Numerical techniques for large dimensional dynamical systems

Recent Trends in Nonlinear Science 2015

Examples

Contents

1 Small problems

1.1 A two-bar truss arc

Consider the truss arch shown in Fig. 1, where a load of $p\,\mathrm{N}$ is appied on the upper node. The bars a natural length of l_0 and their lower nodes are fixed at a distance of $2l$ meters. The bars behave like an elastic spring with spring constant equal to $k\,\mathrm{N/m}$. The (x, y) coordinates (in metres) of the top node can be found as the solution of the following system

$$\begin{aligned} kl_0\Big(\frac{x+l}{r_1}+\frac{x-l}{r_2}\Big)-2kx &= 0,\\ kl_0\Big(\frac{y}{r_1}+\frac{y}{r_2}\Big)-2ky-p &= 0, \end{aligned} \tag{1}$$

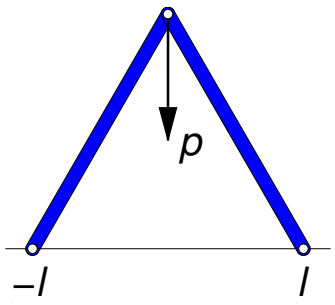
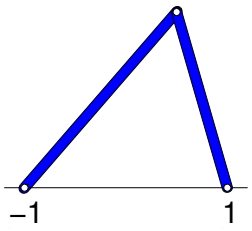
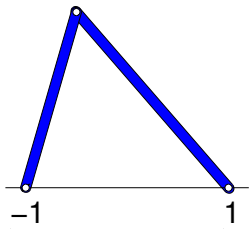


Figure 1: A two-bar truss arc.



Configuration corresponding to u



Configuration corresponding to Su

Figure 2: The effect of the symmetry (3) on the dimensionless system.

where

$$r_1 = \sqrt{l^2 + 2lx + x^2 + y^2}, \qquad \text{y} \quad r_2 = \sqrt{l^2 - 2lx + x^2 + y^2},$$

Dividing both equations by kl and introducing the dimensionless variables

$$u_1 = x/l, \qquad u_2 = y/l, \qquad \lambda = p/(kl), \qquad \text{y} \quad \mu = l_0/l$$

(notice that $\mu > 1$) system of equations (1) can be rewritten as

$$f(u, \lambda) = 0,$$

where, for every $\mu > 0$,

$$f(u, \lambda) = g(u, \lambda, \mu) = \begin{bmatrix} \mu\Big(\dfrac{u_1 + 1}{d_1} + \dfrac{u_1 - 1}{d_2}\Big) - 2u_1 \\ \mu\Big(\dfrac{u_2}{d_1} + \dfrac{u_2}{d_2}\Big) - 2u_2 - \lambda \end{bmatrix}, \tag{2}$$

and

$$d_1 = \sqrt{1 + 2u_1 + u_1^2 + u_2^2}, \qquad \text{y} \quad d_2 = \sqrt{1 - 2u_1 + u_1^2 + u_2^2}.$$

The arc does not necessarily have to be in a symmetric configuration. In fact, the function f is equivariant under the symmetry S given by

$$Su = \begin{bmatrix} -u_1 \\ u_2 \end{bmatrix}. \tag{3}$$

An example of the effect of this symmetry can be seen in Fig. 2.

The configuration of the possible equilibria varies with the natural length μ. Fig. 3 shows the equilbria for positive λ when $\mu = \sqrt{2}$ (bars forming a 45 degree angle with the ground). Stable equilbria are marked in blue and unstable in red. We see that there are a symmetric brach which undergoes a saddle-node bifurcation, stable equilbria being those of a higher top point than unstable ones. There are two non symmetric branches which are unstable. Each of them is transformed from the other by the simmetry S. Their extreme points are $(\pm\mu, 0)$ and $(1, 0)$.

For larger bars, though, the disposition of the branches is more elaborate. In Fig. 4, the equilibria for positive λ when $\mu = 1/\cos(7\pi/16)$ are shown. Again stable equilibria are marked in blue. The equilibria marked in green correspond to thos wher the Jacobian of f with respect to u

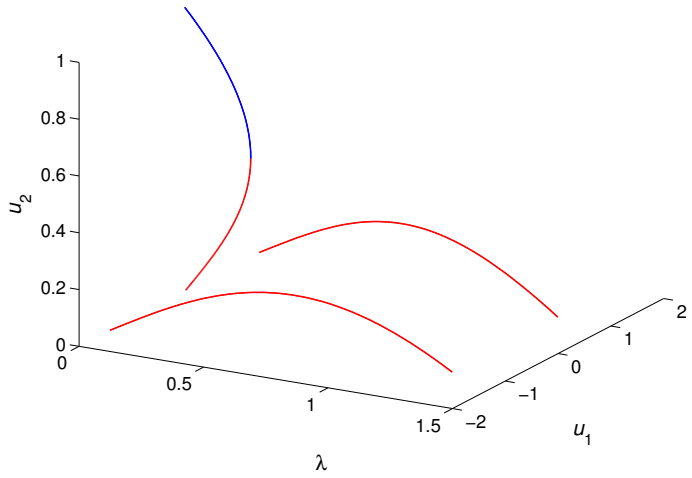


Figure 3: Truss arc equilibria for $\mu = \sqrt{2}$. Blue: stable equilbria. Red: unstable equilibria.

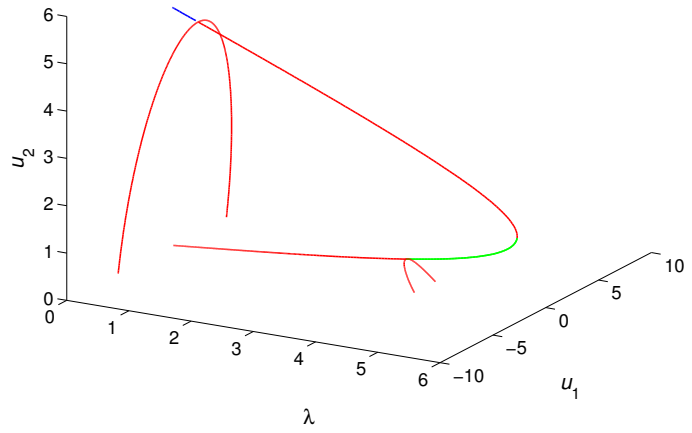


Figure 4: Truss arc equilibria for $\mu = 1/\cos(7\pi/16)$. Blue: stable equilbria. Red: unstable equilibria. Greeen: unstable equilibria with two positive eigenvalues in f_u.

with two positive eigenvalues, whereas those marked in red have just one positive eigenvalue. Fig. 4 We notice that there is a much smaller length of stable equilibria, and that the non symmetric branches intersect the symmetric one in simmetry breaking bifurcations. Can you figure out the cofiguration for values of μ between $\sqrt{2}$ and $1/\cos(7\pi/16)$?

In what follows we will denote by $e_1, \ldots, e_m$ the *coordinate vectors* in $\mathbb{R}^m$. With this notation we notice that quantities d_1 y d_2 can be rewritten as

$$d_1 = \|u + e_1\|\,, \qquad d_2 = \|u - e_1\|\,,$$

Similarly, f in (2) can be rewritten as

$$f(u,\lambda) = g(u,\lambda,\mu) = \mu\Big(\frac{1}{d_1}(u+e_1) + \frac{1}{d_2}(u-e_1)\Big) - 2u - \lambda e_2. \tag{4}$$

Furthermore, sine the differentiala of d_1 and d_2 with respect to u is

$$\partial_u d_1 = [\partial_{u_1} d_1, \partial_{u_2} d_1] = \frac{1}{d_1}(u+e_1)^T, \qquad \partial_u d_2 = [\partial_{u_1} d_2, \partial_{u_2} d_2] = \frac{1}{d_2}(u-e_1)^T,$$

we can write the differential of f with respect to u as

$$f_u = \begin{bmatrix} \partial_{u_1} f_1 & \partial_{u_2} f_1 \\ \partial_{u_1} f_1 & \partial_{u_2} f_1 \end{bmatrix} = \mu\Big(\Big(\frac{1}{d_1} + \frac{1}{d_2}\Big)I - \frac{1}{d_1^3}(u+e_1)(u+e_1)^T - \frac{1}{d_2^3}(u-e_1)(u-e_1)^T\Big) - 2I,$$

where I denotes the identity matrix. Notice also that $(u \pm e_1)(u \pm e_1)^T$ is the matrix

$$(u \pm e_1)(u \pm e_1)^T = \begin{bmatrix} u_1 \pm 1 \\ u_2 \end{bmatrix} [u_1 \pm 1, u_2] = \begin{bmatrix} (u_1 \pm 1)^2 & (u_1 \pm 1)u_2 \\ u_2(u_1 \pm 1) & u_2^2 \end{bmatrix}.$$

Thus two MATAB or OCTAVE function seturning f and f_u could be as follows.

```
function f=funp(u,params)
 % FUNP returns the force acting on the top node
 %            in the truss arc, for xy coordinates u,
 %            dimensionles load lambda=params(1)
 %            and natural bar-length mu=params(2);

 lambda=params(1); mu=params(2);
 % Coordinate vectors in e1 and e2
 e1=[1; 0]; e2=[0; 1];
 y1=u+e1; y2=u-e1; d1=norm(y1); d2=norm(y2);
 r=( y1/d1 + y2/d2);
 f=-lambda*e2 -2*u + mu*r;
```

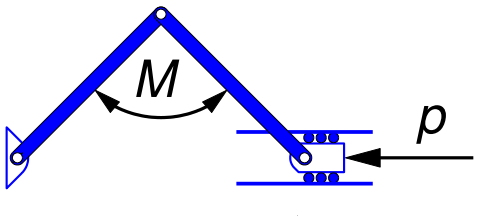


Figure 5: Buckling of two bars joined by a torsional spring

```
function J=jfunp(u,params)
  % JFUNP returns the Jacobian matrix wrt u of function funp,

  lambda=params(1); mu=params(2);
  % Coordinate vectors in e1 and e2
  e1=[1; 0]; e2=[0; 1];
  y1=u+e1; y2=u-e1; d1=norm(y1); d2=norm(y2);
  r=( y1/d1 + y2/d2);
  %f=-lambda*e2 -2*u + mu*r;
  I=eye(2); % Matlab and Octave's command eye
            % return the identity matrix
  J=(-2 + mu*(1/d1 + 1/d2))*I -  mu*((y1*y1')/(d1^3)...
      + (y2*y2')/(d2^3));
```

1.2 A gradual aproach to Euler buckling

Consider two rigid bars joined by a torsional spring of constant M, pinned in one end and subject to a horizontal force on the other as indicated in Fig. 5 The potential energy, in dimensionless quantities after dividing by the rotational spring constant and the sum of the length of the two bars is

$$V(\theta) = \frac{1}{2}(\pi - \theta)^2) - p(1 - 2\cos((\pi - \theta)/2)),$$

where θ is the angle between the two bars and p is the dimensionless force (See e. g., [5, § I.1].

This system can be considered with more bars as indicated in Fig. 6[h] For a system with N bars of dimensionless length $1/N$ the potential energy is given by

$$V(\theta_1, \ldots, \theta_N) = \sum_{j=2}^{N} \frac{1}{2}(\theta_j - \theta_{j-1})^2 - p\left(1 - \frac{1}{N}\sum_{j-1}^{N} \cos(\theta_j)\right), \tag{5}$$

where θ_j is the angle between the j-th bar and the horizontal line, as shown in Fig. 6. Notice that these angles must satisfy the constraint

$$\frac{1}{N}\sum_{j=1}^{N} \sin(\theta_j) = 0. \tag{6}$$

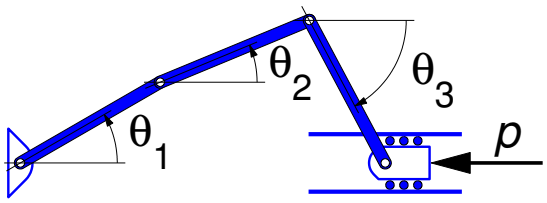


Figure 6: Three-bar buckling

[b]

The equilibria are the solution of the system of equations

$$\partial_{\theta_j} V - \lambda \frac{1}{N} \cos(\theta_j) = 0, \qquad j = 1, \ldots, N$$

where λ is the Lagrange multiplier of restriction (6), that is

$$-\theta_{j-1} + 2\theta_j - \theta_{j+1} - p\frac{1}{N}\sin(\theta_j) - \lambda \frac{1}{N} \cos(\theta_j) = 0, \qquad j = 1, \ldots, N, \tag{7}$$

where, for convenience in writing the equations, we have introduced two additional variables

$$\theta_0 = 0, \qquad \theta_{N+1} = 0. \tag{8}$$

Notice that there are $N + 1$ unknowns, the N angles $\theta_1, \ldots, \theta_N$ and the Lagrange multiplier λ which are the solution or the system of $N + 1$ equations given by the restriction (6) and the N equations (9).

If we increase the spring constant with N, so that istead of (7) we have

$$-N(\theta_{j-1} - 2\theta_j + \theta_{j+1}) - p\frac{1}{N}\sin(\theta_j) - \lambda \frac{1}{N} \cos(\theta_j) = 0, \qquad j = 1, \ldots, N, \tag{9}$$

then, when $N \to \infty$, this systems models the equilibrium of a long and slender bar inextensible bar subject to a compressive load, a problem that was resolved by Euler [1, § 1.13.1]. In this case, Euler shown that the trivial solution was only stable for $p < \pi^2$, and that indeed, bifurcation from the trivial solution occurred at $p = n^2\pi^2$, $n = 1, 2, \ldots$. In the case of the system (9) together with the (6), the trivial solution

$$\theta_1 = \ldots, \theta_N = 0, \qquad \lambda = 0,$$

also exists for all values of lambda, and bifurcations from the trivial solution occur when

$$p_k = 2N^2(1 - \cos(\pi \frac{k}{N})), \qquad k = 1, \ldots, N - 1.$$

Observe that

$$p_1 = \pi^2 + O(N^{-2}).$$

Fig. 7 shows the x-coordinate of the last point in the set of bars (where load p is applied) for the solutions in branch bifurcating at $p = p_1$ for $N = 20$, for $p \leq 50$. As before, stable solutions are

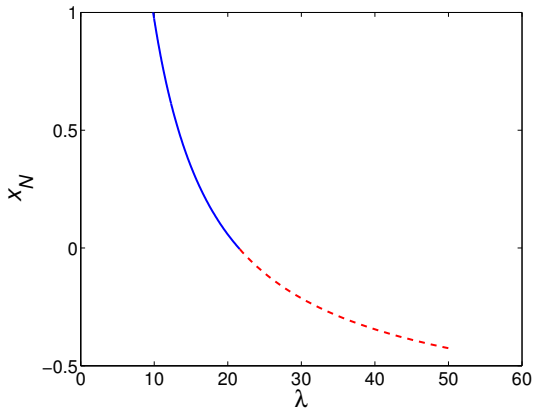


Figure 7: x-coordinate of endpoint of solutions of branch bifurcating at p_1, for $N = 20$, for $p \leq 50$. Stable solutions in blue, unstable in red.

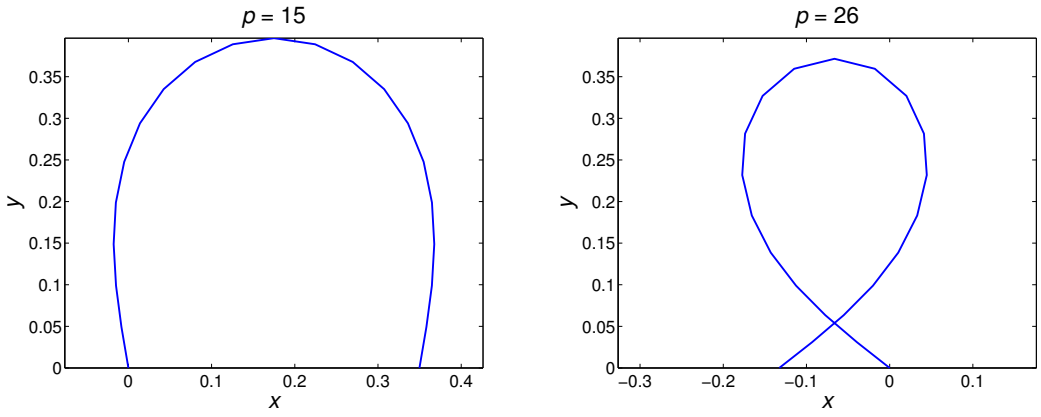


Figure 8: Two solution on branch bifurcating at p_1. Left figure corresponds to a stable solution, and right figure to an unstable one.

represented in blue and unstable in red. Notice that solutions are unstable as soon as x_N becames negative. An example of stable and unstable solutions on this branch can be found in Fig. 8.

Finally, we show two MATLAB or OCTAVE functions computing the left-hand side of sistem of equations (9) and (6).

```
function [f,fp]=fun_Euler(u,p)
% FUN_EULER returns in f minus the gradient of the potential
%     potential energy in bucklihg bars minus the lagrange
%     multiplier times the gradit of the restriction.
%  INPUT
%    u array of length N+1 with the angles
%        in U(1:N) and the Lagrange multiplier
%        in U(N+1)
%    p dimensionless force

N=length(u)-1; lambda=u(N+1);
% lambda is the  Lagrange multipliear of the restriction
%
%    sum from n=1 up to n=N sin(u_n)/N = 0
%
fd=[u(1)-u(2);-diff(diff(u(1:N))); u(N)-u(N-1)]*N;
sines=sin(u(1:N)); cosines=cos(u(1:N));
fs=(p/N)*sines-(lambda/N)*cosines;
f=[fd-fs; mean(sines)];

function [J,fp]=jfun_Euler(u,p)
% JFUN_EULER computes the Jacobian matrix
%            of function fun_Euler.
%  INPUT
%    u array of length N+1 with the angles
%        in U(1:N) and the Lagrange multiplier
%        in U(N+1)
%    p dimensionless force
%
%  OUTPUT
%    J (N+1)x(N+1) array with the Jacobian matrix
%        of function fun_Euler.
%    fp array the same size as u with fhe partial
%       derivative of fun_Euler wrt p.

N=length(u)-1; lambda=u(N+1);
% lambda is the  Lagrange multipliear of the restriction
%
%    sum from n=1 up to n=N sin(u_n)/N = 0
%
```

```
sines=sin(u(1:N)); cosines=cos(u(1:N));

e=ones(N,1); D=spdiags(N*[-e 2*e -e],-1:1,N,N);
D(1,1)=-D(1,2); D(N,N)=-D(N,N-1);
% Jacobian matrix is stored as a sparse matrix
% since it is a tridiagonal matrix.
D=D-(1/N)*spdiags(p*cosines+lambda*sines,0,N,N);
J=[D,cosines/N;(cosines')/N,0];

fp=[-sines/N; 0];
```

2 Convection in fluid-saturated porous media

2.1 Problem and equations

Consider a rectangular box filled with fluid-saturated porous material and heated from below and

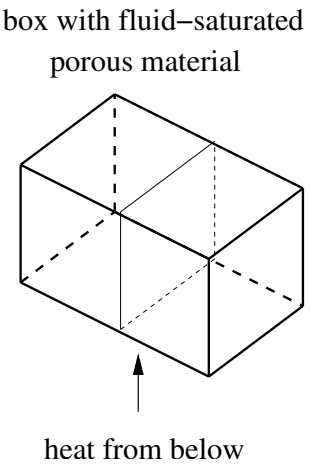


Figure 9: Schematic representation of the convection problem.

with insulated horizontal walls. The temperature may vary linearly with height and the fluid remain at rest if the difference in temperature between the top and bottom walls in small, or convection can be developed and temperature distribution may have more complicated patterns. If one of the horizontal dimensions is much larger than the other, there are some two-dimensional configurations which can be computed by considering a cross-section, as shown in Fig. 2.1, where the setting is as shown in Fig. 10.

After adimensionalization, the deviation $u(t, x, y)$ from the linear temperature distribution $T_0(1-y)$ and the fluid velocity $v(t, x, y)$ and pressure $p(t, x, y)$, is the solution of the following partial differential equation (PDE) set in the domain

$$\Omega = [0, 1] \times [0, 1],$$

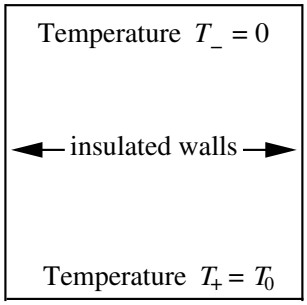


Figure 10: Schematic representation of the convection problem on a cross-section.

$$\begin{aligned} u_t + \sqrt{\mu}\,\mathbf{v}\cdot(\nabla u - \mathbf{e}_2) &= \Delta u, \\ -\nabla p - \mathbf{v} + \sqrt{\mu} u \mathbf{e}_2 &= 0, \\ \nabla\cdot\mathbf{v} &= 0, \\ \nabla u\cdot n &= 0, \qquad x = 0,1, \quad y\in[0,1], \\ u &= 0, \qquad x\in[0,1], \quad y = 0,1, \\ \mathbf{v}\cdot n &= 0, \qquad (y,z)\in\partial\Omega, \end{aligned} \tag{10}$$

where n represents the outward normal vector, $e_2 = [0,1]^T$ is the vertical unit vector, and $\mu = Ra$, the Rayleigh number, is a scalar parameter (see e.g., [9], [7], [8] and the references cited therein).

Observe that the eigenfunctions of the Laplacian operator subject to the boundary conditions imposed on u in (10) are

$$p_{j,k}(x,y) = \cos(\pi j x)\sin(\pi k y), \qquad j = 0,1,\dots, \quad k = 1,2,\dots. \tag{11}$$

It is well-known that the eignenfunctions of the Laplacian operator form an orthogonal set in $L^2(\Omega)$,. In fact, we have

$$\int_\Omega p_{j,k}(x,y)p_{l,m}(x,y)\,dx\,dy = \begin{cases} 0, & \text{if } (j,k)\neq(l,m), \\ 1/4 & \text{otherwise.} \end{cases}$$

Being the functions $p_{j,k}$ an orthogonal set in $L^2(\Omega)$, we can express

$$u(x,y,t) = \sum_{j=0,k=1}^{\infty} \hat{u}_{j,k}(t)p_{j,k}(x,y). \tag{12}$$

Similary, and due to the velocity v begin divergence-free, it can be expressed as

$$\mathbf{v}(x,y,t) = \sum_{j,k=1}^{\infty} \hat{v}_{j,k}(t)\frac{2}{\sqrt{j^2+k^2}}\begin{bmatrix} -kq_{j,k}(x,y) \\ jp_{j,k}(x,y) \end{bmatrix}, \tag{13}$$

where

$$q_{j,k}(x,y) = \sin(\pi j x)\cos(\pi k y), \qquad j = 1,2\dots, k = 0,1,\dots.$$

It is straightforward to check that, with the boundary conditions imposed on v, divergence-free functions are orthogonal in $L^2(\Omega)$ to gradients, so that projecting the second equation in (10) onto the divergence-free functions we have that

$$\mathbf{v} = \mathbf{v}(u),$$

and, indeed,

$$\hat{v}_{j,k} = \frac{j\sqrt{\mu}}{2\sqrt{j^2+k^2}}\hat{u}_{j,k}, \qquad j,k = 1,\dots, \tag{14}$$

and, thus,

$$\mathbf{v} = \mathbf{v}(u) = \sqrt{\mu}\sum_{j,k=1}^{\infty} \hat{u}_{j,k}(t)\frac{j}{j^2+k^2}\begin{bmatrix} -kq_{j,k}(x,y) \\ jp_{j,k}(x,y) \end{bmatrix}, \tag{15}$$

Steady-state solutions of (10) are then solutions of

$$f(u,\mu) \equiv -\Delta u + \sqrt{\mu}\,\mathbf{v}(u)\cdot(\nabla u - \mathbf{e}_2) = 0, \qquad \text{in } \Omega \tag{16}$$

subject to

$$\nabla u \cdot n = 0, \qquad x = 0,1, \quad y \in [0,1], \tag{17}$$

$$u = 0, \qquad x \in [0,1], \quad y = 0,1, \tag{18}$$

and time dependent solutions are solutions of

$$u_t + f(u,\mu) = 0,$$

subject also to the boundary conditions (17–18). In terms of the Fourier coefficients $\hat{u}_{j,k}$ this equation can be written as the *infinite-dimensional* dynamical system

$$\frac{d\hat{u}_{j,k}}{dt} + \pi^2(j^2+k^2)\hat{u}_{j,k} + \sqrt{\mu}(\mathbf{v}\cdot\widehat{(\nabla u - \mathbf{e}_2)})_{j,k} = 0, \qquad j = 0,1,\dots, \quad k = 1,2,\dots, \tag{19}$$

where $(\mathbf{v}\cdot\widehat{(\nabla u - \mathbf{e}_2)})_{j,k}$ stands for the (j,k)-Fourier coefficient of $\mathbf{v}\cdot(\nabla u - \mathbf{e}_2)$, given by

$$(\mathbf{v}\cdot\widehat{(\nabla u - \mathbf{e}_2)})_{j,k} = \frac{(p_{j,k}, \mathbf{v}\cdot(\nabla u - \mathbf{e}_2))}{\|p_{j,k}\|^2}, \qquad j = 0,1,\dots, \quad k = 1,2,\dots, \tag{20}$$

where, here and in the sequel, $(\cdot,\cdot)$ denotes the standard inner product in $L^2(\Omega)$,

$$(v,w) = \int_\Omega v(x,y)w(x,y)\,dx\,dy$$

and $\|\cdot\|$ its associated norm. Later on we well see how the Fourier coefficients $(\mathbf{v}\cdot\widehat{(\nabla u - \mathbf{e}_2)})_{j,k}$ are computed (rather, approximated) in practice.

It is easy to check that the left hand side of (16) is equivariant by the group of symmetries generated by S_y and S_z

$$S_x u(x,y) = u(1-x,y), \qquad S_y u(x,y) = -u(x,1-y). \tag{21}$$

In addition, for $p = 2,3,\dots$, the subspaces

$$\mathcal{Y}_p = \text{span}\{p_{pl,k}, \quad l = 0,1,\dots, \quad k = 1,2,\dots\}$$

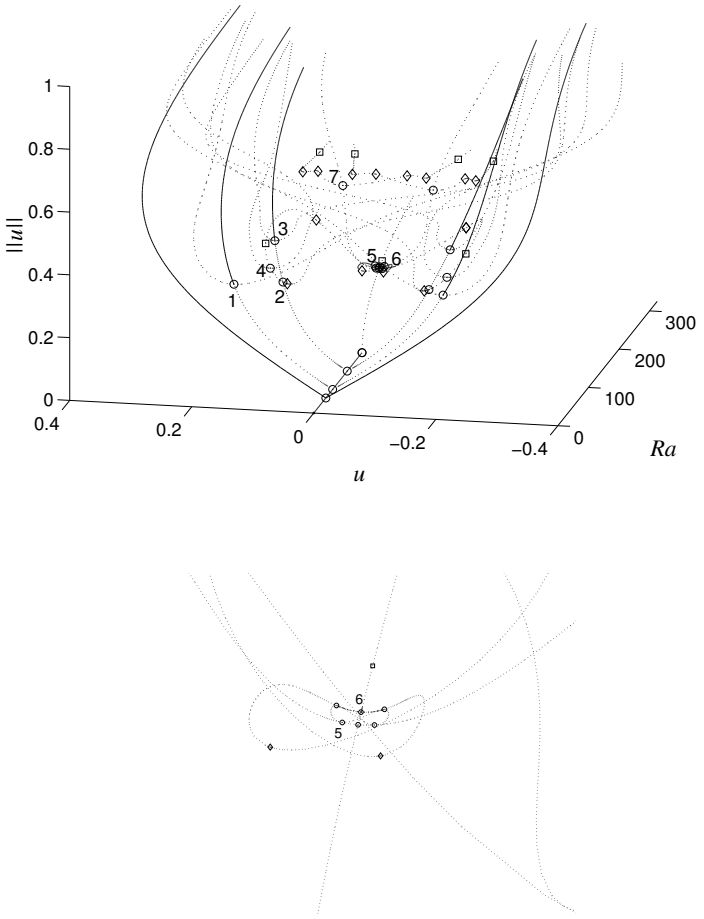


Figure 11: Bifurcation diagram of system (16), $\mu = Ra$ the Rayleigh number. Bifurcation points: $\circ$, pitchfork or transcritical; $\diamond$, saddle-node; $\square$, Hopf bifurcation (Top: Full diagram. Bottom: detail of bifurcation points 5 and 6).

are invariant by f and have the translational invariance $T_p^s u(x,y) = u(x+\frac{2s}{p}, y)$, for $s = 1, \ldots, p-1$ (see e.g. [9]).

Fig. 11 which is taken from [4], shows a bifurcation diagram of the equation (16), that is a representation of the branches of its steady-state solutions parameter μ for $0 \leq \mu \leq 325$. A two-dimensional version of this diagram can be found for example in [9]. The vertical axis represents the L^2 norm of the solution u and the transversal axis the value of u at the centre of the left wall, that is $u(0, 1/2)$.

Fig. 12 shows two stable configurations (temperature contour levels on the left and streamlines on the right) for Ra = 100. They correspond to the first primary branch (top) and to the second primary branch after bifurcation point labeled with 1 in Fig. 11.

It is easy to check that the the Fréchet derivative of $f(u, \mu)$ when u is the trivial solution $u = 0$

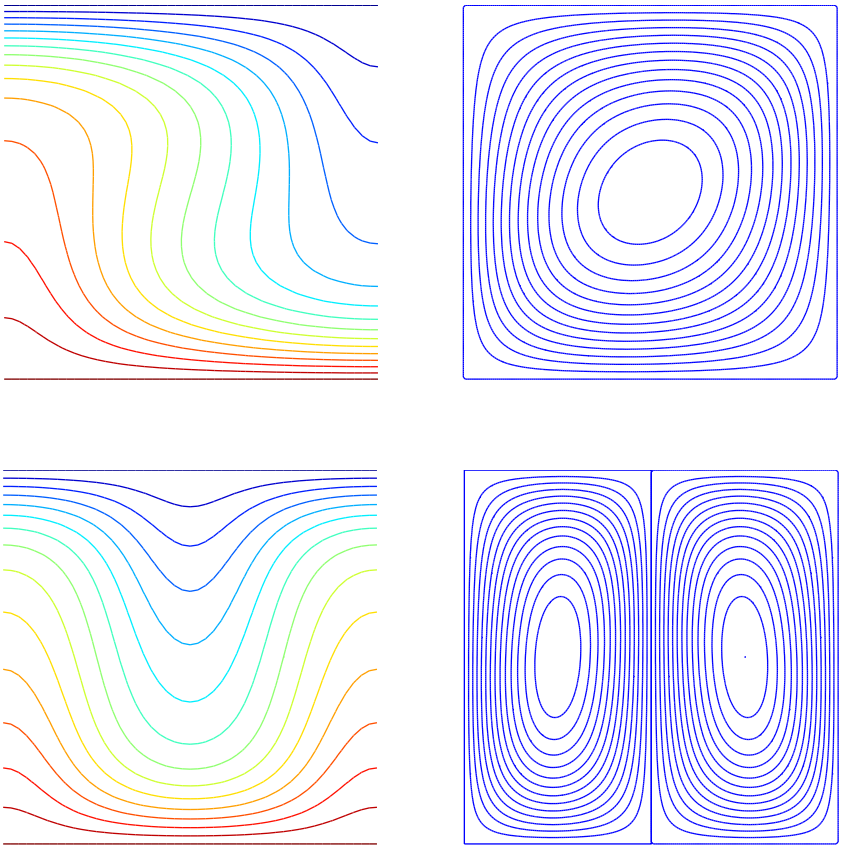


Figure 12: Temperature contour lines (left) and streamlines (right) for stable configurations on first primary branch (top) and second primary branch for Ra = 100

.

acting of function

$$\varphi = \sum_{j=0,k=1}^{\infty} \hat{\varphi}_{j,k}(t) p_{j,k}.$$

satisfying

$$\begin{aligned} \nabla\varphi \cdot n &= 0, \qquad x = 0, 1, \quad y \in [0,1], \\ \varphi &= 0, \qquad x \in [0,1], \quad y = 0, 1, \end{aligned}$$

is given by

$$f_u(0,\mu)\varphi = -\Delta\varphi - \sqrt{\mu} v(\varphi) \cdot e_2,$$

where $\mathbf{v} = \mathbf{v}(\varphi)$ and $p = p(\varphi)$ are the solution of

$$\begin{aligned} -\nabla p - \mathbf{v} + \sqrt{\mu}\varphi \mathbf{e}_2 &= 0, \\ \nabla \cdot \mathbf{v} &= 0, \\ \mathbf{v} \cdot n &= 0, \qquad (y,z) \in \partial\Omega, \end{aligned}$$

In terms of the Fourier coefficients the Fréchet derivative at $u = 0$ is given by

$$(\widehat{f_u(0,\mu)\varphi})_{j,k} = \pi^2(j^2+k^2)\hat{\varphi}_{j,k} - \mu\frac{j^2}{j^2+k^2}\hat{\varphi}_{j,k}, \qquad j = 0, 1, \dots, \quad k = 1, 2, \dots .$$

Thus, $f_u(0,\mu)$ has a zero eigenvalue for the following values of μ

$$\mu_{j,k} = \pi^2\frac{(j^2+k^2)^2}{j^2}, \qquad j, k = 1, 2, \dots,$$

the corresponding (unit-norm) eigenfunction being

$$\frac{1}{2}p_{j,k},$$

We now check that this values correspond to branching points (in fact, they correspond to pitchfork bifurcations) by checking that the derivative $f_\mu(0,\mu)$ of f with respect to μ at $u = 0$ is

$$f_\mu(0,\mu) = 0, \tag{22}$$

so that $f_\mu(0,\mu)$ is in the range of the Fréchet derivative $f_u(0,\mu)$. For this purpose, we need first to compute the derivative $\mathbf{v}_\mu(0,\mu)$ of the velocity $\mathbf{v}$ at $u = 0$. Taking derivatives with respect to μ in the second equation in (10) we have

$$-\nabla p_\mu - \mathbf{v}_\mu + \frac{1}{2\sqrt{\mu}}u\mathbf{e}_2 = 0,$$

which, arguing as before, implies

$$\mathbf{v}_\mu = \frac{1}{2\mu}\mathbf{v}.$$

(We reach this result also by taking derivatives with respect to μ in (15)).

Taking derivatives with respec to μ in the expression of f in (16) we have the derivative $f_\mu(0,\mu)$ of f with respect to μ is given by is

$$\begin{aligned} f_\mu(u,\mu) &= \frac{1}{2\sqrt{\mu}}\mathbf{v}\cdot(\nabla u - \mathbf{e}_2) + \sqrt{\mu}\mathbf{v}_\mu\cdot(\nabla u - \mathbf{e}_2) = \frac{1}{2\sqrt{\mu}}\mathbf{v}\cdot(\nabla u - \mathbf{e}_2) - \frac{1}{2\sqrt{\mu}}\mathbf{v}\cdot(\nabla u - \mathbf{e}_2) \\ &= \frac{1}{\sqrt{\mu}}\mathbf{v}\cdot(\nabla u - \mathbf{e}_2). \end{aligned} \tag{23}$$

Since for $u = 0$ we have that $\mathbf{v} = 0$, it follows that (22) holds.

We comment on another important property of this example. It is not difficult to show that Δu satisfies the same boundary conditions as u, and, furthermore, it is no difficult to show by induction that $\delta^j u$ also satisfies the same boundary conditions as u. As a consequence, and it is possible to show (see e. g. [6]) that the Fourier coefficients of the solutions decay exponentially fast, that is,

$$|\hat{u}_{j,k}| \leq Ce^{-\gamma(j+k)},$$

for some $C > 0$ and $\gamma > 0$. An example can be seen in Fig. 13, where we show against the wave number $\lambda = \pi\sqrt{j^2+k^2}$ the absolute value $|\hat{u}_{j,k}|$ of the non null fourier coefficients of the solutions depicted in Fig. 12.

This property of exponential decay of Fourier coefficients of solutions make equation (16) particularly well-suited to be discretized by a spectral method, as we explain in the following section.

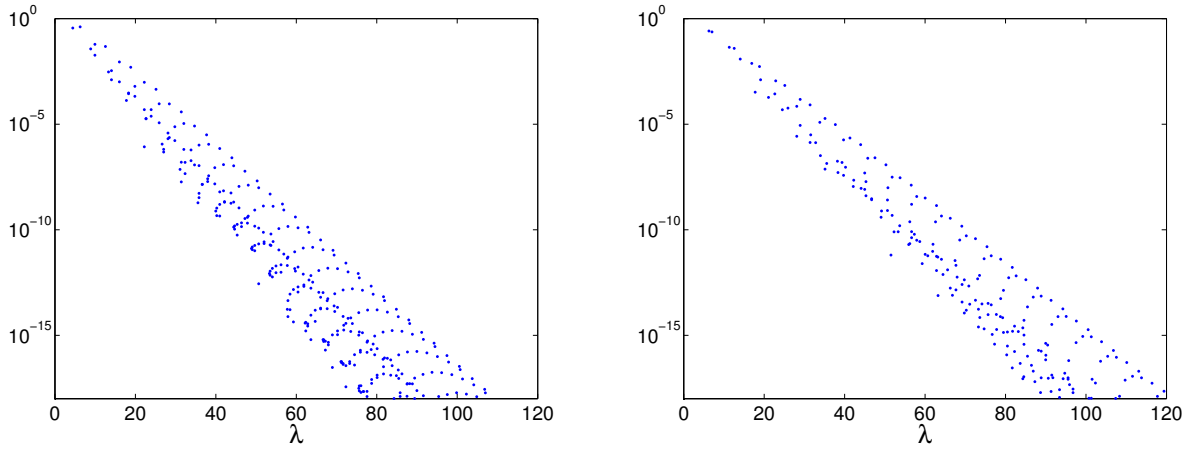


Figure 13: Value of $|\hat{u}_{j,k}|$ against wave number $\lambda = \pi\sqrt{j^2+k^2}$ for solutions depicted in Fig. 12. Left, primary branch; right, secondary branch.

2.2 Spectral Discretization

To approximate solutions of we consider, for a given positive integer N approximations U_N of the form

$$U_N(t,x,y) = \sum_{j=0}^{N}\sum_{k=1}^{N-1} U_{j,k}(t)p_{j,k}(x,y),$$

where the functions $p_{j,k}$ are those defined in (11) Observe that, contrary to the solution u of (16), the approximation is, for every t a linear combination of a *finite* number of eigenfunctions $p_{j,k}$. The coefficients $U_{j,k}$ of pproximations U_N are required to satisfy

$$\frac{d\hat{U}_{j,k}}{dt} + \pi^2(j^2+k^2)\hat{U}_{j,k} + \sqrt{\mu}(\mathbf{V}_N \cdot \widehat{(\nabla U_N - \mathbf{e}_2)})_{j,k} = 0, \qquad j = 0,1,\ldots,N, \quad k = 1,2,\ldots,N-1, \tag{24}$$

where $\mathbf{V}_N$ is the velocity given by

$$\mathbf{V}_N = \sqrt{\mu}\sum_{j,k=1}^{N-1} \hat{U}_{j,k}(t)\frac{j}{j^2+k^2}\begin{bmatrix} -kq_{j,k}(x,y) \\ jp_{j,k}(x,y) \end{bmatrix}, \tag{25}$$

and $(\mathbf{V}_N \cdot \widehat{(\nabla U_N - \mathbf{e}_2)})_{j,k}$ standas for the Fourier coefficient of $\mathbf{V}_N \cdot (\nabla U_N - \mathbf{e}_2)$

$$(\mathbf{V}_N \cdot \widehat{(\nabla U_N - \mathbf{e}_2)})_{j,k} = \frac{(p_{j,k}, \mathbf{V}_N \cdot (\nabla U - \mathbf{e}_2))}{\|p_{j,k}\|^2}, \qquad j = 0,\ldots,N, \quad k = 1,\ldots,N-1. \tag{26}$$

Observe that equations (24) and (26) are (19) and (20), respectively, but with u, $\hat{u}_{j,k}$ and $\mathbf{v}$ replaced by U_N, $\hat{U}_{j,k}$ and $\mathbf{V}_N$, respectively. Thus one may be led to think that $\hat{u}_{j,k} = \hat{U}_{j,k}$, for $j = 0,\ldots,N$ and $k = 1,\ldots,N-1$, or, in othewords, that U_N is the truncation of the Fourier expansion to u. This is not true in general. The reason is that the Fourier coefficients of the nonlinear terms differ in general

$$(\mathbf{v} \cdot \widehat{(\nabla u - \mathbf{e}_2)})_{j,k} \neq (\mathbf{V}_N \cdot \widehat{(\nabla U_N - \mathbf{e}_2)})_{j,k}, \qquad j = 0,\ldots,N, \quad k = 1,\ldots,N-1.$$

Notice that whereas the coefficients on the left hand side above depend on an infinite number of Fourier modes, those on the right hand side depend on just $N^2 - 1$ coefficients, so that, they are not equal in general.

If we denote by

$$\mathcal{S}_N = \text{span}(p_{j,k}, \quad j = 0, \ldots, N, \quad k = 1, \ldots, N-1),$$

and by P_N the orthogonal projection of $L^2(\Omega)$ onto $\mathcal{S}_N$ the equations (24) can be written as

$$\frac{d}{dt}U_N - \Delta U - \sqrt{\mu}P_N\big(\mathbf{V}_N \cdot (\nabla U_N - \mathbf{e}_2)\big) = 0, \tag{27}$$

and that, in view of (25), the velocity V_N is the solution of

$$\begin{aligned} -\nabla P_N - \mathbf{V}_N + \sqrt{\mu}U_N\mathbf{e}_2 = 0, \qquad & \text{in } \Omega, \\ \mathbf{V}_N \cdot n = 0, \qquad & \text{on } \partial\Omega, \end{aligned}$$

where P_n is some function of the form

$$P_N(x, y) = \sum_{j,k=0}^{N} \hat{P}_{j,k} \cos(\pi j x) \cos(\pi k y).$$

The discretization here presented is an example of a Fourier spectral discretization. More details and further information can be found for example in [2], [10]

2.3 Computation of Fourier coefficients of nonlinear terms

We address in these section how to compute the Fourier coefficients

$$\hat{C}_{j,k} = (\mathbf{V}_N \cdot \widehat{(\nabla U_N} - \mathbf{e}_2))_{j,k}, \qquad j = 0, \ldots, N+1, \quad k = 1, \ldots, N-1.$$

It will be down in a very easy and simple manner using collocation and discrete sine and cosine transforms.

In the sequel, given a function

$$g = \sum_{j=0}^{N} \sum_{k=1}^{N-1} \hat{f}_{j,k} p_{j,k} \in \mathcal{S}_N$$

we will denote by $\underline{\hat{f}}$ the vector of its Fourier coefficients

$$\underline{\hat{f}} = \begin{bmatrix} \hat{f}_{0,1} \\ \hat{f}_{1,1} \\ \vdots \\ \hat{f}_{N,N-1} \\ \hat{f}_{N+1,N-1} \end{bmatrix}.$$

Also, for a postive integer N we consider the grid in Ω given by

$$\mathcal{G}_N = \{(x_l, y_m) \mid 0 \le l, m \le N\},$$

where

$$x_l = \frac{l}{N}, \quad y_m = \frac{m}{N}, \qquad 0 \le l, m \le N.$$

Given a (continuous) function w defined on Ωwe denote by $\underline{w}$ the vector in $\mathbf{R}^{(N+1)^2}$ of its restriction to the grid $\mathcal{G}_N$ that is

$$\underline{w} = \begin{bmatrix} w_{0,0} \\ w_{1,0} \\ \vdots \\ w_{N,N+1} \\ w_{N+1,N+1} \end{bmatrix},$$

were, for simplicity we denote

$$w_{l,m} = w(x_l, y_m), \qquad 0 \le l, m \le N.$$

For our computations we will use the grid restrictions of the eigenfunctions $p_{j,k}$. We start by noticing an important difference between the eigenfunctions and their grid restrictions:

$$p_{j,k} \neq 0 \quad \text{for } k \neq 0, \qquad \text{but} \qquad \underline{p}_{j,k} \neq 0, \qquad \text{for } k \neq 0, \pm N, \pm 2N, \dots. \tag{28}$$

We comment on two important properties that the grid restrictions of the eigenfunctions $p_{j,k}$ have.

i) *Orthogonality* In a similar way to the the eigenfunctions $p_{j,k}$ being orthogonal to one another, so it is the case of its grid restrictions

$$(\underline{p}_{j,k}, \underline{p}_{r,s})_N = 0, \qquad \text{if } (j,k) \neq (r,s) \bmod(N)$$

where $(\cdot,\cdot)_N$ is the inner product given by

$$(\underline{f}, \underline{g}) = \sum_{l=0}^{N}{}'' \sum_{m=0}^{N}{}'' f_{l,m} g_{l,m}$$

where the two commas mean that the first and last term are halved, that is

$$\sum_{l=0}^{N}{}'' w_j = \frac{1}{3} w_0 + \sum_{j=0}^{N-1} w_j + \frac{1}{2} w_N.$$

This property is the a consequence of a similar property for the Fourier modes

$$\phi_{j,k} = \exp(2\pi i(jx + ky)), \qquad j,k = 0, \pm 1 \pm 2, \dots, \tag{29}$$

for wich

$$\sum_{l,m=0}^{N-1} \phi_{j,k}(x_l, y_m) \phi_{r,s}(x_l, y_m) = 0, \qquad \text{if } (j,k) \neq (r,s) \bmod(N),$$

being the cosines and sines the real and imaginary parts of the Fourier modes.

ii) *Aliasing* Being the restrictions $\underline{p}_{j,k} \in \mathbb{R}^{(N+1)^2}$ orthogonal to one another, and being $\mathbb{R}^{(N+1)^2}$ a space of finite dimension equal to $(N+1)^2$ they cannot be different. In fact we have

$$\underline{p}_{j+nN,k} = \begin{cases} \underline{p}_{j,k}, & n \text{ even}, \\ \underline{p}_{N-j,k}, & n \text{ odd}, \end{cases} \qquad \underline{p}_{j,k+nN} = \begin{cases} \underline{p}_{j,k}, & n \text{ even}, \\ -\underline{p}_{j,N-k}, & n \text{ odd}. \end{cases}$$

(don't panic, we will not be bothered by this). This property is the a consequence of a similar property for the Fourier modes $\phi_{j,k}$, for which

$$\underline{\phi}_{j,k} = \underline{\phi}_{r,s}, \qquad \text{if } (j,k) \neq (r,s) \bmod(N).$$

As a consequence, if a function g is a linear combination of the first $N^2 - 1$ modes

$$g = \sum_{j=0}^{N} \sum_{k=1}^{N-1} \hat{g}_{j,k},$$

we can find its coefficients by simple inner products

$$\hat{g}_{j,k} = \frac{(\underline{p}_{j,k}, \underline{g})_N}{\|\underline{p}_{j,k}\|_N^2}, \tag{30}$$

where $\|\cdot\|_N$ denotes the norm associated to the inner product $(\cdot,\cdot)_N$. Let us mention that an easy calculation shows that

$$\|\underline{p}_{j,k}\|_N^2 = \begin{cases} N^2/2, & j = 0, N, \\ N^2/4, & j \neq 0, N, \end{cases} \qquad k = 1, \ldots, N-1.$$

Formula (30) can be used to compute the Fourier coefficients of a function in the space $\mathcal{S}_N$. Unfortunately, we have that although $U_N \in \mathcal{S}$, in general, $\mathbf{V}_N(\nabla U_N - \mathbf{e}_2) \notin \mathcal{S}_N$. Hoever, from the expression of $\mathbf{V}^N$ in (25), and taking into account that

$$\nabla U_N = \pi \sum_{j=0}^{N} \sum_{j=0}^{N} \hat{U}_{j,k} \begin{bmatrix} -j \sin(\pi j x) \sin(\pi k y) \\ k \cos(\pi j x) \cos(\pi k y) \end{bmatrix}$$

we deduce that $\mathbf{V}_N(\nabla U_N - \mathbf{e}_2)$ will be a linear combination of functions of the form

$$\sin(\pi j x) \sin(\pi l x) \cos(\pi k y) \sin(\pi m y), \qquad j, l = 0, 1 \ldots, N, \quad k, m = 1, \ldots, N-1.$$

and

$$\cos(\pi j x) \cos(\pi l x) \sin(\pi k y) \cos(\pi m y), \qquad j, l = 0, 1 \ldots, N, \quad k, m = 1, \ldots, N-1.$$

But recalling the trigonometric formulas

$$\begin{aligned} \sin(\alpha)\sin(\beta) &= \frac{1}{2}\big(\cos(\alpha-\beta) - \cos(\alpha+\beta)\big), \\ \cos(\alpha)\cos(\beta) &= \frac{1}{2}\big(\cos(\alpha-\beta) + \cos(\alpha+\beta)\big), \\ \sin(\alpha)\cos(\beta) &= \frac{1}{2}\big(\sin(\alpha+\beta) + \sin(\alpha-\beta)\big), \\ \cos(\alpha)\sin(\beta) &= \frac{1}{2}\big(\sin(\alpha+\beta) - \sin(\alpha-\beta)\big), \end{aligned}$$

we conclude that

$$U_N \in \mathcal{S}_N \quad \Rightarrow \quad \mathbf{V}_N(\nabla U_N - \mathbf{e}_2) \in \mathcal{S}_{2N}$$

which allow us to obtain the Fourier coefficients of $\mathbf{V}_N \cdot (\nabla U_N - \mathbf{e}_2)$ by using formula (30) if we evaluate the functions $p_{g,k}$, $\mathbf{V}_N$ and U_N in the grid $\mathcal{G}_{2N}$. Thus a simple procedure (because it involves standard operations) to obtain the Fourier coefficients of $\mathbf{V}_N \cdot (\nabla U_N - \mathbf{e}_2)$ is as follows.

Step 1 Obtain the Fourier coefficients

$$-j\pi\hat{U}_{j,k}, \quad \text{and} \quad k\pi\hat{U}_{j,k}, \qquad j = 0, \ldots, N, \qquad k = 1, \ldots, N-1$$

of $\partial_x U_N$ and $\partial_y U_N$, respectively.

Step 2 Obtain the vectors of nodal values $\underline{\partial_x U}_N$ and $\underline{\partial_y U}_N$ on the grid $\mathcal{G}_{2N}$

Step 3 Obtain the Fourier coefficients

$$\sqrt{\mu}\frac{-jk}{j^2+k^2}\hat{U}_{j,k}, \quad \text{and} \quad \sqrt{\mu}\frac{j^2}{j^2+k^2}\hat{U}_{j,k}, \qquad j = 0, \ldots, N, \qquad k = 1, \ldots, N-1$$

of the two components $V_N^{(x)}$ and $V_N^{(y)}$, respectively, of the velocity $\mathbf{V}_N$.

Step 4 Obtain the vectors of nodal values $\underline{V^{(x)}}_N$ and $\underline{V^{(y)}}_N$ on the grid $\mathcal{G}_{2N}$.

Step 5 Obtain the vector of nodal values $\underline{\mathbf{V}_N \cdot (\nabla U_N - \mathbf{e}_2)}$ on the grid $\mathcal{G}_{2N}$, by multiplying and summing componentwise the vectors obtained in steps 2 and 4.

Step 6 Use formula (30) to obtain the $4N^2 - 1$ Fourier coefficients of $\underline{\mathbf{V}_N \cdot (\nabla U_N - \mathbf{e}_2)}$, and discard those corresponding to $j > N$ or $k \geq N$.

Le us mention that Steps 2 and 4 are standard numerical procedures wich can be carried out by the inverses of the discrete sine and cosine transforms, as well as step 5, which can be carried out by discrete sine and cosine transforms.

This transforms and their inverses are usually implemented in many software packages by means of the *fast cosine and sine transform*, which are easily computed by means of the *fast Fourier transform* (FFT) algorithm. For example in MATLAB and OCTAVE, a variant of the FFT algorithm and its inverse is available in the commands `fft` and `ifft`.

The FFT algorithm allows us to obtain the $N^2 - 1$ Fourier coefficients in (30) (or the $4N^2 - 1$ coefficients in Step 6) in a number of flops proportional to $N^2 \log(N)$ (resp. $4N^2 \log(N)$) instead of N^4 (resp. $16N^4$) flops that would require computing the coefficients using inner products.

At the price of allowing an (usually negligible) error equal of size equal to $|\hat{u}_{N,k}|^2$ on the coefficients $\hat{u}_{N,k}$, $k = 1, \ldots, N-1$ the value $2N$ can be replaced by $3N/2$ (withe the corresponding saving in computational cost), in what is known as *the 3/2 rule* (see e. g. [2].

References

[1] D. Bigoni, *Nolinear Solid Mechanics. Bifurcation Theory and Material Instability*, Cambridge University Press, Cambridge, 2012.

[2] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral methods. Fundamentals in single domains*, Springer-Verlag, Berlin, 2006.

[3] N. Cousin-Rittemard and I. Gruais, *On the connection of isolated branches of a bifur- cation diagram: the truss arch system*, Dynamical System : an international journal, 24 (2009) 315–341. (Also available at https://hal.archives-ouvertes.fr/hal-00833032).

[4] B. García-Archilla, J Sánchez and C. Sim, *Krylov methods and determinants for detecting bifurcations in one parameter dependent partial differential equations*, BIT 46 (2006), no. 4, 731757.

[5] M. Golubitsky and D. G. Schaeffer, *Singularities and Groups in Bifurcation Theory. Volume I*, Springer-Verlag, New-York, 1985.

[6] M. D. Graham, R. H. Steen and E. S. Titi, *Computational efficiency and approximate Iinertial manifolds for a Bénard Convection System*, J. Nonlinear Sci., 3 (1993), 153–167

[7] P.H. Steen, *Pattern selection for finite-amplitude convection states in boxes of porous media*, J. Fluid Mech., 136 (1983), pp. 219–241.

[8] P.H. Steen, *Container geometry and the transition to unsteady Bénard convection in porous media*, Phys. Fluids, 29 (1986), pp. 925–933.

[9] R. S. Riley and K. H. Winters, *Modal exchange mechanism in Lapwood convection*, J. Fluid. Mech., 204 (1989), pp. 325–358.

[10] L. N. Trefethen, *Spectral methods in MATLAB*, SIAM, Philadelphia, 2000.