71–79.

[4] A.V. Getling, Rayleigh–Bénard Convection: Structures and Dynamics. Advanced Series in Nonlinear Dynamics, **11**, World Scientific, 1998.

[5] V.I. Iudovich, On the origin of convection. J. Appl. Math. Mech., **30** (1966), 1193–1199.

[6] D.D. Joseph, On the stability of the Boussinesq equations. Arch. Rational Mech. Anal., **20** (1965), 59–71.

[7] Y. Kagei and W. von Wahl, The Eckhaus criterion for convection roll solutions of the Oberbeck–Boussinesq equations. Int. J. Non-linear Mechanics, **32** (1997), 563–620.

[8] T. Kawanago, A symmetry-breaking bifurcation theorem and some related theorems applicable to maps having unbounded derivatives. Japan J. Indust. Appl. Math, **21** (2004), 57–74.

[9] F. Kikuchi and L. Xuefeng, Determination of the Babuska–Aziz constant for the linear triangular finite element. Japan J. Ind. Appl. Math., **23** (2006), 75–82.

[10] M.-N. Kim, M.T. Nakao, Y. Watanabe and T. Nishida, A numerical verification method of bifurcating solutions for 3-dimensional Rayleigh–Bénard problems. Numer. Math., **111** (2009), 389–406.

[11] O. Knüppel, PROFIL/BIAS—A fast interval library. Computing, **53** (1994), 277–287, `http://www.ti3.tu-harburg.de/Software/PROFILEnglisch.html`.

[12] R. Krishnamurti, Some further studies on the transition to turbulent convection. J. Fluid Mech., **60** (1973), 285–303.

[13] K. Nagatou, N. Yamamoto and M.T. Nakao, An approach to the numerical verification of solutions for nonlinear elliptic problems with local uniqueness. Numer. Funct. Anal. Optim., **20** (1999), 543–565.

[14] K. Nagatou, K. Hashimoto and M.T. Nakao, Numerical verification of stationary solutions for Navier–Stokes problems. J. Comput. Appl. Math., **199** (2007), 424–431.

[15] M.T. Nakao, A numerical approach to the proof of existence of solutions for elliptic problems. Japan J. Appl. Math., **5** (1988), 313–332.

[16] M.T. Nakao, N. Yamamoto and S. Kimura, On best constant in the optimal error stimates for the $H_0^1$-projection into piecewise polynomial spaces. Journal of Approximation Theory, **93**, (1998), 491–500.

[17] M.T. Nakao, Numerical verification methods for solutions of ordinary and partial differential equations. Numer. Funct. Anal. Optim., **22** (2001), 321–356.

[18] M.T. Nakao, K. Hashimoto and Y. Watanabe, A numerical method to verify the invertibility of linear elliptic operators with applications to nonlinear problems. Computing, **75** (2005), 1–14.

[19] M.T. Nakao, Y. Watanabe, N. Yamamoto and T. Nishida, Some computer assisted proofs for solutions of the heat convection problems. Reliable Computing, **9** (2003), 359–372.

[20] M.T. Nakao and Y. Watanabe, An efficient approach to the numerical verification for solutions of elliptic differential equations. Numer. Algor., **37** (2004), 311–323.

[21] T. Nishida, T. Ikeda and H. Yoshihara, Pattern formation of heat convection problems. Proceedings of the International Symposium on Mathematical Modeling and Numerical Simulation in Continuum Mechanics, T. Miyoshi et al. (eds.), Lecture Notes in Computational Science and Engineering, **19**, Springer-Verlag, 2002, 155–167.

[22] M. Plum, Explicit $H_2$-estimates and pointwise bounds for solutions of second-order elliptic boundary value problems. J. Math. Anal. Appl., **165** (1992), 36–61.

[23] P.H. Rabinowitz, Existence and nonuniqueness of rectangular solutions of the Bénard problem. Arch. Rational Mech. Anal., **29** (1968), 32–57.

[24] J.W.S. Rayleigh, On convection currents in a horizontal layer of fluid, when the higher temperature is on the under side. Phil. Mag., **32** (1916), 529–546; Sci. Papers, **6**, 432–446.

[25] S.M. Rump, On the solution of interval linear systems. Computing, **47** (1992), 337–353.

[26] S.M. Rump, A note on epsilon-inflation. Reliable Computing, **4** (1998), 371–375.

[27] Y. Watanabe, N. Yamamoto, M.T. Nakao and T. Nishida, A numerical verification of nontrivial solutions for the heat convection problem. J. Math. Fluid Mech., **6** (2004), 1–20.

[28] Y. Watanabe, A computer-assisted proof for the Kolmogorov flows of incompressible viscous fluid. J. Comput. Appl. Math., **223** (2009), 953–966.

[29] N. Yamamoto, A numerical verification method for solutions of boundary value problems with local uniqueness by Banach's fixed point theorem. SIAM J. Numer. Anal., **35** (1998), 2004–2013.