MATHEMATICS OF COMPUTATION
Volume 76, Number 259, July 2007, Pages 1341–1356
S 0025-5718(07)01945-X
Article electronically published on March 7, 2007

PARABOLIC AND HYPERBOLIC CONTOURS FOR COMPUTING THE BROMWICH INTEGRAL

J. A. C. WEIDEMAN AND L. N. TREFETHEN

Abstract. Some of the most effective methods for the numerical inversion of the Laplace transform are based on the approximation of the Bromwich contour integral. The accuracy of these methods often hinges on a good choice of contour, and several such contours have been proposed in the literature. Here we analyze two recently proposed contours, namely a parabola and a hyperbola. Using a representative model problem, we determine estimates for the optimal parameters that define these contours. An application to a fractional diffusion equation is presented.

1. Introduction

Since the advent of modern numerical analysis the numerical inversion of the Laplace transform has been an active area of research. Recently, however, a marked increase in activity has occurred, partly in connection with the solution of linear fractional differential equations. These equations describe phenomena such as anomalous diffusion in fields like mathematical finance and mathematical biology, two areas of applied mathematics that are particularly active at the present time. This renewed interest brought variations on old themes, and in this paper we analyze some of these new developments.

The methods for inverting the Laplace transform that we shall consider are all based on numerical integration of the Bromwich complex contour integral

$$f(t) = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} e^{zt} F(z)\, dz, \qquad \sigma > \sigma_0. \tag{1.1}$$

Here $F(z)$ is the transform that needs to be inverted, σ_0 is the convergence abscissa, and $\sigma > \sigma_0$. This means that all the singularities of $F(z)$ lie in the open half-plane $\operatorname{Re} z < \sigma$.

As it stands, the integral (1.1) is not well-suited for numerical integration. First, the exponential factor is highly oscillatory on the Bromwich line, $z = \sigma + iy$, $-\infty < y < \infty$. Second, the transform $F(z)$ typically decays slowly as $|y| \to \infty$. Integrals that are both oscillatory and slowly decaying on unbounded domains are often hard to compute.

Received by the editor December 9, 2005.
2000 *Mathematics Subject Classification.* Primary 65D30, 44A10.
Key words and phrases. Laplace transform, Talbot's method, trapezoidal rule, fractional differential equation.
The first author was supported by the National Research Foundation in South Africa under grant FA2005032300018.

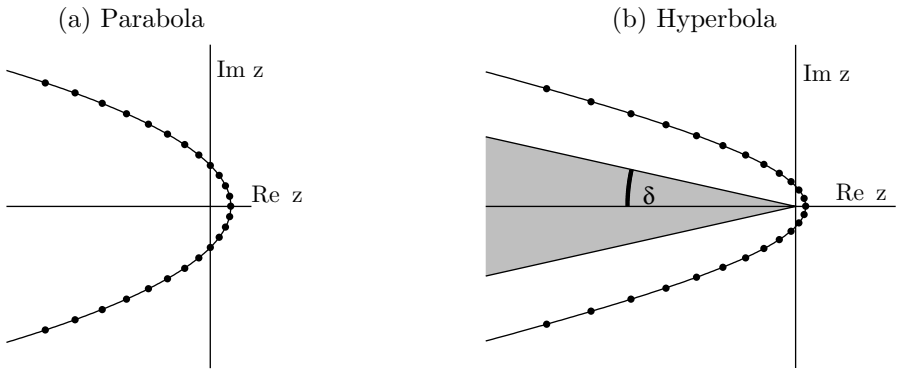

FIGURE 1. The contours (1.2) and (1.3). The hyperbolic contour encloses the shaded sectorial region whenever $\alpha < \pi/2 - \delta$; cf. (1.4).

One strategy for circumventing the slow decay is due to Talbot [10], who suggested that the Bromwich line be deformed into a contour that begins and ends in the left half-plane, such that $\mathrm{Re}\, z \to -\infty$ at each end. On such a contour, the exponential factor in (1.1) forces a rapid decay of the integrand as $\mathrm{Re}\, z \to -\infty$. This makes the integral well suited for approximation by the trapezoidal or midpoint rules [5]. Owing to Cauchy's theorem, such a deformation of contour is permissible as long as no singularities are traversed in the process, and provided that $|F(z)| \to 0$ uniformly in $\mathrm{Re}\, z \le \sigma_0$ as $|z| \to \infty$. (It should be clear, however, that this technique cannot be applied to all transforms. For example, if the singularities of $F(z)$ have an unbounded imaginary part, Talbot's method may fail.)

Talbot's original contour has a cotangent shape, and is rather complicated to analyze; see [7], [10], [14]. Two simpler types of contours have been proposed recently, namely parabolas [1] and hyperbolas [2], [6], [9]; see Figure 1. The parabola is parameterized by

$$z = \mu\Big(iu + 1\Big)^2, \qquad -\infty < u < \infty. \tag{1.2}$$

(In fact, a more general parabola was used in [1], but we shall argue that for the model problem discussed here, it is sufficient to consider this form of the contour.) We consider the hyperbola in the form used in [2], namely

$$z = \mu\Big(1 + \sin\big(iu - \alpha\big)\Big), \qquad -\infty < u < \infty. \tag{1.3}$$

In Section 4 we verify that this is indeed a hyperbola, with asymptotes

$$y = \pm \cot(\alpha)(\mu - x), \qquad z = x + iy. \tag{1.4}$$

In both cases the positive parameter μ controls the width of the contour. The hyperbola contains the additional free parameter α, which determines its asymptotic angle. In this respect the hyperbola is superior to both the parabola and Talbot's contour—it allows for singularities with an unbounded imaginary part, provided they lie in a sectorial region around the negative real axis; see Figure 1. (Note that the contours shown in Figure 1 are schematic and not intended to be optimal. Actual optimal contours can be seen in Figures 2 and 5 below.)

On either of the contours (1.2), (1.3) the Bromwich integral (1.1) becomes

$$f(t) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} e^{z(u)t} F\big(z(u)\big) z'(u)\, du. \tag{1.5}$$

Approximation by the trapezoidal rule, with uniform node spacing h, yields the approximation $f_{h;N}(t) \approx f(t)$, where

$$f_{h;N}(t) = \frac{h}{2\pi i} \sum_{k=-N}^{N} e^{z(u_k)t}\, F\big(z(u_k)\big)\, z'(u_k), \qquad u_k \equiv kh. \tag{1.6}$$

In this formula there are some free parameters, notably the geometric parameter μ and the node spacing h. The hyperbola contains the additional parameter α. Our main objective in this article is to derive estimates for these parameters that will ensure, for given t and N, a small error

$$E_{h;N}(t) = f(t) - f_{h;N}(t).$$

Of course, the accuracy of the method is closely related to the singularities of the transform $F(z)$, but analyzing all possible singularity distributions is impractical. To make the analysis manageable, we shall focus on a model problem, namely

$$e^{\lambda t} = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} \frac{e^{zt}}{z-\lambda}\, dz, \qquad \lambda < 0. \tag{1.7}$$

That is, we are primarily concerned with singularities on the negative real axis, a choice motivated by the solution of parabolic PDEs. When solving such PDEs, the analogous formula is

$$\exp(At)\, \boldsymbol{u}_0 = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} e^{zt} \Big((zI - A)^{-1} \boldsymbol{u}_0 \Big)\, dz, \tag{1.8}$$

where A is typically a $J \times J$ symmetric negative definite matrix approximation to the Laplacian, and $\boldsymbol{u}_0$ is a $J \times 1$ vector representing the initial condition. The scalar λ in (1.7) represents a typical eigenvalue of A.

The outline of the paper is as follows. In the next section, we recall a basic error estimate for trapezoidal approximations. This estimate is applied in Section 3 (parabola) and Section 4 (hyperbola) to derive formulas for the optimal parameters μ, h, and α, as well as estimates of the corresponding convergence rates. An application to a time-fractional diffusion equation is presented in Section 5. Comparisons with related methods for inverting the Laplace transform are made in Section 6.

2. Error estimates

Our error analysis is based on well-known error estimates for the trapezoidal rule on the real line. Consider, in analogy with (1.5)–(1.6), the absolutely convergent integral

$$I = \int_{-\infty}^{\infty} g(u)\, du,$$

and its infinite and finite trapezoidal approximations

$$I_h = h \sum_{k=-\infty}^{\infty} g(kh), \qquad I_{h;N} = h \sum_{k=-N}^{N} g(kh).$$

The quantity $DE = |I - I_h|$ is often referred to as the discretization error, whereas $TE = |I_h - I_{h;N}|$ is the truncation error.

Estimating the discretization error is a well-known exercise in numerical analysis. Proofs are typically based on the residue theorem or the theory of Fourier transforms, and examples of each technique can be found, respectively, in [5] and [8, p. 345]. The only difference here is that we shall take our function $g(u)$ to be complex-valued, so its singularities need not be symmetric with respect to the real axis as is usually the case. Nevertheless, the standard proofs remain essentially the same, and therefore we omit the proof of the following theorem.

Theorem 2.1. *Let $w = u + iv$, with u and v real. Suppose $g(w)$ is analytic in the strip $-d < v < c$, for some $c > 0$, $d > 0$, with $g(w) \to 0$ uniformly as $|w| \to \infty$ in that strip. Suppose further that for some $M_+ > 0$, $M_- > 0$ the function $g(w)$ satisfies*

$$\int_{-\infty}^{\infty} |g(u + ir)|\, du \le M_+, \qquad \int_{-\infty}^{\infty} |g(u - is)|\, du \le M_-,$$

for all $0 < r < c$, $0 < s < d$. Then

$$|I - I_h| \le DE_+ + DE_-,$$

where

$$DE_+ = \frac{M_+}{e^{2\pi c/h} - 1}, \qquad DE_- = \frac{M_-}{e^{2\pi d/h} - 1}.$$

Remark 2.2. In the literature the function g is typically real-valued, $c = d$, $M_+ = M_- = M$, say, and the error estimate reduces to the more familiar

$$|I - I_h| \le \frac{2M}{e^{2\pi c/h} - 1}.$$

Remark 2.3. In Section 3 we encounter the situation $c < \infty$ but $d = \infty$. In this case

$$DE_+ = O\big(e^{-2\pi c/h}\big), \qquad DE_- = O\big(M_-(d) e^{-2\pi d/h}\big), \qquad h \to 0. \tag{2.1}$$

To derive a good estimate for DE_-, an optimal value of d should be determined by balancing the growth of $M_-(d)$ with the decay of $e^{-2\pi d/h}$ as $d \to \infty$.

As for the truncation error, if one assumes that $g(u)$ decays rapidly as $u \to \pm\infty$, the truncation error can be approximated by the magnitude of the last term retained in the trapezoidal sum. That is, for fixed h,

$$TE = O\big(|g(hN)|\big), \qquad N \to \infty. \tag{2.2}$$

In the next section, we shall determine optimal parameters for the parabolic contour (1.2), by asymptotically matching DE_+, DE_-, and TE. This gives two equations from which the two free parameters, μ and h, can be determined.

3. The parabola

Consider the conformal map

$$z = \mu\Big(iw + 1\Big)^2, \tag{3.1}$$

which reduces to (1.2) when w is real. Let $c > 0$ and consider the line

$$w = u + ic, \qquad -\infty < u < \infty. \tag{3.2}$$

Its image under the map (3.1) is

$$z = \mu\Big((1-c)^2 - u^2\Big) + 2\,i\,\mu\,u\,\Big(1-c\Big). \tag{3.3}$$

According to the theory of Section 2, we need to find a strip of analyticity as wide as possible, i.e., we would like to maximize c.

As c is increased from 0, the parabola (3.3) closes and degenerates into the negative real axis when c reaches the value 1. In Theorem 2.1 we therefore let $c = 1$ and according to the first formula in (2.1) the associated error is given by

$$DE_+ = O\Big(e^{-2\pi/h}\Big), \qquad h \to 0. \tag{3.4}$$

Turning to the lower half-plane, we consider for $d > 0$,

$$w = u - id, \qquad z = \mu\Big((1+d)^2 - u^2\Big) + 2\,i\,\mu\,u\,\Big(1+d\Big), \qquad -\infty < u < \infty.$$

As d increases from 0, the parabola widens and moves to the right. At the same time, the factor e^{zt} grows and contributes a factor $e^{\mu(1+d)^2 t}$ to the quantity $M_-(d)$ in (2.1); therefore

$$DE_- = O\Big(e^{\mu(1+d)^2 t - 2\pi d/h}\Big),$$

for all $d > 0$. Define

$$p(d) = \mu(1+d)^2 t - 2\pi d/h;$$

then the best choice of d is determined by setting $p'(d) = 0$. This yields

$$d = \frac{\pi}{\mu t h} - 1 \tag{3.5}$$

(which is positive for h sufficiently small). With this d the best estimate of DE_- becomes

$$DE_- = O\Big(e^{-\pi^2/(\mu t h^2) + 2\pi/h}\Big), \qquad h \to 0. \tag{3.6}$$

Turning to the truncation error, we use (2.2) to estimate

$$TE = O\big(e^{\mu t(1-(hN)^2)}\big), \qquad N \to \infty. \tag{3.7}$$

Asymptotically balancing DE_+, DE_-, and TE yields the two equations

$$-\frac{2\pi}{h} = -\frac{\pi^2}{\mu t h^2} + \frac{2\pi}{h} = \mu t(1-(hN)^2). \tag{3.8}$$

Solving them, one obtains expressions for the optimal parameters:

$$h_* = \frac{3}{N}, \qquad \mu_* = \frac{\pi}{12}\frac{N}{t}. \tag{3.9}$$

The corresponding convergence rate is geometric, namely

$$E_N = O\big(e^{-2\pi N/3}\big) \approx O\big(8.12^{-N}\big), \qquad N \to \infty. \tag{3.10}$$

Figure 2 shows the optimal strip of analyticity in the case $N = 12$ and $t = 1$. The lines w_+ and w_- in the figure represent $w = u + i$ and $w = u - di$. Their images in the z-plane are denoted by z_+ and z_-.

To test the validity of (3.9), we have implemented the method (1.6) on the model problem (1.7). In Figure 3 we show a contour plot of the numerically computed (absolute) error, in the (μ, h) parameter plane. Darker shades represent higher accuracy.

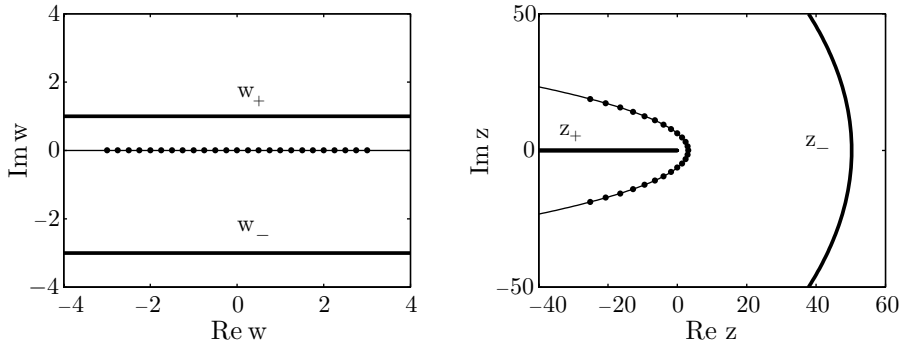

FIGURE 2. The strip of analyticity in the w-plane for the parabolic contour, and its image in the z-plane. This represents the optimal configuration in the case $N = 12$ and $t = 1$.

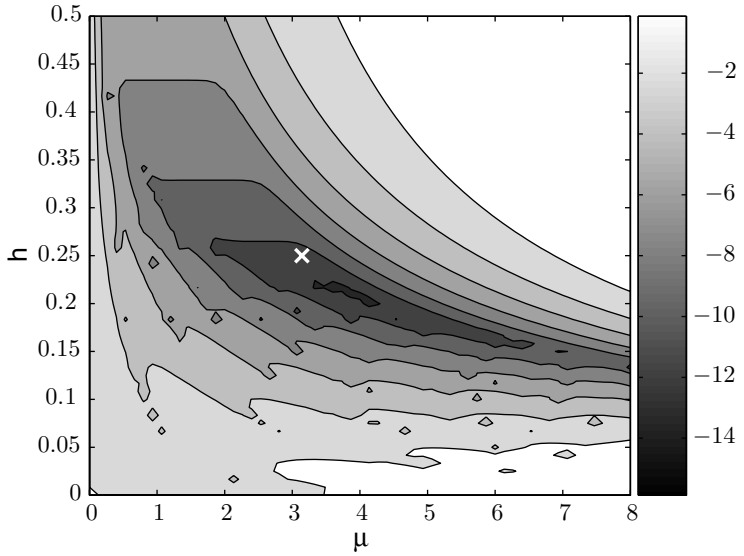

FIGURE 3. Contour plot of $\log_{10}|f(t) - f_{h;N}(t)|$, where the test problem is defined by (1.7), with $\lambda = 0$, $t = 1$. Here the parabolic contour is used, with $N = 12$. The cross represents the parameter estimates (3.9).

Figure 3 lends credibility to our error model. In the lower part of the figure (i.e., small h) the truncation error, TE, dominates. In the northwest section of the figure the error DE_+ dominates; note that this error is independent of μ and hence the plot shows horizontal contour lines in this area. In the north and northeast sections of the plot the error DE_- dominates. The white cross marks the spot defined by (μ_*, h_*), and is seen to be close to the actual optimal point.

It is important to note that the parameter estimates (3.9) depend on t. This means that for each new value of t the transform $F(z)$ will be sampled on a different contour. For a scalar problem such as (1.7) this is not a major issue, but for a

matrix problem such as (1.8) this could be inefficient. Ideally one would like to fix parameters and then use a single contour for various values of t. In the case of the classic Talbot contour and its hyperbolic variant, suggestions for finding such a contour were put forward, respectively, in [7] and [3]. In the case of the parabola we argue as follows.

By examining the error estimates (3.4), (3.6), and (3.7), one observes that only the latter two depend on t. The discretization error DE_- increases with t, whereas the truncation error TE decreases with t. If we wish to maintain a small absolute error on an interval $t_0 \le t \le t_1$, say, we therefore require in place of (3.8) that

$$-\frac{2\pi}{h} = -\frac{\pi^2}{\mu t_1 h^2} + \frac{2\pi}{h} = \mu t_0 (1 - (hN)^2). \tag{3.11}$$

The solution of these two equations is

$$h_* = \frac{\sqrt{8\Lambda + 1}}{N}, \qquad \mu_* = \frac{\pi}{4\sqrt{8\Lambda+1}} \frac{N}{t_1}, \tag{3.12}$$

where we have defined

$$\Lambda = \frac{t_1}{t_0}.$$

The corresponding convergence rate is again geometric, but with decreasing decay rate as Λ increases from 1:

$$E_N = O\Big(e^{-\left(2\pi/\sqrt{8\Lambda+1}\right) N}\Big), \qquad N \to \infty. \tag{3.13}$$

Note that these parameter and error estimates reduce to (3.9) and (3.10) in the case $\Lambda = 1$.

Figure 4 represents an empirical check on the accuracy of the estimate (3.13). We have plotted, for $t_0 = 1$ and various Λ, the quantity

$$M_N = \max_{t_0 \le t \le \Lambda t_0} |f(t) - f_{h;N}(t)|, \tag{3.14}$$

as a function of N. The test problem was (1.7), with $\lambda = -1$. The dash-dot lines represent the theoretical error estimate (3.13), and agree well with the actual convergence rate shown as the solid curves.

In Figure 4, note that the error curves eventually start to increase as N gets larger. This is a consequence of roundoff error and characteristic of the ill-conditioning associated with the inversion of the Laplace transform. (These results, and those elsewhere in the paper, were computed in MATLAB with machine precision about 2.2×10^{-16}.) See also Remark 4.2 below.

To conclude this section, we consider whether it might be advantageous to consider a more general parabola of the form

$$z = \sigma + \mu(iw + \alpha)^2.$$

First, the parameter α cannot improve the accuracy. This is clear from the equivalent expression $z = \sigma + \mu\alpha^2(iw/\alpha + 1)^2$, which shows that α may be absorbed into the parameter μ and the step-size h. Second, the parameter σ will contribute factors of $e^{\sigma t}$ to the error estimates derived above. This can only reduce the error if $\sigma < 0$, but then the contour intersects the negative real axis. This is unacceptable when solving problems where the precise locations of the λ's are not known.

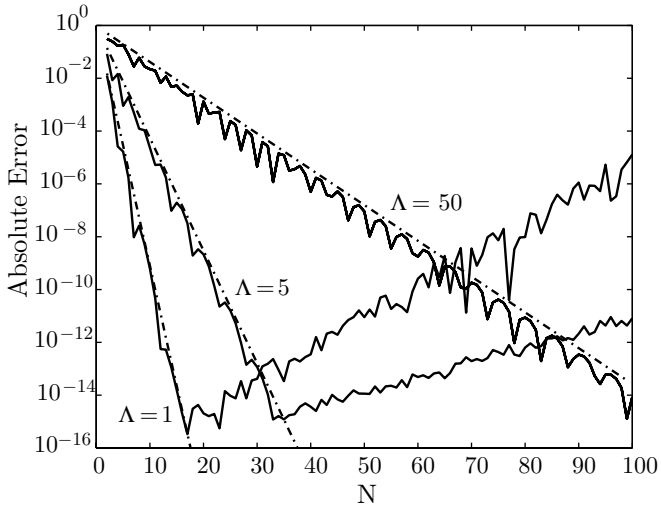

FIGURE 4. Plot of the maximum absolute error on $[t_0, \Lambda t_0]$; cf. (3.14). The parabolic contour is used, with parameter estimates (3.12). The solid curves represent the actual errors, and the dash-dot lines the theoretical convergence estimate (3.13).

4. The hyperbola

Next, we consider

$$z = \mu\Big(1 + \sin\big(iw - \alpha\big)\Big), \tag{4.1}$$

which is identical to (1.3) when w is real. We use essentially the same analysis as in the previous section.

The image of the horizontal line $w = u + ic$ is

$$z = \mu\,\Big(1 - \sin(\alpha + c)\cosh(u)\Big) + i\,\mu\,\cos(\alpha + c)\sinh(u),$$

which can be expressed as the hyperbola

$$\Big(\frac{\mu - x}{\sin(\alpha + c)}\Big)^2 - \Big(\frac{y}{\cos(\alpha + c)}\Big)^2 = \mu^2, \qquad z = x + iy.$$

When $c = 0$ the left branch of this hyperbola is the contour of integration as shown in Figure 1(b).

When c increases from 0, the hyperbola closes and degenerates into the negative real axis when c reaches the value $\pi/2 - \alpha$. The associated error expression is

$$DE_+ = O(e^{-2\pi(\pi/2-\alpha)/h}), \qquad h \to 0.$$

We examine the lower half of the w-plane by considering $c < 0$. When c decreases from 0 the hyperbola widens until it reaches the limiting value $c = -\alpha$, at which point the image in the z-plane becomes a vertical line. For Theorem 2.1 to remain valid, one cannot decrease c beyond this point, since the factor e^{zt} then grows unboundedly. Therefore we take $c = -\alpha$, or $d = \alpha$ in the formula for DE_- in (2.1).

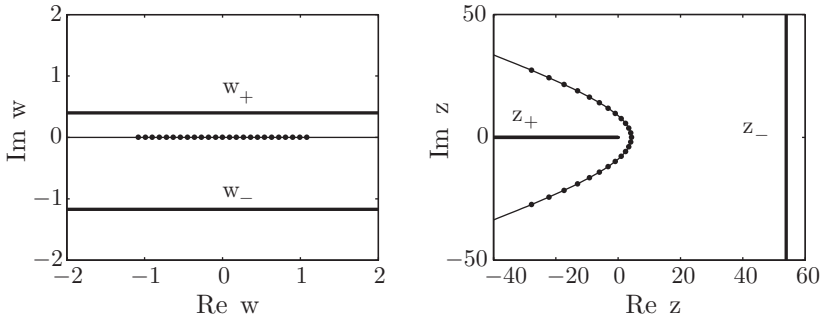

FIGURE 5. Same as Figure 2, but for the hyperbolic contour.

In addition, $M_-(d) = O(e^{\mu t})$, and the associated error formula is

$$DE_- = O\big(e^{\mu t - 2\pi\alpha/h}\big), \qquad h \to 0.$$

The truncation error is estimated in the usual manner as

$$TE = O\big(e^{\mu t(1-\sin(\alpha)\cosh(hN))}\big), \qquad N \to \infty.$$

The strip of analyticity used in deriving these estimates is shown in Figure 5.

An asymptotic balance of the three quantities DE_+, DE_-, and TE yields the two equations

$$-\frac{2\pi(\pi/2-\alpha)}{h} = \mu t - \frac{2\pi\alpha}{h} = \mu t\Big(1-\sin(\alpha)\cosh(hN)\Big). \tag{4.2}$$

Considering α to be the free variable, one can solve for the parameters h and μ

$$h = \frac{A(\alpha)}{N}, \qquad \mu = \frac{4\pi\alpha - \pi^2}{A(\alpha)}\frac{N}{t}, \tag{4.3}$$

where

$$A(\alpha) = \cosh^{-1}\Big(\frac{2\alpha}{(4\alpha-\pi)\sin\alpha}\Big).$$

Provided α is in $(\pi/4, \pi/2)$, the above choices of h and μ provide, once again, a geometric convergence rate as $N \to \infty$, namely

$$E_N = O(e^{-B(\alpha)N}), \qquad \text{where} \qquad B(\alpha) = \frac{\pi^2 - 2\pi\alpha}{A(\alpha)}. \tag{4.4}$$

The function $B(\alpha)$ has a unique local maximum in $(\pi/4, \pi/2)$, say at $\alpha = \alpha_*$, which we have computed numerically to four places as

$$\alpha_* = 1.1721. \tag{4.5}$$

The corresponding optimal parameters h and μ are

$$h_* = \frac{1.0818}{N}, \qquad \mu_* = 4.4921\frac{N}{t}. \tag{4.6}$$

The optimal decay rate is $B(\alpha_*) = 2.3157$, which predicts a convergence rate of about

$$E_N = O(e^{-2.32N}) = O(10.2^{-N}), \qquad N \to \infty. \tag{4.7}$$

This is faster convergence than the corresponding rate (3.10) that can be achieved on the parabolic contour.

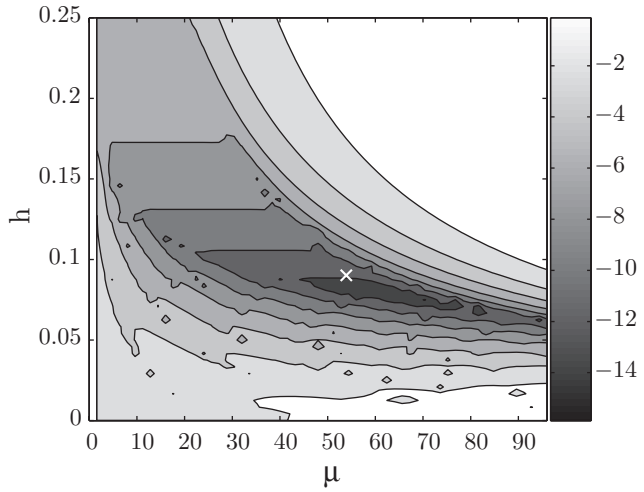

FIGURE 6. Same as Figure 3, but for the hyperbolic contour, with $\alpha = 1.1721$. The cross represents the parameter estimates (4.6).

In Figure 6 we show a contour plot of the error in the (μ, h) parameter plane. As in the case of Figure 3, the numerical results confirm the validity of the theoretical error model to a good degree. The estimated optimal parameters (μ_*, h_*) (indicated by the cross) are indeed close to the actual minimum point.

In the event that the solution needs to be computed in an interval $[t_0, t_1]$, with $t_1 = \Lambda t_0$, the same analysis as in Section 3 can be used; cf. eqs. (3.11)–(3.13). Eq. (4.2) is replaced by

$$-\frac{2\pi(\pi/2-\alpha)}{h} = \mu t_1 - \frac{2\pi\alpha}{h} = \mu t_0\Big(1 - \sin(\alpha)\cosh(hN)\Big), \tag{4.8}$$

the solution of which is given by

$$h = \frac{A(\alpha)}{N}, \qquad \mu = \frac{4\pi\alpha - \pi^2}{A(\alpha)}\frac{N}{t_1}. \tag{4.9}$$

The definition of $A(\alpha)$ is now modified to

$$A(\alpha) = \cosh^{-1}\left(\frac{(\pi-2\alpha)\Lambda + 4\alpha - \pi}{(4\alpha-\pi)\sin\alpha}\right),$$

and the estimated convergence rate is given by (4.4), with

$$B(\alpha) = \frac{\pi^2 - 2\pi\alpha}{\cosh^{-1}\left(\frac{(\pi-2\alpha)\Lambda+4\alpha-\pi}{(4\alpha-\pi)\sin\alpha}\right)}. \tag{4.10}$$

For a given $\Lambda \geq 1$, the maximum of $B(\alpha)$ can be computed numerically. Results for a few values of Λ are given in Table 1.

These numbers are tested in Figure 7. The solid curves are the actual errors, and the dash-dot curves the estimated convergence rates $O\big(e^{-B(\alpha)N}\big)$, with $B(\alpha)$ given in the final column of the table.

What if the singularities lie off the real axis, e.g., in the sectorial region shown in Figure 1(b)? It is easy to see that the above estimates for the truncation error TE and the discretization error $DE\_$ are unaffected by this change. To obtain an

TABLE 1. Optimal parameters for the hyperbola for approximation on an interval $[t_0, \Lambda t_0]$.

Λ	α	$hN\ (= A(\alpha))$	$(\mu t_1)/N$	$B(\alpha)$
1	1.1721	1.0818	4.4921	2.3157
2	1.1431	1.5280	2.9417	1.7587
5	1.0791	2.4580	1.5013	1.2570
50	0.9381	5.5582	0.3452	0.7152

estimate for DE_+, one cannot increase the value of c beyond $\pi/2 - \alpha - \delta$, for at this point the image of the horizontal line w_+ in the w-plane becomes a hyperbola z_+ in the z-plane with asymptotes parallel to the edges of the shaded sector in Figure 1(b). The estimate for DE_+ should therefore be modified to

$$DE_+ = O(e^{-2\pi(\pi/2-\alpha-\delta)/h}).$$

With the left side of (4.8) modified accordingly, one can solve as before for h and μ:

$$h = \frac{A(\alpha)}{N}, \qquad \mu = \frac{4\pi\alpha - \pi^2 + 2\pi\delta}{A(\alpha)} \frac{N}{t_1}, \tag{4.11}$$

where

$$A(\alpha) = \cosh^{-1}\left(\frac{(\pi - 2\alpha - 2\delta)\Lambda + (4\alpha - \pi + 2\delta)}{(4\alpha - \pi + 2\delta)\sin\alpha}\right).$$

The estimated decay rate of the error is

$$B(\alpha) = \frac{\pi^2 - 2\pi\alpha - 2\pi\delta}{A(\alpha)}. \tag{4.12}$$

For given Λ and δ, the function $B(\alpha)$ can be maximized to find the optimal α.

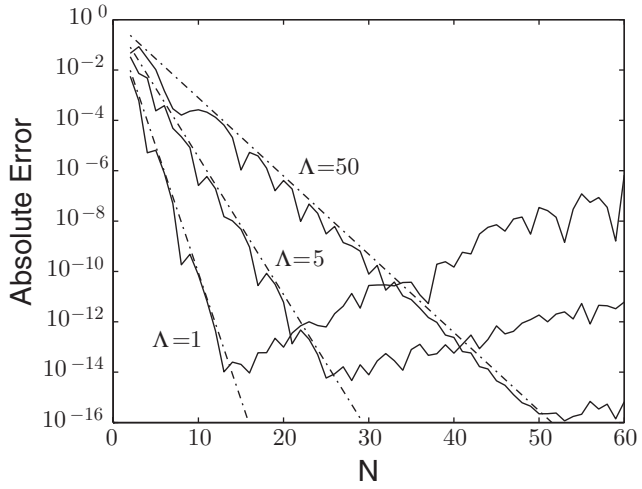

FIGURE 7. Same as Figure 4, but the hyperbolic contour is used, with parameter estimates given in Table 1. Note that the horizontal scale is different in the two figures.

We point out that this analysis related to the sectorial region is based on a worst-case scenario. It will only work well if the singularities are located essentially everywhere on the edges of the sectorial region. If only a few singularities relatively close to the origin are present, for example, it should be possible to increase c beyond the value $\pi/2-\alpha-\delta$, to obtain a better estimate of the convergence rate. In this case the optimal parameters will depend on the singularity distribution as well as the value of t and is best handled on a case-by-case basis. We shall not pursue this here.

Remark 4.1. Results similar to those presented in Figure 7 can be found in [3], and a few remarks on the connection to our work are in order. The hyperbolic contour (1.3) was introduced in [2] and further analyzed in [3]. Parameter estimates derived in the latter reference are similar, but not identical, to our estimates (4.9), namely

$$h=\frac{a(\theta)}{N},\qquad \mu=\frac{2\pi d(1-\theta)}{a(\theta)}\frac{N}{t_1},$$

where

$$a(\theta)=\cosh^{-1}\Big(\frac{\Lambda}{(1-\theta)\sin\alpha}\Big). \tag{4.13}$$

Here d and θ are parameters that satisfy $0<d<\pi/2-\alpha$ (in the case of singularities on the negative real axis), and $0<\theta<1$. No explicit formulas nor guidelines for choosing α and d optimally were provided in [3], but optimal values for θ were obtained numerically. Comparing the analysis of [3] with what we have done here, the distinction is evident. In [3] the authors used a symmetric strip of analyticity (the parameter d is, in fact, the half-width of that strip). In our analysis we utilized the fact that a non-symmetric strip of analyticity offers an additional degree of freedom that can be exploited. Indeed, when one compares Figure 7 to the corresponding figure in [3], one sees that the convergence rates observed here are significantly better.

Remark 4.2. An interesting idea pursued in [3] was to minimize the effects of ill-conditioning by having the parameter θ depend on N. In this manner the authors were able to obtain error curves that saturate around the numerical round-off level, i.e., the error growth observed in Figures 4 and 7 was avoided. The explicit dependence of θ on N was not determined however. We shall not pursue the matter of conditioning here.

5. An application

One application of the type mentioned in the Introduction is the time-fractional diffusion equation [4]

$$D_t^\beta u=u_{xx},\qquad 0\le x\le\pi. \tag{5.1}$$

The operator D_t^β is Caputo's fractional derivative, defined for $0<\beta<1$ by

$$D_t^\beta f(t)=\frac{1}{\Gamma(1-\beta)}\int_0^t\frac{f'(s)}{(t-s)^\beta}\,ds.$$

As $\beta\to 1$, the fractional derivative approaches the ordinary derivative, and (5.1) reduces to the standard diffusion equation $u_t=u_{xx}$. With $0<\beta<1$, the PDE is classified as sub-diffusive [4].

We supplement (5.1) with boundary conditions

$$u(0,t) = 0, \qquad u(\pi,t) = 0, \tag{5.2}$$

and an initial condition

$$u(x,0) = \sin x. \tag{5.3}$$

By introducing a suitable grid and a discrete approximation to the second derivative plus boundary conditions (5.2), the problem can be semi-discretized as

$$D_t^\beta \, \boldsymbol{u} = A \, \boldsymbol{u}, \qquad \boldsymbol{u}(t) = \boldsymbol{u}_0. \tag{5.4}$$

For A, we shall use the Chebyshev spectral collocation second derivative matrix of order $J \times J$; see [11], [15]. Below we shall use $J = 16$, which is sufficient to resolve the spatial profile to within machine precision.

After taking a Laplace transform, the problem reduces to computing

$$\boldsymbol{u}(t) = \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} e^{zt} \boldsymbol{F}(z) \, dz, \tag{5.5}$$

where

$$\big(zI - z^{1-\beta} A\big) \boldsymbol{F}(z) = \boldsymbol{u}_0. \tag{5.6}$$

The method (1.6) may be applied to the integral (5.5), at the cost of solving the $J \times J$ (full) linear system (5.6) at each node $z(kh)$ of the trapezoidal rule. Because of symmetry, however, only $N + 1$ system solves rather than $2N + 1$ are actually required.

The analytical solution to the semi-discrete problem (5.4) can be expressed in terms of the Mittag-Leffler function [4]

$$\boldsymbol{u}(t) = E_\beta\big(At^\beta\big) \, \boldsymbol{u}_0, \qquad \text{where} \qquad E_\beta(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\beta k + 1)}. \tag{5.7}$$

As $\beta \to 1$, this reduces to $\boldsymbol{u}(t) = \exp\big(At\big)\boldsymbol{u}_0$, which is the matrix exponential solution to the semi-discrete diffusion equation. On the other hand, if $\beta = 1/2$ the Mittag-Leffler function reduces to a form involving the complementary error function, and the solution to (5.1)–(5.3) is then given by

$$u(x,t) = e^t \operatorname{erfc}(\sqrt{t}) \, \sin x. \tag{5.8}$$

Using this as a reference solution, we intend to confirm the theory of Sections 3 and 4.

Note that the eigenvalues of A lie on the negative real axis [16]. This means that the transform in (5.5) has branch-point singularities on that axis. Our analysis of Sections 3–4 used only the location of the singularities, not their type, and therefore we expect the estimates of those sections to remain valid. This is indeed the case, as confirmed in Figure 8.

In that figure, we have plotted, as a function of N, the quantity

$$M_N = \max_{t_0 \le t \le \Lambda t_0} \|u(x,t) - u_N(x,t)\|_2, \tag{5.9}$$

where $u(x,t)$ is the exact solution (5.8), and $u_N(x,t)$ is the numerical solution computed by approximating (5.5)–(5.6). The L_2-norm in (5.9) is approximated by Clenshaw-Curtis quadrature, the natural rule for the Chebyshev collocation method. We reiterate that the errors shown in the figure come primarily from the

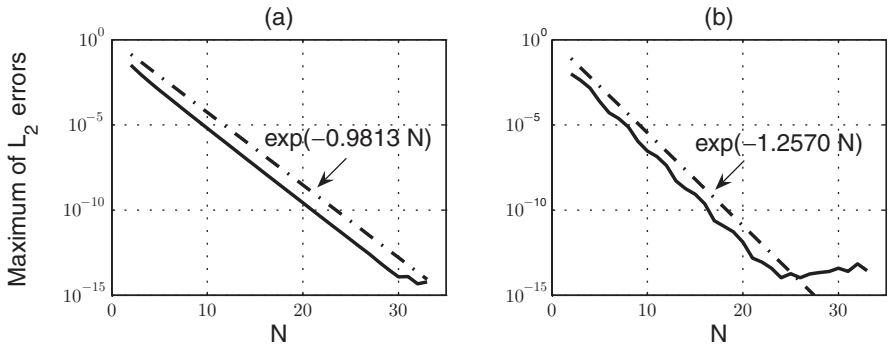

FIGURE 8. The maximum L_2-error on $0.5 \le t \le 2.5$, as defined by the quantity M_N in (5.9). In (a) the parabolic contour is used with parameter values (3.12), and in (b) the hyperbolic contour with parameter values given in Table 1. Solid curves represent actual errors and dash-dot lines their theoretical estimates.

Laplace transform approximation (i.e., the "time integration"); the space variable is treated exactly for all practical purposes.

The solid curves in Figure 8 represent the computed values of M_N, whereas the dash-dot lines represent the theoretical convergence estimates (3.13) and (4.4), with the decay rate of the latter given in Table 1. Parameter values were $t_0 = 0.5$ and $\Lambda = 5$.

The good agreement between theory and practice is evident in the figure, as is the rapid convergence for both the parabola and the hyperbola. To put the rapid convergence into perspective: Suppose one wishes to solve the problem (5.1)–(5.3) on the interval $0.5 \le t \le 2.5$ uniformly to within an L_2-error of, say, 10^{-10}. According to the figure, this will require a value of $N \approx 21$ (resp. $N \approx 16$) for the parabolic (resp. hyperbolic) contour. The required accuracy can therefore be attained at the relatively small cost of solving 22 (resp. 17) linear systems of size $J \times J$, plus the asymptotically negligible cost of computing a vector sum of the form (1.6) for selected values of t in the interval of interest.

Similar convergence curves that focused on a fixed value of t (i.e., $\Lambda = 1$) were presented earlier in [13].

6. Conclusions

We analyzed Talbot's method for inverting the Laplace transform on two types of contours, one parabolic and the other hyperbolic. Optimal parameter choices for a representative model problem were derived. The optimal convergence rates thus derived favor the hyperbolic contour over the parabolic one.

Our analysis included the cases where the problem has to be solved at a single value of t, as well as on an interval $[t_0, t_1]$. In addition, we considered the situation for which the hyperbolic contour was devised, namely where the singularities lie off the real axis but in a sectorial region about the negative real axis; cf. (4.11)–(4.12).

As an alternative to the contours analyzed here, one should mention the original Talbot contour [10]. Optimal parameters of the type derived here were obtained in [14]. It was also shown, however, that the convergence of the classic Talbot

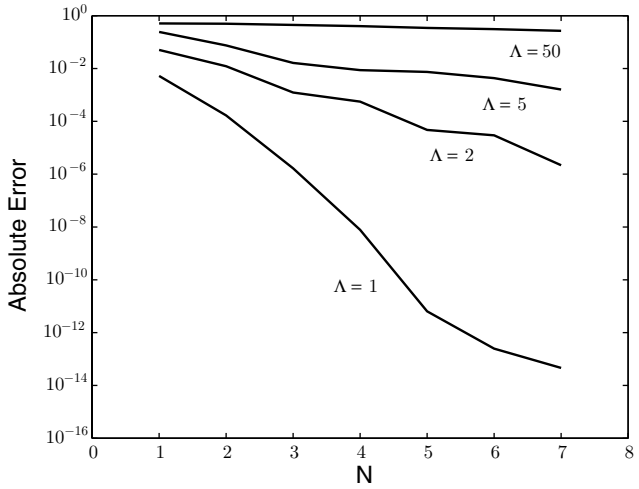

FIGURE 9. Same as Figures 4 and 7, but the best rational approximation method of [12] is used. The error is defined by (6.1).

method can be improved dramatically by a simple modification. At present we are investigating the optimal parameters and convergence rate that can be achieved on this modified Talbot contour, and results will be reported elsewhere.

A second alternative, introduced in [12], is the method based on best rational approximation to e^z on the negative real axis. When computing solutions at a single value of t, this method converges about twice as fast as the methods discussed here. When solutions are computed on an interval $t_0 \le t \le \Lambda t_0$, however, this method loses its superiority as Λ gets large; see [12]. For a different illustration of the same property, we refer to Figure 9.

In that figure, we have plotted the quantity

$$M_N = \min_{t_0 \le t_* \le \Lambda t_0} \max_{t_0 \le t \le \Lambda t_0} |f(t) - f_{h;N}(t)|, \tag{6.1}$$

as a function of N. To make these errors commensurable with those in the rest of the paper, we have used a rational approximation of type $(2N-1, 2N)$ (and not $(N-1, N)$ as in [12]). In (6.1), t_* is the point in $[t_0, \Lambda t_0]$ where coefficients of the rational approximation are calculated; see [12] for details.

The test problem of Figure 9 was the same as the one of Figures 4 and 7. Comparing these three figures (and keeping in mind that the horizontal scales are all different) the superiority of the best rational approximation method is evident when $\Lambda = 1$. It is also clear, however, that the other two methods, and in particular the one based on the hyperbolic contour, surpass it quickly as Λ increases.

ACKNOWLEDGMENTS

Much of this work was done while the first author was on sabbatical leave at the Computing Laboratory of Oxford University, as a Visiting Fellow of Exeter College. Kevin Burrage brought the time-fractional diffusion equation (5.1) to our attention. Helpful conversations with Patrick Kano, María López-Fernández, César Palencia, Achim Schädle, and Thomas Schmelzer are also gratefully acknowledged.

References

1. I. P. Gavrilyuk and V. L. Makarov, *Exponentially convergent parallel discretization methods for the first order evolution equations*, Comput. Methods Appl. Math. **1** (2001), no. 4, 333–355. MR1892950 (2003f:65174)
2. M. López-Fernández and C. Palencia, *On the numerical inversion of the Laplace transform of certain holomorphic mappings*, Appl. Numer. Math. **51** (2004), no. 2-3, 289–303. MR2091405 (2005e:65210)
3. M. López-Fernández, C. Palencia, and A. Schädle, *A spectral order method for inverting sectorial Laplace transforms*, SIAM J. Numer. Anal. **44** (2006), no. 3, 1332–1350 (electronic). MR2231867
4. F. Mainardi, G. Pagnini, and R. K. Saxena, *Fox H functions in fractional diffusion*, J. Comput. Appl. Math. **178** (2005), no. 1-2, 321–331. MR2127888 (2005m:26008)
5. E. Martensen, *Zur numerischen Auswertung uneigentlicher Integrale*, Z. Angew. Math. Mech. **48** (1968), T83–T85. MR0256565 (41:1221)
6. W. McLean and V. Thomée, *Time discretization of an evolution equation via Laplace transforms*, IMA J. Numer. Anal. **24** (2004), no. 3, 439–463. MR2068831 (2005d:47072)
7. M. Rizzardi, *A modification of Talbot's method for the simultaneous approximation of several values of the inverse Laplace transform*, ACM Trans. Math. Software **21** (1995), no. 4, 347–371. MR1364695 (96k:65084)
8. H.-R. Schwarz, *Numerical analysis*, John Wiley & Sons Ltd., Chichester, 1989. MR1005534 (90g:65003)
9. D. Sheen, I.H. Sloan, and V. Thomée, *A parallel method for time discretization of parabolic equations based on Laplace transformation and quadrature*, IMA J. Numer. Anal. **23** (2003), no. 2, 269–299. MR1975267 (2004b:65161)
10. A. Talbot, *The accurate numerical inversion of Laplace transforms*, J. Inst. Math. Appl. **23** (1979), 97–120. MR526286 (80c:65244)
11. L. N. Trefethen, *Spectral methods in MATLAB*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000. MR1776072 (2001c:65001)
12. L. N. Trefethen, J. A. C. Weideman, and T. Schmelzer, *Talbot quadratures and rational approximations*, BIT **46** (2006), no. 3, 653–670. MR2265580
13. J. A. C. Weideman, *Computing special functions via inverse Laplace transforms*, International Conference on Numerical Analysis and Applied Mathematics 2005 (Rhodes) (T.E. Simos, G. Psihoyios, and Ch. Tsitouras, eds.), Wiley-VCH, 2005, pp. 702–704.
14. ———, *Optimizing Talbot's contours for the inversion of the Laplace transform*, SIAM J. Numer. Anal. **44** (2006), no. 6, 2342–2362.
15. J. A. C. Weideman and S. C. Reddy, *A MATLAB differentiation matrix suite*, ACM Trans. Math. Software **26** (2000), no. 4, 465–519. MR1939962 (2003g:65004)
16. J. A. C. Weideman and L. N. Trefethen, *The eigenvalues of second-order spectral differentiation matrices*, SIAM J. Numer. Anal. **25** (1988), no. 6, 1279–1298. MR972454 (90a:65228)

Department of Applied Mathematics, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa

E-mail address: weideman@dip.sun.ac.za

Oxford University Computing Laboratory, Wolfson Bldg., Parks Road, Oxford OX1 3QD, United Kingdom

E-mail address: LNT@comlab.ox.ac.uk