COMPUTATIONAL AND APPLIED PHYSICS PROGRAM

Thermal convection in rotating spherical shells

Ferran Garcia Gonzalez

Thesis report presented to obtain the Doctor Degree of the Universitat Politècnica de Catalunya under the supervision of Dr. Juan Sánchez Umbría and Dr. Marta Net Marcé.

Barcelona, 2012

Abstract

The study of thermal convection in rotating spherical geometry is fundamental to explain many geophysical and astrophysical phenomena such as the generation of the magnetic fields, or the differential rotation observed in the atmospheres of the major planets. The difficulties associated with the experimental studies enhance the importance of the three-dimensional numerical simulations, such as those presented in this dissertation.

In order to obtain the evolution equations, the Boussinesq approximation is applied to the mass, momentum and energy conservation equations, which are rewritten in terms of toroidal and poloidal potentials. Together with the temperature field, they are expanded in spherical harmonics over the sphere, and in the radial direction a collocation method is used. Semi-implicit schemes, based in backward differentiation formulae (IMEX-BDF), implemented with order and time step control (VSVO), are used for time integration.

Applying non-slip boundary conditions with internal heating and very low Prandtl numbers (ratio between the thermal diffusive and the viscous time scales), one of the first exhaustive analysis of the linear stability of the conductive state has been performed. In addition, the existence of preferred polar antisymmetric modes at the onset of convection for high rotation rates has been described.

A study of the efficiency of different high order time integration schemes, either with fixed time-step or VSVO, has been carried out. In our own time evolution codes we apply the IMEX-BDF formulae with an explicit treatment of the nonlinear terms of the equations. The use of *matrix-free* methods allows the implicit treatment of the Coriolis term, and makes the implementation of a step and order control easier. The results show that the use of high order methods, especially those with time-step and order control, increase the efficiency of the time integration, and allows to obtain more accurate solutions.

Finally, at low Prandtl number, and with non-slip boundary conditions, the nonlinear dynamics is deeply explored by means of temporal evolutions. The type of solutions is described, and the nonlinear mean flow properties are studied. Using parameters as close as possible to those of the Earth's outer core, the numerical simulations are compared with laboratory experiments and realistic measurements.

Acknowledgements

Soria, I learned a lot about Thermodynamics and specially Numerical Linear Algebra used to solve Fluid Dynamics problems. Manel Soria was who introduced to me, for the first time, the Krylov methods to solve linear systems.

Finally I would like to dedicate this thesis to my family, my father and my mother who taught me the importance of enjoying while learning anything, and have devoted their lives to my education. Without them this work, and many more things couldn't be possible. Also to my sister, who is an example for me, since we were children we were surrounded by many books, and in our free time we were always in contact with the nature, in my grandfather farm's or staying in camping along the Spanish countryside. I am absolutely sure that this is the reason for my curiosity in physical sciences.

Finally, I dedicate this work also to Miriam, my couple, who has patiently withstood the many times I was 'locked' within my work, and helped me by listening and suggesting continuously. She has made our free time unforgettable, and she has cheered me up whenever I was low down.

Contents

Chapter 1

Introduction

The study of thermal convection in rotating spherical geometries is important for understanding the dynamics of several astrophysical and geophysical phenomena. For the latter, in the case of the fluid interiors of the giant planets, the competition between rotational and buoyancy effects determine the nature of convection, which is the main responsible of transporting the planet's internal heat to its surface. In the case of Jupiter, the Galileo probe gives support to the idea that large-scale zonal winds are part of a deep global circulation [9], which extends far away the shallow layer heated by solar insulation. When the rotation controls the dynamics, the convection tends to occur in the form of quasi-geostrophic columns aligned in the direction of the rotation axis. The Reynolds stresses produced by the columnar convection can drive zonal flows that can take on a banded character similar to that observed at the cloud level on Jupiter [29]. Unlike this, when the buoyancy forces dominate over the Coriolis forces, the rotational constraint can be broken by the non-linear advection and the convection is then fully three-dimensional and strongly chaotic. This would be the case of the convection in the interiors of planets much larger and massive than Jupiter [102].

On the other hand, the idea that magnetic fields of cosmic bodies are generated by convection of electrically conducting fluids in their interiors, is widely accepted, specially since Riga and Karlsruhe experiments [37, 38, 78]. For the Earth's outer core, geodynamo models have been recently constructed by Glatzmaier and Roberts, among others [46, 65, 51, 50]. The authors reproduced by means of numerical simulations the intensity of the geomagnetic field, explained its dipolar structure with a dipole axis nearly aligned with the Earth's rotation axis, and why the field occasionally undergoes dipole reversals. However, in order to stabilise the numerical solutions, these models make use of hyperviscosity which introduces an offsetting of the rotational constraint altering the basic dynamics of the geodynamo [123]. In the Earth's liquid core there are mainly two sources of buoyancy which could maintain convection, namely, temperature and compositional gradients. These gradients are generated during the crystallisation of heavy elements like iron alloys in the inner-outer core boundary. The volatile elements such as sulfur or oxy-

gen diminish the density in this boundary, and the latent heat is transfered to the outer core. From the Earth's age, [74] have estimated that the compositional buoyancy forces supplies nearly 80% of the necessary power to maintain convection, while the other 20% is supplied by the temperature gradients.

The difficulty of addressing these problems, even without consider the generation of magnetic field, is well known. They are three-dimensional problems whose discretization leads to systems of very large dimension $O(10^6)$ even for moderate ranges of parameters, and therefore far away from the estimates corresponding to those of the cosmic bodies. That is, from $10^{15} - 10^{35}$ for the dimensionless Taylor number Ta defined as the square of the rotation in viscous time units, and the Rayleigh number Ra, the ratio between the change in gravitational potential energy and dissipated viscous and thermal energy, around the previous range. Other parameters are more attainable but difficult to determine with precision, for instance the Prandtl number, the ratio between the thermal and viscous diffusive time scales, $\sigma = 10^{-2} - 1$, or the Lewis number, the ratio between the thermal and the compositional diffusive time scales, $\tau = 10^{-3} - 10^{-1}$. Moreover, for the real problems, some of these dimensionless numbers are difficult to estimate.

Apart of the above mentioned difficulties, in a real planet or satellite, the liquid part of the core may not rotate uniformly with a constant angular velocity, the container of the core fluid, the mantle, may not be perfectly spherical with a uniform boundary [60], the core fluid may not be incompressible, and convective motions may be driven by a combination of different buoyancies [53]. However, it is widely believed that these details are of secondary importance, that is, to a first approximation, they do not influence the dynamics. Accordingly, in this work we will use an spherical homogeneous fluid employed by a majority of authors, and then the secondary aspects of the problem are not considered.

The possible interrelationships between convection, rotation, and the generation of magnetic fields by the dynamo process motivated Chandrasekhar's study [24]. There it was shown that an imposed magnetic field can release the constraint imposed by the rotation on the convection. He solved the linear stability problem and determined Ra as a function of σ and Ta.

In the late 1960s, Roberts and Busse [94, 14] started to find the asymptotic dependence of the onset of convection in fast rotating self-gravity fluid spheres in the limits $E \ll 1$ and $\sigma/E \gg 1$, $(E^{-1} = Ta^{1/2})$. From this preliminary work, it was assumed that, at very high Taylor numbers, the flow is strongly nonaxisymmetric and the dynamics must be dominated by rotation and, consequently, the Taylor-Proudman theorem holds in any geometry. That is, the dynamics of the fluid is essentially independent of the coordinate in the direction of the axis of rotation. They also found that the columnar rolls are highly localised in a neighbourhood of a critical cylinder.

The first experimental results at moderate and high Prandtl number and high rotation rates confirmed that the dominant convective modes are columns travelling in the azimuthal direction, see [16, 17]. In these experiments, the convection is forced by ex-

ternally applying a radial heating and the centrifugal force emulates the gravitational field. A subsequent experiment in the Spacelab laboratory [56], in absence of vertical gravity, confirmed these results. In the Spacelab, the gravity was simulated by polarising a dielectric fluid with a very intense electric field.

The asymptotic theory of the onset of convection at high Ta has been adjusted several times. Weakly nonlinear calculations [108] indicated that the theory of Roberts and Busse underestimated the critical Rayleigh number Ra_c, although the stability analysis and three-dimensional simulations in spherical geometry of K. Zhang [118, 120] at moderate Prandtl numbers, confirmed the power law for the onset of convection given by the theory. Only in 2004, E. Dormy *et al.* have developed the definitive asymptotic theory of the onset of columnar convection at moderate Prandtl number and with internal ($Ra = Ra_i$) or differential ($Ra = Ra_e$) heating. However, with internal heating and very low Prandtl number, K. Zhang in [121] showed that the dominant modes are cells confined in the equator at the outer boundary. These structures were not observed in the laboratory because of the difficulty of visualising low Prandtl number fluids.

In the late eighties, the experimental evidence of the approximate validity of the Proudman-Taylor theorem, led researchers to use a quasi-geostrophic model of convection in a rotating annulus with radial gravity and inclined lids [18, 101, 82, 59, 86, 87, 42] to study the differential rotation and zonal mean flows observed by the Voyager spacecraft in the major planets, or to study how these mean flows and the mean heat transport scale with the parameters. Using the quasi-geostrophic approximation, the analysis can be reduced from three to two spatial dimensions. Many features of the spherical case can be found in the cylindrical case because of the strong tendency towards two-dimensionality induced by the Coriolis force.

Fully three-dimensional calculations always limit the range of parameters that can be reached [3, 114, 106]. For this reason, although the first attempts date back to the seventies, the most important advances have been achieved over the last twenty years with the increasing power of computers. Nowadays, the impossibility to solve successfully the problem without limitations opens two additional ways of study. The first is to use two-dimensional models (quasi-geostrophic) [7, 77, 88, 89, 111, 42] and the second calculate three-dimensional solutions with enforced symmetries and/or fixed parity to achieve simulations with parameter values as high as possible. The existing theoretical models [19, 126, 90] are based on the results of these calculations and also, in laboratory experiments. The closest related to this study are those made with liquid gallium ($\sigma = 0.025$) [6, 41], liquid sodium ($\sigma = 0.01$) [21, 104], water ($\sigma = 7$) [6, 8, 41] and magnetizable fluids to achieve radial force [95]. Some information has been extracted from direct observations. Such is the case of [81], which deals with the existence of a polar vortex, from observations of the variations in the Earth's magnetic field since 1870.

In a problem that involves many parameters (Rayleigh number, Prandtl number, Taylor number and the radius ratio of the spheres), it is difficult to summarise in a few lines the current state of the issue. For good reviews on the subject see for instance [128, 15].

Although DNS of the quasi-geostrophic models can be easily performed providing a lot of information, only the results of three-dimensional simulations, which are directly comparable with physical observations and related with our calculations, are commented.

From [115, 106, 114, 113, 29] among others, with rotation rates giving a Ta in the range $10^6 < Ta < 10^{10}$, and $\sigma \approx 1$, it is known that when inertial waves lose their stability, quasiperiodic flows, which modulate the amplitude of the oscillations or produce spatial modulations, appear. At higher Rayleigh numbers, the flow simulations have chaotic spatio-temporal dependence that lead to localised convection. This occurs when the differential rotation becomes so intense that by shear effects the convection is inhibited in the bulk of the layer. In the fluid regions where there is no convection, the conductive state tends to be restored. The complex time dependencies favour the heat transport by convection.

In particular, with differential heating, the dependence of the time-averaged properties such as the Rossby and the poloidal Rossby numbers, $\overline{Ro} = \overline{Re}E$, $\overline{Ro_p}$, the kinetic energy density and non zonal kinetic energy density ratio, $\overline{K/K^{nz}}$ and the modified Nusselt number $\overline{Nu^*} = \overline{Nu}E\sigma^{-1}$, on the modified Rayleigh number $Ra^* = RaE^2\sigma^{-1}$ is known for $E \to 0$, where the Reynolds number $\overline{Re} = \sqrt{2K}$ remain defined as the square root of the mean quadratic velocity:

$$\overline{Ro} = 0.65(Ra^*)^{1/5}, \qquad\qquad \overline{Ro_p} = 0.54(Ra^*)^{2/5}, \qquad (1.1)$$

$$\overline{K/K_{nz}} = 0.36(Ra^*)^{-2/5}, \qquad\qquad \overline{Nu^*} = 0.077(Ra^*)^{5/9}. \qquad (1.2)$$

At low Prandtl numbers the transition to states of chaotic time dependence occurs near the onset of convection with increasing Rayleigh number. The cost of these simulations means that in this case, most information comes from laboratory experiments and DNS using the quasi-geostrophic approximation. For example the results of [6, 104, 41] have never been contrasted with three-dimensional DNS.

The dissertation is organised in two parts. The first part (Chapters 2 and 3) deals with the onset of convection while the second part (Chapters 4 and 5) comprises the study of the nonlinear flows. The understanding of the onset of convection allows to explore the finite amplitude problem in a more selected way. To achieve these objectives a linear and nonlinear codes of thermal convection in rotating spherical geometry were written and checked with several papers [1, 121, 73, 106, 30].

The first part of our study has been published in [79] (Chapter 2) and [40] (Chapter 3). In Chapter 2, the transition between the spiralling columnar modes (SC) and the outer equatorially attached (OEA) modes in the parameter space (σ, E) using non-slip boundary conditions and internal heating $(Ra = Ra_i)$ for a thick $\eta = 0.2$ spherical shell, is analysed. The study comprises:

- Localisation of the transition between mono and multicellular modes preferred at

very low $\sigma = 0.005$. Visualisation of the patterns of convection of the multicellular modes.

- Clarification of the origin of the transition between prograde and retrograde convective modes for the spiralling columnar modes, preferred at $\sigma = 0.1$ and low E.

- Localisation of the transition OEA/SC depending on the Prandtl number ($\sigma \in [0.005, 0.03]$) at low E. Comparison between the OEA and the SC patterns of convection.

- Linear stability analisys near the transition OEA/SC ($\sigma = 0.025$).

- Asymptotic study in which the potential laws of the critical parameters depending on E are obtained for $\sigma = 0.005, 0.01, 0.025$ corresponding to OEA modes, to SC modes, and to modes near the transition OEA/SC, respectively.

In Chapter 3, the linear stability analysis of flows, which are mainly concentrated near the inner boundary in a moderately thick spherical shell $\eta = 0.4$, is performed. This part contains:

- Study of the equatorially antisymmetric modes preferred at $\sigma = 0.01$ and low E, with both non-slip or stress-free boundary conditions, and internal heating ($Ra = Ra_i$). These modes are located in the polar regions.

- Study of the equatorially symmetric modes preferred at $\sigma = 0.1$ and low E, using non-slip boundary conditions and differential heating ($Ra = Ra_e$). These modes are located in high latitude regions.

- Asymptotic study in which the potential laws of the critical parameters are obtained as a function of E, for the two cases considered before. Aside their differences, similar power laws are obtained.

The second part of the dissertation (Chapters 4 and 5), includes the optimisation of the time integration methods, and the nonlinear study of the dependence of the physical properties and of the flow patterns on the Rayleigh number. The parameters are taken as similar as possible to the Earth's outer core and previous experimental results.

The optimisation of the temporal integration of the nonlinear problem using multistep linear methods, such as backward differentiation and extrapolation formulae (IMEX-BDF) of order $k = 2, .., 6$, is performed in Chapter 4 and has been published in [39]. It comprises:

- Implementation of a time integration code using IMEX-BDF with fixed time step, by considering the diagonal part of the Coriolis term and the rest of linear terms implicit, and the rest of Coriolis terms and nonlinear terms explicit.

10

- Resolution of the obtained block diagonal linear systems using adapted LU subroutines.

- Implementation of a time integration code using IMEX-BDF with fixed time step by considering the linear terms and the upper tridiagonal Coriolis terms (resp. lower) implicitly and the rest of the Coriolis and nonlinear terms explicitly, alternatively at each time step. The resolution of the upper triangular block (resp. lower) linear systems are performed using backward (resp. forward) substitution using the previous LU factorisations.

- Krylov techniques are used to implement a IMEX-BDF fixed time step based code, by considering implicit the Coriolis term. A study of the efficiency of the use of different preconditioners and tolerances is performed in the iterative resolution of the block tridiagonal linear system.

- Use of the Krylov techniques in order to implement the three previous codes with order and step size control.

- Implementation of the necessary subroutines for the use of DLSODPK library in order to use a method based on the fully implicit BDF.

- Comparative study of the efficiency of the previous codes to integrate a travelling wave at moderate and low E. The differences between fixed and variable time step methods are deeply analysed. The influence of the type of solution and of the physical and integration parameters on the temporal integrations is also addressed.

Finally, the nonlinear properties and patterns of the flow are studied in Chapter 5 for $\eta = 0.35$, $\sigma = 0.1$ and $E \leq 10^{-4}$, non-slip boundary conditions and differential heating as a function of the Rayleigh number covering from weak to strong supercritical regimes. The main items of this part are:

- Study of the first bifurcated quasiperiodic solutions from the travelling waves, as well as other families of quasiperiodic solutions at higher Ra.

- Description of time complex solutions with intermittent behaviour at moderate Ra, which are found close to the quasiperiodic regime.

- Study of the changes of the time-averaged properties, and the location of convection for a large sequence of solutions with increasing Ra.

- Obtention of the scaling laws of Eqs. (1.1) and (1.2), with non-slip boundary conditions and low Prandtl number $\sigma = 0.1$.

- Comparison of the numerical scalings, obtained from the three-dimensional DNS, with previous numerical [29] as well as experimental results [6, 104, 41].

11

- Use of the scalings obtained to make predictions of physical properties of the Earth's outer core.

Finally, in Chapter 6 the results exposed along this dissertation are discussed and summarised, including brief comments on the further work.

Chapter 2

On the onset of low Prandtl number convection in rotating spherical shells: non-slip boundary conditions

On the onset of low Prandtl number convection in rotating spherical shells: non-slip boundary conditions

ABSTRACT

Accurate numerical computations of the onset of thermal convection in wide rotating spherical shells are presented. Low Prandtl number (σ) fluids, and non-slip boundary conditions are considered. It is shown that at small Ekman numbers (E), and very low σ values, the well known equatorially-trapped patterns of convection are superseded by multicellular outer-equatorially-attached modes. As a result, the convection spreads to higher latitudes affecting the body of the fluid, and increasing the internal viscous dissipation. Then, from $E < 10^{-5}$, the critical Rayleigh number (R_c) fulfils a power-law dependence $R_c \sim E^{-4/3}$, as happens for moderate and high Prandtl numbers. However, the critical precession frequency ($|\omega_c|$) and the critical azimuthal wavenumber (m_c) increase discontinuously, jumping when there is a change of the radial and latitudinal structure of the preferred eigenfunction. In addition, the transition between spiralling columnar (SC), and outer-equatorially-attached (OEA) modes in the (σ, E) space is studied. The evolution of the instability mechanisms with the parameters prevents that multicellular modes become selected from $\sigma \gtrsim 0.023$. As a result, and in agreement with other authors, the spiralling columnar patterns of convection are already preferred at the Prandtl number of the liquid metals. It is also found that, out of the rapidly rotating limit, the prograde antisymmetric (with respect to the equator) modes of small m_c can be preferred at the onset of the primary instability.

14

2.1   Introduction

The first attempts to find the asymptotic dependence of the non-axisymmetric onset of thermal convection in self-gravitating, and internally heated fluid spheres were of [94], and [14]. In these early papers the scaling laws $R_c \sim E^{-4/3}$, $|\omega_c| \sim E^{-2/3}$, $m_c \sim E^{-1/3}$, the symmetry, and the azimuthal drifting of the marginal columnar modes of convection at large rotation rates were established. These modes were localised inside a cylinder centered at a critical interior radius. Later, the nonlocal perturbation analysis of [108] showed that small disturbances, at values of the Rayleigh number just above R_c given by the Roberts-Busse's theory, decay exponentially as $t \to \infty$. The weak point of their analysis comes from using a local theory that neglects the dependence of the boundary inclination with the distance to the axis of rotation.

The numerical calculations of [130], and [118] revealed that, in rapidly rotating spherical shells, the preferred modes of convection are strongly dependent on the Prandtl and Ekman numbers. At fast rotation rates, and moderate and large Prandtl numbers, they are thermal Rossby waves in the form of spiralling columns, while, at low Prandtl numbers, they are waves attached to the equator (equatorially-trapped waves).

A revised asymptotic theory for the spiralling columnar convection was proposed by [117]. Following the guidelines of [108], he improved the critical parameters for $\sigma > 0.1$ by extending the solution of [94], and [14] onto a complex plane of the radial coordinate of the point where the convection amplitude attains its maximum, but maintaining the limit of small inclination of the boundaries. The projection of the solution onto the real axis gives the inclination of the columns with respect to the radial direction. His calculations agree well with the numerical results. Recently, [66] have determined the leading-order value of R_c, and the first-order correction for the columnar convection in spheres with internal heating, and stress-free boundary conditions. They have generalised the work of [117] by considering the full spherical problem. For the same problem, [19] have found explicit expressions for the dependence of R_c on m_c. On the other hand, [32] have completed the asymptotic theory for the onset of columnar convection in spherical shells. The differential heating problem, for which there are no heat sources in the fluid, is also studied. Then, the critical mode of convection localises around the inner radius because, according to the authors, the conduction temperature gradient ($\propto r^{-2}$) decreases very rapidly by increasing r. In this work the corrections due to Ekman suction are included.

In the nineties, a lot of theoretical and numerical papers were devoted to study the spiralling columnar type of convection, mainly with stress-free boundary conditions, since it is relevant to astrophysical problems. The dependence of R_c, ω_c, and m_c on E and σ, and the influence of the Ekman boundary layers on the onset of convection in spherical shells were studied by [122]. For fluids of $\sigma \geq \mathcal{O}(1)$, their numerical results show that the extra dissipation in the Ekman boundary layers generated near the rigid boundaries destabilises the flow, and also reduces $|\omega_c|$, and m_c. However, for $\sigma < \mathcal{O}(1)$, and moderate E, the thermal dissipation dominates over the viscous dissipation, and the boundary layers

15

stabilise the flow.

Low Prandtl number problems are important in magnetohydrodynamics since they model planetary metallic cores. The importance of equatorial drifts in the Earth's core was stated by [34]. They showed a westward drift of the equatorial patches of the radial magnetic field from observations of the magnetic field flux over 400 years. This phenomenon may be due to either waves or advection by the underlying flow. The last estimations of the thermal and magnetic Prandtl numbers for the Earth's outer core are $\sigma \approx 0.1$ and 10^{-6}, respectively. They are based on the value of the viscosity determined by [116], and on the electrical resistivity determined by [103], at the physical pressure and temperature conditions of the Earth's outer core. The study of thermal convection with liquid metals in these conditions is very difficult. The numerical computations require very high resolutions to approach the extremely low Ekman numbers of the cores, and to solve the thin Ekman boundary layers near the rigid boundaries. For this reason there have been very few attempts to compute the marginal modes in this range of parameters. Frequently, as in [86], [89], and [41], a quasi-geostrophic model is used to approximate the real problem.

According to [119, 120], the equatorially-trapped modes are quasi-inertial waves, i.e., at leading order are solutions of the Poincaré equation. Moreover, [3] found that, at moderate rotation rates and stress-free boundary conditions, the dominant modes remain attached to the outer equator, but can be multicellular before the transition to spiralling columns at low Ekman numbers. They supplied a sketch of this transition in the (σ, E) plane. In the work of [89], valid for $E \lesssim \sigma$, it is seen that in cylindrical annuli with curved lids the critical modes at small σ are also quasi-inertial, and multicellular. In addition, they have demonstrated by a perturbation analysis, that R_c decreases when the number of vortices of the streamfunction in the radial direction increases, because it depends on a balance between dissipation and buoyancy. The increase of power consumed by viscous dissipation is overcome by a much more efficient release of power developed by the buoyancy forces. This effect is due essentially to the fall of the precession frequency of the wave when the number of cells increases.

There are neither accurate experiments on the onset of convection using low Prandtl number fluids. The only experimental study is that of [63], although the apparatus is a cylindrical annulus of conical lids. By using mercury ($\sigma = 0.025$) he measured low-frequency rotating waves for Rayleigh numbers of order E^{-2}. However, it is important to notice that, in the nonlinear regime, [6], and [41] have found experimental retrograde zonal flows (generated by Reynolds stresses) storing most of the kinetic energy of the fluid.

The first study of the asymptotic dependence of R_c and ω_c in the small Ekman number limit, for the equatorially-trapped convection in a self-gravitating fluid sphere, including solutions of the spherical boundary layer, is that of [121]. His perturbation analysis is valid only for very low Prandtl numbers, namely $\sigma << Em^{5/2}$. For small but finite σ, no simple power-law dependence can be derived from this perturbation theory. Recently, [124],

16

and [125] have proposed a new method for the study of the asymptotic solutions in a rapidly rotating sphere, which does not assume any asymptotic scaling for $E << 1$. Then it is valid for $0 \leq \sigma/E < \infty$, i.e., also includes moderately small and high Prandtl numbers. The main contribution of these papers is to unify the inertial-oscillation and the convective instability problems for both stress-free and non-slip boundary conditions. The eigenfunctions are written either as a single or as a superposition of quasi-geostrophic inertial modes, and it is shown that the viscous coupling of these modes causes the spiralling structure of convection. The analytical results are compared with those obtained from a numerical code for a wide spherical shell displaying a good agreement in any considered case.

This paper seeks to help to understand the onset of thermal convection at small Prandtl numbers between rigid spherical boundaries. Specifically, the dependence of R_c, ω_c and m_c on E and σ, and the features of the preferred modes of convection in the parameter space are studied. The small value of the radius ratio $\eta = 0.2$ will be maintained along the paper for three reasons. Up till now, the known preferred modes of convection were spiralling columnar or equatorially attached for any η value, so $\eta = 0.2$ facilitates the calculations because the number of relevant marginal stability curves is small, and they are much more separated than for higher η values. In the second place, it concerns to geophysical applications, because the Earth's outer core is a wide spherical shell, so the study is restricted to $0.005 \leq \sigma \leq 0.1$. Finally, it facilitates the comparison with preceding papers.

In Sec. 2.2 we introduce the formulation of the problem, and the numerical method used to find the leading spectra of the linearised equations. We also include some of the tests made in order to check our new computational code. In Sec. 2.3 the marginal stability curves as a function of E and σ, and the structure of the preferred patterns of convection are analysed. A comparison between these results and previous asymptotic and numerical analysis, in the low Prandtl number limit, is also established. Finally, in Sec. 2.4 the paper ends with some brief comments on the results obtained.

2.2 Mathematical model and Numerical method

A spherical shell rotating about an axis of symmetry with constant angular velocity $\mathbf{\Omega} = \Omega\mathbf{k}$, and radial gravity $\mathbf{g} = -\gamma\mathbf{r}$, where γ is a constant, and $\mathbf{r}$ the position vector, is considered. The gap width is $d \equiv r_o - r_i$, where r_i and r_o are the inner and outer radii. We use the same formulation of the problem as in [106], which allows to study at once internal and differential heating. The non-dimensional units are d for the distance, $\nu^2/\gamma\alpha d^4$ for the temperature, and d^2/ν for the time; ν being the kinematic viscosity, and α the thermal expansion coefficient.

The velocity field is written in terms of toroidal and poloidal potentials

$$\mathbf{v} = \nabla \times (\Psi\mathbf{r}) + \nabla \times \nabla \times (\Phi\mathbf{r}), \tag{2.1}$$

consequently, the linearised equations for both potentials, and the temperature perturbation Θ from the conduction state ($\mathbf{v} = \mathbf{0}, T = T_c(r)$), with $r = ||\mathbf{r}||_2$, are

$$\left[(\partial_t - \nabla^2)L_2 - 2E^{-1}\partial_\varphi\right]\Psi = -2E^{-1}Q\Phi, \tag{2.2}$$

$$\left[(\partial_t - \nabla^2)L_2 - 2E^{-1}\partial_\varphi\right]\nabla^2\Phi + L_2\Theta = 2E^{-1}Q\Psi, \tag{2.3}$$

$$\left(\sigma\partial_t - \nabla^2\right)\Theta - \left(R_i + R_e\eta(1-\eta)^{-2}r^{-3}\right)L_2\Phi = 0. \tag{2.4}$$

The parameters of the problem are the internal R_i, and external R_e, Rayleigh numbers, the Prandtl number σ, the Ekman number E, and the radius ratio η,

$$R_i = \frac{\gamma\alpha q d^6}{3c_p\kappa^2\nu}, \quad R_e = \frac{\gamma\alpha\Delta T d^4}{\kappa\nu}, \quad E = \frac{\nu}{\Omega d^2}, \quad \sigma = \frac{\nu}{\kappa}, \quad \eta = \frac{r_i}{r_o}, \tag{2.5}$$

κ being the thermal diffusivity, c_p the specific heat at constant pressure, q the rate of heat due to internal sources per unit mass, and ΔT the difference of temperature between the inner and outer boundaries due only to differential heating. The operators L_2 and Q are defined by $L_2 \equiv -r^2\nabla^2+\partial_r(r^2\partial_r)$, $Q \equiv r\cos\theta\nabla^2-(L_2+r\partial_r)(\cos\theta\partial_r-r^{-1}\sin\theta\partial_\theta)$, (r,θ,φ) being the spherical coordinates, with θ measuring the colatitude. In non-dimensional units the conduction state reads

$$T_c(r) = T_0 - (R_i/2\sigma)r^2 + (R_e\eta/\sigma(1-\eta)^2)/r. \tag{2.6}$$

Non-slip perfect thermally conducting boundaries

$$\Phi = \partial_r\Phi = \Psi = \Theta = 0 \tag{2.7}$$

are used, but for testing purposes also stress-free $\Phi = \partial_{rr}^2\Phi = \partial_r(\Psi/r) = \Theta = 0$, boundary conditions are implemented.

The system is $\mathcal{SO}(2) \times \mathcal{Z}_2$-equivariant, $\mathcal{SO}(2)$ generated by azimuthal rotations, and $\mathcal{Z}_2$ by reflections with respect to the equatorial plane, i.e., the actions

$$\begin{aligned}\varphi &\to \varphi + \varphi_o, \quad u \to u, \quad \text{and}\\ \theta &\to \pi - \theta, \quad\quad u \to \zeta u,\end{aligned} \tag{2.8}$$

with $u = (\Psi, \Phi, \Theta)$, and $\zeta u = (-\Psi, \Phi, \Theta)$, leave the system invariant. In terms of the velocity field the same holds with $u = (v_r, v_\theta, v_\varphi, \Theta)$, and $\zeta u = (v_r, -v_\theta, v_\varphi, \Theta)$

The equations for $X = (\Phi, \Psi, \Theta)$ are solved by expanding the eigenfunctions in spherical harmonic series up to degree L, namely

$$X(t, r, \theta, \varphi) = \sum_{l=0}^{L}\sum_{m=-l}^{l} X_l^m(r, t)Y_l^m(\theta, \varphi), \tag{2.9}$$

with $Y_l^m(\theta, \varphi) = P_l^m(\cos\theta)e^{im\varphi}$, and P_l^m being the normalised associated Legendre functions of degree l and order m. In the radial direction, a collocation method on a Gauss-Lobatto mesh is employed.

In order to find the critical parameters we use a new computational code, which allows to find the leading eigenvalues with a good resolution by means of an iterative method. The linearised equations (2.2)-(2.4), (2.7) are separated into their azimuthal Fourier coefficients. Let the equations for a given azimuthal mode m be written as

$$\dot{\mathbf{x}}_m = \mathcal{A}_m \mathbf{x}_m, \qquad (2.10)$$

where $\mathbf{x}_m$ are all the values at the collocation mesh of the amplitudes of the variables in spherical harmonics of order m. We compute the eigenvalues of $\mathcal{A}_m$ by evolving the equation (2.10) a time interval t. Its solution with initial condition $\mathbf{x}_m^0$ is $\exp(t\mathcal{A}_m)\mathbf{x}_m^0$. The eigenvalues μ of the linear map $\exp(t\mathcal{A}_m)$ are related to the eigenvalues λ of $\mathcal{A}_m$ by $\mu = \exp(t\lambda)$. This transformation maps eigenvalues of maximum real part of $\mathcal{A}_m$ to multipliers of largest modulus of $\exp(t\mathcal{A}_m)$. To find the latter we employ subspace iteration or Arnoldi algorithms (see [72]). The integration of (2.10) is performed by a multi-step BDF-extrapolation formulae (BDF meaning Backward Differentiation Formulae), with initial conditions obtained by a Runge-Kutta method. It is important to notice that an explicit treatment of the term in Q allows to decouple the left hand side of equations (2.2)-(2.4) for each degree l. The time interval t is selected to be as short as possible to reduce the cost of the evaluation of $\exp(t\mathcal{A}_m)\mathbf{x}_m^0$, but large enough to separate the eigenvalues to make their convergence fast.

The onset of convection breaks the axisymmetry of the conduction state. Then, according to [33], it is a Hopf bifurcation giving rise to a wave travelling in the azimuthal direction. Typically, the critical solutions maintain the $\mathcal{Z}_2$ symmetry with respect to the equatorial plane, i. e., they are symmetric with respect to the equator. This implies that only associated Legendre functions with odd $(m + l)$ for Ψ, and even for Φ and Θ contribute to the linear problem. Then, the number of unknowns can be halved if needed. However, it will be seen that at low rotation rates also antisymmetric modes can be selected at onset. Moreover, from moderate rotation rates, the axisymmetric ($m = 0$) modes of convection are waves travelling from the equator or mid-latitudes to the poles. Although they are non-preferred linearly, nonlinear interactions involving these modes could give rise to waves of this type at higher Rayleigh numbers.

Table 2.1 shows the convergence of the numerical method by increasing the number of radial points N_r, and the maximal degree L of the associated Legendre functions, for $E = 3.16 \times 10^{-6}$. All the results of table 2.1 correspond to a fourth order time integration method. The tolerances required are kept at each run of the code. However, we have checked, by changing the time step and the order of the method, that we do not introduce numerical instabilities that could lead to erroneous critical parameters. In the table, R_{i_c} means the critical internal Rayleigh number. The negative frequencies give positive drifting velocities $c = -\omega_c/m$, i. e., in this case, the waves precess in the prograde direction.

We have also checked our code working with non-slip boundary conditions with [1] for differential heating, with [121] for internal heating, with [73] for a non-rotating shell, also

19

$N_r \times L$	R_{i_c}	$-\omega_c$
40×65	8836856.9	16513.198
40×85	8850865.2	16513.533
50×85	8644084.1	16545.639
50×100	8645078.5	16545.794
60×85	8627974.3	16547.513
60×100	8629021.4	16547.655
70×85	8627670.2	16547.753

Table 2.1: Convergence of the iterative method for $\eta = 0.2$, $\sigma = 0.005$, $E = 3.16 \times 10^{-6}$, $m = 5$ and $R_e = 0$, with a fourth order time integrator.

with internal heating, and finally, with [106] for internal heating and stress-free boundary conditions. The satisfactory results of the first comparison are plotted in figure 2.1, for various η values. The critical Rayleigh number $(R_{e_c})_A$ used in [1] is related with the ours by $(R_{e_c})_A = R_{e_c}/(1-\eta)$. The full circles and crosses correspond to their and our numerical results, respectively, and the solid line is their fitting to the solutions. Our results also agree very well with those presented in the first figure of [73] with tiny differences less than 0.06% for the two η of that figure. We do not know exactly the resolution employed in [106], but, the differences are less than 1.3% in R_{i_c}, and 0.2% in ω_c for $\sigma = 0.025$. However, we have found discrepancies greater then 13% in R_{i_c} with [121], which will be explained below.

From now on only $\eta = 0.2$, and internal heating is considered, so $R_e = 0$, and $R = R_i$ is defined.

2.3 Numerical results

2.3.1 Dependence of very low Prandtl number flows on the Ekman number

First of all, the stability of the conduction state is studied for $\sigma = 0.005$, the very low Prandtl number case computed in [121]. A mesh $N_r \times L = 50 \times 85$ is used, because it guarantees relative errors below 0.2% in the critical Rayleigh numbers at least up to $E = 3.16 \times 10^{-6}$ (see table 2.1), and minimises the computing time. Figure 2.2(a-c) shows the critical Rayleigh number R_c^m, and the critical precession frequency $|\omega_c^m|$ versus E, for each neutral stability curve of azimuthal wavenumber m. In order to facilitate the comparison with the mentioned author, and to display well the change of modes, E and R_c^m are rescaled in the figure in accordance with Zhang (Z) units. The relations are $E_Z = E(1 - \eta)^2$ and $(R^m)_Z = R^m E/(1 - \eta)^4$.
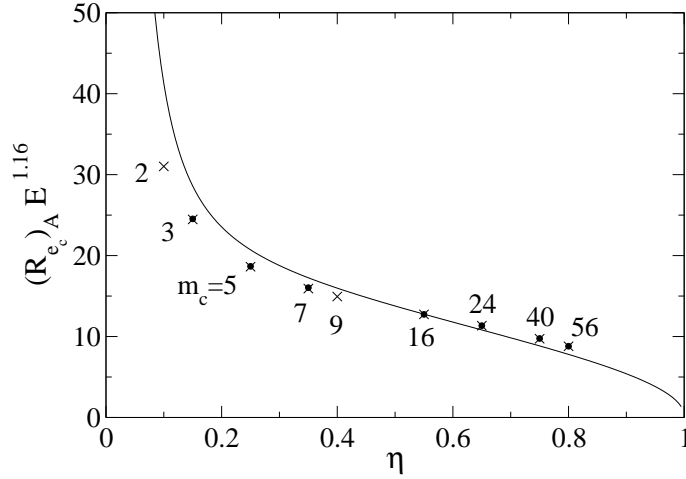
Figure 2.1: Critical Rayleigh number $(R_{e_c})_A E^{1.16}$ versus the radius ratio η for $\sigma = 1$, and $E = 10^{-4}$. The integer numbers denote the critical azimuthal wavenumber, m_c, of the preferred modes of convection for each η computed.

The cusps in each curve of figure 2.2(a) and its enlargement (c) indicate crossings between curves of the same wavenumber m. In figure 2.2(a), for $E_Z > 3.6 \times 10^{-4}$ the preferred $m = 1$ patterns of convection are symmetric retrograde ($\omega_c > 0$) body modes, in the sense that they fill the shell up to the polar regions. Then, they are superseded by prograde waves ($\omega_c^m < 0$) of $m = 2, 3, 4$, whose shape is weakly influenced by rotation. When the Ekman number is decreased, they become more and more confined to the equatorial region, and at $E_Z \lesssim 3 \times 10^{-5}$ these critical eigenfunctions can be clearly recognised as the equatorially-trapped modes (see figure 2.3) described for the first time in [130]. From this point, the crossings of curves of same m produce decreases of m_c, which lead to critical eigenfunctions with the same m_c, and relative extrema inside the arms of the spirals (see figure 2.4). The components v_r and v_φ have extrema on the equatorial plane, and v_θ and v_φ on spherical projections very close to the outer boundary. So, these eigenfunctions display multicellular patterns of convection. From now on, we will refer to their shape on the equatorial plane as *radial structure*, that on spherical projections as *latitudinal structure*, and the local vortices on these projections as *latitudinal cells*. This fact is displayed in figure 2.2(c), where modes with $m = 5, 6, 7$ dominate successively twice in a short range of E_Z. In any case, the main contribution to the latitudinal structure of the dominant modes corresponds to the associated Legendre functions with $l = m_c$ for Φ and Θ, and $l = m_c + 1$ for Ψ. The only difference seems to be the relative weight of the associated Legendre functions of higher degree.

Figure 2.2(b) shows $|\omega_c|$ in a logarithmic scale versus E_Z. The jumps also indicate the changes between radial and latitudinal patterns of convection, which can maintain

21

(a)

100

$(R_c^m)_Z$

75

50

25

m=12
11
10
9
8
7
6
5
4
3
0
2
1

10^{-6} 10^{-5} 10^{-4} 10^{-3}

E_Z

(b)

10^5

$|\omega_c^m|$

10^4

10^3

10^2

1
0
2
3
4
5

10^{-6} 10^{-5} 10^{-4} 10^{-3}

E_Z

(c)

70

$(R_c^m)_Z$

30

m=7
6   5
7
6
5
4   4

2×10^{-6} $E_Z 1\times10^{-5}$ 3×10^{-5}

(d)

100

$(R^m)_Z$

20

7

5

m=3

10^{-5} E_Z 10^{-4}
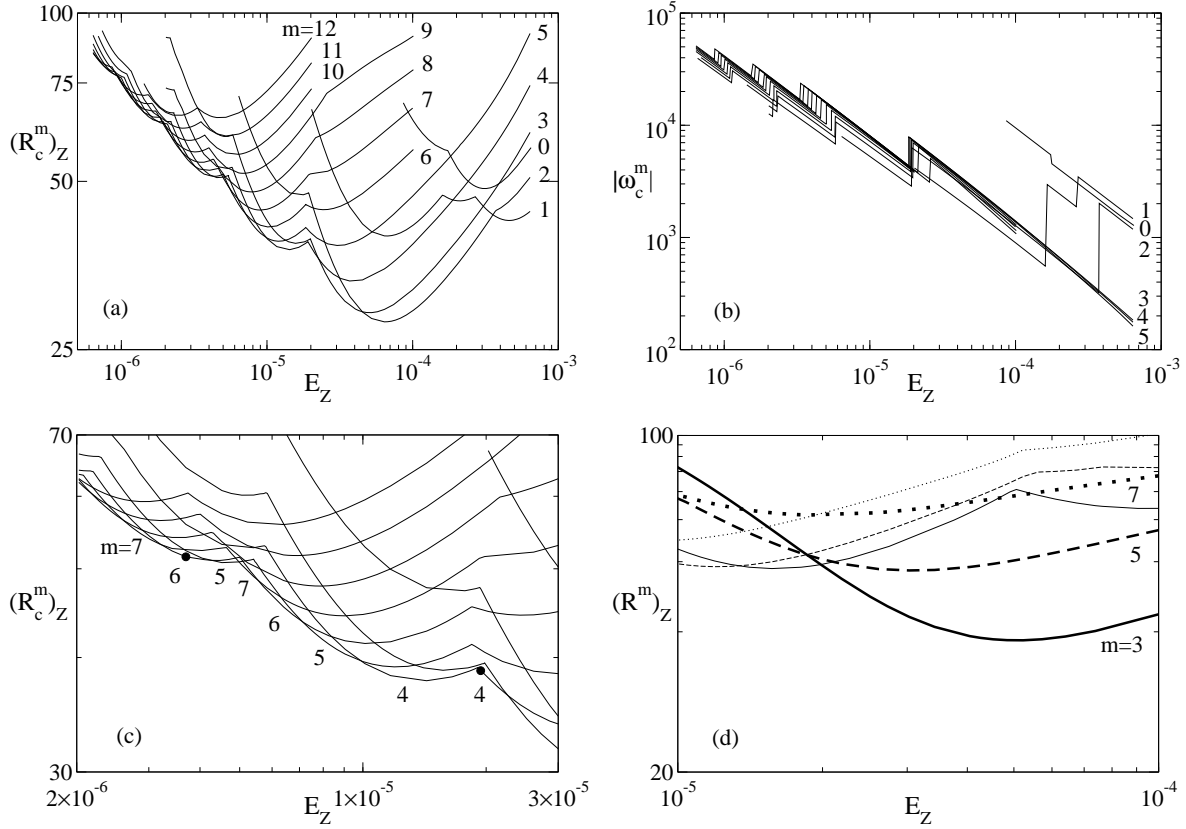
Figure 2.2: (a) The critical Rayleigh number $(R_c^m)_Z$, and (b) the critical precession frequency $|\omega_c^m|$, of each critical mode with azimuthal wavenumber m, plotted versus the Ekman number E_Z for $\sigma = 0.005$. (c) Detail of (a) showing pairs of different preferred modes with same wavenumber m. The full circles indicate the parameters of figures 2.3-2.4. (d) The two lowest marginal stability curves for $m = 3$ (thick and thin solid lines), $m = 5$ (thick and thin dashed lines), and $m = 7$ (thick and thin dotted lines) azimuthal modes. The $m = 3, 5, 7$ thick lines were also calculated before by [121].

the azimuthal wavenumber. When E decreases, the waves drift slower after every jump. Moreover, under the influence of rotation, the vortices tend to spiral eastward, and to split forming cells attached to the outer boundary of the fluid. From now on, we will employ the term outer-equatorially-attached (OEA) modes to describe both the equatorially-trapped patterns (mono-cellular) and the multicellular equatorially-attached patterns, to distinguish them from the well known spiralling columnar (SC) modes preferred at moderate and large Prandtl numbers.

In the work of [117] the localised columnar convection is explained from the dispersion relation of the wave. There, it is seen that the existence of a columnar mode requires a strong buoyancy force counterbalanced by the kinematic diffusivity, and then the thermal diffusivity acts to isolate the columns. However, without an effective kinematic diffusivity ($\sigma < \mathcal{O}(1)$), the pattern becomes a spiralling column. We have checked that when the multicellular modes become selected the viscous dissipation in the body of the fluid increases, but the buoyancy force decreases significantly, so for these solutions the preceding balance is not fulfilled. Nevertheless, our result agree with the conclusions of [125]. They demonstrate that the tendency of the multicellular modes to spiralling is just due to the viscous coupling of different quasi-geostrophic inertial waves.

Figure 2.2(d) shows the two lowest marginal stability curves for $m = 3, 5, 7$, in the range of E_Z of figure 2.1(b) of [121]. The figure 2.1(b) illustrated the comparison between the $m = 3, 5, 7$ curves found in his numerical analysis for $\eta = 0.2$ with those obtained by a perturbation method, but for a full sphere. Our thick marginal curves agree very well with those found in Zhang's paper, but they do not give R_c^m in the full interval. As can be seen in figure 2.2(d), from $E_Z < 1.98 \times 10^{-5}$ they are superseded by the thin lines (not found before), which correspond to multicellular modes. Namely, in the spherical shell the thick curves, which correspond to equatorially-trapped modes of one cell, only give the correct R_c^m down to $E_Z = 1.98 \times 10^{-5}$. From this point multicellular OEA modes become preferred. This explains the important differences we found in R_c^m when comparing both results at high rotation rates in order to check our code.

Figures 2.3, and 2.4 show the contour plots of two preferred eigenfunctions. The first is an $m = 4$ equatorially-trapped mode preferred at moderate E_Z, and the second an $m = 6$ mode, selected as sample of the multicellular patterns of convection preferred at small Ekman numbers. Three projections of the components of the velocity field (v_r, v_θ, v_φ), and the kinetic energy density are displayed in each figure (see figure captions). The contour plots of the temperature perturbation resemble very much those of the radial velocity, and they are not plotted. The (a) projections are taken on spheres of $r = r_i + 0.70d$ and $r = r_i + 0.47d$ for v_r, and $r = r_i + 0.98d$ for the rest, although all of them are plotted of the same size. In this way all the projections are taken nearly cutting the extrema of every scalar field. The projections on spherical surfaces of smaller radius loose some of the cells extending only to low latitudes. All (b) projections are taken on the equatorial plane, and all (c) meridional sections cut more or less by the centre of a v_r vortex. The scale of greys is the same for each row of contour plots in each figure. White means

Figure 2.3: Preferred equatorially-trapped mode of convection with azimuthal wavenumber $m = 4$, plotted at $E_Z = 1.92 \times 10^{-5}$ for $\sigma = 0.005$. In (a-c) contour plots of the radial velocity field on a sphere, on the equatorial plane, and on a meridional section, respectively. In (d-f) idem (a-c) for the colatitudinal velocity. In (g-i) idem (a-c) for the azimuthal velocity. In (j-l) idem (a-c) for the kinetic energy density.
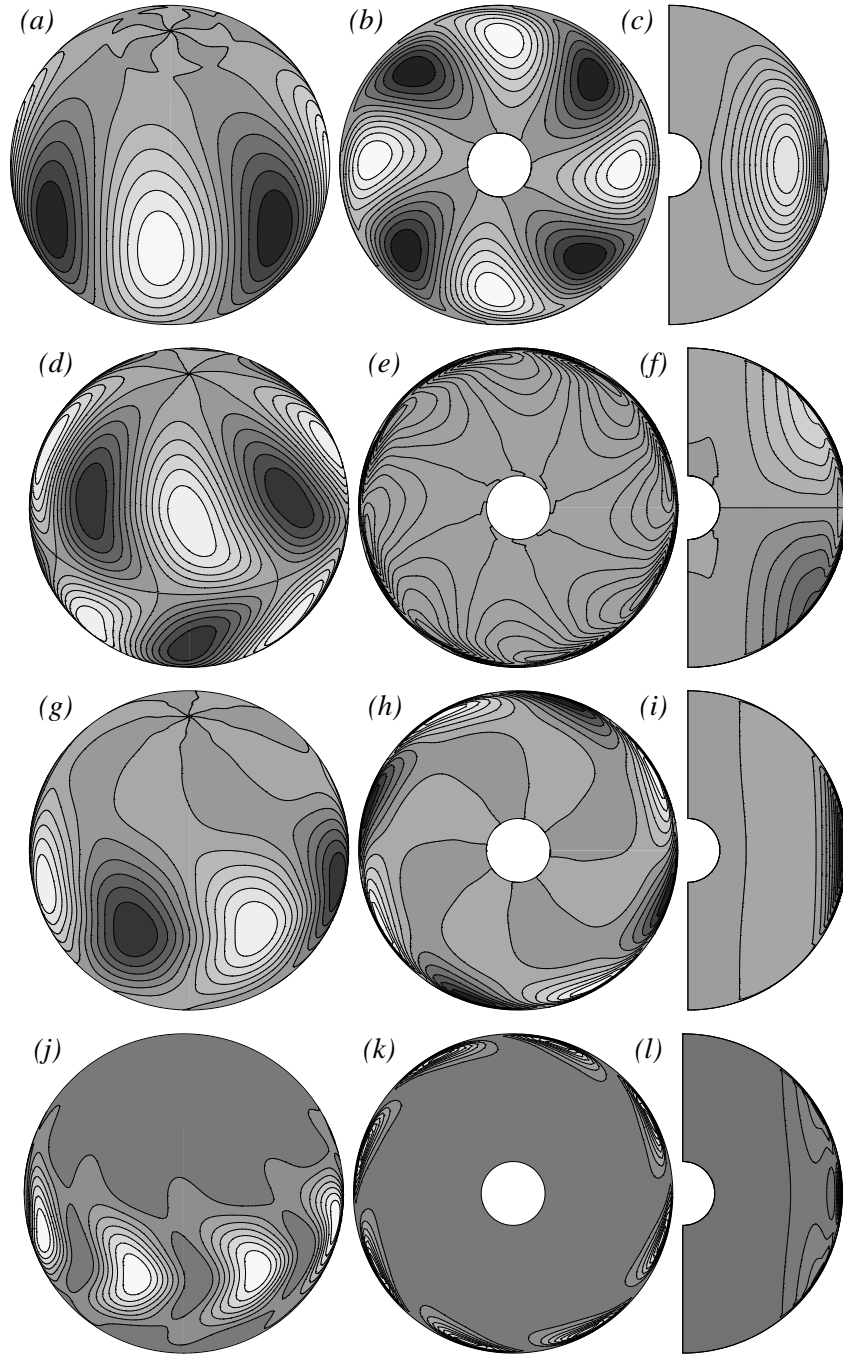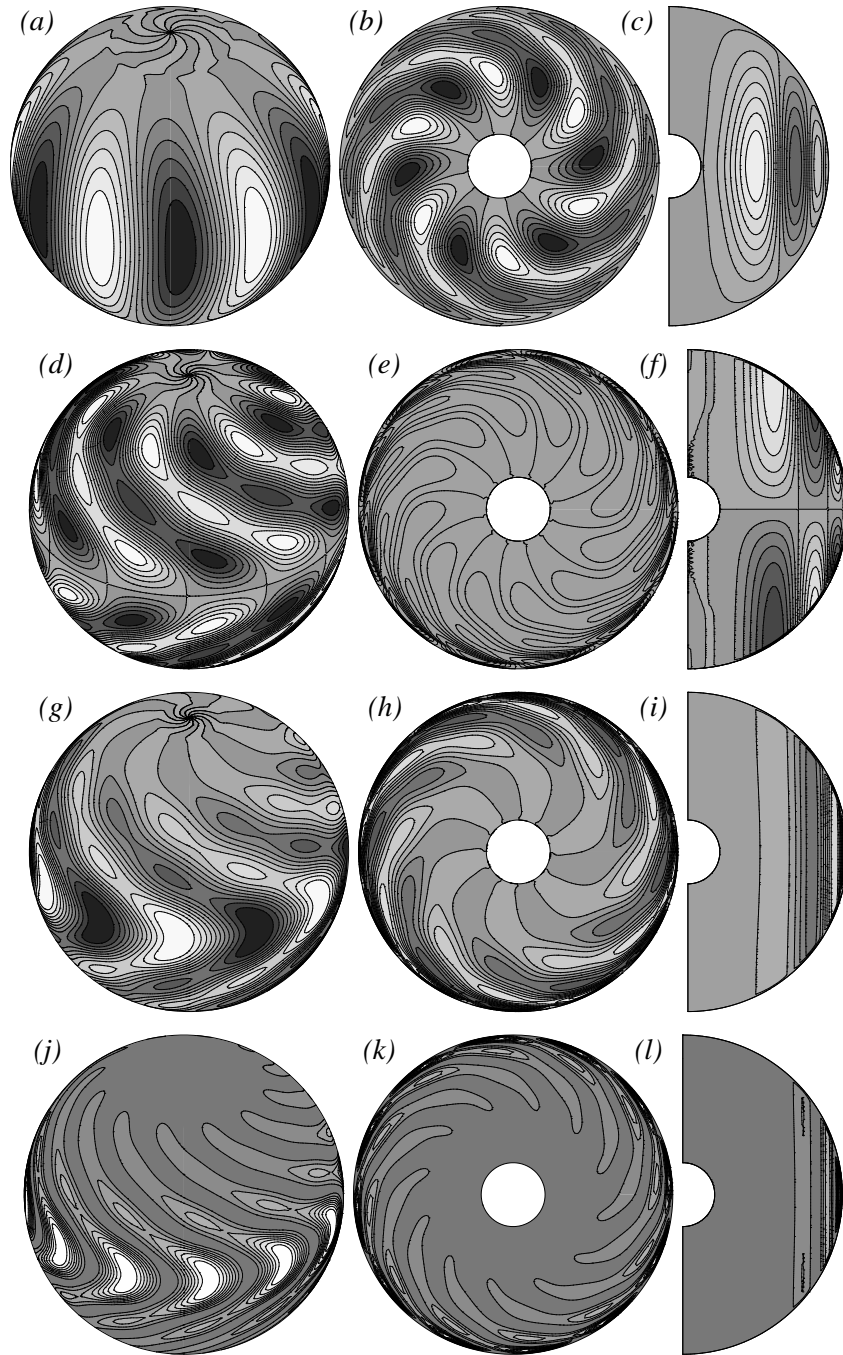
Figure 2.4: Preferred tri-cellular mode of convection with azimuthal wavenumber $m = 6$, plotted at $E_Z = 3.69 \times 10^{-6}$ for $\sigma = 0.005$. Same contour plots as in figure 2.3.

positive components of the velocity field. Notice that $v_\theta = 0$ in the equatorial plane, so the contour lines in plots labelled (e) are residuals.

Figure 2.3 corresponds to the preferred eigenfunction at $E_Z = 1.92 \times 10^{-5}$, $(R_c^4)_Z = 38.90$ of figure 2.2(c). In [119] it is shown that the inertial waves extent to a characteristic latitude $\mathcal{L} = (2/m)^{1/2}$, m being the azimuthal wavenumber. Since this latitude is very small, the presence of an inner sphere is irrelevant for wave motions with $m > \pi(1 + \eta)/2(1 - \eta)$, where η is the radius ratio of the spherical shell. We have measured the extent of the thermal-inertial waves in the equatorial zone following the criterion given for the solutions of the Poincaré equation, i.e., we consider an exponential decrease of the waves with latitude, and look for the latitude $\mathcal{L}$ at which the maximum of v_φ diminishes down to v_φ/e. Numerically we found $\mathcal{L} = (90^o - \theta) \approx 29^o$. Although $E = 3.00 \times 10^{-5}$ is not extremely high, the waves are much more confined than the theoretical prediction. In the case of figure 2.3, theoretically for $m = 4$, $\mathcal{L} = 40.5^o$. The discrepancy is probably due to the inclusion of the buoyancy term in our linear analysis.

For $E_Z < 1.9 \times 10^{-5}$ multicellular patterns of convection become preferred. When the Ekman number is decreased the convection spreads to high latitudes without detaching from the equator. The spiralling flow tends to split forming a multicellular spiral in the boundary layer. The $m = 6$ tri-cellular pattern found at $E_Z = 3.70 \times 10^{-6}$ is shown in figure 2.4. In this case, the convection affects up to $(90^o - \theta) \approx 44^o$, the main Θ cell moves towards the inner boundary. We think that the bi-cellular mode (not shown here) was also found numerically in [121] (see its figure 2.3), but interpreted as the spiralling columnar mode preferred at higher Prandtl numbers.

The meridional sections display the z-dependence of the components of the velocity field. Except in the very thin Ekman boundary layers, v_φ fulfils the Taylor-Proudman constraint, because it is a component orthogonal to the axis of rotation. We have checked that at high rotation rates this becomes the strongest component on the spherical surfaces near the outer boundary. In consequence, the contour plots of kinetic energy density of figure 2.4(j-l) resemble those of v_φ, and the energy is also almost z-independent. The perturbations from bi-dimensionality mainly come from v_r and v_θ.

2.3.2 Dependence of low Prandtl number flows on the Ekman number

In order to study the dependence of the low Prandtl number flows on the Ekman number, including the transition between prograde and retrograde modes, $\sigma = 0.1$ has been selected.

Figure 2.5 displays the results of the linear stability analysis as a function of E (see figure caption). If $E < 10^{-4}$ the regime dominated by the Coriolis force is fully developed, and, in agreement with [118], and [117] prograde SC modes of convection become preferred. So, from this point, the flow remains mono-cellular down to the lowest E values
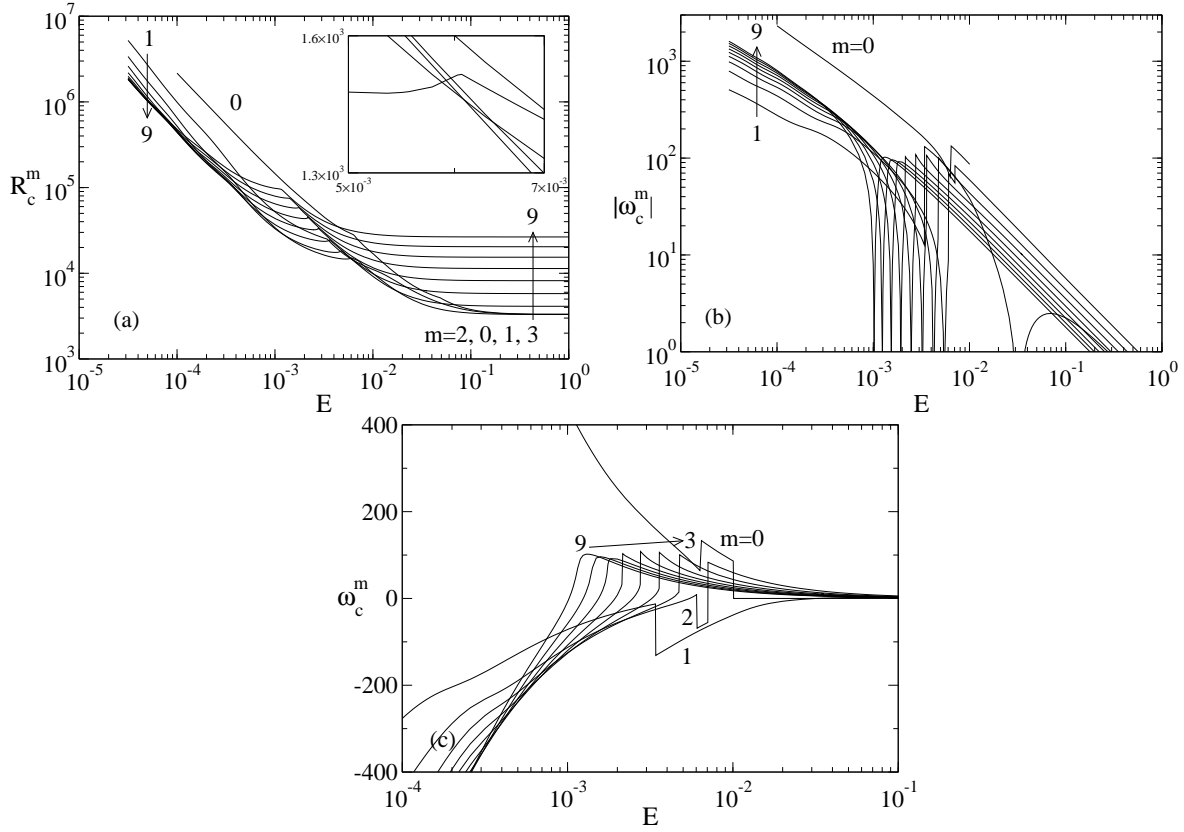
Figure 2.5: (a) The critical Rayleigh number R_c^m, and (b) the critical precession frequency $|\omega_c^m|$, of each critical mode of azimuthal wavenumber m, plotted versus E for $\sigma = 0.1$. (c) Detail of (b) in linear vertical scale showing the jump to spiralling columnar modes of convection.

explored. In this case, the vortices detach from the equator when the rotation rate is increased, avoiding the formation of extremely strong Ekman layers near $\theta = \pi/2$, and move towards the inner boundary, i.e., the Ekman layers only affect mid- and high-latitudes. For instance, at $E = 3.78 \times 10^{-5}$ the SC modes have the maximum at $r \approx 0.6r_o$, and the velocity field near the outer surface affects latitudes higher than $(90^o - \theta) \approx 13^o$. It is important to notice that these values vary a lot with the parameters E, σ, and η.

Out from the asymptotic regime, the neutral stability curves of $|\omega_c^m|$ and ω_c^m versus E (figures 2.5(b,c)), show clearly three different types of transitions among modes of the same m. For $E > 3.43 \times 10^{-3}$ and $m < 3$, there is a competition between antisymmetric prograde and symmetric retrograde body modes. The former can also be preferred at the onset of convection, as happens with the $m = 1$ mode in the interval $5.88 \times 10^{-3} < E < 6.10 \times 10^{-3}$, located just before the big cusp of the envelope of figure 2.5(a) (see also the detail). We
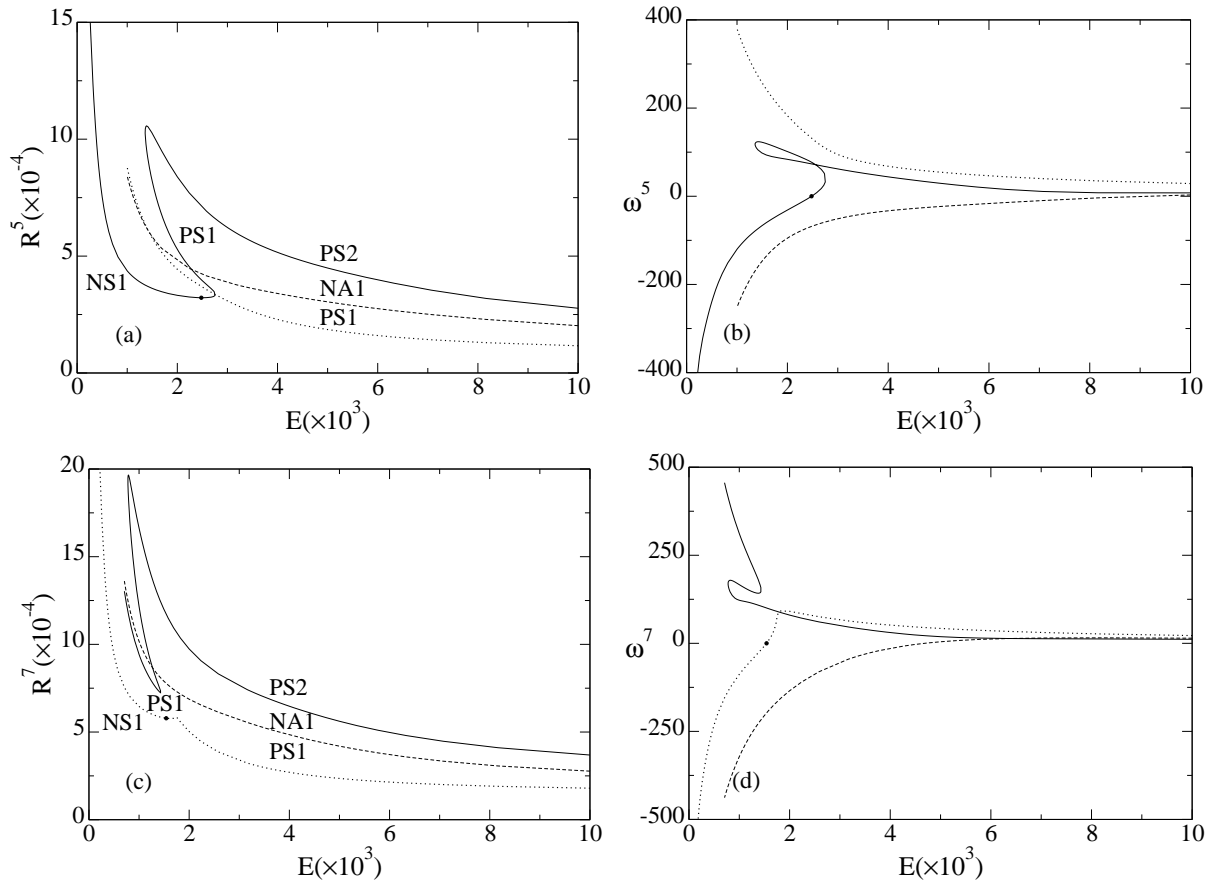
Figure 2.6: (a) The Rayleigh number R^5, and (b) the precession frequency ω^5 of the three lowest marginal stability curves of azimuthal wavenumber $m = 5$, plotted versus E for $\sigma = 0.1$. In (c) and (d) R^7 and ω^7 for the three lowest curves of $m = 7$, respectively. The meaning of labels PS1, PS2, NS1, and NA1 is summarised in Table 2.2.

Acronyms	Symmetry	Sign of ω	Number of latitudinal cells
PS1	**S**ymmetric	**P**ositive (retrograde)	1
PS2	**S**ymmetric	**P**ositive (retrograde)	2
NS1	**S**ymmetric	**N**egative (prograde)	1
NA1	**A**ntisymmetric	**N**egative (prograde)	1

Table 2.2: Acronyms for the leading modes found at high Ekman number

have compared the balance of terms of the equations for the preferred antisymmetric mode, and for the neighbouring symmetric solutions of higher and lower E. The only difference found is a slight decrease of the viscous diffusive and buoyancy terms in the antisymmetric modes.

If $m > 3$ only symmetric modes become selected. For $3 \leq m \leq 6$ the cusp of each curve in figure 2.5(a) is the intersection of two curves of symmetric retrograde modes, although one of them becomes prograde when E is decreased. This change of sign of the precession frequency corresponds to the 'fall' of $|\omega_c^m|$ to zero in figure 2.5(b). The crossings can also be seen as jumps of ω_c^m in figure 2.5(c). For $m > 6$ the jumps disappear. To see in detail why this happens, the three lowest neutral stability curves with $m = 5$ and $m = 7$ are plotted in figure 2.6. The type of line gives the correspondence between the (E, R^m) and (E, ω^m) curves. Moreover, the meaning of the acronyms used in the labels of the curves, and, from now on, in the text is summarised in table 2.2.

Following the solid curve in figure 2.6(a) from high to low E values, the body modes of the PS2 curve transform their spatial structure continuously, becoming PS1 modes after the first fold, and NS1 after the second. These NS1 modes are the SC preferred at high rotation rates. The jump of the $m = 5$ critical curve in figure 2.5(c) is associated to the intersection of the dotted PS1, and the solid NS1 curves at $E = 2.76 \times 10^{-3}$ in figure 2.6(a). Notice that at this point both are mono-cellular symmetric solutions of positive precession frequency (PS1). The change of sign of ω_c takes place at $E = 2.48 \times 10^{-3}$, marked with a full circle in the figure, close to the crossing.

For $m = 7$ there is a qualitative change in the neutral stability curves. By comparing figure 2.6(a) with figure 2.6(c) it is easy to recognise an interchange of curves. In the second figure, the right part of PS1 remains connected with the left part of NS1, forming the first lower marginal curve, and the left part of PS1 (dotted in figure 2.6(a)) remains connected with the first part of NS1, forming the solid line of figure 2.6(c), which has two very narrow folds. This phenomenology was also described for stress-free boundary conditions in the pioneering numerical work of [130].

The interchange of like-parity modes is analysed in detail for the onset of oscillatory binary fluid convection in [11]. According to these authors, the interchange takes place at a resonant double-Hopf bifurcation (surely hidden by the critical non-resonant double-Hopf bifurcations), where the crossing neutral curves become tangent. At this point both

modes will have the same frequency. This explains why along the critical stability curves (like those of $m = 7, 8, 9$ in figure 2.5(c)) there are cusps without jumps in the frequency, or abrupt changes in the eigenfunctions. The values of m (3 and 6 in this case) which separate the different behaviours of the neutral curves depend on η. A resonant double-Hopf bifurcation has codimension-three, and to determine the point of tangency without changing m, a third parameter (η or σ) must be moved.

The continuous transformation of the eigenfunctions along the folds of the curves of symmetric modes, and the interaction of modes described, explain why never multicellular modes are preferred at moderate and high σ values. Unlike what happens for $\sigma = 0.005$, for $\sigma = 0.1$ the crossing of the lowest curves with the same m takes place after the transformation, and then, all of them are NS1 (SC). For instance, the mono-cellular $m = 7$ NS1 (SC) mode, which becomes critical at $E \approx 5.77 \times 10^{-5}$, comes from a bi-cellular PS2 mode.

2.3.3 Prandtl number dependence

In order to study the transition between the OEA and the SC patterns of convection, the envelope of the marginal stability curves (σ, R_c^m) is shown in figure 2.7(a-c) for four E values in the small Ekman number limit ($3.16 \times 10^{-6} < E < 10^{-4}$). In addition, we include Table 2.3 with the parameters and features of the bi-critical points where changes of types of preferred eigenfunctions (not just changes of m) take place. The table contains additional points calculated for $E = 4.47 \times 10^{-6}$ and $E = 1.41 \times 10^{-6}$ without rounding off new curves up to $\sigma = 0.1$. Figure 2.7(d) contains a diagram showing the transitions of the table. From the double-Hopf OEA/SC bifurcation, the SC modes detach from the equator, moving towards a cylindrical radius $r \in (r_i, r_o)$, while the OEA modes remain equatorially-trapped when σ decreases.

Figure 2.7(b) displays small increasing, and big decreasing jumps in $|\omega_c|$. The small correspond either to jumps between OEA modes of the same number of latitudinal cells, or between SC modes. In both cases m_c increases in one unit. This is the normal situation. In contrast, the big jumps are due either to jumps between OEA modes of different number of latitudinal cells, between OEA and SC modes, or between SC modes of decreasing m_c, see figure 2.7(c). In order to help to follow them, the discrete points are joined with a dotted line.

In figure 2.7(d) it can be seen that there is a direct transition between the equatorially-trapped (OEA1) and SC modes only near $E = 10^{-4}$, i.e., just at the beginning of the small Ekman number limit. For lower E, the transition is directly from multicellular to SC modes. Moreover, we have been unable to find any simple law defining the OEA/SC transition with non-slip boundaries, unlike in [3] for stress-free boundary conditions.

At moderate E, the OEA/SC transition occurs near the Prandtl number of some usual experimental fluids, namely for $\sigma < 0.023$. This result agrees with the experimental results of [6] for liquid gallium ($\sigma = 0.027$). They suggest that the nonlinear flows keep
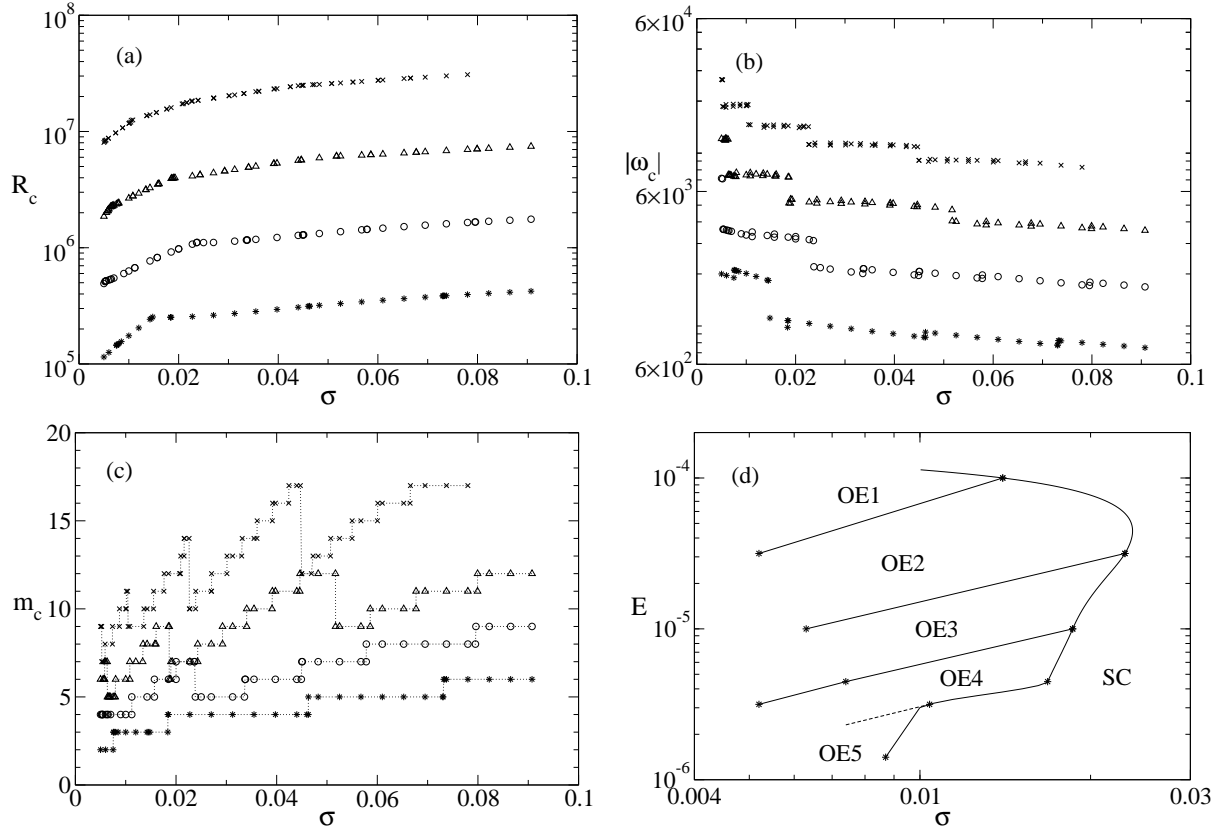
Figure 2.7: (a) The critical Rayleigh number R_c, (b) the critical precession frequency $|\omega_c|$, (c) the critical wavenumber m_c, plotted versus σ. From bottom to top, $*$, $\circ$, $\triangle$, and $\times$ mean $E = 10^{-4}, 3.16 \times 10^{-5}, 10^{-5}$, and 3.16×10^{-6}, respectively, and (d) diagram showing the transitions among the different modes of convection. The stars correspond to calculated points.
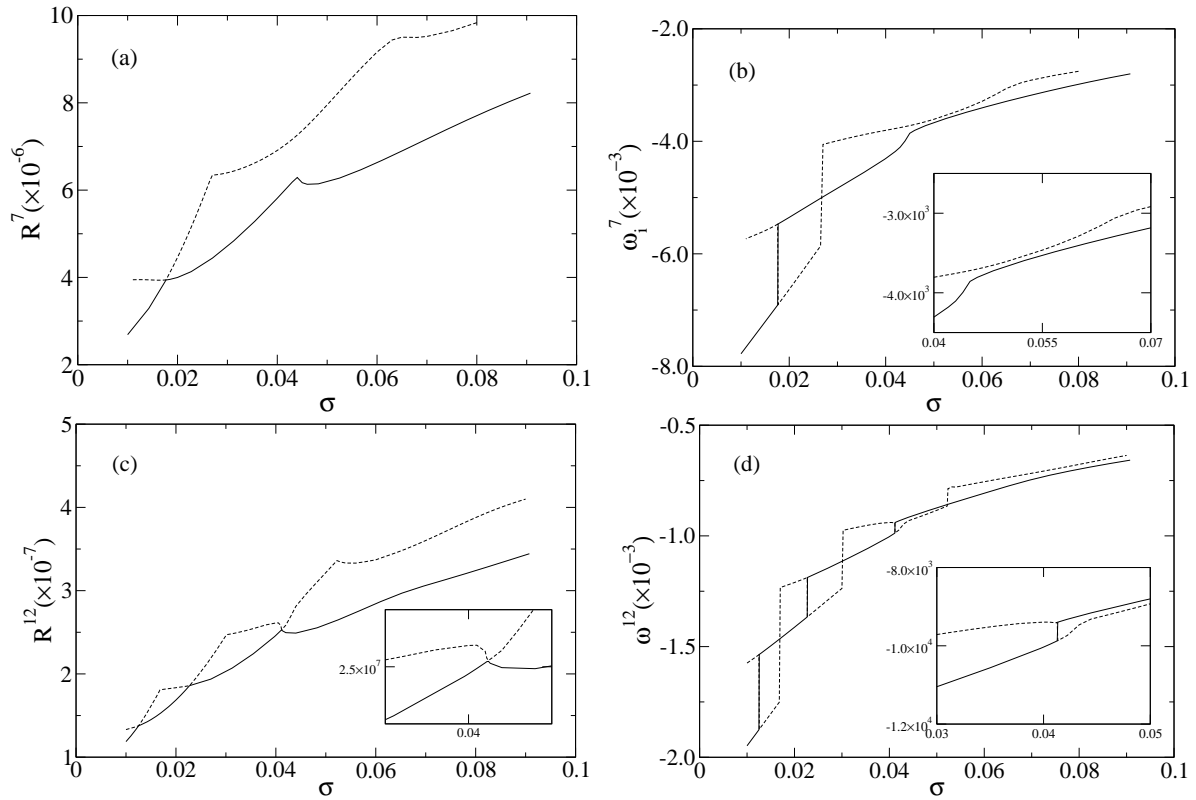
Figure 2.8: (a, c) The Rayleigh number R^7, and R^{12}, and (b, d) the precession frequencies ω^7, and ω^{12}, respectively, of the two lowest envelopes of the neutral stability curves of $m = 7$, and $m = 12$, plotted versus σ, for $E = 1 \times 10^{-5}$ and $E = 3.16 \times 10^{-6}$, respectively. The solid lines give R_c^m, and the dashed lines correspond to the second mode which becomes linearly unstable.

JUMP	$E \times 10^5$	σ	$R_c \times 10^{-5}$	$|\omega_c| \times 10^{-3}$	m_c
OEA1/OEA2	3.16	0.0052	5.12	7.11/3.63	4/4
OEA2/OEA3	1.00	0.0063	22.4	12.0/7.49	7/5
OEA3/OEA4	0.447	0.0074	66.0	18.2/12.9	9/7
OEA3/OEA4	0.316	0.0052	83.3	26.5/18.6	9/7
OEA1/SC	10.0	0.014	2.47	1.83/1.11	3/3
OEA2/SC	3.16	0.023	11.1	3.12/2.20	7/7
OEA3/SC	1.00	0.0186	39.5	7.25/5.14	9/6
OEA4/SC	0.447	0.0168	101.6	12.6/9.56	11/8
OEA4/SC	0.316	0.0104	120.	19.1/14.6	11/9
OEA5/SC	0.141	0.0087	306.	34.9/27.8	13/11
SC/SC	1.00	0.051	61.4	4.69/4.05	12/9
SC/SC	0.316	0.023	182.	14.2/11.2	14/10
SC/SC	0.316	0.045	250.	10.8/9.09	17/12

Table 2.3: Parameters and critical values of the bi-critical points of figure 2.7 giving rise to jumps of decreasing $|\omega_c|$. The number after OEA indicates the number of latitudinal cells near the outer boundary.

traces of the SC Rossby waves for Ekman numbers of order 10^{-7}, in a spherical shell of $\eta = 0.35$. The location of the transition point can also explain why [124] found OEA critical modes for $\sigma = 0.023$, a very small η, and E order 10^{-6}. Near the OEA/SC transition a small variation of the parameters and/or the boundary conditions can vary the type of preferred eigenfunction.

In order to understand the jumps among SC modes of decreasing m (see figure 2.7(c)), and why the eigenfunctions after the jumps still have the same structure, the two lowest envelopes of the neutral curves with $m = 7$ for $E = 10^{-5}$, and $m = 12$ for $E = 3.16 \times 10^{-6}$ are plotted in figure 2.8 as functions of σ. In figure 2.8(a) it can be seen how the solid curve develops a cusp, and the dashed line a bump, near $\sigma = 0.043$. This causes an abrupt decrease of R^7 along the solid line for larger σ. The frequency curves of figure 2.8(b) get also distorted (see detail), becoming nearly tangent close to $\sigma = 0.05$. We have checked that other curves of low m have the same behaviour. In contrast, figure 2.8(c,d) show the shape of the curves for higher m. It can be seen that close to $\sigma = 0.04$ the two curves have already reconnected in the way explained in subsection 2.3.2. The details in figures 2.8(c,d) show traces of the fold developed by one of the curves before becoming tangent, and the almost equal frequencies of both modes after the bifurcation.

Since for $\sigma > 0.023$ the neutral stability curves are more or less folded, the different degree of deformation of the curves depending on m, propitiates that curves of low m dominate suddenly again, after others of higher m. This explains the fall of m_c and $|\omega_c|$ in figures 2.7(b,c) at the SC/SC jumps. In addition, at the resonant double-Hopf

σ	Interval $(E \times 10^4)$	$E \times 10^5$	ε_{R_c} (%)	ε_{ω_c} (%)
0.1	$(0.35, 1)$	3.500	2.5×10^{-5}	2.1×10^{-3}
0.025	$(0.01, 0.1)$	0.100	5.0×10^{-1}	8.6×10^{-2}
0.005	$(0.01, 0.1)$	0.316	7.8×10^{-2}	9.7×10^{-3}
0.005	$(0.01, 0.1)$	0.112	2.6	2.6×10^{-1}
0.005	$(0.01, 0.1)$	0.100	2.9	3.4×10^{-1}

Table 2.4: Prandtl number, interval considered for adjusting the asymptotic laws of figure 2.9, E of the point in the interval at which the error due to truncation is maximal, and the corresponding numerical relative errors (ε%) of R_c and ω_c.

bifurcations neutral curves of same m interchange their properties, and then when this m reappears as preferred (in a non-resonant double-Hopf bifurcation) the same structure is seen again along the critical stability curve. In this case, for a given m, it is possible to capture the resonant double-Hopf bifurcation by moving η and/or E, in addition to σ.

2.3.4 The asymptotic limit

Figure 2.9 shows the critical Rayleigh number R_c, the precession frequency $|\omega_c|$, and the critical wavenumber m_c versus E, for the two fluids studied in the latter section (see figure caption). For completeness, we include the results of the linear stability analysis of a fluid of $\sigma = 0.025$ (mercury/liquid gallium), which, in addition of being near the OEA/SC border, has experimental interest in magnetohydrodynamics. To study experimentally this transition it is also worth mentioning the usefulness of the liquid sodium. With $\sigma = 0.01$, it is below the critical value $\sigma = 0.023$, but still near the competition between OEA/SC modes at lower E (see figure 2.7(d)).

For $\sigma = 0.025$, and high E values, the critical waves have $\omega_c > 0$, but the antisymmetric $m = 1$ mode with $\omega_c < 0$ is preferred in $3.13 \times 10^{-3} < E < 3.63 \times 10^{-3}$. According to the preceding results, the preferred patterns of convection of fluids with $\sigma \approx 0.023$ can be either OEA or SC modes depending on E. For $E \approx 4.47 \times 10^{-5}$ it is really difficult to distinguish from the contour plots the type of preferred eigenfunction. In fact, to decide it, it is necessary to know its position in the (σ, R_c) plane of figure 2.7. For $E < 3.16 \times 10^{-5}$ all the critical patterns are SC.

In figure 2.9(a) the power laws of the numerical fittings for low E are included in our non-dimensional units. The intervals used for the fitting, and the maximal numerical relative errors due to the numerical truncation of the data of figure 2.9 are given in Table 2.4. They are estimated by recomputing the points with 70 radial, and 110 latitudinal points. In general an over-truncation overestimates R_c and underestimates ω_c. However, as can be seen in the table, we have only introduced relative errors below 3% for R_c with $\sigma = 0.005$ and $E \approx 10^{-6}$. With these values, the estimated error for the exponent of the
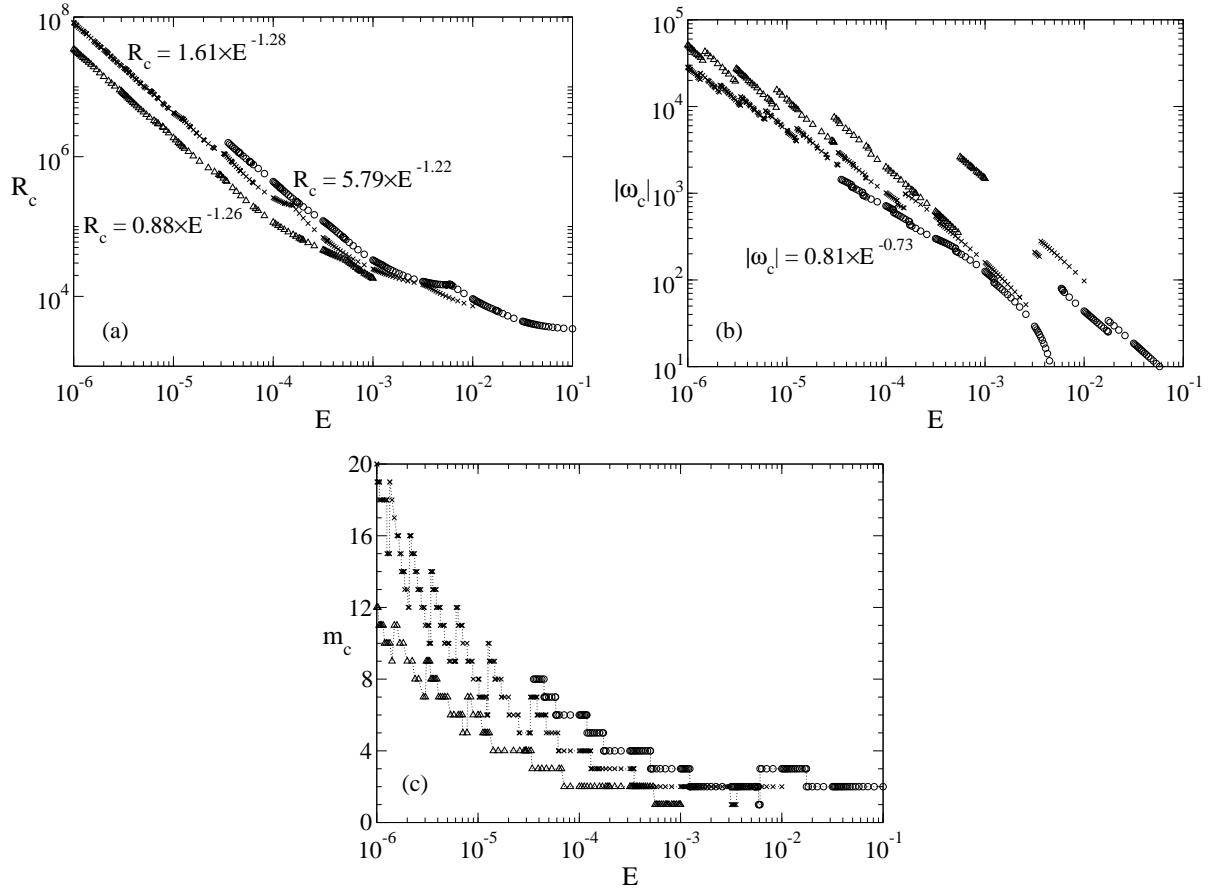
Figure 2.9: (a) The critical Rayleigh number R_c, (b) the corresponding precession frequency $|\omega_c|$, and (c) the critical azimuthal wavenumber m_c versus E. The labels near the curves give the power laws estimated from the numerical results. The symbols $\triangle$, $\times$, and $\circ$ indicate $\sigma = 0.005, 0.025$, and 0.1, respectively.

power law of R_c is 1.3%, and 17% for the prefactor. For $\sigma = 0.025$, 60 radial and 100 latitudinal points are used everywhere, so the error of the exponent is less than 0.2%, and that of the prefactor is 1.7%. For $\sigma = 0.1$ the truncation error is even smaller, because the curve is cut at higher E.

In any case the power of R_c is not far from $-4/3$, the leading order of the asymptotic expansions of [94], [14], and [32] (see introduction) for moderate Prandtl numbers. In contrast, the dependence of $|\omega_c|$ with E is not so clear. For $\sigma = 0.005$ the precession frequency diminishes sharply when a new pattern of convection is selected, but keeping the previous slope (see figure 2.9(b)). The jumps are smaller as E is decreased. Perhaps this fact indicates that for ω, the asymptotic limit is not fully reached. Although the fitting of each straight segment gives approximately a 1.10 power, exceeding very much $-2/3$ given by the asymptotic limits, the law $|\omega_c| = 5.33 \times E^{-0.66}$ is found by considering the interval $10^{-6} < E < 10^{-5}$. The jumps for $E \leq 3.0 \times 10^{-5}$ along this curve are between OEA modes. By increasing σ the height of the jumps in the same range of E is smaller. For $\sigma = 0.025$ and $E < 4.47 \times 10^{-5}$, there are some E for which it is very difficult to classify the eigenfunctions. As in the preceding case, the fitting for $E < 10^{-5}$ gives $|\omega_c| = 2.12 \times E^{-0.68}$. Finally for $\sigma = 0.1$, the power law dependence of $|\omega_c|$ is included, since for $E < 4.47 \times 10^{-3}$ the dominant modes correspond to a same pattern of eigenfunctions (SC), and the small jumps of increasing $|\omega_c|$ correspond to changes of increasing m_c.

The dependence of m_c on E is plotted in figure 2.9(c). Again the discrete points are joined with a dotted lined. By neglecting the discontinuities, we have checked that the potential law $m_c \sim E^{-1/3}$ of the theoretical expansions also fits well to our numerical values. We have found $m_c = 0.15 \times E^{-0.31}$ for $\sigma = 0.005$, $m_c = 0.16 \times E^{-0.34}$ for $\sigma = 0.025$, both for $E < 10^{-5}$, and $m_c = 0.30 \times E^{-0.32}$ for $\sigma = 0.1$ and $E < 10^{-4}$.

2.4   Concluding remarks

The detailed numerical computations of this work clarify previous results, and also show some unexpected aspects of the thermal convection at low Prandtl numbers, σ.

Far from the small Ekman number limit ($E > 3.16 \times 10^{-3}$), the antisymmetric prograde waves with low m_c can be preferred; m_c depending on η. For instance, $m_c = 1$ for $\eta = 0.2$, and in other computations with $\eta = 0.35$ and differential heating (not displayed here), $m_c = 2$.

For $\eta = 0.2$, and a wide range of E (see table 2.3), the transition OEA/SC takes place near $\sigma = 0.023$, namely, it occurs near the Prandtl number of the experimental liquid metals in normal laboratory conditions. At this value both modes become unstable almost at once, and consequently nonlinear chaotic dynamics may be expected at or near R_c. So, the numerical and experimental study of these regimes appears as an interesting objective easy to reach in the near future. By decreasing E the transition to SC modes takes place at

lower σ values, this means that at the very low Ekman numbers of the planetary cores it is very unlikely that convection can excite the outer-equatorially-attached modes. However, they could be excited by other mechanisms, or become relevant due to the action of the magnetic field on the equatorial waves.

The location of the transition point could explain why [106] detect temporal-chaotic states near the onset of convection with $\eta = 0.4$ and $\sigma = 0.025$, and a transition to stable travelling waves followed by amplitude vacillations for moderate and large σ values. However, the nonlinear results of [101], [59], [87] and [88] among others, obtained for cylindrical annuli of large radius ratios, suggest that in thin spherical shells, the preceding behaviour could have another reason. Sideband instabilities leading to spatially modulated waves, and mean zonal flow/resonance instabilities at moderate/low σ are the mechanisms responsible for the instability of the thermal Rossby waves. At low σ, the region of stability of the waves comes away from the neutral stability curve, and becomes narrower until disappearing for very low σ values, so stable complex time-dependent flows develop at the onset of convection.

Our results are restricted to wide spherical shells, and it could be interesting (but very expensive) to see how the transition OEA/SC moves by narrowing the gap. Other linear calculations made for $\eta = 0.35$ down to $E = 1.20 \times 10^{-7}$ with differential heating, show that the dominant modes in the small-Ekman-number limit are also columnar for $\sigma = 0.1$, but attached to the inner boundary.

For σ greater than approximately 0.023, the modulation of the eigenfunctions along the folds of the neutral stability curves (σ, R^m) explains the bi-critical bifurcations labelled SC/SC. When two neutral stability curves become tangent, the eigenfunctions at the tangency point have the same structure. Moreover, a sudden decrease in the Rayleigh number takes place (see figure 2.8). Consequently, if a neutral curve becomes preferred after the tangency, at a higher σ, the new preferred modes look like those that were preferred before (but for the azimuthal wavenumber).

For very low σ, the equatorially trapped mono-cellular modes are soon superseded by multicellular modes when E decreases. Therefore, a direct transition between equatorially trapped and spiralling columnar modes exists only in a narrow range of Ekman numbers, at the beginning of the rapidly rotating regime. This result agrees qualitatively with the sketch of [89].They also found that in rotating annuli at low E and σ the transition is between multicellular and spiralling columnar modes. The z-dependence of the modes of convection inside the fluid shell is stronger in spherical than in cylindrical geometry. When the multicellular patterns become selected, the convection tends to fill the shell, spiralling from the body of the fluid by the effect of the rotation, but without detaching from the outer equatorial boundary. At the same time, small vortices, which remain stuck to the outer boundary, appear. By comparing the structure of these modes with those found by [3], and [124] with stress-free boundary conditions, it seems that the non-slip boundaries tend to inhibit the formation of such vortices, in the sense that they are smaller and more confined.

The numerical potential laws obtained with non-slip boundary conditions and low Prandtl numbers agree well on average with the leading order of the theoretical asymptotic expansions of [94], [14] and [32]. However, the jumps of the precession frequencies, for low σ values, can indicate that even at $E = 10^{-6}$ the asymptotic limit for ω_c is not fully reached.

Acknowledgements

Chapter 3

Antisymmetric polar modes of thermal convection in rotating fluid spherical shells at high Taylor numbers

Antisymmetric polar modes of thermal convection in rotating fluid spherical shells at high Taylor numbers

ABSTRACT

The onset of thermal convection in a rotating spherical shell of intermediate radius ratio $\eta = 0.4$ is studied numerically for Taylor numbers $Ta \geq 10^{11}$ and the Prandtl number of the liquid sodium ($\sigma = 0.01$). For the first time, it is shown that at very high Taylor numbers the first unstable mode can be antisymmetric with respect to the equator, and confined inside a cylinder tangent to the inner sphere at the equator (polar mode). The exponent of the power law determined from the asymptotic dependence of the critical Rayleigh number for very high Ta is 0.57, lower than 2/3, given theoretically for the spiralling columnar modes, and than 0.63, found numerically for the outer equatorially-attached modes.

Ferran Garcia, Juan Sánchez and Marta Net.

The thermal convection in fluid spherical shells is a fundamental problem in geophysics and astrophysics. For instance, the Earth's magnetic field is generated in its interior by convection driven by thermal and compositional buoyancy. In this way, many of the dynamo features are predetermined by the properties of convection. The large-scale zonal winds observed in the surface of Jupiter at mid- and low-latitudes and of Saturn seem to be maintained by deep convection [5]. In the last twenty years, a great quantity of experimental, theoretical and numerical studies, devoted to improve the understanding of the basic mechanisms which govern the convection in spherical geometry, have appeared. The introductory sections of [124, 32, 42, 41, 79], among others, provide good reviews of the state of the art on this subject.

The theoretical paper [14] established that the critical mode for the onset of thermal convection in self-gravitating, and internally heated fluid spheres was a columnar travelling wave, localised around a critical radius $r_i < r_c < r_o$, and symmetric by reflections with respect to the equator. In the inequality r_i and r_o mean the inner and outer radii of the shell. This critical mode fulfils $(v_r, v_\theta, v_\varphi)(r, \theta, \varphi) = (v_r, -v_\theta, v_\varphi)(r, \pi - \theta, \varphi)$, and $\Theta(r, \theta, \varphi) = \Theta(r, \pi - \theta, \varphi)$, where $(v_r, v_\theta, v_\varphi)$ is the velocity field in spherical coordinates (θ measuring the colatitude), and Θ the temperature perturbation of the conduction state. The validity of this symmetry has been confirmed numerically by several authors, by comparing with the modes obtained for fluid spherical shells of small $\eta = r_i/r_o$, and also in laboratory experiments [23]. Since then, symmetric columnar solutions have been assumed in most of the studies devoted to find or to improve the asymptotic dependence of the onset of instability in spherical shells, independently of its η and of the Prandtl number σ of the fluid. However, there are some ranges of parameters where the latest asymptotic theories in fluid spheres [66] and spherical shells [32] do not apply. For instance, the latter does not fit properly for a small range of η around 0.48 and $\sigma = 1$. As far as we know the reason remains unknown. Moreover, they do not cover small-σ fluids because in fluid spheres and small-η spherical shells the convection sets in with a symmetric pattern, but outer equatorially-attached and multicellular [124, 125, 79]. So, in a problem with four parameters it is feasible that other kind of modes become preferred, mainly at very high Ta.

The numerical linear stability analysis of the conduction state presented in this *letter* shows, for the first time, the existence of antisymmetric modes of convection preferred at high Ta in fluid shells of small σ and moderate η, independently of the boundary conditions applied. For these solutions, the velocity field and Θ fulfil $(v_r, v_\theta, v_\varphi)(r, \theta, \varphi) = (-v_r, v_\theta, -v_\varphi)(r, \pi - \theta, \varphi)$, and $\Theta(r, \theta, \varphi) = -\Theta(r, \pi - \theta, \varphi)$. Antisymmetric equatorially trapped inertial waves (solutions of the Poincaré equation) were calculated before by [119], and recently [79] have found numerically preferred antisymmetric thermal Rossby modes filling the shell at moderate Ta.

The Boussinesq approximation of the mass conservation, linear momentum and energy equations in the rotating frame of reference,

$$
\begin{aligned}
\nabla \cdot \mathbf{v} &= 0, \\
(\partial_t + \mathbf{v} \cdot \nabla)\,\mathbf{v} &= -\nabla\pi + \nabla^2\mathbf{v} - 2Ta^{1/2}\mathbf{k} \times \mathbf{v} + \Theta\mathbf{r}, \\
\sigma\,(\partial_t + \mathbf{v} \cdot \nabla)\,\Theta &= \nabla^2\Theta + (R + \eta(1-\eta)^{-2}r^{-3}R')\,\mathbf{r} \cdot \mathbf{v},
\end{aligned}
$$

are solved for a fluid spherical shell rotating about an axis of symmetry with constant angular velocity $\boldsymbol{\Omega} = \Omega\mathbf{k}$, subject to internal or differential heating and radial gravity $\mathbf{g} = -\gamma\mathbf{r}$, where γ is a constant, and $\mathbf{r}$ the position vector. Stress-free ($v_r = \partial_r(v_\theta/r) = \partial_r(v_\varphi/r) = 0$) or non-slip ($v_r = v_\theta = v_\varphi = 0$), and perfectly conducting ($\Theta = 0$) boundaries are employed. Notice that in the formulation the centrifugal force is neglected since in the Earth's outer core and in the major planets $\Omega^2/\gamma \ll 1$, and that the same units as in [106] are used. The non-dimensional Taylor, internal (R) and external (R') Rayleigh and Prandtl numbers are

$$
Ta^{1/2} = \frac{\Omega d^2}{\nu}, \quad R = \frac{q\gamma\alpha d^6}{3c_p\kappa^2\nu}, \quad R' = \frac{\gamma\alpha\Delta T d^4}{\kappa\nu}, \quad \sigma = \frac{\nu}{\kappa},
$$

and the conduction state is given by $\mathbf{v} = 0$, and $T_c(r) = T_0 - (R/2\sigma)r^2 + (R'\eta/\sigma(1-\eta)^2)/r$. In the above definitions α means the thermal expansion coefficient, ν the kinematic viscosity, κ the thermal diffusivity, c_p the specific heat at constant pressure, q the rate of heat due to internal sources per unit mass, ΔT the difference of temperature between the inner and outer boundaries due only to differential heating, and $d = r_o - r_i$ the gap width.

The solutions up to $Ta = 10^{12}$ are computed with the method described in [79], by using 60 radial points, and spherical harmonics of maximal degree 100, but we have checked that, at $Ta = 5 \times 10^{11}$, an increase of the resolution to 80 by 160 leads to maximal differences of 1.2% in the critical Rayleigh number of the preferred modes of azimuthal wavenumber m, R_c^m, and of 0.02% in the critical precession frequency ω_c^m. From now on $R' = 0$, unless it is said explicitly.

Fig. 3.1, computed with stress-free boundary conditions, shows the dependence of R_c^m, and ω_c^m on $Ta > 10^{11}$, for $m = 12, \cdots, 19$. Negative precession frequencies mean that the waves travel in the prograde direction with phase speed $c_m = -\omega_c^m/m$. The changes of slope in Fig. 3.1(a) and jumps of Fig. 3.1(b) correspond to the crossing of symmetric outer equatorially-attached modes [120] and the antisymmetric modes of Fig. 3.2 (see figure caption for the meaning of each plot), which spread from the inner to the outer boundaries, but confined inside a cylinder tangent to the inner sphere at the equator (polar modes). The kinetic energy density is almost z-independent, as it happens with the spiralling columnar and the equatorially-attached modes, although, for the polar solutions, v_φ also depends strongly on z, but its maximum value is less than half that of the other components of the velocity field. The contour plots of Θ are not shown because

42
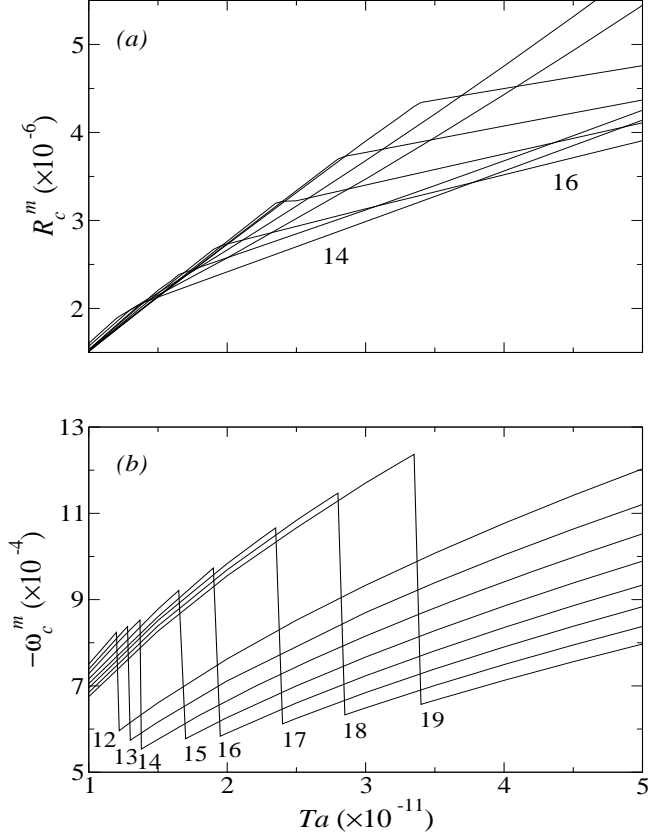
Figure 3.1: (a) The critical Rayleigh number R_c^m, and (b) the critical precession frequency $-\omega_c^m$, of each critical mode of azimuthal wavenumber $m = 12, \cdots, 19$, plotted versus Ta, for $\eta = 0.4$, $\sigma = 0.01$.

they resemble very much those of v_r. The main difference is that they are a few degrees out of phase in the φ coordinate.

The envelope of the curves of R_c^m versus Ta gives the critical Rayleigh number R_c, and the preferred pattern of convection. In the range of Ta of Fig. 3.1, the dominant modes are $m = 17, 18, 14, 16$, successively. At $(Ta, R_c) = (1.49 \times 10^{11}, 2.13 \times 10^6)$ the $m = 18$ symmetric equatorial mode is superseded by the $m = 14$ antisymmetric polar mode with a decrease of a 31% in $-\omega_c$. We have checked that with non-slip boundaries there is a jump from spiralling columnar to antisymmetric polar modes in the same range of Ta. The eigenfunctions look like that of Fig. 3.2, with very thin Ekman layers for v_θ and v_φ that stabilize the fluid, rising R_c by 140% at $Ta = 5 \times 10^{11}$, although the critical precession frequency, ω_c, decreases hardly by 1.4%. Fig. 3.2 is calculated for stress-free boundary conditions. White means the largest positive velocity and kinetic energy, and the background grey means $\mathbf{v} = 0$ in any case.
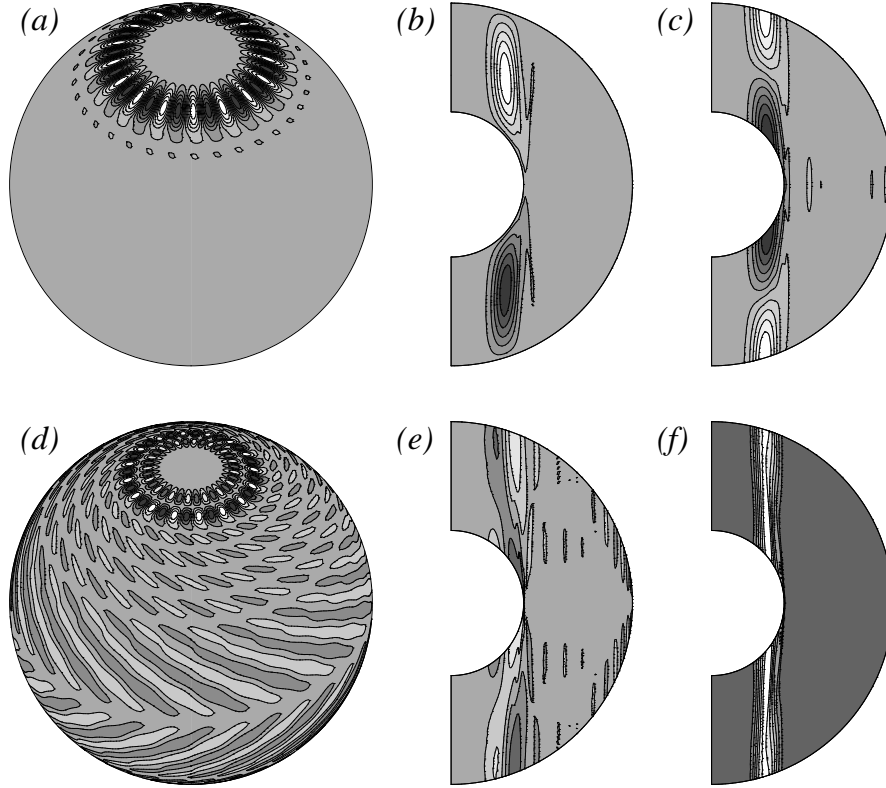
Figure 3.2: Preferred antisymmetric polar mode of convection with azimuthal wavenumber $m = 16$ at $Ta = 4 \times 10^{11}$, for $\eta = 0.4$, $\sigma = 0.01$. (a, d) Contour plots of the radial and azimuthal velocities on spheres of radii $r = r_i + 0.50d$ and $r \simeq r_i + 0.99d$, respectively. (b, c, e, f) Contour plots of the radial, colatitudinal and azimuthal velocities, and of the kinetic energy density on meridional sections, respectively.

To calculate the asymptotic dependence of the critical values, we have increased the degree of the harmonics gradually up to 150 with Ta to maintain the errors into the range determined previously. Figure 3.3 shows the results. The solid line corresponds to the potential fitting $R_c = 0.78 \times Ta^{0.57}$, and, although it is not included in the figure because of the jumps among modes, that for $-\omega_c$ gives $\omega_c = -14.21 \times Ta^{0.33}$. The power 0.57 is clearly lower than 2/3, the leading order of the asymptotic expansions given theoretically by [94, 14, 32] for the spiralling columnar modes, and than 0.63, found numerically in [79] for the outer equatorially-attached modes. However, it is important to notice that this characteristic is not exclusive of the antisymmetric convection. The same power rules for the inner modes attached to the external part of the tangent cylinder. See for instance Fig. 3.4 computed with the parameters of the outer Earth's core, i.e, with non-slip boundary conditions and differential heating ($R = 0$) for $\eta = 0.35$ and $\sigma = 0.1$.

Figure 3.3: The critical Rayleigh number R_c, and the critical precession frequency $-\omega_c$ plotted versus Ta, for $\eta = 0.4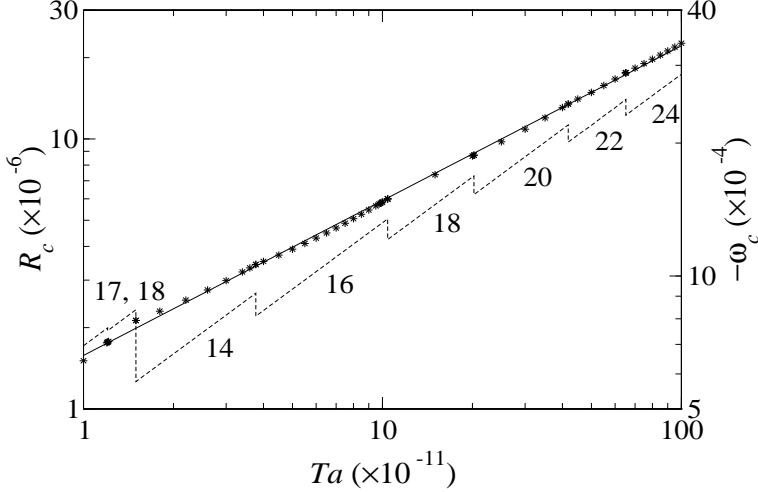$, $\sigma = 0.01$. The stars correspond to the computed R_c values, the solid line to its potential fitting, and the dashed line to $-\omega_c$. The left OY-axis scales R_c and the right OY-axis $-\omega_c$.

In both cases the flow is restricted to a small portion of the fluid at high latitudes. In contrast to what happen with the equatorially-attached modes of the same number of radial cells [79], or with the spiralling columnar modes, in Fig. 3.3 by increasing Ta, $-\omega_c$ decreases when m increases. In addition, only even modes become preferred. Probably, this fact is due simply to the geometry, since in the case of Fig. 3.4 the preferred modes are odd for $Ta > 10^{13}$. Despite these differences, all of them fulfill the power law $\omega_c \propto Ta^{1/3}$.

Fig. 3.5 shows the internal Rayleigh number R^m and ω^m versus Ta for the two lowest envelopes of the neutral stability curves of $m = 14$ and $m = 16$, again with $\eta = 0.4$ and $\sigma = 0.01$. The solid lines correspond to $m = 14$, and the dashed to $m = 16$. Notice that the thick lines, which are the lowest envelopes for each mode, give the R_c^m and ω_c^m of Fig. 3.1. The thin lines refer to the second modes in becoming unstable and contributing to the convection. At $Ta = 10^{11}$ the second eigenfunction is already antisymmetric for $m = 14$ (thin solid line), but symmetric and bicelullar for $m = 16$ (thin dashed line). The curve of bicelullar modes cross that of antisymmetric modes at $Ta = 1.28 \times 10^{11}$. In both cases ($m = 14, 16$), by increasing R^m the new interchange of eigenfunctions takes place in the opposite way than following the thick curves, i.e, the antisymmetric polar modes are superseded by those equatorially trapped [121]. However, the latter are soon replaced by polar symmetric modes (see Fig. 3.6), at $Ta \simeq 1.78 \times 10^{11}$ for $m = 14$, and at $Ta \simeq 2.27 \times 10^{11}$ for $m = 16$. We have checked that it is so at least up to $Ta = 10^{13}$. The two types of polar solutions have the kinetic energy density confined between $\theta \simeq 13^o$ and $\theta \simeq 21^o$, remaining almost z-independent. In addition their phase speeds differ less than
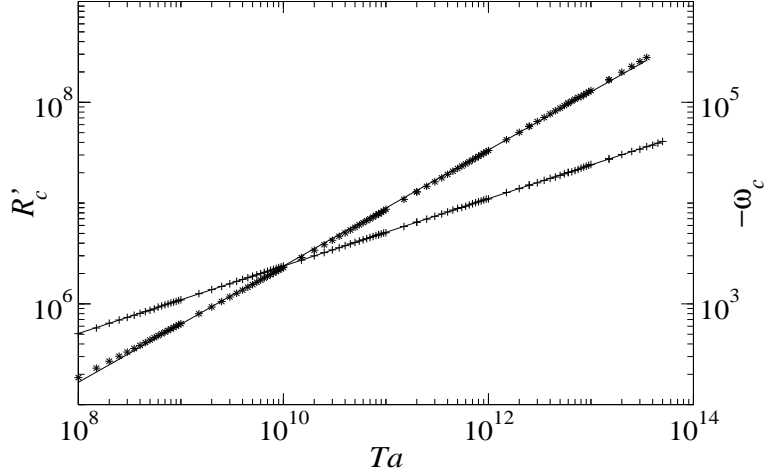
Figure 3.4: The critical Rayleigh number R'_c, and the critical precession frequency $-\omega_c$ plotted versus Ta, for $\eta = 0.35$, $\sigma = 0.1$, non-slip boundary conditions and differential heating. The stars and plus symbols correspond to the computed R'_c and $-\omega_c$ values, respectively, and the solid lines to their potential fitting. The left OY-axis scales R'_c and the right OY-axis $-\omega_c$.

0.5%, but the Rayleigh number of the symmetric solutions increases by 12%.

In this *letter* we have reported the existence of preferred polar antisymmetric modes of thermal convection, and determined numerically the power law dependence of R_c with Ta. Moreover, from the preceding results we can conclude that, although we have not found preferred symmetric polar modes, it is possible they become critical for other parameters. Furthermore, the same could happen with other families of modes, not shown in this *letter*, but found at not very high R. For instance, the third and fourth preferred eigenfunctions for $m = 18$ at $Ta = 2 \times 10^{12}$ are antisymmetric and symmetric polar bicellular modes, respectively. We have also detected antisymmetric equatorially trapped modes similar to the inertial waves of [119], and tri and quadricellular modes like those described in [3].

Assuming the equatorial symmetry in the study of low-σ convection in spherical geometries of moderate η can lead to wrong solutions, although they can be indistinguishable from the real flows when they are observed from a pole (compare Fig. 3.2 and Fig. 3.6).

The non-preferred modes of convection may also contribute to nonlinear flows. The superposition of the polar symmetric and antisymmetric nonlinear waves may give rise to coherent polygonal structures without equatorial symmetry, of the type observed in the north pole of Saturn.

46
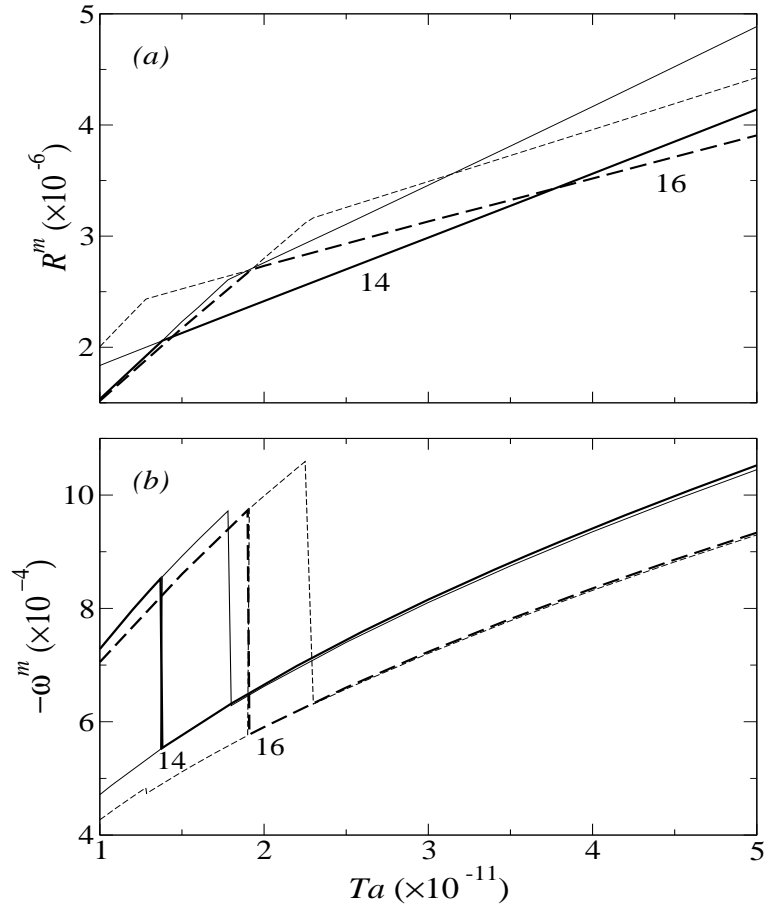
Figure 3.5: (a) The Rayleigh number R^m, and (b) the precession frequency $-\omega^m$, of the first and second envelopes of the neutral stability curves of $m = 14$, and $m = 16$, plotted versus Ta, for $\eta = 0.4$, $\sigma = 0.01$.
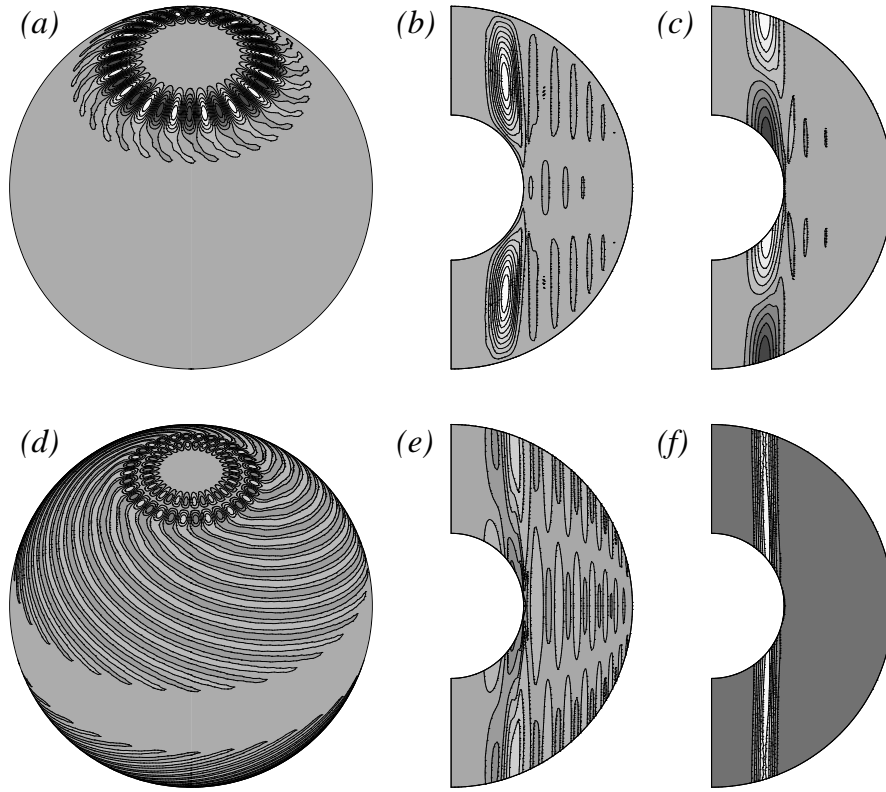
Figure 3.6: Symmetric polar mode of convection with azimuthal wavenumber $m = 16$ at $Ta = 4 \times 10^{11}$, for $\eta = 0.4$, $\sigma = 0.01$. Same contour plots as in Fig. 3.2.

Finally, the preceding results motivate the study of the *laminar* nonlinear dynamics of the polar flows, and their implication in the generation of planetary and stellar magnetic fields. The polar convection is known to grow in strongly nonlinear regimes [81, 29]. However, our preliminary nonlinear results indicate the existence of polar antisymmetric flows at very low Rayleigh numbers.

Acknowledgments

Chapter 4

A comparison of high-order time integrators for thermal convection in rotating spherical shells

A comparison of high-order time integrators for thermal convection in rotating spherical shells

ABSTRACT

A numerical study of several time integration methods for solving the three-dimensional Boussinesq thermal convection equations in rotating spherical shells is presented. Implicit and semi-implicit time integration techniques based on backward differentiation and extrapolation formulae are considered. The use of Krylov techniques allows the implicit treatment of the Coriolis term with low storage requirements. The codes are validated with a known benchmark, and their efficiency is studied. The results show that the use of high order methods, especially those with time step and order control, increase the efficiency of the time integration, and allows to obtain more accurate solutions.

Ferran Garcia, Marta Net, Bosco García-Archilla and Juan Sánchez.

4.1 Introduction

The study of the thermal convection in rotating spherical geometries is fundamental to explain many geophysical and astrophysical phenomena, such as the generation of the magnetic fields, or the differential rotation observed in the atmosphere of the major planets. The difficulties related to the experimental studies enhance the importance of three-dimensional numerical simulations in these fields. For this reason the development and improvement of the numerical techniques is basic for this research.

Most of the numerical papers dealing with spherical geometries (see [69, 112, 76, 29], among others), employ second order time integrators to find periodic, quasi-periodic, or even chaotic attractors by direct numerical simulation (DNS). The exclusive use of time integrations is not sufficient to establish the origin of the laminar flows and their dependence on the parameters, specially in the case of subcritical or multicritical transitions. In these situations pseudo-arclength continuation methods [68, 98, 99], and the linear stability analysis of the time dependent solutions [100, 40] provide powerful instruments to clarify the dynamics. In addition, the computation of Liapunov exponents is used to assess if a system is chaotic. All these tools have been historically applied to low-dimensional systems, but the increase in computational power, and the advances in numerical linear algebra have motivated their use in high dimensional systems. However, these techniques require to find accurate solutions, and, in finding these, higher order time integration methods may prove advantageous. In fact, even if one is only interested in low precision time evolutions, high order methods can sometimes be more efficient than low order ones.

For the spatial discretization of the equations on the sphere, many of the above mentioned codes use pseudo-spectral methods based on spherical harmonics basis functions, which provide high accurate solutions with relatively few grid points [20]. These methods are based on transformations from the spectral to the physical space [83]. The calculation of the quadratic terms, appearing in the truncated equations, is performed in the physical space.

For time-integration, some authors [44, 112, 30], among others, use semi-implicit integration methods, namely, they treat the linear terms (except the Coriolis term) implicitly with a Crank-Nicholson scheme, and the non-linear and the Coriolis terms explicitly with an Adams-Bashforth method [45]. In [112], a mixed Euler and integrating factor technique is also considered. Others, as in [62] use a modified second order accurate Runge-Kutta method. In all these methods the linear systems of equations to be solved at every step can be separated into spherical harmonic components, which can be solved independently, so that only a set of small linear systems must be solved at each time step. At high rotation rates, though, when the Coriolis force dominates over the viscous forces, stability considerations lead to extremely small time steps. To avoid this situation, the Coriolis term can be treated implicitly [45, 62]. Then the equations for a given spherical harmonic of order m are coupled for all degrees l. This implies that a large block-banded linear system has to be solved for each order m. As it was described in [62], solving them by

means of direct factorization methods becomes prohibitive if relatively high resolutions are employed.

In this paper we study the efficiency of several time integration methods, with the Coriolis term treated either as semi-implicit or as fully implicit, giving rise to the different algorithms presented. We show that (preconditioned) Krylov or Arnoldi iterative methods can be used to solve the Coriolis-implicit linear system, reducing considerably the amount of memory required.

The time integration methods we use are based on the backward differentiation formulae (BDF) implemented both with constant stepsize and variable stepsize and variable order (VSVO). The BDF are widely used in very stiff problems [55]. Being fully-implicit methods, they are usually implemented with a Newton-Krylov method for solving the non-linear equations [61]. In real geophysical and astrophysical flows, however, the non-linear terms are very costly to evaluate, and thus, in our time integration codes, we apply semi-implicit backward differentiation-extrapolation (IMEX-BDF) formulae, where the non-linear terms of the equations are treated explicitly.

The rest of the article is organized as follows. In Section 4.2, the formulation of the problem and the spatial discretization of the equations are introduced. In Section 4.3, the time discretization schemes and some details of their implementation are described. Section 4.4 contains a comparison with a previous benchmark [30]. In Section 4.5 the results of the study are presented, that is, the efficiency of the time integrators is compared and analyzed, the differences between the constant stepsize methods are shown, and the study of the implicit and semi-implicit variable stepsize and variable order methods is reported. Finally, the paper closes in Section 4.6 with a brief summary of the main conclusions.

4.2   Mathematical model and spatial discretization

We consider the thermal convection of a spherical fluid shell differentially heated, rotating about an axis of symmetry with constant angular velocity $\mathbf{\Omega} = \Omega\mathbf{k}$, and subject to radial gravity $\mathbf{g} = -\gamma\mathbf{r}$, where γ is a constant and $\mathbf{r}$ the position vector. The mass, momentum and energy equations are written, in a rotating frame of reference with angular velocity $\mathbf{\Omega}$, using the same formulation and non-dimensional units as in [106]. The units are the gap width, $d = r_o - r_i$, for the distance, $\nu^2/\gamma\alpha d^4$ for the temperature, and d^2/ν for the time, r_i and r_o being the inner and outer radii, respectively, ν the kinematic viscosity, and α the thermal expansion coefficient.

The velocity field is expressed in terms of toroidal, Ψ, and poloidal, Φ, potentials

$$\mathbf{v} = \nabla \times (\Psi\mathbf{r}) + \nabla \times \nabla \times (\Phi\mathbf{r}). \tag{4.1}$$

Consequently, the equations for both potentials, and the temperature perturbation, $\Theta =$

$T - T_c$, from the conduction state $\mathbf{v} = \mathbf{0}, T = T_c(r)$, with $r = ||\mathbf{r}||_2$, are

$$\left[(\partial_t - \nabla^2)L_2 - 2E^{-1}\partial_\varphi\right]\Psi = -2E^{-1}\mathcal{Q}\Phi - \mathbf{r} \cdot \nabla \times (\boldsymbol{\omega} \times \mathbf{v}), \tag{4.2}$$

$$\left[(\partial_t - \nabla^2)L_2 - 2E^{-1}\partial_\varphi\right]\nabla^2\Phi + L_2\Theta = 2E^{-1}\mathcal{Q}\Psi + \mathbf{r} \cdot \nabla \times \nabla \times (\boldsymbol{\omega} \times \mathbf{v}), \tag{4.3}$$

$$\left(\sigma\partial_t - \nabla^2\right)\Theta - R\eta(1-\eta)^{-2}r^{-3}L_2\Phi = -\sigma(\mathbf{v} \cdot \nabla)\Theta, \tag{4.4}$$

where $\boldsymbol{\omega} = \nabla \times \mathbf{v}$ is the vorticity.

The parameters of the problem are the Rayleigh number R, the Prandtl number σ, the Ekman number E, and the radius ratio η. They are defined by

$$R = \frac{\gamma\alpha\Delta T d^4}{\kappa\nu}, \quad E = \frac{\nu}{\Omega d^2}, \quad \sigma = \frac{\nu}{\kappa}, \quad \eta = \frac{r_i}{r_o}, \tag{4.5}$$

where κ is the thermal diffusivity, and ΔT the difference of temperature between the inner and outer boundaries.

The operators L_2 and $\mathcal{Q}$ are defined by $L_2 \equiv -r^2\nabla^2 + \partial_r(r^2\partial_r)$, $\mathcal{Q} \equiv r\cos\theta\nabla^2 - (L_2 + r\partial_r)(\cos\theta\partial_r - r^{-1}\sin\theta\partial_\theta)$, (r, θ, φ) being the spherical coordinates, with θ measuring the colatitude, and φ the longitude. In non-dimensional units the conduction state reads $T_c(r) = T_0 + R\eta/\sigma(1-\eta)^2r$.

Non-slip perfect thermally conducting boundaries $\Phi = \partial_r\Phi = \Psi = \Theta = 0$ are used.

A standard treatment of the spatial dependence of the equations is used, so we comment only on the basic points (see e.g., [44, 112], for details). The functions $X = (\Psi, \Phi, \Theta)$ are expanded in spherical harmonic series up to degree L, namely

$$X(t, r, \theta, \varphi) = \sum_{l=0}^{L}\sum_{m=-l}^{l} X_l^m(r,t)Y_l^m(\theta, \varphi), \tag{4.6}$$

with $\Psi_l^{-m} = \overline{\Psi_l^m}$, $\Phi_l^{-m} = \overline{\Phi_l^m}$, $\Theta_l^{-m} = \overline{\Theta_l^m}$, $\Psi_0^0 = \Phi_0^0 = 0$ to uniquely determine the two scalar potentials, and $Y_l^m(\theta, \varphi) = P_l^m(\cos\theta)e^{im\varphi}$, P_l^m being the normalised associated Legendre functions of degree l and order m. The equations (4.2)-(4.4) written for the complex coefficients become

$$\partial_t\Psi_l^m = \mathcal{D}_l\Psi_l^m + \frac{1}{l(l+1)}\left[2E^{-1}\left(im\Psi_l^m - [\mathcal{Q}\Phi]_l^m\right) - [\mathbf{r} \cdot \nabla \times (\boldsymbol{\omega} \times \mathbf{v})]_l^m\right], \tag{4.7}$$

$$\partial_t\mathcal{D}_l\Phi_l^m = \mathcal{D}_l^2\Phi_l^m - \Theta_l^m + \frac{1}{l(l+1)}\left[2E^{-1}\left(im\mathcal{D}_l\Phi_l^m + [\mathcal{Q}\Psi]_l^m\right)\right.$$
$$\left. + [\mathbf{r} \cdot \nabla \times \nabla \times (\boldsymbol{\omega} \times \mathbf{v})]_l^m\right], \tag{4.8}$$

$$\partial_t\Theta_l^m = \sigma^{-1}\mathcal{D}_l\Theta_l^m + \sigma^{-1}l(l+1)R\eta(1-\eta)^{-2}r^{-3}\Phi_l^m - [(\mathbf{v} \cdot \nabla)\Theta]_l^m, \tag{4.9}$$

with boundary conditions

$$\Psi_l^m = \Phi_l^m = \partial_r\Phi_l^m = \Theta_l^m = 0. \tag{4.10}$$

The spherical harmonic coefficients of the operator $\mathcal{Q} = \mathcal{Q}^u + \mathcal{Q}^l$ are

$$[\mathcal{Q}^u f]_l^m = -l(l+2)c_{l+1}^m D_{l+2}^+ f_{l+1}^m, \quad [\mathcal{Q}^l f]_l^m = -(l-1)(l+1)c_l^m D_{1-l}^+ f_{l-1}^m, \qquad (4.11)$$

$$\text{with} \quad D_l^+ = \partial_r + \frac{l}{r}, \quad c_l^m = \left(\frac{l^2 - m^2}{4l^2 - 1}\right)^{1/2}, \text{ and } \quad \mathcal{D}_l = \partial_{rr}^2 + \frac{2}{r}\partial_r - \frac{l(l+1)}{r^2}. \quad (4.12)$$

In the radial direction, a collocation method on a Gauss-Lobatto mesh of $N_r + 1$ points is employed ($N_r - 1$ being the inner number of points).

The spherical harmonic coefficients of the non-linear terms in equations (4.7-4.9) are obtained following [44]. The velocity and vorticity fields are computed first on a collocation mesh in the three coordinates (r, θ, φ) with the help of dealiased Legendre and fast Fourier transforms [13]. The cross product is computed on the mesh, and, finally, transformed to the spectral space. The computation of the coefficients of the non-linear terms of equation (4.9) requires the evaluation of the inner product $(\mathbf{v} \cdot \nabla)\Theta$ on the collocation mesh, and then to transform back to the spectral space.

The mode $m = l = 0$ is only nonzero for Θ, and the amplitudes for $m = 0$ are real. With these considerations, a large system of ordinary differential equations of dimension $N = (3L^2 + 6L + 1)(N_r - 1)$ must be integrated in time.

4.3   Time integration methods

In order to simplify the notation, equations (4.7-4.9) are written in the form

$$\mathcal{L}_0 \dot{u} = \mathcal{L}u + \mathcal{N}(u), \qquad (4.13)$$

where $u = (\Psi_l^m(r_i), \Phi_l^m(r_i), \Theta_l^m(r_i))$ is the vector containing the values of the spherical harmonic coefficients at the inner radial collocation points, and $\mathcal{L}_0$ and $\mathcal{L}$ are linear operators including the boundary conditions. The former is invertible. It is the identity for Ψ_l^m and Θ_l^m, and the operator $\mathcal{D}_l$ for Φ_l^m (see the time derivatives in eqs. (4.7–4.9)). The operator $\mathcal{L}$ includes the terms $\mathcal{D}_l\Psi_l^m$, $\mathcal{D}_l^2\Phi_l^m - \Theta_l^m$, and $\sigma^{-1}\mathcal{D}_l\Theta_l^m + \sigma^{-1}l(l+1)R\eta(1-\eta)^{-2}r^{-3}\Phi_l^m$ of eqs. (4.7), (4.8) and (4.9), respectively, in any of the schemes used, and part of the Coriolis terms to be specified below. The operator $\mathcal{N}$, which will be treated explicitly in the IMEX-BDF formulae, will always contain the non-linear, and the rest of the Coriolis terms.

Time integration methods look for approximations $u^n \approx u(t_n)$ to solutions of (4.13) at time levels $t_n = t_{n-1} + \Delta t_{n-1}$, $n = 1, 2, \ldots$, . The IMEX-BDF formulae mentioned before are related to the backward differentiation formulae (BDF) [31, 55]. These are collocation multistep methods, which obtain u^{n+1} from the previous approximations u^{n-j}, $j = 0, 1, \ldots, k - 1$, k being the number of steps in the formula, in the way we now describe. Consider first the interpolating polynomial $q_{n,k}$ of degree at most k, such that $q_{n,k}(t_{n-j}) = u^{n-j}$, for $j = -1, 0, \ldots, k - 1$, where the unknown u^{n+1} is included. The
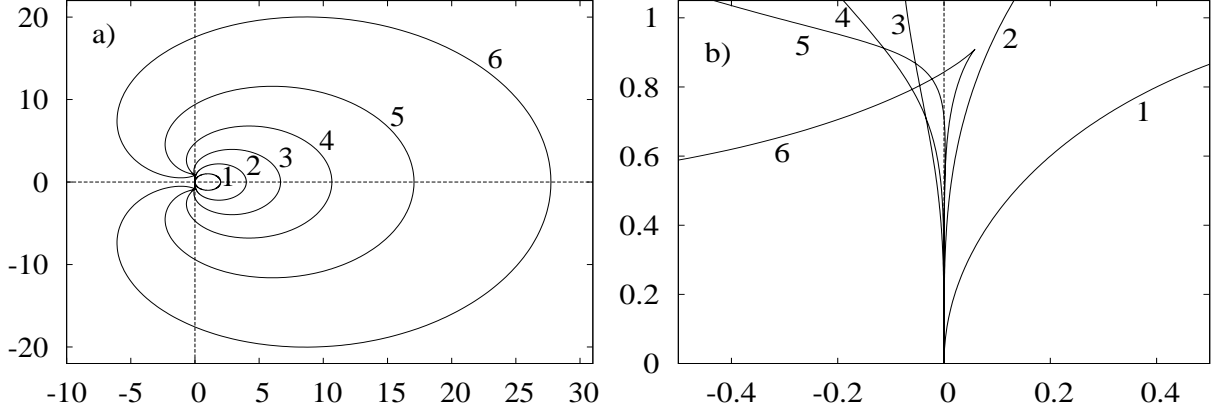
Figure 4.1: a) Absolute stability regions of the BDF up to order six with constant stepsize. The regions are the exterior of the closed curves. b) Detail close to the origin.

value of the latter is determined by requiring that $q_{n,k}$ satisfies the differential equation at $t = t_{n+1}$, that is, $\mathcal{L}_0 \dot{q}_{n,k}(t_{n+1}) = \mathcal{L}u^{n+1} + \mathcal{N}(u^{n+1})$. In order to avoid solving a full non-linear problem at every step, in the IMEX-BDF formula [67] the term $\mathcal{N}(u^{n+1})$ is replaced by the extrapolated value

$$\mathcal{L}_0 \dot{q}_{n,k}(t_{n+1}) = \mathcal{L}u^{n+1} + p_{n,k-1}(t_{n+1}), \tag{4.14}$$

where $p_{n,k-1}$ is the interpolating polynomial of degree at most $k-1$, such that $p_{n,k-1}(t_{n-j}) = \mathcal{N}(u^{n-j})$, for $j = 0, 1, \ldots, k-1$. It is useful to express $q_{n,k}(t)$ as the sum $q_{n,k}(t) = q_{n,k}^0(t) + u^{n+1} l_{n,k}(t)$, where $l_{n,k}$ is the polynomial of degree at most k taking value 1 at t_{n+1}, and 0 at t_{n-j}, for $j = 0, 1, \ldots, k-1$. Consequently $q_{n,k}^0$ coincides with $q_{n,k}$ at t_{n-j}, for $j = 0, 1, \ldots, k-1$, and takes value 0 at t_{n+1}. Then, the linear system to be solved in order to find u^{n+1} can be expressed as

$$\left(\mathcal{I} - \frac{\Delta t_n}{\gamma_0(n)} \mathcal{L}_0^{-1} \mathcal{L} \right) u^{n+1} = \frac{\Delta t_n}{\gamma_0(n)} \mathcal{L}_0^{-1} p_{n,k-1}(t_{n+1}) - \frac{\dot{q}_{n,k}^0(t_{n+1})}{\dot{l}_{n,k}(t_{n+1})}, \tag{4.15}$$

where $\mathcal{I}$ is the identity operator, and $\gamma_0(n) = \dot{l}_{n,k}(t_{n+1}) \Delta t_n$. The practical computation of $\gamma_0(n)$, $\dot{q}_{n,k}^0(t_{n+1})/\dot{l}_{n,k}(t_{n+1})$, etc. can be done systematically and easily by following ideas in [54, **S** III.5].

When the time steps are constant ($\Delta t_n = \Delta t$) the expressions (4.14) and (4.15) become simpler. For example, (4.15) becomes

$$\left(\mathcal{I} - \frac{\Delta t}{\gamma_0} \mathcal{L}_0^{-1} \mathcal{L} \right) u^{n+1} = \sum_{i=0}^{k-1} \frac{\alpha_i}{\gamma_0} u^{n-i} + \sum_{i=0}^{k-1} \frac{\beta_i \Delta t}{\gamma_0} \mathcal{L}_0^{-1} \mathcal{N}(u^{n-i}), \tag{4.16}$$

where the coefficients α_i, β_j and γ_0 do not depend on n, and are listed, for instance, in [98]. It is well known that, if the step number k satisfies that $k \leq 6$, and the time steps are

constant or the ratios $\Delta t_n/\Delta t_{n-1}$ are kept adequately bounded (see e.g. [54, **S** III.3]), the BDF formulae are 0-stable, i.e., they propagate the errors in a stable way. If $k \geq 7$ they are unstable and useless for time integration. It is also known that for $k \leq 6$ the k-step formula is convergent of order k, that is, the errors $u(t_n) - u^n$ are $O((\max_{0 \leq j \leq n-1} \Delta t_j)^k)$. Therefore the k-step formula is also termed the k-th order formula. Their regions of absolute stability are shown in Fig. 4.1. A point $z = \lambda \Delta t$ is in the absolute stability region of a numerical time integration method, if when it is applied to the equation $y' = \lambda y$, $\lambda \in \mathbb{C}$, with fixed stepsize Δt, the numerical solution y_n is bounded as $n \to \infty$. For the BDF formulas, the regions are the exterior of the closed curves shown in Fig. 4.1a. Fig. 4.1(b) shows a detail of them close to the origin. Their knowledge helps to understand the behaviours described in Section 4.5.

Notice that contrary to the variable stepsize case (4.15), the matrix in the system to be solved at every step in (4.16) is constant for all n, being this property one of the main attractions of constant stepsize formulae. On the other hand, the main reason to use variable stepsizes is to maintain local (time-integration) errors below a desired tolerance, with the aim of maintaining a given accuracy along all the time interval considered. Changing the stepsize can also be combined with using formulae of different orders (step numbers) k, while maintaining accuracy. In this way, the integration can be started with $k = 1$ (and small Δt_0), when the lack of previously computed values precludes the use of higher order formulae, and then increase the order (and the stepsize) as the integration advances and previous approximations u^{n-j} are available. The strategy of variable stepsize and variable order (VSVO) codes we have used is outlined in the following paragraph. Let us mention first that for the fixed-stepsize codes, the starting values u^j, $j = 1, \ldots, k-1$ are obtained by time integration from t_{j-1} to t_j with a VSVO code and sufficiently small tolerances ε^a and ε^r (see (4.18) below for the meaning of these values).

As explained in [54, **S** III.7] the local error of the k-step BDF can be estimated as $\Delta t_n [\dot{q}_{n,k+1}(t_{n+1}) - \dot{q}_{n,k}(t_n)]$, so that, in a similar way, that of the k-step IMEX-BDF formula can be estimated as

$$\text{EST}^n(k) = \Delta t_n \left[\mathcal{L}_0^{-1} \big(p_{n,k}(t_{n+1}) - p_{n,k-1}(t_{n+1}) \big) - \big(\dot{q}_{n,k+1}(t_{n+1}) - \dot{q}_{n,k}(t_n) \big) \right]. \quad (4.17)$$

The corresponding factor for the stepsize, i.e., the factor by which Δt_n is multiplied to obtain the next time stepsize Δt_{n+1}, is

$$f^n(k) = min \left(f_{max}(k), \max_{i=1,\ldots,N} \left(\frac{\varepsilon^r |u_i^{n+1}| + \varepsilon^a}{\text{EST}_i^n(k)} \right)^{1/(k+1)} \right), \quad (4.18)$$

where ε^a and ε^r are, respectively, the tolerances below which the absolute and relative values of the local (time discretization) errors are required, i indexes the components of the vectors, and $f_{max}(k)$ are upper bounds for the time step ratio $\Delta t_{n+1}/\Delta t_n$, which are used in practical implementations of the BDF methods to enhance their stability properties (see

[54, **S** III.5]). They depend on the order k, and we have taken $f_{max}(k) = 5, 4, 3, 2, 1.6, 1.4$ for $k = 1, \ldots, 6$, after diverse numerical experiments.

If $f^n(k) < 1$, it is assumed that the local error is unacceptably large, and the computed value u^{n+1} is rejected and recomputed with a smaller stepsize, namely, $\Delta t'_n = 0.9 f^n(k) \Delta t_n$ (0.9 being a safety factor). If, on the contrary, $f^n(k) > 1$, u^{n+1} is considered sufficiently accurate, so that the integration can proceed to find a new approximation u_{n+2}. Before doing so, the estimations $\text{EST}^n(k-1)$ and $\text{EST}^n(k+1)$ of the local error of the $(k-1)$-step and $(k+1)$-step formulae are obtained, together with their corresponding factors $f^n(k-1)$ and $f^n(k + 1)$. The value $k' = k - 1, k$, or $k + 1$ corresponding to the largest factor is chosen for the next step (always with k' not larger than a previously fixed maximum allowed order), and u_{n+2} is computed with the k'-step formula, and with stepsize equal to $\Delta t_{n+1} = 0.9 f^n(k') \Delta t_n$.

Whether the k-step IMEX-BDF formula is used with variable or constant stepsizes, the linear systems (4.15) or (4.16) have to be solved to compute u^{n+1}. In the sequel, we will represent these systems as

$$\mathcal{H} u^{n+1} = v^n, \tag{4.19}$$

$\mathcal{H}$ being the radial discretization of $\mathcal{I} - h\mathcal{L}_0^{-1}\mathcal{L}$ with $h = \Delta t_n/\gamma_0(n)$ or $h = \Delta t/\gamma_0$, and we comment now on their effective solution. We first notice that once v^n is evaluated, equation (4.19) decouples for each azimuthal wave number m, thus, at every time step, $L + 1$ linear systems of the form

$$H^m U^m = V^m, \quad m = 0, ..., L \tag{4.20}$$

have to be solved. The vectors U^m and V^m contain, respectively, the unknowns and the right hand side of (4.19) with azimuthal wave number m. The structure of the matrices H^m depend on which terms of eqs. (4.7-4.9) are included in the operator $\mathcal{L}$, that is, on which terms are treated implicitly (of course they depend also on the way the unknowns are ordered). We describe now the main features of these matrices for all the methods tested. They are summarised in tables 4.1 and 4.2. The full details have been left to the Appendix.

The inclusion in $\mathcal{L}$ of the diagonal parts of the Coriolis term (from now on referred to as $\mathcal{C}^d$) containing $im\Psi_l^m$ and $im\Phi_l^m$ (see eqs. (4.7-4.8)), and of $\mathcal{Q}$ in $\mathcal{N}$, gives block-diagonal matrices H^m, whose memory requirements are $O(N_r^2(L + 1)(L + 2))$ due to the triangular truncation (4.6), namely, quadratic in the spherical harmonic truncation order. From now on, the time discretization with this treatment of the operators will be called the $\mathcal{Q}$-explicit method.

Several authors (see [44, 112], and [62] in the low rotation method section, among others) take the Coriolis term fully explicit in their codes, then the linear systems (4.19) decouple also for the spherical harmonic degree l, and all the H^m are equal. They have

58

block-diagonal structure, each block depending on l. This means that matrices of dimension N_r^2 must be inverted using LU factorisations for each degree l. This approach is efficient in memory requirements since only a storage $O(N_r^2(L+1))$ is needed. However, at low Ekman and high Rayleigh numbers (those with realistic significance) very small time steps are needed to satisfy the CFL condition.

Taking into account that, in principle, the more implicit the method is, the larger the stepsize can be, and that, for the problem we are considering, the evaluation of the non-linear terms is expensive in CPU time, it could be computationally efficient to include other parts of the Coriolis term in the linear operator $\mathcal{L}$.

By adding $\mathcal{Q}^u$ or $\mathcal{Q}^l$ (see eq. 4.11) to $\mathcal{L}$, the matrices H^m become upper or lower block-Hessenberg matrices, respectively. They can be solved with the same memory requirements and number of operations as the Q-explicit method, because $\mathcal{Q}^u$ and $\mathcal{Q}^l$ must be evaluated in any case. The systems (4.20) are solved by using backward or forward block substitution, by inverting the same diagonal submatrices as in the Q-explicit method. In order to implement this possibility in a symmetric way, the two options are used alternately, that is, one step is performed with $\mathcal{Q}^u$ implicit and $\mathcal{Q}^l$ explicit, and vice-versa in the following step. From now on, this time discretization will be called the Q-splitting method.

By setting $\mathcal{Q}$ totally implicit the operator $\mathcal{N}$ only includes the non-linear terms. Using this scheme the matrices H^m become block-tridiagonal. A direct block method for solving these linear systems involves a big amount of memory storage. A LU factorisation requires the storage of block matrices with a memory size $O(N_r^2(L-m+1)^2)$ for each $m = 0, \cdots, L$, due to the change of the inner block structure during the Gaussian elimination. Therefore, the total amount of memory needed is cubic in the spherical harmonic truncation parameter L. The storage of these matrices becomes prohibitive for high resolutions when dealing with direct solvers [45, 62]. However it will be seen that *matrix-free* methods based on Krylov techniques [96], GMRES [97] in our case, can be used efficiently, to solve the block-tridiagonal linear systems with the same memory requirements as in the Q-explicit case. The increase of the cost in solving the linear systems (4.20) may be offset by the increase of the time stepsize. The initial approximation for the solution of the linear system is obtained by extrapolation from the previous steps, as is usually done in the implementations of the BDF. From now on this method will be called the Q-implicit method.

The key point for using Krylov methods efficiently is to have a good preconditioner. It has to resemble as much as possible the original matrix, and the linear system with the preconditioning matrix must be easy to solve. We have used the block-diagonal, or the upper or lower block-triangular parts of H^m as preconditioners. All of them can be solved by an adapted direct LU method. We have concluded that adding the non-diagonal blocks does not improve significantly the convergence, and then only the results for the block-diagonal preconditioner will be reported.

As it is well-known, time integration with a constant stepsize can be unnecessarily

expensive because the stepsize has to be taken small enough to cope with possible fast transients. VSVO methods [54], though, do not suffer from this shortcoming. The price to pay is that the matrices of the linear systems to be solved at every step depend on the current time step, which changes during the integration. In order to avoid doing a massive number of factorisations, Krylov methods can be used with a preconditioning matrix depending on a fixed time stepsize Δt^*. At each step, the number of iterations needed by the iterative method will (partly) depend on how close the current Δt is to Δt^*. When the convergence of the iterative linear solver degrades the preconditioning matrix can be updated with the current time step. This is the strategy we have followed for the all the VSVO algorithms employed (see the Appendix for more details). In the case of the IMEX-BDF methods, the preconditioner is updated if the number of iterations of the linear solver exceeds a certain number (10 in all the calculations described).

All the semi-implicit methods described before have been implemented with constant stepsize, and with variable time stepsize and order using our own codes (except the Q-splitting VSVO method, which has been used only with constant stepsizes). From now on, the VSVO implementations of the Q-explicit and Q-implicit methods will be called Q-explicit VSVO and Q-implicit VSVO, respectively.

The last option considered is to take also the non-linear terms implicitly with a VSVO formulation of the BDF. This leads to the solution of a non-linear system of equations at each step. Due to the stiffness of the problem, it cannot be efficiently solved by fixed point iteration, and it is then solved by an inexact Newton's method. At each Newton's iteration, linear systems of similar structure to (4.19), but now including the Jacobian of the non-linear operator $\mathcal{N}$, have to be solved. The matrix products by the Jacobian are computed by finite differencing, and the linear systems are solved by a Krylov method. See [61] for further details. From now on this method will be called fully implicit method. We used the DLSODPK code of the ODEPACK package [61]. In this case the matrices depend not only on the current time step, but also on the current solution. As before, they can be preconditioned by the block-diagonal, or by the upper or lower block-triangular matrices computed with a fixed time step Δt^*. Again, a criterion for recomputing the LU factorisations must be established. In our implementation, the preconditioner is updated only if the ratio between the current Δt and Δt^* exceeds 10 or is smaller than 1/10.

Observe that the fully implicit method requires one evaluation of the nonlinear terms per iteration of the linear solver, and a further one per iteration in Newton's method. Since, in this particular and other computational fluid dynamics problems the evaluation of the non-linear terms is expensive, semi-implicit are likely to outperform fully-implicit methods despite the smaller stepsizes Δt they may be forced to take due to their poorer stability properties.

Finally it must be stressed that besides preconditioning techniques, the efficiency of Newton-Krylov methods can be improved by using weighted norms in all error-like vectors, i.e., using vectorial tolerances to control the convergence. This has proved useful in our calculations only for the fully implicit method (see [61] for details of the implementation).

The weights we have used are given in Section 4.5.

Method	Terms in $\mathcal{L}$	Terms in $\mathcal{N}$	Symbol, line
Q-explicit	$\mathcal{C}^d$	$\mathcal{Q}^l$ and $\mathcal{Q}^u$	+, solid
Q-splitting	$\mathcal{C}^d$, $\mathcal{Q}^l$ and $\mathcal{Q}^u$ alternately	$\mathcal{Q}^u$ and $\mathcal{Q}^l$ alternately	×, dotted
Q-implicit	$\mathcal{C}^d$, $\mathcal{Q}^l$ and $\mathcal{Q}^u$		*, dashed
Q-explicit VSVO	$\mathcal{C}^d$	$\mathcal{Q}^l$ and $\mathcal{Q}^u$	+, solid
Q-implicit VSVO	$\mathcal{C}^d$, $\mathcal{Q}^l$ and $\mathcal{Q}^u$		*, dashed
DLSODPK	fully implicit		∘, dash-dotted

Table 4.1: Summary of methods tested for the equation $\mathcal{L}_0\dot{u} = \mathcal{L}u + \mathcal{N}(u)$. The second and third columns indicate which part of the Coriolis terms are treated implicitly (included in $\mathcal{L}$) and explicitly (included in $\mathcal{N}$), respectively. The operators $\mathcal{C}^d$, $\mathcal{Q}^l$ and $\mathcal{Q}^u$ represent, respectively, the diagonal part of the Coriolis terms, and the lower and upper parts of the operator $\mathcal{Q}$. The last column contains the symbols and lines used in Figs. 4.2 to 4.4 to label each method.

Method	Step and order	Linear solver	Preconditioner
Q-explicit	fixed	block-diag. LU	
Q-splitting	fixed	back and forward block substitution	
Q-implicit	fixed	GMRES	block-diag. LU
Q-explicit VSVO	variable	GMRES	block-diag. LU
Q-implicit VSVO	variable	GMRES	block-diag. LU
DLSODPK	variable	GMRES	block-diag. LU

Table 4.2: Summary of solvers used for the linear systems with matrix $\mathcal{H} = \mathcal{I} - h\mathcal{L}_0^{-1}\mathcal{L}$ and $h = \Delta t_n/\gamma_0(n)$ or $h = \Delta t/\gamma_0$ (see eqs. (4.15-4.16)). The first column indicates if the time step and the order are fixed or variable, the second indicates the linear solver, and the third the preconditioner used if the solver is iterative.

Tables 4.1 and 4.2 summarise, for future reference, the main features of the methods tested, and the linear solvers and preconditioner used for each one.

4.4 Comparison with the benchmark

To check our new non-linear code we have compared our results with the non-magnetic data of [30], where several research groups present the results of a benchmark study for a

convection-driven dynamo problem in a rotating spherical shell. Although the benchmark deals with the convergence and accuracy of several spatial discretizations of the problem, and not with the time integration, we have used the same parameters because they are well known to the community of people working on convection in spherical geometry, and because the kind of computation we are interested in (computation of periodic orbits and their stability, for instance) are always close to regimes which are not yet chaotic. Therefore, a transient to a periodic regime such as that of [30] is a good test for our purposes.

The values of the parameters in [30] are $\eta = 0.35$, $\sigma = 1$, $E = 10^{-3}$ and $R_B = 100$, R_B being the Rayleigh number defined in [30]. Taking into account the different non-dimensional units of both formulations, it turns out that $R = (1 - \eta)\sigma Ta^{1/2}R_B$, so $R = 65000$. In addition, the relation between the temperatures is $T = R\sigma^{-1}T_B$, T_B being the temperature scaled as in [30]. The initial condition recommended is velocity $\mathbf{v} = \mathbf{0}$, and temperature

$$T_B(r, \theta, \varphi) = \frac{r_i r_o}{r} - r_i + \frac{2A}{\sqrt{2\pi}}(1 - x^2)^3 P_m^m(\theta) \cos m\varphi, \qquad (4.21)$$

with $A = 0.1$, $x = 2r - r_i - r_o$, $m = 4$, and $P_m^m(\theta) = \sqrt{(2m+1)!!/2(2m)!!} \sin^m \theta$ the normalised associated Legendre function of order and degree m. By starting from it the solution tends, after an abrupt transient, to an azimuthal travelling wave of wave number $m = 4$.

In table 4 of [30] the authors provide the values and the estimated errors of some data corresponding to the reference solution, which can be considered that obtained with the pseudo-spectral code of Christensen, Wicht and Glatzmaier (CWG). They are the precession frequency ω, the mean kinetic energy density

$$E_{kin} = \frac{1}{2V_s} \int_{V_s} v^2 dV, \qquad (4.22)$$

V_s being the volume of the fluid shell, and the temperature T_B and the azimuthal velocity v_φ at the point $r = (r_i + r_o)/2$, $\theta = \pi/2$, whose φ-coordinate is given by $v_r(r, \theta, \varphi) = 0$ and $\partial_\varphi v_r(r, \theta, \varphi) > 0$. Their values are

$$\omega = 0.1824 \pm 0.0050, \quad E_{kin} = 58.348 \pm 0.050, \quad T_B = 0.42812 \pm 0.00012,$$
$$\text{and } v_\varphi = -10.1571 \pm 0.0020. \quad (4.23)$$

We have integrated with the spatial resolution which corresponds to the code by CWG, i.e., with $N_r = 24$, and a triangular truncation of the spherical harmonics (4.6) with $L = 32$. The total number of equations is $N = 75095$. From now on, the test case with this resolution, and the parameters of the benchmark will be called C_1.

Most of the groups in [30] used a second order Crank-Nicholson scheme for the implicit terms, and a second order Adams-Bashforth method for the explicit ones. Therefore, to

62

compare our solutions, it is enough to integrate using order $k = 2$. However we have checked that all the solutions described in section 4.5, for the C_1 case, give values inside the intervals (4.23), although in the case of VSVO methods only when local error tolerances were taken smaller than 10^{-2}. These results are not surprising because our codes are based on the same spatial discretization than those of the CWG group, who obtained the benchmark reference data, and we use at least second-order time integrators.

4.5   Results

To compare the time integration schemes presented in Sect. 4.3 we integrate the C_1 case of the previous section, and a second test also with $\eta = 0.35$ and $\sigma = 1$, but with $E = 10^{-4}$, $R = 800000$, $m = 7$, $N_r = 48$ and $L = 63$ (from now on it will be called the C_2 case). The total number of equations is now $N = 577442$. The higher resolution is needed to properly solve the problem with a lower Ekman number. The solution tends now to a travelling wave with $m = 7$. In both tests $R \simeq 1.78R_c$ has been selected to be at the same relative distance to the critical Rayleigh number, R_c, at which the convection sets in.

Since at the initial condition (4.21) the norm $||\partial_t u||_2$ is extremely large, there is a strong initial transient. This produces large irregular errors when the equations are integrated with a constant stepsize method. Therefore the very initial transient has been discarded to make our comparisons, and all the test runs of the time-stepping codes are started with the same initial condition obtained after the solution has smoothed by a short time integration t_0. To obtain it, the Q-implicit VSVO method with very low tolerances has been used. Then the system (4.7-4.10) is evolved from the new initial condition at time t_0 to a final time t_f, at which the solution, $u(t)$, is close to the periodic regime. As the limit solution is a wave, $||u||_2$ tends to a constant. In the C_1 case, $t_0 = 0.053$ and $t_f = 0.253$, and, in the C_2 case, $t_0 = 0.0112$ and $t_f = 0.12$. The time t_0 has been selected as the first time at which $||u(t_0)||_2 \simeq ||u(t_f)||_2$. After t_0, the norm of u increases until it reaches a maximum, and then finally decreases to a constant value close to $||u(t_f)||_2$.

To check the efficiency of the different schemes the relation between the relative error, and the run time in seconds, rt, is studied. The former is defined as

$$\varepsilon(u) = \frac{||u - u_r||_2}{||u_r||_2}, \tag{4.24}$$

where $u = u(t_f)$ is the solution we want to check, and $u_r = u_r(t_f)$ is an accurate reference solution obtained with the Q-implicit VSVO method, with tolerances for the relative and absolute errors set to $\varepsilon^r = \varepsilon^a = 10^{-14}$ and 10^{-13} in the C_1 and C_2 tests, respectively.

The decrease of the relative error (4.24) is achieved by decreasing the stepsize in the case of fixed stepsize methods, or by decreasing the tolerances for the local errors in the case of the VSVO methods. In our implementation of the semi-implicit VSVO methods we demand $f^n(k)$ in (4.18) to satisfy $f^n(k) < 1$. In the test we show below we have taken

$\varepsilon^a = \varepsilon^r$, since we checked that other choices of ε^a do not significantly alter the efficiency of the methods.

On the other hand, the fully implicit DLSODPK can use component-dependent tolerances ε_i^a for each component e_i of the estimated local error. In fact, if N is the size of the vector $u = (u_1, \ldots, u_N)^T$, DLSODPK requires

$$\left(\frac{1}{N} \sum_{i=1}^{N} \left(\frac{|e_i|}{\varepsilon_i^r |u_i| + \varepsilon_i^a} \right)^2 \right)^{1/2} < 1, \tag{4.25}$$

which is in general less demanding than having $f^n(k) < 1$ as in our semi-implicit methods. For the tests we present below, we have taken the ε_i^a proportional to ε^r in the following way: $\varepsilon_i^a = f_\psi \varepsilon^r$, $\varepsilon_i^a = f_\phi \varepsilon^r$ and $\varepsilon_i^a = f_\Theta \varepsilon^r$, for the blocks corresponding to the variables ψ, ϕ and Θ, respectively. The factors used are $f_\psi = 10$, $f_\phi = 1$, and $f_\Theta = 10^4$, which correspond to the orders of magnitude of the variables. We checked that the efficiency of the method was improved only marginally if other values of ε_i^a were used.

All the calculations have been performed on a cluster of Pentium IV personal computers using a single processor at 3.4 GHz, with 2 Mb of cache memory. The FFT routines used to evaluate the non-linear terms are those of the FFTW3 library [36]. The action of the linear operators required by the linear solvers, the computation of the non-linear terms, and the Legendre transforms have been implemented as matrix-matrix products using the GotoBLAS library [48] to increase the efficiency of the codes. The block structure of some of the matrices is used to minimise the number of operations. For instance, when matrix products by H^m in any of its variants are required, they are computed by using matrix-matrix subroutines for their sub-blocks (see the Appendix). It is easy to see that each sub-matrix of dimension $N_r - 1$ of H^m always multiplies the real and imaginary parts of the same spherical harmonic coefficient of one of the functions Ψ, Φ or Θ. These products and any other for which it is possible, are grouped to optimise the memory transfers.

Figs. 4.2(a) and (b) show the dependence of $\log \varepsilon(u)$ against the logarithm of the constant time step Δt for the cases C_1 and C_2, respectively. The order of the methods can be seen in the slopes of the curves, and it is indicated beside. For every order k in Figs. 4.2(a) and (b), the largest values of Δt in each curve are those in the limit of (absolute) stability of the corresponding k-th order formula, that is, the largest value in the increasing sequences of Δt which we tried that did not produced an overflow due to instability. The negative slopes at the left of the $k = 5$ curves correspond to the accumulation of rounding errors, which grows as Δt decreases. We have checked that, for very low Δt, the slope corresponds to a power -1, as is expected when the accumulated rounding error is proportional to the number of time steps, i.e., to Δt^{-1} (See e.g., [58, S1.4-2]). The combination of the Crank-Nicholson and Adams-Bashforth schemes used by some authors would be comparable to the $k = 2$ case. Although the coefficient of the leading term of the local truncation error is lower for the Crank-Nicholson method

(1/12) than for the second order BDF formula (2/9), the former requires an additional evaluation of the linear terms of the equations. Moreover if the Crank-Nicholson method is used with fixed time step, oscillations due to eigenvalues of large and negative real parts inherent to the scheme are not damped [4, **S**3.6]. To solve this, usually, either an average between consecutive steps is done from time to time, or the first steps of the integration are performed with a smaller time-step size.

With respect to the size of the error for a given Δt, we notice both in Figs. 4.2(a) and (b) that the Q-implicit method is in general more accurate than the other two, more markedly so in Fig. 4.2(b). With respect to the largest stepsize the methods can use for a given order k, Fig. 4.2(a) shows that, in the C_1 case, the Q-implicit method can take larger stepsizes than the other two methods for all the orders. In the C_2 case, as the Ekman number is smaller, the Q-term starts to be dominant, and, as expected, the relative differences of the upper limit of Δt among the methods are larger. Surprisingly, and as opposed to the C_1 case, for the orders $k = 3$ and $k = 4$ in the C_2 case, the Q-splitting method can use larger stepsizes than the Q-implicit. This may be partially explained as follows. For the smaller Ekman number, the eigenvalues λ of the Q-term are close to the imaginary axis and their imaginary parts grow as E decreases (see [119, 79, 40]). Then, if Δt is not very small, the product $\lambda\Delta t$ is outside the absolute stability region of orders $k = 3$ and $k = 4$ (see Fig. 4.1(b)). This may render the Q-implicit unstable and, hence, limit the range of available stepsizes for these orders. Except for this two cases, the upper limit of Δt is lower for the Q-splitting method, and even lower for the Q-explicit. Notice also that in fig. 4.2(b) the slope of the longest curve corresponds to the Q-implicit method with order $k = 2$. For this order, the stability region of the second order BDF contains a larger portion of the imaginary axis, and this may allow the method to take significantly larger stepsizes.

Figs.4.2(c) and (d) show the dependence of $\log \varepsilon(u)$ against $\log \varepsilon^r$ for the VSVO methods (recall $\varepsilon^r = \varepsilon^a$) . These two plots show that, for a given ε^r, the ratio of the error $\varepsilon(u)$ of DLSODPK over that of the Q-implicit VSVO method is, in average, $O(\sqrt{N})$. This is because our error control is much more restrictive than that of DLSODPK as was explained before. The figures also indicate the high accuracy of the reference solution, since in this plot we compare it with that obtained by a standard integrator, and $\varepsilon(u)$ can be made smaller than 10^{-12}. On the other hand, the fact that the errors of the Q-explicit are slightly larger than those of the Q-implicit method simply indicates that the latter is a more accurate method.

Fig. 4.3(a-d) contains the efficiency curves for all the methods considered. The logarithm of the relative error $\varepsilon(u)$ is plotted against the run time in seconds rt. Figs.4.3(a) and (b) show the results for the constant time stepsize methods of orders 2 to 5, and Figs.4.3(c) and (d) those for the VSVO codes together with the constant time stepsize Q-splitting method for comparison purposes. As before, plots (a) and (c) correspond to the C_1 case, and (b) and (d) to the C_2 case. Each point in Fig. 4.3 has its corresponding one in Fig. 4.2, but notice that the smaller the values of Δt and ε^r in Fig. 4.2, the larger
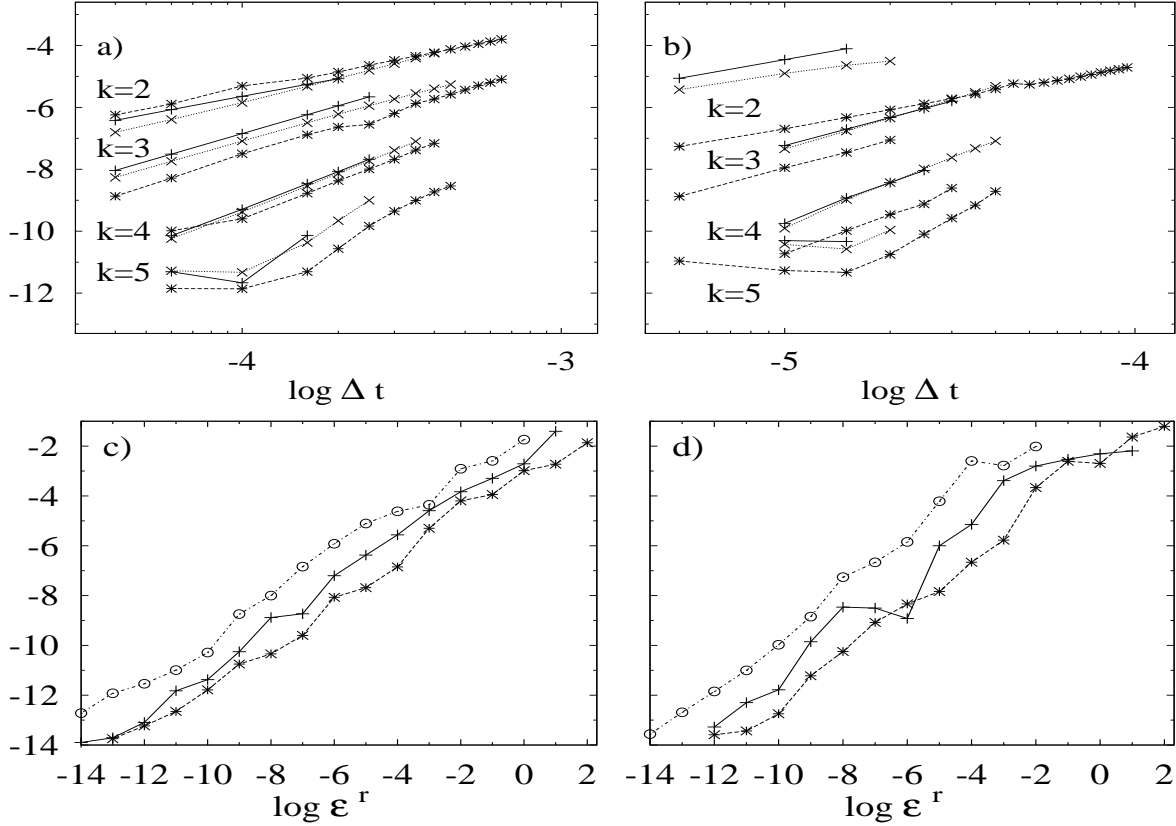
65

Figure 4.2: (a) $\log \varepsilon(u)$ versus $\log \Delta t$ for constant stepsize integration, k from 2 to 5, in the C_1 case. (c) $\log \varepsilon(u)$ versus $\log \varepsilon^r$ for the VSVO methods in the C_1 case. (b), (d) Same as (a) and (c), respectively, for the C_2 case. The symbols mean: Q-explicit ($+$, solid line), Q-splitting ($\times$, dotted line), Q-implicit ($*$, dashed line), Q-explicit VSVO ($+$, solid line), Q-implicit VSVO ($*$, dashed line) and DLSODPK ($\circ$, dash-dotted line). The vertical axis is the same in each row of plots.

the corresponding run times in Fig. 4.3.

In order to understand the results, the computational cost of the most expensive processes involved in the time integration with constant stepsize has been considered in the formulae defined below. The initial integration with a VSVO method has not been included, but it is negligible because the integration time, $t_f - t_0$, was chosen to be large enough. The computational cost of a single step of the Q-explicit and the Q-splitting methods are essentially the same, so only the former will be described.

The run time of the integration with the Q-explicit and Q-implicit methods, during the interval $[t_0, t_f]$, and using time steps $\Delta t_{Q\,exp}$ and $\Delta t_{Q\,imp}$, respectively, can be estimated

by

$$C_{Q\,exp} = N_{S\,exp}(C_{NL} + C_Q + C_{LS}),$$
$$C_{Q\,imp} = N_{S\,imp}(C_{NL} + (N_{GMR} + 1)(C_{MP} + C_{LS})),$$

where C_{NL}, C_Q, C_{LS}, and C_{MP} are, respectively, the costs of the evaluation of the non-linear terms, the evaluation of the operator $\mathcal{Q}$, the solution of the block-diagonal linear systems (4.20) in the case of taking $\mathcal{Q}$ explicitly, and the matrix products by the block-tridiagonal matrices H^m when $\mathcal{Q}$ is taken implicitly. The integer N_{GMR} is the average number of iterations performed by the linear iterative solver (GMRES in our case). Each iteration involves a matrix product and an application of the preconditioner, that is, a solution with the block-diagonal matrix. An additional unit is added to N_{GMR} to take into account the initial evaluation of the residual, and the final preconditioner application. The numbers of steps are $N_{S\,exp} = (t_f - t_0)/\Delta t_{Q\,exp}$, and $N_{S\,imp} = (t_f - t_0)/\Delta t_{Q\,imp}$. As C_{NL} is the highest cost, the rest of them are written as a fraction of it as $C_Q = f_Q C_{NL}$, $C_{LS} = f_{LS} C_{NL}$, and $C_{MP} = f_{MP} C_{NL}$. In this way

$$C_{Q\,exp} = N_{S\,exp}C_{NL}(1 + f_Q + f_{LS}), \tag{4.26}$$
$$C_{Q\,imp} = N_{S\,imp}C_{NL}(1 + (N_{GMR} + 1)(f_{MP} + f_{LS})). \tag{4.27}$$

The cost C_{NL}, in seconds, and the factors f_Q, f_{LS}, and f_{MP} only depend on the spatial resolution employed, and not on the parameters of the problem. Their ranges of variation are $0.1 \le C_{NL} \le 3.42$, $0.043 \le f_Q \le 0.053$, $0.22 \le f_{LS} \le 0.33$, and $0.19 \le f_{MP} \le 0.33$ when the truncation parameters vary in the intervals $24 \le N_r \le 70$, and $32 \le L \le 90$. We have checked that the differences between the real run time, and the estimation obtained by the above formulae are so small that they model accurately the run time of the integrations.

Figs.4.3(a) and (b) show that for approximately $\varepsilon(u) < 10^{-9}$, the $k = 5$ method is the most efficient. The corresponding curves cannot be extended to the left (higher Δt) due to stability reasons. If $\varepsilon(u) > 10^{-9}$ the most efficient methods vary from order 2 to 4 depending on the error required.

In the ranges of run time for which all the methods of a given order are available, and for a given run time, all the methods of the same order have similar efficiency, except for the case C_2 with $k = 2$ (see Fig.4.3(b)).

For a given constant stepsize, the $\mathcal{Q}$-explicit and the $\mathcal{Q}$-splitting methods of all the orders have almost the same computational cost, and therefore the higher order methods should be preferred.

At first sight it could seem from (4.26-4.27), and the values of the factors given above, that the computational cost of the $\mathcal{Q}$-implicit method is much higher than for the others, and that it does not depend on the order. However, as the order increases, the predictions of the solution at the end of each step, based on extrapolation using the order of the
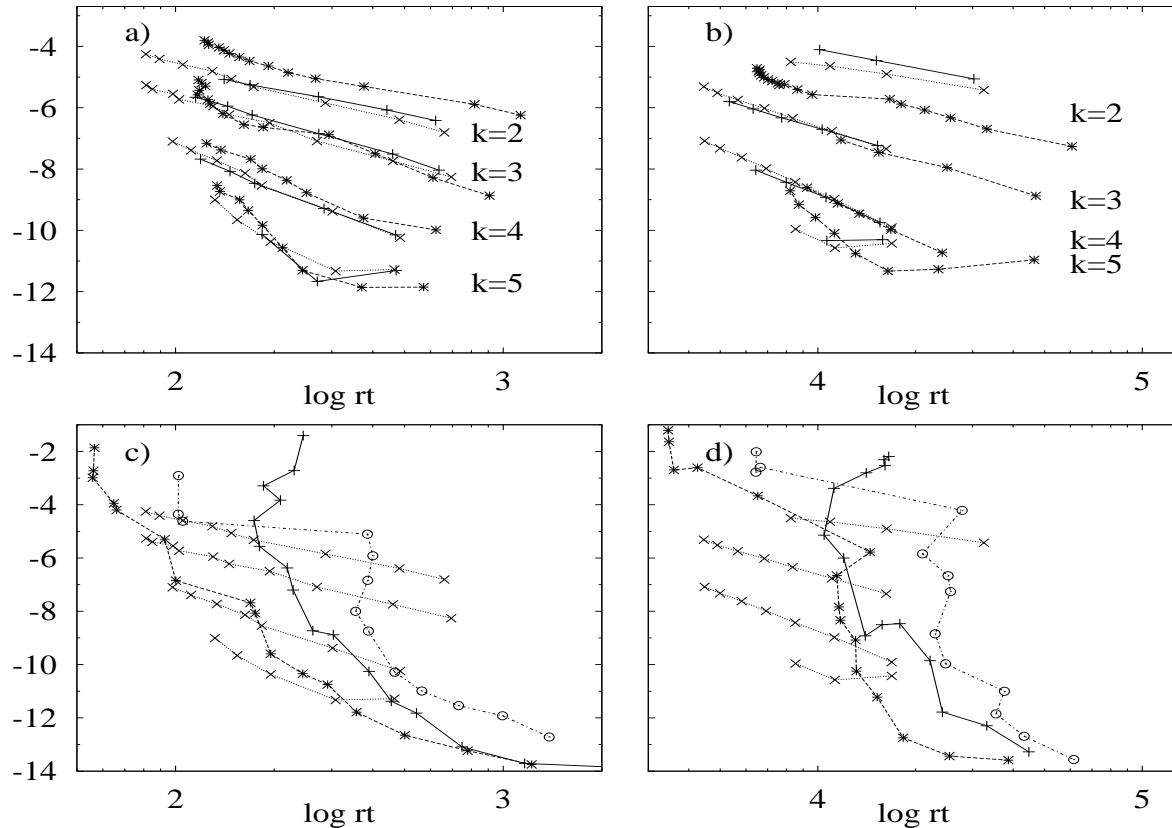
Figure 4.3: (a) $\log \varepsilon(u)$ versus the logarithm of the run time, $\log rt$, for constant stepsize integration, k from 2 to 5, in the C_1 case. (c) Same as (a) for the Q-splitting and VSVO methods. (b), (d) Same as (a) and (c), respectively, in the C_2 case. The symbols mean: Q-explicit ($+$, solid line), Q-splitting ($\times$, dotted line), Q-implicit ($*$, dashed line), Q-explicit VSVO ($+$, solid line), Q-implicit VSVO ($*$, dashed line), and DLSODPK ($\circ$, dash-dotted line). The vertical axis is the same in each row of plots.

integrator, are better, and then the number of iterations N_{GMR} to solve the linear system during the corrections is lower. Moreover, the stepsize required for a given $\varepsilon(u)$ is larger, and then N_{Simp} is lower than for the other methods. All this makes its computational cost comparable to the rest of constant stepsize methods.

In the low-cost regions (left of the plots), not all the methods are available due to the constraints of stability. If very small errors are not required, the Q-splitting method, with orders 2 and 3 in the C_1 case, and 3 and 4 in the C_2 case, is the most efficient of the constant stepsize methods. Determining which are the optimal order and stepsize requires some experiments. If many computations are going to be performed with a fixed spatial resolution, and in a relatively small region of parameters, it is worth making these initial experiments. As stated before the cost of the Q-explicit and the Q-splitting methods
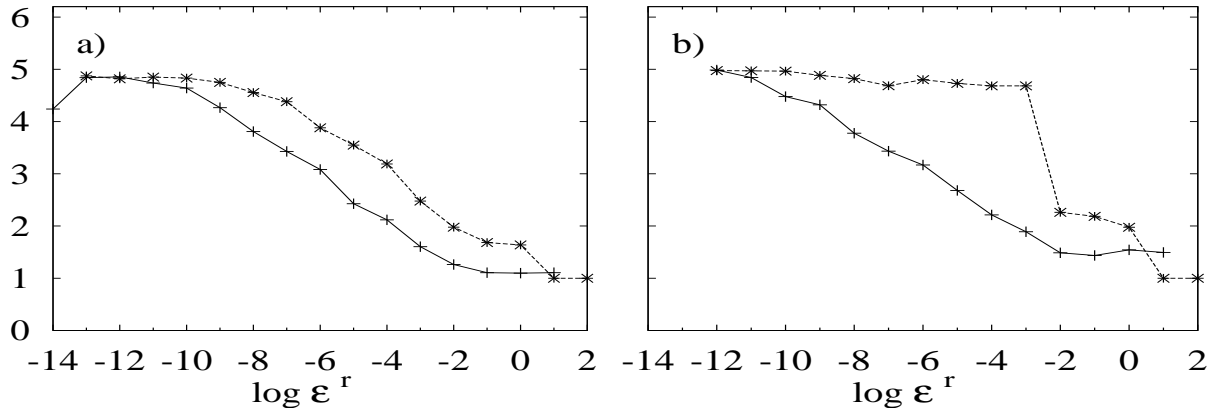
Figure 4.4: (a), (b) The average order, k_{av}, versus the logarithm of the relative tolerance, $\log \varepsilon^r$, for the semi-implicit VSVO methods, and the C_1 and C_2 cases, respectively. The symbols mean: Q-explicit VSVO ($+$, solid line), and Q-implicit VSVO ($*$, dashed line). The vertical axis is the same for the two plots.

is essentially the same, but the latter has shown to be more stable, allowing for larger stepsizes, and hence, better efficiency.

In Figs.4.3(c) and (d) the VSVO methods are compared with the Q-splitting of orders from 2 to 5, which is the best of the constant stepsize methods, except in some regions of very low error. The fully implicit method using DLSODPK is always more expensive than the Q-implicit VSVO methods because each iteration of the linear solver, and of the Newton's method requires an expensive evaluation of the non-linear terms. The Q-explicit VSVO method is also less expensive than DLSODPK, except for the higher $\varepsilon(u)$, for which the cost of the former increases due to the increase of the number of rejections, and of iterations of the linear solver. The factor f_{LS} indicates that an evaluation of the non-linear terms is between 3 and 4 times more expensive than a block-diagonal linear solving. Therefore DLSODPK will never be more efficient than the Q-implicit VSVO method, unless the former could take much larger time steps. Our numerical experiments suggest that this is not the case, especially for very low values of the Ekman number, where very rapid time scales must be resolved to maintain the accuracy.

Fig.4.3(c) shows that the Q-implicit VSVO method is the best option for the C_1 case. Some constant stepsize methods can be more efficient, but at the price of doing some previous experiments to find out the optimal stepsize. In the C_2 case (Fig.4.3(d)), there are more eigenvalues of larger imaginary part, and small and negative real part, as explained before. Therefore, in order to keep the local error below a given tolerance, the Q-implicit VSVO method is forced to use the BDF of order 2 in the left part of the curve shown in Fig. 4.3(d), and 5 in the right part, depending on the tolerance, as can be seen in Fig.4.4(b), where the average order, k_{av}, is plotted versus the relative

tolerance, ε^r, for the semi-implicit VSVO methods. The only curve with a jump is that corresponding to the Q-implicit VSVO method in the C_2 case. At the jump the order increases, the stepsize becomes larger, and the cost for a given error decreases. The jump reflects the shape of the stability regions of the BDF for constant stepsizes (shown in Fig. 4.1) which, as commented above, for orders 3 and 4 contain a smaller portion of the imaginary axis around to the origin. This abrupt transition between the two parts prevents the Q-implicit VSVO method from being as efficient as the Q-splitting method in the region of intermediate errors.

4.6 Conclusions

We have compared three time-integration semi-implicit methods, both with constant stepsize and VSVO implementations (this last option only for two of them), and a fully implicit VSVO method. The semi-implicit methods differ among them in the treatment of the Coriolis term of the equation, which, at very low Ekman numbers (E), dominate the dynamics.

In view of the difficulties of treating the Coriolis term implicitly, it could be thought that writing the equations in an inertial frame of reference could be competitive. In this case there are two possibilities. The first consists in writing the solution of (4.13) as $u = u_h + u_{bc}$, u_h and u_{bc} satisfying, respectively, homogeneous and non-homogeneous boundary conditions, the latter coming from the solid-body rotation of the two spheres. The function u_{bc} can be expressed in terms only of the spherical harmonic Y_1^0 for the potential Ψ and would be proportional to E^{-1}. After substituting u into (4.13) some new linear terms would appear coming from the non-linear operators with a structure at least as complicated as that of the Coriolis terms. These new terms would be dominant for low E, and one would be again interested in treating them implicitly to cope with Courant restrictions of the time step. The second option consists in using non-homogeneous boundary conditions for u. This affects only the boundary conditions for the coefficient Ψ_1^0 of the spherical harmonic Y_1^0 in the expansion of Ψ. In this case E enters not only in the boundary conditions, but also hidden in the nonlinear terms preventing the use of semi-implicit methods. There is another reason to avoid the equations in the inertial frame. The solutions after the first bifurcation from the conductive state are azimuthal waves of high azimuthal wave number, and drift frequency ω usually positive and below that of the spheres Ω. Then, integrating in the inertial frame implies coping with the typical limitation of the time step of high oscillatory problems, no matter the order of the method employed (see [4, **S**3.7]). In our case, if one is interested in computations close to this waves, it could be beneficial to write the equations for the deviations from the waves.

We have shown that when integrating the thermal convection equations of fast rotating fluid spherical shells, it is possible to handle the Coriolis term, and even the non-linear term, implicitly, if iterative Krylov methods are used to solve the corresponding linear

systems. This is due to the low memory requirements of these iterative linear solvers. We have shown that taking the Coriolis term implicitly results in a more efficient time integration for low values of the Ekman number. On the other hand, for the high resolutions needed to resolve the Ekman boundary layers, using adapted LU decompositions to solve the linear systems in semi-implicit or fully implicit methods is not viable, due to the prohibitive amounts of memory this would require.

We have checked our new time integration codes with a known benchmark [30]. We have found that to get relative errors of size 10^{-3}, of the order of those of the benchmark results, the most efficient methods are the fully implicit DLSODPK and the Q-implicit VSVO, with $\epsilon^a = 10^{-2}$.

The results in the present paper show that high order methods can be used not only to compute high-accurate solutions, but to obtain the same accuracy as with lower order ones but more efficiently. In practice, the most efficient method depends on the value of the Ekman number E, on the precision wanted, and even on the type of solution. For instance, if one is just interested in obtaining smooth solutions by DNS, the best choice is to implement a plain Q-explicit or, better, a third or fourth order Q-splitting method. However, if the time integration is part of a continuation process, and/or one is interested in calculating the stability of the solutions, small time integration errors must be obtained. In this case, the Q-implicit VSVO method will probably be the most efficient option. Moreover, since the lowest run times correspond to the Q-implicit VSVO method with large tolerances, this method may also be useful for passing long uninteresting transients, where having a control of the time stepsize might be important.

It must be pointed out that, since the results obtained depend on the solution integrated, and in this case it was smooth, the constant stepsize methods have made accurate computations with relatively large stepsizes. However, had the selected test solution been irregular, including for instance repeated transients [70] or bursts [110], it is reasonable to assume that in order to obtain similar accuracy the constant stepsize methods would have had to use a very small step, and then the Q-implicit VSVO method would have been the most efficient method under any circumstances.

With regard to possible parallel implementations, the main difference between the methods is in the relative number of evaluations of the linear and non-linear terms. Evaluating or solving the linear terms is highly parallelizable and is not difficult to implement. The evaluation of the non-linear terms is completely different. Their parallelisation is complex, mainly due to the triangular expansion in spherical harmonics. In the fully implicit method, actions of the full Jacobian are required, which imply evaluations of the non-linear terms or their linearizations. Therefore it would be more sensitive to any inefficiency in the parallelisation process. The rest of the methods would experience very similar speedups if implemented in parallel because they require a similar number of evaluation of the non-linear terms.

Finally, the results we have obtained are also relevant in other situations, when the partial differential equations to be solved have linear diffusion terms, which can be treated

implicitly because the resulting linear systems are relatively easy to solve, and other terms expensive to compute, and whose inclusion in a fully implicit scheme is impossible or very difficult.

Acknowledgements

4.7   Appendix

The structure of the matrices H^m appearing in eq. (4.20) is detailed here. Recall that $\mathcal{H}u^{n+1} = v^n$, with $\mathcal{H} = \mathcal{I} - h\mathcal{L}_0^{-1}\mathcal{L}$ and $h = \Delta t_n/\gamma_0(n)$ or $h = \Delta t/\gamma_0$, is the system to be solved to obtain the solution at time t_{n+1}, u^{n+1}, once the right hand side of eq. (4.15) or (4.16), v^n, has been computed. When this equation is separated into its azimuthal components the systems $H^m U^m = V^m$, $m = 0, ..., L$ are obtained.

If the functions Ψ, Φ, and Θ are expanded in spherical harmonics as in eq. (4.6), we define the vectors $U_0^0 = (\mathcal{R}\Theta_0^0)^\top$, $U_l^0 = (\mathcal{R}\Psi_l^0, \mathcal{R}\Phi_l^0, \mathcal{R}\Theta_l^0)^\top$, and $U_l^m = (\mathcal{R}\Psi_l^m, \mathcal{I}\Psi_l^m, \mathcal{R}\Phi_l^m, \mathcal{I}\Phi_l^m, \mathcal{R}\Theta_l^m, \mathcal{I}\Theta_l^m)^\top$ if $m \neq 0$, where $\mathcal{R}$ and $\mathcal{I}$ indicate real and imaginary parts, respectively. Then the vector of all amplitudes for a given order m, U^m, and the vector of all unknowns, u, are

$$U^m = (U_m^m, U_{m+1}^m, \cdots, U_{L-1}^m, U_L^m)^\top, \text{ and } u = (U^0, U^1, \cdots, U^{L-1}, U^L)^\top.$$

Suppose now that all the linear terms, included $\mathcal{C}^d$, $\mathcal{Q}^l$ and $\mathcal{Q}^u$, are treated implicitly in the operator $\mathcal{L}$. With the ordering given to U^m, the matrix H_m has dimension $6(L - m + 1)(N_r - 1)$ and the block-tridiagonal structure

$$H^m = \begin{pmatrix} A_m^m & B_m^m & 0 & ... & 0 \\ C_{m+1}^m & A_{m+1}^m & B_{m+1}^m & ... & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & ... & 0 & B_{L-1}^m \\ 0 & ... & 0 & C_L^m & A_L^m \end{pmatrix}, \tag{4.28}$$

72

where the blocks of dimension $6(N_r - 1)$, which also have sub-block structure, are

$$A_l^m = \begin{pmatrix} (A_l)^\psi & \alpha_l^m I & 0 & 0 & 0 & 0 \\ -\alpha_l^m I & (A_l)^\psi & 0 & 0 & 0 & 0 \\ 0 & 0 & (A_l)^\phi & \alpha_l^m I & (E_l)^\phi & 0 \\ 0 & 0 & -\alpha_l^m I & (A_l)^\phi & 0 & (E_l)^\phi \\ 0 & 0 & F_l & 0 & (A_l)^\Theta & 0 \\ 0 & 0 & 0 & F_l & 0 & (A_l)^\Theta \end{pmatrix}, \qquad (4.29)$$

$$B_l^m = \begin{pmatrix} 0 & 0 & (B_l^m)^\phi & 0 & 0 & 0 \\ 0 & 0 & 0 & (B_l^m)^\phi & 0 & 0 \\ (B_l^m)^\psi & 0 & 0 & 0 & 0 & 0 \\ 0 & (B_l^m)^\psi & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad (4.30)$$

and C_l^m has the same structure as B_l^m with $(B_l^m)^\phi$ substituted by $(C_l^m)^\phi$, etc. The sub-blocks of dimension $N_r - 1$ are

$$(A_l)^\psi = I - h(D_l)^\psi, \qquad (A_l)^\phi = I - h(D_l^{-1})^\phi (D_l^2)^\phi, \qquad (A_l)^\Theta = I - (h/\sigma)(D_l)^\Theta,$$

$$(B_l^m)^\psi = -\frac{(l+2)}{(l+1)} 2hE^{-1}c_{l+1}^m (D_l^{-1})^\phi (D_{l+2}^+)^\psi, \qquad (B_l^m)^\phi = -\frac{(l+2)}{(l+1)} 2hE^{-1}c_{l+1}^m (D_{l+2}^+)^\phi,$$

$$(C_l^m)^\psi = -\frac{(l-1)}{l} 2hE^{-1}c_l^m (D_l^{-1})^\phi (D_{1-l}^+)^\psi, \qquad (C_l^m)^\phi = -\frac{(l-1)}{l} 2hE^{-1}c_l^m (D_{1-l}^+)^\phi.$$

$(E_l)^\phi = h(D_l^\phi)^{-1}$, $F_l = -(h/\sigma)l(l+1)R\eta(1-\eta)^{-2}Diag(r_i^{-3})$, $\alpha_l^m = 2hE^{-1}m/l(l+1)$, and I is the identity matrix of dimension $N_r - 1$. The superscripts ψ, ϕ or Θ over a radial operator denote to which function they are applied. The corresponding boundary conditions are different and are included in the matrices.

The diagonals of H^m with blocks B_l^m and C_l^m are the discretization of the operators $\mathcal{Q}^d$ and $\mathcal{Q}^u$, respectively, and the matrices A_l^m include that of the rest of the linear terms. Therefore, the matrices H^m of eq. (4.28) correspond to the $\mathcal{Q}$-implicit or the $\mathcal{Q}$-implicit VSVO methods, depending on the value of h. If all the non-diagonal blocks B_l^m and C_l^m are set to zero, the resulting block-diagonal matrices correspond to the $\mathcal{Q}$-explicit or $\mathcal{Q}$-explicit VSVO algorithms. If only B_l^m or C_l^m are set to zero, H^m are the lower or upper block-Hessenberg matrices of the $\mathcal{Q}$-splitting method.

The LU decomposition is performed only for the diagonal blocks A_l^m, either to solve the block-diagonal or block-Hessenberg systems in the $\mathcal{Q}$-explicit or $\mathcal{Q}$-splitting methods, or to be used as preconditioners for the iterative solvers in the rest of algorithms (see table 4.2). As can be seen in eq. (4.29), the A_l^m can be separated in two sub-blocks

73

of dimensions $2(N_r - 1)$ and $4(N_r - 1)$, for which a tailored block-LU decomposition is performed.

When the block-diagonal is used as preconditioner, a frozen value h^* is set instead of the exact h. In the case of the Q-implicit method with fixed time stepsize, Δt is constant and known in advance and then $h^* = h$. For the VSVO methods, the value of h^* is selected at the first step, and the corresponding block-LU decomposition is computed. The updating of h^* and the preconditioner are performed, as explained in Section 4.3, when the linear solver requires too many iterations to converge or h^* is very different from h, depending on the method employed. The number of updates during the integrations used to be very small. In most of our calculations it was below 10. The higher values corresponded to the higher tolerances for the local errors. In these cases the range of Δt used during the time evolution was wider and this forced more updates. Therefore, except for the calculations for which high order methods are not required, the time spent in preparing the preconditioner was negligible.