Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **514,** 1258–1265 (2022) https://doi.org/10.1093/mnras/stac1418
Advance Access publication 2022 May 23

On the numerical integration of an explicit solution of the homologous collapse's radial evolution in time

Manuel Calvo,★ Antonio Elipe ★ and Luis Rández★
IUMA, Universidad de Zaragoza, C/ Pedro Cerbuna 12, E-50009 Zaragoza, Spain

Accepted 2022 May 17. Received 2022 May 12; in original form 2022 January 12

ABSTRACT

In a recent paper, Slepian and Philcox derive an explicit solution of the homologous collapse from rest of a uniform density sphere under its self-gravity as a function of time. Their solution is given in terms of two curvilinear integrals along a suitable Jordan contour; in practice, it must be approximated by a quadrature rule. The aim of this paper is to examine how the choice of the contour and the quadrature rule affects the accuracy and the efficiency of this integral solution approximation. More precisely, after a study of the complex roots of a transcendental equation that relates time with the variable, some alternative Jordan contours that turn out to be more convenient are proposed. Then, by using as quadrature rule the composite trapezoidal rule because of its reliability and spectral convergence accuracy, some numerical experiments are presented to show that the combination of contours and quadrature rule allows us to obtain numerical results with high accuracy and low computational cost.

Key words: methods: analytical – methods: numerical – large-scale structure of Universe.

1 INTRODUCTION

The problem of the gravitational collapse from rest of a cool, self-gravitating cloud has been widely used in astrophysics to model the formation of stars and galaxies (Tomita 1969; Kippenhahn & Weigert 1990; Dyson & Williams 1997; Choudhuri 1998). The simplest case is to consider the cloud spherical, although other models considering a uniform spheroid show that an oblate spheroid tends toward a disc, and a prolate spheroid toward a spindle (Lin, Mestel & Shu 1965).

In the simplest case (Lin, Mestel & Shu 1965), the equation of motion is

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = -\frac{Gm}{r^2}, \tag{1}$$

where m is the mass of the sphere contained within radius r, and G is Newton's constant. Assuming the sphere has a uniform density at the initial time $t = 0$, the radius r is a function of time only and the sphere contracts homologously, and so stays uniform.

The above second-order differential equation (1) can be reduced to a first order by the change

$$v(r(t)) = \frac{\mathrm{d}r}{\mathrm{d}t}, \tag{2}$$

resulting in

$$\frac{1}{2}\mathrm{d}(v^2) = -\frac{Gm}{r^2}\mathrm{d}r \implies \left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 = 2Gm\left(\frac{1}{r} - \frac{1}{R}\right), \tag{3}$$

with $R \equiv r(t = 0)$ is the radius of the initial sphere at $t = 0$. Then,

$$\mathrm{d}t = -\left(\frac{8\pi G R^2 \rho_0}{3}\right)^{-1/2}\left(\frac{r}{R-r}\right)^{1/2}\mathrm{d}r, \tag{4}$$

where ρ_0 is the mean density of the sphere at the initial time.

This quadrature can be easily computed by making (Lin et al. 1965), for instance, the change

$$r = R\cos^2\theta, \quad \theta \in [0, \pi/2), \tag{5}$$

and the solution is

$$2\theta + \sin 2\theta = \left(\frac{32G\pi\rho_0}{3}\right)^{1/2} t. \tag{6}$$

By integrating the above equation (4) from $r = R$ to $r = 0$ [equivalently from $\theta = 0$ to $\theta = \pi/2$ in (8)], we obtain the time of the collapse via free-fall as

$$t_{\mathrm{ff}} = \left(\frac{3\pi}{32\,G\rho_0}\right)^{1/2}, \tag{7}$$

at which the sphere has collapsed into a point.

Thus, equation (6) can by written as

$$2\theta + \sin 2\theta = (\pi/t_{\mathrm{ff}})\,t. \tag{8}$$

Hence, to know the radius $r = r(\theta(t))$ of the collapsing sphere at a certain instant t from equation (5), it is necessary to solve the trascendental equation (8). The situation is quite similar to the two-body problem (see e.g. Danby 1988), in which it is necessary to solve Kepler's equation $E - e\sin E = n(t - t_0)$ in order to know the radius $r = a(1 - e\cos E)$.

Slepian & Philcox (2021) obtain an explicit formula to solve equation (8); the method they use was discovered by Ullisch (2020) in finding the closed-form solution of the so-called 'Geometric Goat

★ E-mail: calvo@unizar.es (MC); elipe@unizar.es (AE); randez@unizar.es (LR)

© 2022 The Author(s)
Published by Oxford University Press on behalf of Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article/514/1/1258/6590834 by Universidad de Zaragoza user on 14 September 2022

Problem,' which says: *'A circle containing one acre is cut by another whose centre is on the circumference of the given circle, and the area common to both is one-half acre. Find that radius of the cutting circle'* (Heaton, Holycross & Myers 1894). This method has been successfully used more recently by Philcox, Goodman & Slepian (2021) to obtain an explicit formula for solving Kepler's equation.

In their paper, Slepian & Philcox (2021) derive an explicit form of the radius r as a function of time in terms of two contour integrals in the complex plane. More precisely, they show that $\theta(t)$ can be given explicitly as a quotient of two curvilinear integrals along the circular contour $\mathcal{C} = \{z; |z - \pi/4| = \pi/4 - \epsilon\}$, where ϵ is a small positive constant. Furthermore, they show that these integrals can be approximated to high accuracy by using the FFT of some periodic function so that this approach is faster than standard iterative methods to solve (8). The aim of the present note is twofold.

First of all, after a study of the complex zeros of (8), to propose alternative contours that can be more efficient in practical computations. Secondly to simplify the numerical evaluation of the contour integrals by using some results on the spectral convergence of quadratures of periodic functions.

2 THE INTEGRAL SOLUTION

By introducing the notation

$$u = 2\theta, \qquad \tau = \pi \left(t/t_{\rm ff}\right), \tag{9}$$

the transcendental equation (8) in the unknown u becomes

$$\phi(u) \equiv u + \sin u = \tau, \quad \tau \in [0, \pi]. \tag{10}$$

Since, $\phi(0) = 0$ and $\phi(\pi) = \pi$ and $\phi^{'}(u) > 0$ for all $u \in (0, \pi)$, for all $\tau \in (0, \pi)$, by the inverse function theorem (Apostol 1974) there is a unique $u = \phi^{-1}(\tau) \in (0, \pi)$. Then, as shown in Slepian & Philcox (2021), the solution of (10) can be written in the integral form

$$u = \phi^{-1}(\tau) = \frac{\oint_{\mathcal{C}} \dfrac{z dz}{\phi(z) - \tau}}{\oint_{\mathcal{C}} \dfrac{dz}{\phi(z) - \tau}}, \quad \tau \in (0, \pi), \tag{11}$$

where $\mathcal{C} = \mathcal{C}(\tau)$ is a Jordan curve [that may depend on the parameter $\tau \in (0, \pi)$] in the complex plane enclosing u and such that $\phi(z) \neq \tau$ for all $z \neq u$ in the interior of $\mathcal{C}$.

Note that in view of (5), for each τ the solution u of (11), gives the radius vector $r = r(\theta)$ as a function of u by

$$r = \frac{R}{2}\left[1 + \cos u\right] = \frac{R}{2}\left[1 + \delta_0 \sqrt{1 - (\tau - u)^2}\right] \tag{12}$$

with $\delta_0 = 1$ if $u \in [0, /2]$, or else $\tau \in [0, 1 + \pi/2]$ and $\delta_0 = -1$ for $u > \pi/2$.

We note that (10) implies that

$$\frac{{\rm d}u}{{\rm d}\tau} = \frac{1}{1 + \cos u}, \tag{13}$$

and therefore the inverse of $\phi(u)$ becomes singular when $u = \pi$ or equivalently when $\tau = \pi$. Because of this, some practical problems are expected in the definition of $\phi^{-1}(\tau)$ when τ is close to π.

The theoretical definition (11) of u is independent of the Jordan curve $\mathcal{C} = \mathcal{C}(\tau)$ provided that it satisfies the above condition ($\phi(z) \neq \tau$ for all $z \neq u$ in the interior of $\mathcal{C}(\tau)$). However, the integrals of (11) cannot be solved exactly and must be approximated by a quadrature rule; and then the numerical solution depends both on the quadrature rule and also on the selected Jordan curve. Concerning the quadrature rule we will use here the composite trapezoidal rule with K nodes because in spite of its simplicity it provides high accuracy for moderate values of K. Moreover, this rule has spectral convergence (Trefethen 2000) as a function of K and it is used in the computation of the Fast Fourier Transform (FFT). On the other hand, the accuracy depends on the choice of $\mathcal{C}$, and particularly on the length of $\mathcal{C}$ and as we will see next, smaller length gives better accuracy at fixed number of nodes.

3 THE CHOICE OF THE JORDAN CURVE

As remarked above, for all $\tau \in (0, \pi)$ there is a unique real u such that $\phi(u) = \tau$. Furthermore, Slepian & Philcox (2021) show that by using the argument principle in the rectangular region of the complex plane $(0, \pi) \times i(-M, M)$ for arbitrary $M > 0$ there is a unique zero of $\phi(u) - \tau$ for all $\tau \in (0, \pi)$. Hence for all $\tau \in (0, \pi)$ we may take as $\mathcal{C}(\tau)$ any Jordan curve contained in the complex region $\{z \,|\, {\rm Re}\, z \in (0, \pi)\}$ that includes the unique real solution of $\phi(z) = \tau$. In particular Slepian and Philcox propose as $\mathcal{C}(\tau)$ the circle of center $z = \pi/2$ and radius $\pi/2 - \epsilon$, with ϵ a positive small constant:

$$\mathcal{C}_0(\tau) = \left\{z = \pi/2 + (\pi/2 - \epsilon)\exp(i w), \quad {\rm d}w \in [-\pi, \pi]\right\}. \tag{14}$$

Clearly with this choice the integral expression (11) holds true only for $\tau \in (\phi(\epsilon), \pi - \phi(\epsilon))$ and not all $(0, \pi)$. On the other hand, as remarked above, from a computational point of view it is important for each τ to have Jordan curves $\mathcal{C}(\tau)$ with the smallest length in order to have accurate approximations. Thus, in this section, we will study the location of the complex zeros of $\phi(u) - \tau = 0$ for real τ and we will propose several circles $\mathcal{C}(\tau)$ with $\tau \in (0, \pi)$ that include only one zero of $\phi(u) = \tau$ and have radius as small as possible.

Putting $u = x + i\,y$, (x and y real), u is a zero of (10) if and only if x and y satisfy

$$x + \sin x \;\cosh y = \tau, \qquad y + \cos x \;\sinh y = 0. \tag{15}$$

For $y = 0$, equations (15) are equivalent to $x + \sin x = \tau$, that gives the unique real zero for each $\tau \in (0, \pi)$. For $y \neq 0$, the second equation of (15) defines x as a function of y by

$$x = x(y) = \arccos\left(-\frac{y}{\sinh y}\right) \equiv \arccos\left[g(y)\right], \tag{16}$$

with $g(y) = -y/\sinh y$.

Since $g(y)$ is an even function of y, next we consider only the case $y > 0$.

Now $g(y)$ is a strict monotonic increasing function with

$$\lim_{y\to 0} g(y) = -1, \quad \text{and} \quad \lim_{y\to+\infty} g(y) = 0, \tag{17}$$

consequently for all $y = y^+ > 0$ there exist two symmetric values of $x = \pm x^+$ such that $\cos x^+ = y^+$, and $x^+ \in (\pi/2, \pi)$.

By substituting (16) into (15), we have

$$\tau = \arccos[g(y)] + \sqrt{1 - g(y)^2}\cosh y > \pi, \tag{18}$$

for all $y > 0$, and only

$$\lim_{y\to 0} \arccos[g(y)] + \sqrt{1 - g(y)^2}\cosh y = \pi. \tag{19}$$

In view of the above discussion, we conclude the following:

PROPOSITION 3.1. *Concerning, the complex solutions of the transcendental equation* $\phi(\mathrm{u}) = \tau$ *in the unknown* u *with* $\tau \in \mathbb{R}$ *we have:*

1) For $\tau < 0$, *there are no solutions.*
2) For each $\tau \in [0, \pi]$, *there exists a unique real solution.*

Downloaded from https://academic.oup.com/mnras/article/514/1/1258/6590834 by Universidad de Zaragoza user on 14 September 2022

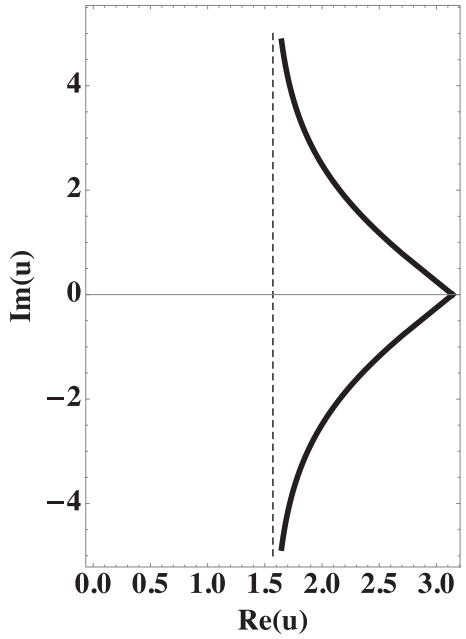

Figure 1. The set of complex zeros of the equation (10): $\phi(u) = \tau$ for $\tau \in (\pi, +\infty)$. Observe that there are two complex conjugate zeros for each $\tau > \pi$, and when $\tau \to \pi$ they tend to $\pi + 0\,i$, whereas when $\tau \to +\infty$, $y \to +\infty$ and they are asymptotic to the line $\mathrm{Re}(u) = \pi/2$.

3) For each $\tau > \pi$, there are two complex conjugate solutions $\mathrm{x}^+ = [\mathrm{x}(\mathrm{y}^+)] + \mathrm{y}^+\,\mathrm{i}$, *and* $\overline{x}^+ = \left[x(y^+)\right] - y^+\,i$, *where* $\mathrm{x}^+ \to \pi$ *when* $\tau \to \pi$ *and* $\mathrm{x}^+ \to (\pi/2, +\infty)$ *when* $\tau \to +\infty$ *and similarly for its conjugate* $\overline{x}^+$.

In Fig. 1, we display the two branches of the curves $x = x(y)$ given by equation (16) for $y = \mathrm{Re}(u) > 0$ and $y < 0$. Since by (16), $x(y) = x(-y)$, for all $y \neq 0$ there are two complex conjugate zeroes of $\phi(u) = \tau$ when $\tau > \pi$. Note that when $\tau \to \pi$ with $\tau > \pi$, then $y \to 0$ and $x(y) \to \pi$, whereas when $\tau \to +\infty$, $y \to +\infty$ and $x(y) \to \pi/2$.

The second point deals with the choice of the Jordan curves $\mathcal{C} = \mathcal{C}(\tau)$. It is known that in the case of circular contours, a reduction of the radius improves the numerical accuracy of the solution. Hence, for the numerical experiments, we will consider two families of circles $\mathcal{C}_1$ and $\mathcal{C}_2$, together with the circle $\mathcal{C}_0$ given in equation (14), which is the one proposed by Slepian & Philcox (2021).

A straightforward choice follows from the two-sided error bound

$$u < \phi(u) = u + \sin u < u + 1, \qquad u \in (0, \pi). \tag{20}$$

Now from $\tau = \phi(u)$ it follows that

$$\tau - 1 < u < \tau, \tag{21}$$

and in view of this two sided-error bound, a possible choice is to take the circle

$$\mathcal{C}_1(\tau) = \left\{ z = (\tau - 1/2) + 1/2 \exp(i\,w),\ \ w \in [-\pi, \pi] \right\}, \tag{22}$$

because for all $\tau \in (0, \pi)$, the contour (22) contains the unique real solution of $\phi(u) = \tau$ and no other complex solution is within it. In fact for all $\tau \in (0, \pi)$, the circle (22) is contained in $\mathrm{Re}(z) < \pi$ and as shown in Proposition 3.1 there is only one real zero and for $\tau < 0$ there are no complex zeroes.

Note that all the circles (22) have radius 1/2, smaller than the radius $\pi/2 - \epsilon$ used in Slepian & Philcox (2021) and therefore a better numerical behaviour is expected for the circle (22).

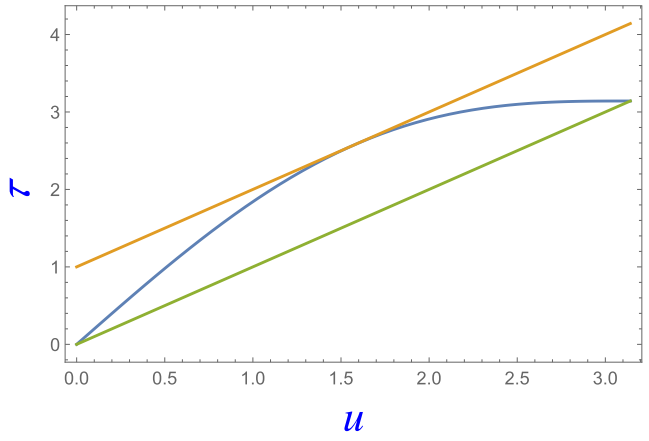

Figure 2. The function $\tau = u + \sin u$ with $u \in (0, \pi)$ is lower-bounded by the chord $\tau = u$ (in green), and an upper bound is the tangent at $u = \pi/2$ (orange). This fact allows us to define the Jordan curves $\mathcal{C}_2$.

Some improvements of (22) can be carried out taking into account that $\phi(u) = u + \sin u$ is a convex function downwards in $u \in (0, \pi)$ and the upper bound given in equation (20) $\phi_+(u) = u + 1$ is the tangent of $\phi(u)$ at $u = \pi/2$ and the lower bound $\phi_-(u) = u$ is the chord between the two end points of $\phi(u)$ at $u = 0$ and $u = \pi$ (see Fig. 2).

Then, we may replace this lower bound by the polygonal joining the points $(0, \phi(0) = 0)$, $(\pi/2, \phi(\pi/2) = \pi/2 + 1 = \tau_1)$, and $(\pi, \phi(\pi) = \pi)$, that is

$$\phi_-(u) = \begin{cases} \left(\dfrac{\pi + 2}{\pi}\right) u & \text{for} \quad u \in (0, \pi/2], \\ \left(\dfrac{\pi - 2}{\pi}\right) u + 2 & \text{for} \quad u \in (\pi/2, \pi]. \end{cases} \tag{23}$$

Hence, we will consider also the Jordan curve $\mathcal{C}_2 = \mathcal{C}_2(\tau)$ defined by

$$\mathcal{C}_2 = \begin{cases} \{z = \mu_1 + \rho_1 \exp(i\,w)\,;\ \tau \in (0, \tau_1)\} \\ \{z = \mu_2 + \rho_2 \exp(i\,w)\,;\ \tau \in [\tau_1, \pi)\} \end{cases} \tag{24}$$

with $w \in [-\pi, \pi]$, $\tau_1 = \pi/2 + 1$, and

$$\mu_1 = \left(\frac{\pi + 1}{\pi + 2}\right)\tau - \frac{1}{2},\ \rho_1 = \frac{1}{2} - \left(\frac{1}{\pi + 2}\right)\tau,$$
$$\mu_2 = \frac{2(\pi - 1)\tau - (3\pi - 2)}{2(\pi - 2)},\ \rho_2 = \frac{2\tau - (\pi + 2)}{2(\pi - 2)}. \tag{25}$$

Furthermore, by using the convexity of $\phi(u)$ in $(0, \pi)$ other improvements can be derived from the above contours $\mathcal{C}_j$, $j = 0, 1, 2$. Consider, for example, the contour $\mathcal{C}_0$; if we split the interval $(0, \pi)$ by some intermediate value $u_1 \in (0, \pi)$, we may take the contour

$$\widetilde{\mathcal{C}}_0 = \widetilde{\mathcal{C}}_0(\tau) = \left\{ z = \mu_j + \rho_j\ \exp(i\,w) \right\}, \quad j = 1, 2, \tag{26}$$

with $w \in [-\pi, \pi]$, and

$$\text{if } \tau \in (0, \tau_1),\ \mu_1 = \phi(u_1)/2,\ \rho_1 = \phi(u_1)/2,$$
$$\text{if } \tau \in (\tau_1, \pi),\ \mu_2 = \frac{\phi(\pi) + \phi(u_1)}{2},\ \rho_2 = \frac{\phi(\pi) - \phi(u_1)}{2}. \tag{27}$$

Thus, if $u_1 = \pi/2$, $\phi_(u_1) = \pi/2 + 1$ and the radius $\rho = \pi/2 - \epsilon$ is reduced to ρ_1 or ρ_2 depending on the value of τ.

Just in case, the reader is interested, in Fig. 3, we present some finer upper and lower bounds of the function $\tau = u + \sin u$, which will serve as a hint in finding new Jordan circles with small radii.

Downloaded from https://academic.oup.com/mnras/article/514/1/1258/6590834 by Universidad de Zaragoza user on 14 September 2022

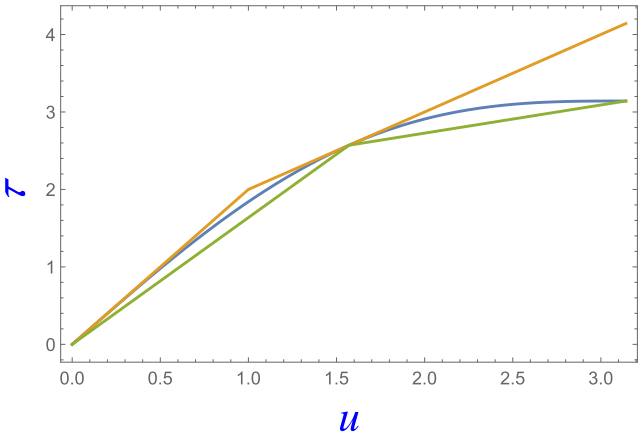

Figure 3. It is possible to find some narrow upper and lower bounds of the function $\tau = u + \sin u$ with $u \in (0, \pi)$, like the polygonal lines in the graphic, which enables defining some Jordan circles with very small radius, in a similar way, as we did to obtain the Jordan curves $\mathcal{C}_2$.

4 THE INTEGRAL EXPRESSION ALONG THE CIRCLES

The integral expression (11) along a Jordan circle

$$\mathcal{C} = \mathcal{C}(\tau) = \{z = \mu + \rho \exp(i w),\, w \in [-\pi, \pi]\}, \tag{28}$$

with some center $\mu = \mu(\tau)$ and radius $\rho = \rho(\tau)$ becomes

$$u = u(\tau) = \mu + \rho \frac{\int_{-\pi}^{\pi} \exp(2iw) G(w;\tau) \mathrm{d}w}{\int_{-\pi}^{\pi} \exp(iw) G(w;\tau) \mathrm{d}w}, \tag{29}$$

with $G(w) = G(w;\tau)$ given by

$$G(w) = \left[\mu - \tau + \rho \exp(iw) + \sin(\mu + \rho\ \exp(iw))\right]^{-1}. \tag{30}$$

Observe that the numerator on the right hand side of (29) is the Fourier coefficient $\widehat{G}(-2;\tau)$ of the 2π periodic function in w, $G(w;\tau)$ corresponding to the wavenumber $k = -2$, and similarly the denominator is the Fourier coefficient $\widehat{G}(-1;\tau)$ of the wavenumber $k = -1$. This suggests using a Fast Fourier Transform (FFT) algorithm to calculate all coefficients $\widehat{G}(k;\tau)$ up to some $|k| \le k_0$ and then to take the corresponding to $k = -2$ and -1 (Ullisch 2020; Slepian & Philcox 2021). However, we will use here a direct calculation of these particular coefficients by using the composite trapezoidal rule that for a given positive integer K approximates the integral

$$I(\Gamma) = \int_0^{\pi} \Gamma(w) \mathrm{d}w, \tag{31}$$

by the discrete sum

$$I_K(\Gamma) = \frac{\pi}{2K}\left[\Gamma(w_0) + 2\sum_{j=1}^{K-1} \Gamma(w_j) + \Gamma(w_K)\right], \tag{32}$$

with $w_j = (\pi/K)j$, with $j = 0, 1, \ldots, K$.

Now taking into account that

$$\begin{aligned} \mathrm{Re}\left[\exp(2iw)\ G(w)\right] &= \mathrm{Re}\left[\exp(-2iw)\ G(-w)\right], \\ \mathrm{Im}\left[\exp(2iw)\ G(w)\right] &= -\mathrm{Im}\left[\exp(-2iw)\ G(-w)\right], \end{aligned} \tag{33}$$

and similarly for the integral of the denominator of (29). The expression (29) becomes

$$u = u(\tau) = \mu + \rho \frac{\int_0^{\pi} \mathrm{Re}\left[\exp(2iw) G(w;\tau)\right] \mathrm{d}w}{\int_0^{\pi} \mathrm{Re}\left[\exp(iw) G(w;\tau)\right] \mathrm{d}w}, \tag{34}$$

Then by using the sumation formula (32), the numerator of (34) is approximated by

$$\mathrm{NI}_K = \frac{\pi}{2K}\left[G(0) + G(\pi) + 2\sum_{j=1}^{K-1} \mathrm{Re}\left[\exp(2iw_j) G(w_j)\right]\right], \tag{35}$$

and similarly for the denominator

$$\mathrm{DI}_K = \frac{\pi}{2K}\left[G(0) - G(\pi) + 2\sum_{j=1}^{K-1} \mathrm{Re}\left[\exp(iw_j) G(w_j)\right]\right]. \tag{36}$$

With this, the approximation of equation (34) is

$$u_K = u_K(\tau) = \mu + \rho\ \frac{\mathrm{NI}_K}{\mathrm{DI}_K}, \tag{37}$$

Decomposing $G(w)^{-1}$ into its real and imaginary part

$$G(w)^{-1} = A(w) + i\ B(w), \tag{38}$$

with

$$\begin{aligned} A(w) &= \mu - \tau + \rho \cos w + \sin(\mu + \rho \cos w)\ \cosh(\rho \sin w), \\ B(w) &= \rho \sin w + \cos(\mu + \rho \cos w)\ \sinh(\rho \cos w), \end{aligned} \tag{39}$$

we have

$$\begin{aligned} \mathrm{NI}_K &= \frac{\pi}{2K}\left[G(0) + G(\pi) + 2\sum_{j=1}^{K-1} \Omega_2(w_j)\right], \\ \mathrm{DI}_K &= \frac{\pi}{2K}\left[G(0) - G(\pi) + 2\sum_{j=1}^{K-1} \Omega_1(w_j)\right], \end{aligned} \tag{40}$$

where

$$\Omega_q(w) = \frac{A(w)\cos(qw) + B(w)\sin(qw)}{A^2(w) + B^2(w)}, \quad q = 1, 2. \tag{41}$$

Observe that according to equation (30)

$$G(0) = [\mu + \rho - \tau + \sin(\mu + \rho)]^{-1} \tag{42}$$

becomes singular when the center μ and the radius ρ of $\mathcal{C}(\tau)$ satisfy that $u = \rho + \mu$ is a solution of $u + \sin u = \phi(u) = \tau$. Also for $G(\pi)$ it follows from (30) that

$$G(\pi) = [\mu - \rho - \tau + \sin(\mu - \rho)]^{-1} \tag{43}$$

is singular when $u = \mu - \rho$ is solution of $\phi(u) = \tau$. Then to avoid inaccuracies in calculations near these singularities, we may replace the approximation (37) by the regularized approximation

$$u_K(\tau) = \mu + \rho\ \frac{\mathrm{WNI}_K}{\mathrm{WDI}_K}, \tag{44}$$

where

$$\begin{aligned} \mathrm{WNI}_K &= \frac{\pi}{2K}\left[G(\pi)^{-1} + G(0)^{-1} + 2W\sum_{j=1}^{K-1} \Omega_2(w_j)\right], \\ \mathrm{WDI}_K &= \frac{\pi}{2K}\left[G(\pi)^{-1} - G(0)^{-1} + 2W\sum_{j=1}^{K-1} \Omega_1(w_j)\right], \end{aligned} \tag{45}$$

Downloaded from https://academic.oup.com/mnras/article/514/1/1258/6590834 by Universidad de Zaragoza user on 14 September 2022

and $W = G(0)^{-1}\, G(\pi)^{-1}$.

We note that for $\tau = \tau_1 = \pi/2 + 1$, the exact solution of (10) is $u_1 = \phi^{-1}(\tau_1) = \pi/2$.

This exact solution is obtained with the approximate solution $u_K(\tau_0)$ for all K, with the Jordan curve (28) with radius $\rho_0 = 1/2$ and center $\mu_0 = \tau_1 - 1/2$. Indeed, it follows from equations (42) and (43) that

$$\begin{aligned} G(0)^{-1} &= \mu_0 + \rho_0 - \tau_1 + \sin(\mu_0 + \rho_0) = \sin\left[\pi/2 + 1\right], \\ G(\pi)^{-1} &= \mu_0 - \rho_0 - \tau_1 + \sin(\mu_0 - \rho_0) = 0, \end{aligned} \tag{46}$$

and then, from relations (44) and (45), we have

$$u_K(\tau_1) = \mu_0 + \rho_0(-1) = \pi/2, \quad \text{for all } K. \tag{47}$$

5 NUMERICAL EXPERIMENTS

In this section, we present the results of some numerical experiments to show the accuracy of the numerical solution (44) of $u_K = u_K(\tau)$ for several integer values of K and the Jordan curves $\mathcal{C}_0$, $\mathcal{C}_1$, and $\mathcal{C}_2$ defined in Eqs. (14), (22), and (24) respectively.

We compute the absolute and relative errors for $\tau \in (0, \pi)$ denoted by

$$\begin{aligned} \text{AbsErr} &= \text{AbsErr}(\tau) = |u(\tau) - u_K(\tau)|, \\ \text{RelErr} &= \text{RelErr}(\tau) = |(u(\tau) - u_K(\tau))/u(\tau)|\,. \end{aligned} \tag{48}$$

As remarked above, for $\tau = \tau_1 = \pi/2 + 1$, we have $u(\tau_1) = u_K(\tau_1) = 0$ for all K and therefore, $\text{AbsErr}(\tau_1) = \text{RelErr}(\tau_1) = 0$. Because of this we study the two subintervals, $\tau \in (0, \tau_1)$ and $\tau \in (\tau_1, \pi)$.

An important point already remarked is the spectral convergence with respect to K derived from the use of the composite trapezoidal quadrature rule with $K + 1$ nodes to approximate the quadratures of the integral solution (11). Roughly speaking, we can say that we have spectral convergence when $\log \text{AbsErr} \approx -\alpha K$, with α a positive quantity independent of K, and $K \to \infty$.

At this point, we recall Theorem 3 (page 33) of Trefethen (2000): If $G(w; \tau)$ can be extended to an analytic function in the complex strip $|\text{Im}\, w| < a$, with $|G(* + iy; \tau)| \le c$ uniformly for all $y \in (-a, a)$, then

$$|\widehat{G}(k;\tau) - \widehat{G}_N(k;\tau)| = \mathcal{O}\Big(\exp(-2(a-\epsilon)N)\Big), \quad N \to \infty \tag{49}$$

for every $\epsilon > 0$.

In our case, the function $G(w; \tau)$ depends on the parameter $\tau \in (0, \pi)$ and therefore $a = a(\tau)$. Furthermore when $\tau \to \pi$, $G(w; \tau)$ becomes singular and therefore when $\tau \to \pi$, then $a(\tau) \to 0$.

In conclusion, in our case we have taken $N = 2K$, $(K \ge 2)$ and instead of $\widehat{G}_N$ we write $\widehat{G}_K$ and then the spectral behaviour means that

$$|\widehat{G}(k;\tau) - \widehat{G}_K(k;\tau)| = \mathcal{O}(\exp(-aK)) \tag{50}$$

with some $a = a(\tau)$ that depends on $\tau \in (0, \pi)$ and $a(\tau) \to 0$ when $\tau \to \pi$.

Then the approximation $u_K(\tau)$ to $u = \phi^{-1}(\tau)$ given by

$$u(\tau) = \mu + \rho\, \frac{\widehat{G}_K(-2;\tau)}{\widehat{G}_K(-1;\tau)}, \tag{51}$$

also satisfies

$$|u(\tau) - u_K(\tau)| = \mathcal{O}(\exp(-a(\tau)K)), \quad K \to \infty, \tag{52}$$

that is the desired spectral convergence.

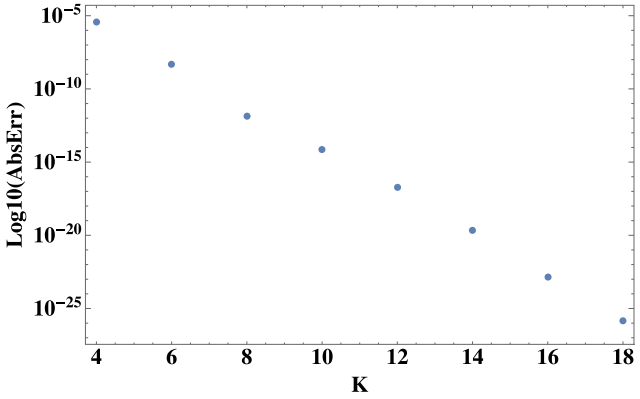

Figure 4. $\log_{10}[\text{AbsErr}_K(\tau = 1.3)]$ as a function of K for $K = 4, 6, \ldots, 18$, and the curve $\mathcal{C}_1$. The plot shows the typical linear behaviour because of the spectral convergence.

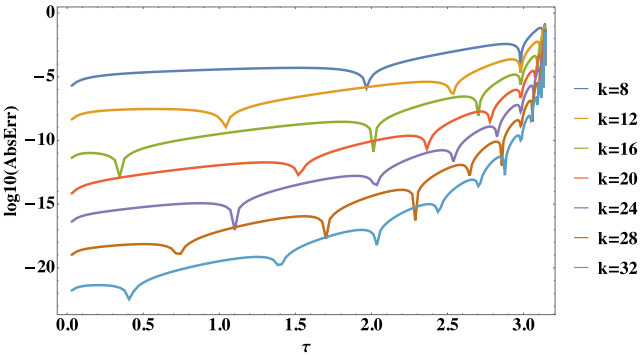

Figure 5. $\log_{10}\text{AbsErr}$ of the numerical solution for $u \in (0, \pi)$ using the contour $\mathcal{C}_0$ for several values of K for $(K = 8, 12, 16, \ldots, 32)$. The sharp down-picks corresponds to those points of the grid in which both numerical and exact solutions almost coincide. We can also see that the number of picks increases with K.

Note that with a standard quadrature of order p

$$\left|\widehat{G}(q) - \widehat{G}_K(q)\right| = \mathcal{O}(K^{-p}), \quad q = -2, -1, \tag{53}$$

and therefore the AbsErr behave as $\mathcal{O}(K^{-p})$.

The spectral accuracy implies that $\log(\text{AbsErr}_K(\tau))$ behaves asymptotically as a linear function of K whereas with standard integrators $\log(\text{AbsErr}_K(\tau))$ behaves as a linear function of $\log K$. In Fig. 4, we display in $\log_{10}$ scale the $\text{AbsErr} = |u(\tau) - u_K(\tau)|$ of the numerical solution u_K computed with the contour $\mathcal{C}_1$ for $K = 4, 6, \ldots, 18$, and for a particular case $\tau = 1.3$. We can observe the linear behaviour because of the spectral convergence.

We remark that instead of showing $\log_{10}\text{AbsErr}$ as a function of $\tau \in (0, \pi)$, we show this quantity as a function of $u = \phi^{-1}(\tau) \in (0, \pi)$, because $\text{AbsErr}(u) = |u - u_K(\phi((\tau))|$ and then, we do not have to solve the transcendental equation $\phi(u) = \tau$ for each value of τ.

In Fig. 5, we plot the $\log_{10}\text{AbsErr}(u)$ using the contour $\mathcal{C}_0$ for several values of K. A first observation is that at some values of K, the $\text{AbsErr}(u)$ shows some numerical discontinuities, that is, particular values of u at which $\log_{10}\text{AbsErr}(u)$ becomes very large negative. This is due to the fact that both numerical and exact solutions coincide at some $\tau = \tau^*$ and therefore $u^* = \phi(\tau^*) = u_K(\tau^*)$ and therefore, $\log_{10}\text{AbsErr}(u) \to -\infty$ when $u \to u^*$. In other words, the numerical solution oscillates around the exact solution and both agree at $\tau =$

Downloaded from https://academic.oup.com/mnras/article/514/1/1258/6590834 by Universidad de Zaragoza user on 14 September 2022

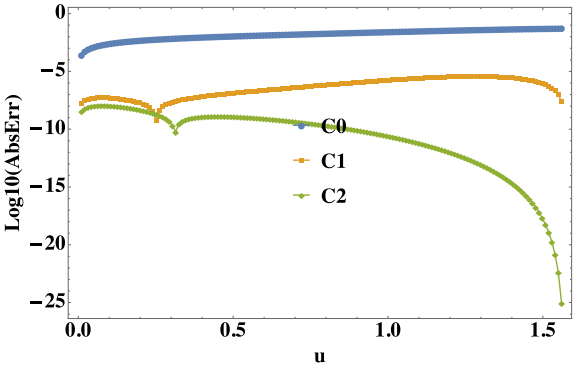

Figure 6. Absolute errors of the approximate solution u_K with $K = 4$ for the contours $\mathcal{C}_0$ (blue), $\mathcal{C}_1$ (orange), and $\mathcal{C}_2$ (green) as a function of u for $u \in (0, \pi/2)$ or, equivalently, for $\tau \in (0, \tau_1 = 1 + \pi/2)$. We see that even for a small value of K we gain significant digits with $\mathcal{C}_1$ in comparison with $\mathcal{C}_0$, and with $\mathcal{C}_2$ in relation to $\mathcal{C}_1$. The same applies to Figs 6–8. Note also that contours $\mathcal{C}_0$ and $\mathcal{C}_1$ give almost constant absolute errors along the interval $0, \pi/2$, whereas the other contour is more accurate near $u \approx \pi/2$.

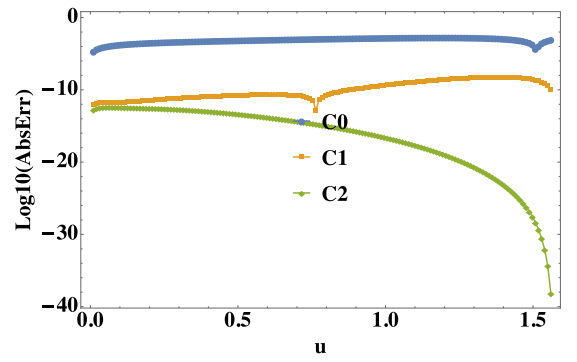

Figure 7. Absolute errors of the approximate solution u_K with $K = 6$ for the contours $\mathcal{C}_0$ (blue), $\mathcal{C}_1$ (orange), and $\mathcal{C}_2$ (green) as a function of u for $u \in (0, \pi/2)$ or, equivalently, for $\tau \in (0, \tau_1 = 1 + \pi/2)$.

τ^*. Moreover, the number of oscillations increases with K, as it can be seen in Fig. 5.

In Fig. 6, we display the $\log_{10}$AbsErr for $K = 4$ and the three Jordan curves $\mathcal{C}_0 = \mathcal{C}(\rho_0, \mu_0)$, $\mathcal{C}_1 = \mathcal{C}(\rho_1, \mu_1)$, and $\mathcal{C}_2 = \mathcal{C}(\rho_2, \mu_2)$. By the same token, these errors are given as a function of $u \in (0, \pi/2)$.

A second remark is that, even with the moderate value $K = 4$, the AbsErr remains smaller for the curve $\mathcal{C}_1$ than for the curve $\mathcal{C}_0$, and along the curve $\mathcal{C}_2$ is much smaller than for $\mathcal{C}_1$, reaching 25 significant digits at the end of the considered interval. Observe that in spite of the fact that all circles in $\mathcal{C}_2$ have the same radius $\rho_0 = 1/2$ for all $\tau \in (0, \tau_2)$ the accuracy depends on τ and it attains a maximum at about $u \simeq 1.5$.

For the circular contours $\mathcal{C}_2$, the radius

$$\rho_2 = \left[\frac{1}{2} - \left(\frac{1}{\pi + 2}\right)\right] \tau \;\to\; 0, \quad \text{as} \quad \tau \to \tau_2 = \pi/2 + 1,$$

or else when $u \to \pi/2$ and this implies that by using $\mathcal{C}_2$ the accuracy improves significantly when τ approaches τ_2. This can be seen clearly in Figs 6–8. In particular, for $K = 8$ the AbsErr reaches 50 digits of precision. The behaviour of the absolute errors is similar for the three considered cases of $K = 4$, 6, 8, for the three different Jordan curves.

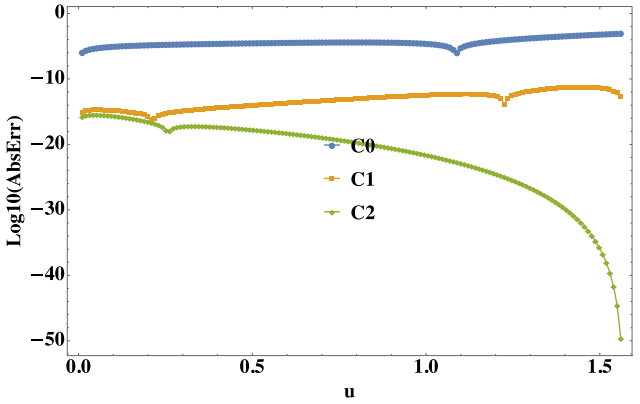

Figure 8. Absolute errors of the approximate solution u_K with $K = 8$ for the contours $\mathcal{C}_0$ (blue), $\mathcal{C}_1$ (orange), and $\mathcal{C}_2$ (green) as a function of u for $u \in (0, \pi/2$ or, equivalently, for $\tau \in (0, \tau_1 = 1 + \pi/2]$.

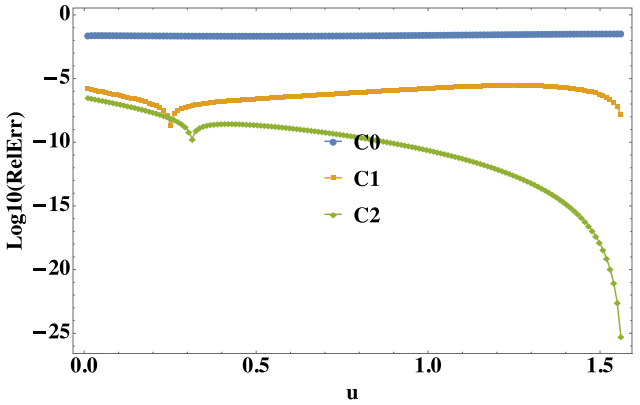

Figure 9. Relative errors of the approximate solution u_K with $K = 4$ for the contours $\mathcal{C}_0$ (blue), $\mathcal{C}_1$ (orange), and $\mathcal{C}_2$ (green) as a function of u for $u \in (0, \pi/2]$ or, equivalently, for $\tau \in (0, \tau_1 = 1 + \pi/2]$. The behaviour of relative error is very similar to the one of the absolute error for the three considered Jordan circles.

In Figs 9 and 10, we display the relative errors for $K = 4$ and 8, and for the three contours $\mathcal{C}_0$, $\mathcal{C}_1$, and $\mathcal{C}_2$ for $u \in (0, \pi/2]$. As it can be seen, they show a similar behaviour as in the case of absolute errors.

However, in the interval $\tau \in (\tau_1 = \pi/2 + 1, \pi)$, the accuracy of the approximate integral solution u_K of $\phi(u) = \tau$ strongly depends on the closeness of τ to π. The reason is that as remarked above $\tau = \pi$ is a singular point in the implicit equation $\phi(u) = \tau$ in the unknown u because $(\mathrm{d}\tau/\mathrm{d}u) \to 0$ when $u \to \pi$, or else when $\tau \to \pi$ and in particular the spectral accuracy with K does not hold. To show these properties of the numerical solution we split the interval $\tau \in (\tau_1, \pi)$ into the two subintervals $\tau \in (\tau_1, \tau_2)$, and $\tau \in (\tau_2, \pi)$, with $\tau_2 = \phi^{-1}(\pi/2 + \pi/3) = 2.617\,993\,78$.

In Fig. 11, we display the AbsErr(τ) for the Jordan curve $\mathcal{C}_1$ for $K = 4$, 8, 16 in the first subinterval $\tau \in (\tau_1, \tau_2)$, or else $u \in (\pi/2, \pi/2 + \pi/3)$. The solutions are very accurate for τ close to τ_1 and their accuracy diminishes when τ is close to τ_2. In any case, the spectral accuracy holds with K.

In Fig. 12, we show the AbsErr(τ) for the second subinterval $\tau \in (\tau_2, \pi)$ or else $u \in (\pi/2 + \pi/3, \pi)$ for $K = 4$, 8, 16. The accuracy

Downloaded from https://academic.oup.com/mnras/article/514/1/1258/6590834 by Universidad de Zaragoza user on 14 September 2022

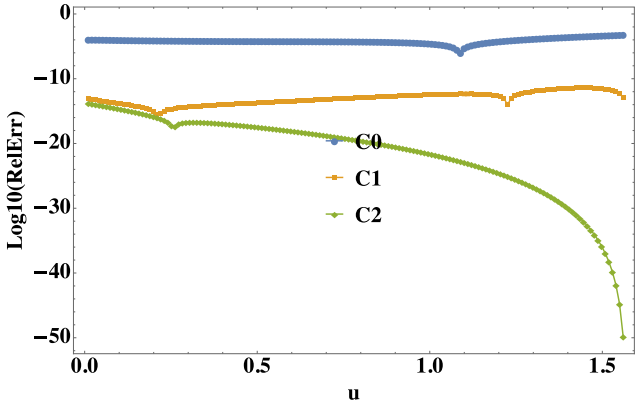

Figure 10. Relative errors of the approximate solution u_K with $K = 8$ for the contours $\mathcal{C}_0$ (blue), $\mathcal{C}_1$ (orange), and $\mathcal{C}_2$ (green) as a function of u for $u \in (0, \pi/2]$ or, equivalently, for $\tau \in (0, \tau_1 = 1 + \pi/2)$.

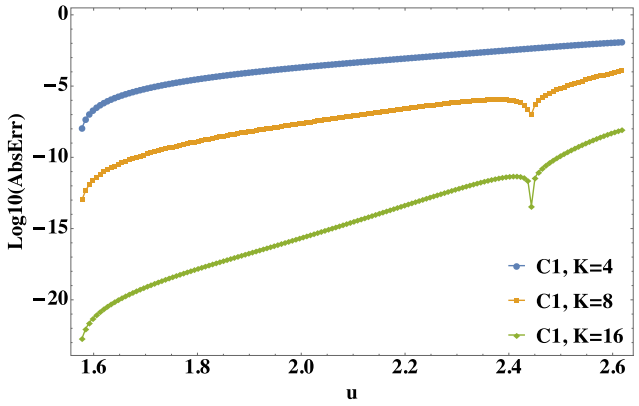

Figure 11. Absolute errors of the approximate solution u_K with $K = 4, 8, 16$ for the contour $\mathcal{C}_1$ as a function of u for $u \in (\pi/2, \pi/2 + \pi/3)$ or, equivalently, for $\tau \in (\tau_1 = 1 + \pi/2, \tau_2 = 3.11799...)$.

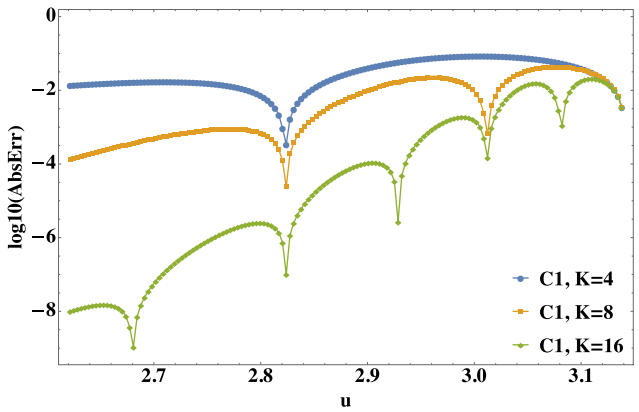

Figure 12. Absolute errors of the approximate solution u_K with $K = 4, 8, 16$ for the contour $\mathcal{C}_1$ as a function of u for $u \in (\pi/2 + \pi/3, \pi)$ or, equivalently, for $\tau \in (\tau_2, \pi)$. Note that in this interval, although the accuracy is better as K increases, eventually there is a lose of accuracy as u approaches to π. The same happens for the other Jordan curves due to singularity of the inverse of equation (10) at $u = \pi$.

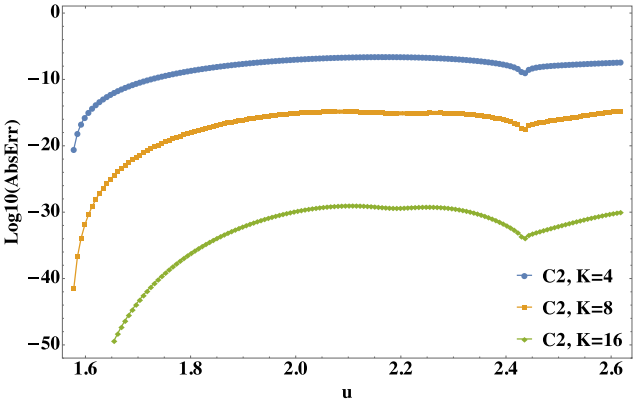

Figure 13. Absolute errors of the approximate solution u_K for $K = 4, 8, 16$ along the Jordan curve $\mathcal{C}_2$ for $u \in (\pi/2, \pi/2 + \pi/3)$ or, equivalently, for $\tau \in (\tau_1, \tau_2 = 3.117999...)$.

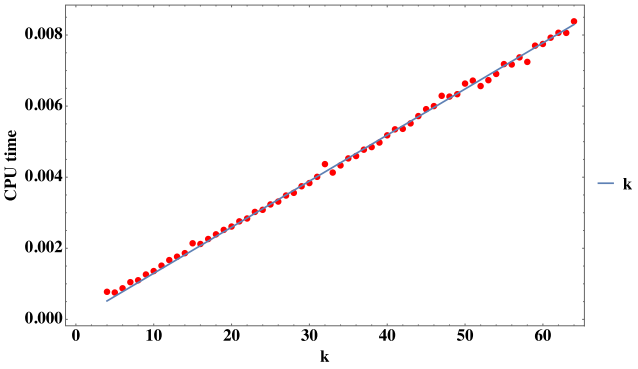

Figure 14. CPU time for the Jordan curve $\mathcal{C}_0$ for $4 \leq K \leq 64$, with its linear fit, which means that the CPU time is linear with K.

of the solutions improves with K but now the max AbsErr does not exhibit spectral accuracy.

In Fig. 13, we present the AbsErr for the contour $\mathcal{C}_2$ and $K = 4$, 8, 16. Note again, for the Jordan curve $\mathcal{C}_1$ the radius is $\rho_0 = 1/2$ for all $\tau \in (\tau_1, \pi)$ whereas for the contour $\mathcal{C}_2$ the radius satisfies $\rho_1 \to 0$ when $\tau \to \tau_1$, and $\rho_1 \to 1/2$ when $\tau \to \pi$. Because of this, the absolute errors (AbsErr) with $\mathcal{C}_2$ are much smaller when τ is close to τ_1, whereas are very similar to those of $\mathcal{C}_1$ when $\tau \to \pi$.

Although our main goal in this paper was to increase the accuracy by reducing the radius of the Jordan cicle in numerically integrating equation 11, it raised a doubt as to whether our method is faster than the one followed by Slepian & Philcox (2021). These authors use the FFT to get coefficients $\widetilde{G}(k; \tau)$ for $k = -1, -2$, and it is known that FFT behaves like $K \log K$. In our case, we computed the CPU needed for the Slepian & Philcox Jordan curve $\mathcal{C}_0$ and for $4 \leq K \leq 64$ and make a linear fit (K) of the resulting points, represented in Fig. 14.

Besides, in order to have a fair CPU comparison, we took the same Jordan curve $\mathcal{C}_0$, and for different values of K ($4 \leq K \leq 128$) we compute in the same computer the solution by using the PHYTON software provided by Slepian & Philcox (2021) and our own PHYTON code (GitHub). The results are displayed in Fig. 15. As we can see, our method is faster that the one of Slepian & Philcox, and the difference increases with K. Our method with K points is almost equivalent to the SP method with $2K$ points; this is due to the symmetry of the function $G(w; \tau)$ that allows to reduce the

Downloaded from https://academic.oup.com/mnras/article/514/1/1258/6590834 by Universidad de Zaragoza user on 14 September 2022

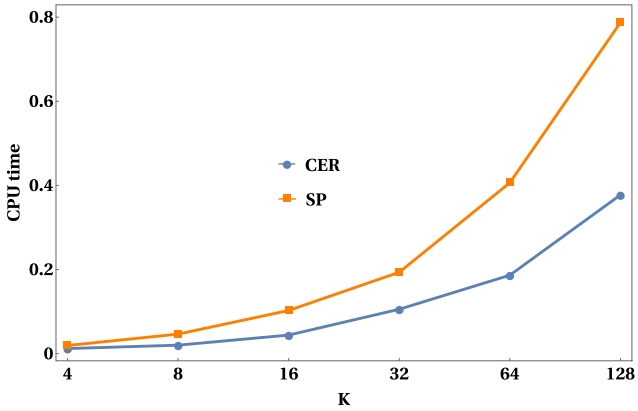

Figure 15. CPU time for the Jordan curve $\mathcal{C}_0$ for $4 \leq K \leq 128$ for both methods computed at the same computer. In orange is the Slepian and Philcox (SP) method, and in blue is the one here proposed (CER).

integration of one period (2π) to a half period (π). Moreover, the CPU in our case is smaller because we only need to compute two Fourier coefficients of the function $G(w;\tau)$, namely $\widehat{G}(-2;\tau)$ and $\widehat{G}(-1;\tau)$, and then, we only have to compute sums. But using the FFT, it is necessary to compute it for $|k| \leq K/2$.

6. CONCLUSIONS

The numerical calculation of the explicit integral solution (11) of the transcendental equation $\phi(u) \equiv u + \sin u = \tau$ of the homologous collapse's radial evolution problem, that relates the parameter u with the physical time t is studied. Since the integral solution depends on the location of the roots of $\phi(u) = \tau$ in the complex plane, a complete study of this point is given in Proposition 3.1. On the other hand, we consider the two main issues that determine the accuracy of the approximate solution: the choice of the Jordan curves in the curvilinear integrals and the quadrature rule to approximate these integrals.

For the quadrature rule, we use the composite trapezoidal rule because of its reliability (recall that it is used in the FFT) and the spectral convergence behaviour with respect to the number of nodes K.

Concerning the Jordan contours, we show that in the case of circles, a reduction of the radius improves the numerical accuracy of the solution. Hence, taking into account the above mentioned Proposition 3.1, for our numerical experiments, we consider two new families of circles $\mathcal{C}_1$, and $\mathcal{C}_2$, and $\mathcal{C}_0$, the one given by Slepian & Philcox (2021).

Finally, we present the results of a number of numerical experiments to show the following items:

(i) The spectral convergence of the numerical solution $u_K(\tau)$, that is, $|u(\tau) - u_K(\tau)| \approx \mathcal{O}(\exp(-\alpha(\tau))\, K)$, with a positive function $\alpha(\tau)$ that tends to zero when $\tau \to \pi^-$.

(ii) The error behaviour $|u(\tau) - u_K(\tau)|$ as a function of K along the interval $\tau \in (0, \pi)$.

(iii) The accuracy is a non-uniform function of $\tau \in (0, \pi)$. In the three considered circles, it is greater near zero, and highly decreases when τ approaches to π.

(iv) The shorter the radius is, the greater is the accuracy, hence circles of $\mathcal{C}_2$ provide the higher accuracy with respect to the other two families of Jordan circles $\mathcal{C}_0$ and $\mathcal{C}_1$.

(v) The computational cost behaves linearly with K, whereas the FFT used by Slepian & Philcox (2021) behaves like $K \log K$. We checked that indeed out method is faster by integrating the same problem with two PHYTON codes, ours and the one of Slepian & Philcox.

(vi) Our method with K points (or data) is in computational cost almost equivalent to the Slepian & Philcox method with $2K$ data. This is due to the fact that we take into account the symmetry of $G(w; tau)$ with respect to w that allows us to use only N points in $[0, \pi]$ instead of the $2K$ data in the period $[0, 2\pi]$.
Notwithstanding the above, since usually the value of K is relatively small, both computing times are similar.

ACKNOWLEDGEMENTS

We are extremely indebted to an annonymous reviewer whose comments and criticisms helped us in improving the manuscript.

This work has been supported by Grants PID2019-109045-GB-C31 and PID2020-117066-GB-I00 funded by MCIN/AEI/10.13039/501100011033 and by the Aragon Government and European Social Fund (groups E24-20R and E41-20R).

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author. A implementation of our code is available on GitHub.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Apostol T. M., 1974, Mathematical Analysis. 2nd ed., Addison-Wesley, Reading, MA
Choudhuri A. R., 2012, The Physics of Fluids and Plasmas, An Introduction for Astrophysicists. Cambridge Univ. Press, Cambridge
Danby J. M. A., 1988, Fundamentals of Celestial Mechanics, 2nd ed., Willmann-Bell, Inc., Richmond, VA
Dyson J. E., Williams D. A., 1997, The Physics of the Interstellar Medium, 2nd ed., IoP Publishing, Boca Raton, FL
Heaton H., Holycross H. W., Myers C. E., 1894, Problem 30. Amer. Math. Monthly, 1:4, 395
Kippenhahn R., Weigert A., 1990, Stellar Structure and Evolution. Springer-Verlag, Berlin
Lin C. C., Mestel L., Shu F. H., 1965, ApJ, 142, 1431
Philcox O. H. E., Goodman J., Slepian Z., 2021, MNRAS, 506, 6111
Slepian Z., Philcox O. H. E., 2021, MNRAS preprint (arXiv:2103.09823)
Tomita K., 1969, Prog. Theor. Phys., 42, 9
Trefethen L. N., 2000, Spectral Methods in MATLAB. SIAM, Philadelphia
Ullisch I., 2020, Math Intelligencer, 42, 12

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article/514/1/1258/6590834 by Universidad de Zaragoza user on 14 September 2022