An improved Talbot method for numerical Laplace transform inversion

BENEDICT DINGFELDER, Zentrum Mathematik, M3, Technische Universität München, Boltzmannstrasse 3, 85748 Garching bei München, Germany (dingfelder@mytum.de)
J.A.C. WEIDEMAN, Department of Mathematical Sciences, University of Stellenbosch, Stellenbosch 7600, South Africa (weideman@sun.ac.za)

The classical Talbot method for the computation of the inverse Laplace transform is improved for the case where the transform is analytic in the complex plane except for the negative real axis. First, by using a truncated Talbot contour rather than the classical contour that goes to infinity in the left half-plane, faster convergence is achieved. Second, a control mechanism for improving numerical stability is introduced. These two features are incorporated into a software code, whose performance is assessed on transforms from tables as well as from actual applications. It is shown that even when the transform has singularities off the negative real axis, rapid convergence can still be achieved in many cases.

Categories and Subject Descriptors: G.1.2 [**Numerical Analysis**]: Approximation, quadrature

General Terms: Inverse Laplace transform, Talbot's method, trapezoidal rule, midpoint rule

Additional Key Words and Phrases: Inverse Laplace transform, Talbot's method, trapezoidal rule, midpoint rule

1. INTRODUCTION

Let $f(t)$ be real-valued and piecewise continuous on $[0,\infty)$ and of exponential order $f = O(e^{\gamma_0 t})$, $t \to \infty$, for some real constant γ_0. Then the Laplace transform of $f(t)$ is defined by

$$F(z) = \int_0^\infty e^{-zt} f(t)\, dt, \qquad \mathrm{Re}\, z > \gamma_0.$$

This paper deals with the numerical solution of the inverse problem, i.e., given $F(z)$, compute $f(t)$ at a specified value of t. The function $F(z)$ is analytic in the half-plane $\mathrm{Re}\, z > \gamma_0$, and if it can be evaluated in the complex plane (as opposed to just on the real axis), this analyticity can and should be exploited by good numerical methods. One such method is numerical quadrature applied to the inverse formula

$$f(t) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{zt} F(z)\, dz, \qquad \mathrm{Re}\, \gamma > \gamma_0, \tag{1}$$

known as the Bromwich integral. Two effective quadrature rules are the simple trapezoidal and midpoint rules, used in combination with contour deformation.

The purpose of the contour deformation is to exploit the exponential factor in (1). In particular, assume the path of integration in (1) can be deformed to a Hankel contour, i.e., a contour whose real part begins at negative infinity in the third quadrant, then winds around all singularities and terminates with the real part again going to negative infinity in the second quadrant. Examples are shown in Figure 1 below. On such contours the exponential factor causes a rapid decay, which makes the integral particularly suitable for approximation by the trapezoidal or midpoint rules. The contour deformation can be justified by Cauchy's theorem, provided the contour remains in the domain of analyticity of $F(z)$. Some mild restrictions on the decay of $F(z)$ in the left half-plane are also required; for sufficient conditions, see [Talbot 1979; Trefethen et al. 2006].

Suppose such a Hankel contour can be parameterized by

$$C:\ z = z(\theta), \qquad -\pi \le \theta \le \pi, \tag{2}$$

where $\mathrm{Re}\, z(\pm\pi) = -\infty$. Then

$$f(t) = \frac{1}{2\pi i} \int_C e^{zt} F(z)\, dz = \frac{1}{2\pi i} \int_{-\pi}^{\pi} e^{z(\theta)t} F(z(\theta)) z'(\theta)\, d\theta. \tag{3}$$

We approximate the latter integral by the N-panel midpoint rule with uniform spacing $h = 2\pi/N$, which yields

$$f(t) \approx \frac{1}{Ni} \sum_{k=1}^{N} e^{z(\theta_k)t} F(z(\theta_k)) z'(\theta_k), \qquad \theta_k = -\pi + \Big(k - \frac{1}{2}\Big)h. \tag{4}$$

If the contour (2) is symmetric with respect to the real axis, and if $F(\overline{z}) = \overline{F(z)}$ (which is the case for $f(t)$ real-valued), then half of the transform evaluations can be saved. That is, one needs to consider only quadrature nodes in the upper (or lower) half-plane. When comparing numerical results it is important to keep this point in mind as several other papers, including [Duffy 1993; Weideman 2006; Weideman and Trefethen 2007], considered quadrature rules with a total of $2N$ nodes.

Popular contours C are the parabola [Butcher 1957; Gavrilyuk and Makarov 2005] and the hyperbola [Sheen et al. 2003; López-Fernández and Palencia 2004; Gavrilyuk and Makarov 2005], as well as the cotangent contour introduced in [Talbot 1979]. Here we consider the Talbot contour in the form

$$z(\theta) = \frac{N}{t}\,\zeta(\theta), \quad \zeta(\theta) = -\sigma + \mu\,\theta\cot(\alpha\theta) + \nu\,i\,\theta, \quad -\pi \le \theta \le \pi, \tag{5}$$

where σ, μ, ν and α are constants to be specified by the user. In the original Talbot contour the parameter α did not appear, i.e., $\alpha = 1$. For reasons outlined below, the modification we propose is to consider $0 < \alpha < 1$. The scaling factor N/t was introduced in the original paper [Talbot 1979] and also proposed in [Weideman 2006]. Assuming a fixed value of t, this is the appropriate scaling for maximizing the convergence rate in the trapezoidal/midpoint approximation as $N \to \infty$.

In the original Talbot contour ($\alpha = 1$), it follows that $\mathrm{Re}\, z(\pm\pi) = -\infty$. If the trapezoidal rule is used, this means the point at infinity is a quadrature node and the corresponding function value is set to zero in the sum (4). As this seems wasteful, we recommend the use of the midpoint rule instead.

The practical issue addressed in this paper is the choice of the parameters σ, μ, ν and α in (5). The main considerations are summarized in Figure 1. The best contour has to strike a balance between passing too close to the singularities, or too far from them in which case the exponential factor in (3) becomes too large. The quadrature nodes on the best contour should also extend just far enough into the left half-plane to reach the desired accuracy but their contributions should not be negligibly small.

In the case $\alpha = 1$, suggestions for parameter values were made in the original paper [Talbot 1979]. These were implemented in ACM TOMS Algorithm 682 [Murli and Rizzardi 1990]. The performance of this algorithm on a set of challenging test problems was assessed in [Duffy 1993]. In Section 4 we show that the modified contour (5) with the parameter choices proposed here can produce the same accuracy but often with significantly fewer terms in the sum (4).

It is unrealistic to try and find optimal contours for all possible transforms $F(z)$. That is why we limit the analysis here to the situation that seems to occur most frequently in applications, namely the case where all singularities of $F(z)$ are located on the negative real axis. We shall see, however, that the contour parameters derived here also work reasonably well when singularities are off the real axis (but not too far, depending on the value of t).

In the case $\alpha = 1$, an optimized set of parameters σ, μ, ν was proposed in [Weideman 2006]. When all singularities of $F(z)$ are located on the negative real axis, the achievable convergence rate is $O(e^{-0.949N})$ as $N \to \infty$ for fixed $t > 0$. Because $\mathrm{Re}\, z(\pm\pi) = -\infty$ in the original Talbot method, however, too many nodes end up far out in the left half-plane where they make no contribution to the midpoint sum (4). We acknowledge [Trefethen 2006] for this observation, and the suggestion to consider a truncated Talbot contour instead, with $0 < \alpha < 1$. In this case we refer to (5) as the modified Talbot contour.

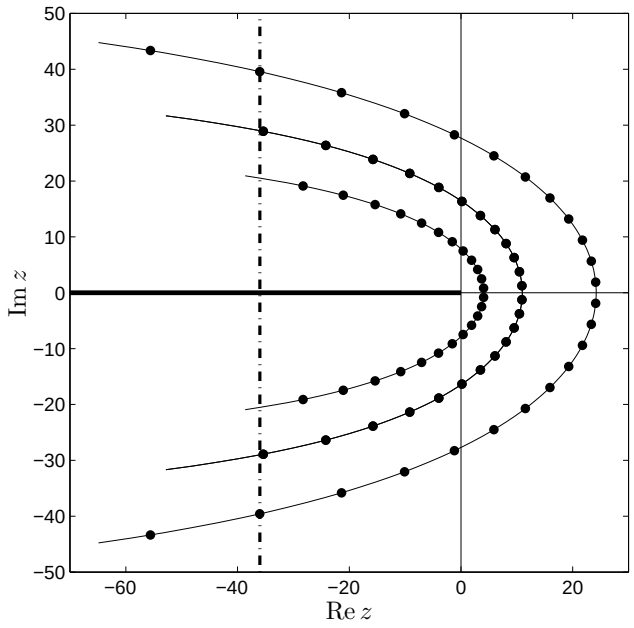

Fig. 1. Three possible Talbot contours (5), each shown with $N = 24$ nodes in the midpoint rule. The middle contour is close to optimal, in the case where $t = 1$ and the singularities of $F(z)$ are located on the negative real axis. Its outermost conjugate pair of nodes reaches right up to the dash-dot curve, where the magnitude of the exponential factor in (3) becomes less than the desired accuracy (here taken to be machine precision, $\epsilon \approx 2.2 \times 10^{-16}$). The inner contour is suboptimal because (a) its outermost two nodes do not reach quite up to the dash-dot curve and (b) it passes too close to the singularities. The outer contour is suboptimal because (a) the contributions of the outermost pair of nodes are less than the desired accuracy and therefore wasteful, and (b) the exponential factor in (3) is too large in the right half-plane.

Optimal parameters σ, μ, ν and α for the modified contour were derived by one of the authors of the present paper, and reported in [Trefethen et al. 2006], later to be included in text books such as [Strang 2007, Sect. 5.3]. Details of the derivation have not been published, however. Having simplified and improved the derivation recently, we now present the details here.

The freedom that the parameter α offers means that the convergence rate $O(e^{-0.949N})$ for the case $\alpha = 1$ can be improved to $O(e^{-1.358N})$, $N \to \infty$, t fixed. This is assuming singularities of $F(z)$ on the negative real axis. By comparison, with the same assumption on the singularity location, the convergence rates achievable by the parabola and hyperbola are, respectively, $O(e^{-1.047N})$ and $O(e^{-1.176N})$; see [Weideman and Trefethen 2007].

It should be noted that the scaling factor N/t in (5) has a few negative consequences. First, because the nodes $z(\theta_k)$ depend on N, function values cannot be re-used as N is increased. The increased convergence rate compensates for this, however. Second, for fixed t the contour can move far into the right half-plane as N is increased. Because of the exponential factor in (1), the terms in the summation can become large and hence suffer from floating-point cancelation errors. We address this issue in Section 3. Third, because the quadrature nodes depend on t as well, the same contour cannot be used for two or more distinct values of t. In this case the method would be inefficient when the transform $F(z)$ is expensive to evaluate. We shall not address the latter issue here, but note that for the case $\alpha = 1$ it was considered in [Rizzardi 1995] and for the parabolic and hyperbolic contours it was considered in [Weideman and Trefethen 2007].

The outline of the paper is as follows: In Section 2 we present the main error analysis that leads to the proposed values of σ, μ, ν and α. The analysis is similar to the one given in [Weideman 2006]. It is slightly more intricate because of the additional parameter α, but at the same time we have made

some improvements that simplify the analysis. A strategy for improving numerical stability for large N is discussed in Section 3. For now it only makes provision for roundoff error control when the singularities of $F(z)$ are located on the negative real axis. Section 4 contains the numerical tests, where we compare the results of the modified contour (5) with parameters proposed here, with Algorithm 682 [Murli and Rizzardi 1990]. Further tests and the details of a code, that implements the proposed method are discussed in Section 5. The code, written in MATLAB [MATLAB 2012], is publicly available for download from the web site [Dingfelder and Weideman 2013].

2. ERROR ESTIMATES

Our basic error estimate is a well-known contour integral formula for the error in the trapezoidal rule approximation of functions that are analytic, periodic, and real-valued; see for example [Kress 1998, Thm. 9.28]. We retain the property of analyticity but extend the theorem as follows: (A) we consider non-periodicity, (B) we consider complex-valued functions, and (C) we consider the midpoint rule rather than the trapezoidal rule. (A) is necessary because the integrand in (3) is not periodic. (B) is necessary because the integrand is complex-valued, and has very different analytic properties in the upper and lower half-planes. (C) is not really necessary, only a matter of convenience. All three of these modifications can be readily incorporated into the proof given in [Kress 1998, Thm. 9.28], which is an elementary application of the residue theorem. We omit the details.

THEOREM 2.1. *Let θ_k be defined as in (4). If $g : D \to \mathbb{C}$ is analytic in the strip $D = \{\theta \in \mathbb{C} : -\pi < \mathrm{Re}\ \theta < \pi$ and $-d < \mathrm{Im}\ \theta < c\}$ where $c, d > 0$, then*

$$\int_{-\pi}^{\pi} g(\theta)\, d\theta - \frac{2\pi}{N} \sum_{k=1}^{N} g(\theta_k) = \mathrm{E}_+(\gamma) + \mathrm{E}_-(\delta).$$

Here

$$\mathrm{E}_+(\gamma) = \frac{1}{2}\left(\int_{-\pi}^{-\pi+i\gamma} + \int_{-\pi+i\gamma}^{\pi+i\gamma} + \int_{\pi+i\gamma}^{\pi} \right)\Big(1 + i\tan(\frac{N\theta}{2})\Big) g(\theta)\, d\theta$$

and

$$\mathrm{E}_-(\delta) = \frac{1}{2}\left(\int_{-\pi}^{-\pi-i\delta} + \int_{-\pi-i\delta}^{\pi-i\delta} + \int_{\pi-i\delta}^{\pi} \right)\Big(1 - i\tan(\frac{N\theta}{2})\Big) g(\theta)\, d\theta$$

for all $0 < \gamma < c$, $0 < \delta < d$ and $N \in \mathbb{N}$.

The contours of integration for computing the error integrals $\mathrm{E}_+(\gamma)$ and $\mathrm{E}_-(\delta)$ are shown in the left diagram of Figure 2. We remark that if $g(\theta)$ is real-valued on the real axis, then $g(\bar{\theta}) = \overline{g(\theta)}$ and d and c can be taken to be equal. When it is moreover 2π-periodic, the contributions on the sides $\mathrm{Re}\,\theta = \pm\pi$ cancel. In that case, if we define $\mathrm{E}(\gamma) = \mathrm{E}_+(\gamma) + \mathrm{E}_-(\gamma)$, then

$$\mathrm{E}(\gamma) = \mathrm{Re} \int_{-\pi+i\gamma}^{\pi+i\gamma} \Big(1 + i\tan(\frac{N\theta}{2})\Big) g(\theta)\, d\theta.$$

By analyzing the behavior of the tangent function in the complex plane, this can be bounded by [Kress 1998, Thm. 9.28]

$$|\,\mathrm{E}(\gamma)| \le \frac{4\pi M}{e^{cN} - 1},$$

where M is a bound on the magnitude of g in the strip D. For a given value of N, two quantities control the size of the error: c, which is the width of the strip of analyticity, and M, which is governed by the growth of g in the complex plane.

Here we apply the error formula of Theorem 2.1 to the case

$$g(\theta) = \frac{1}{2\pi i}\, e^{z(\theta)t} F(z(\theta)) z'(\theta), \qquad -\pi \le \theta \le \pi, \tag{6}$$

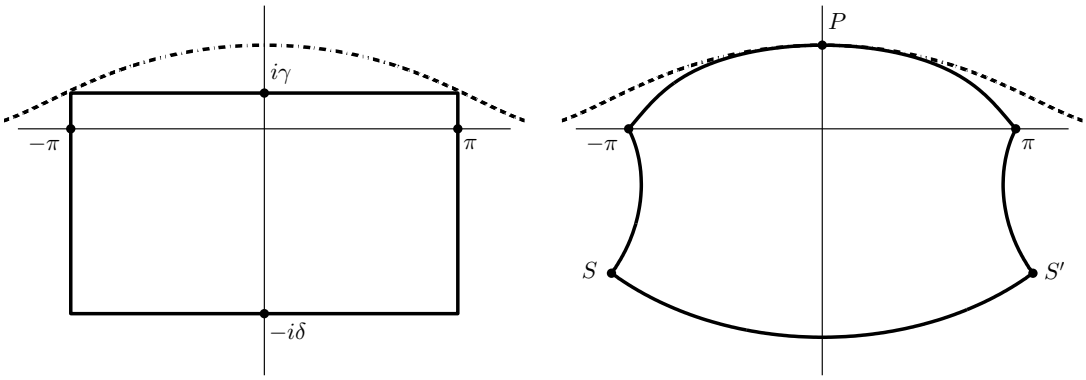

Fig. 2. Left: Rectangular contour in the complex θ-plane for evaluating the error integrals $\mathrm{E}_+(\gamma)$ and $\mathrm{E}_-(\delta)$ of Thm. 2.1. Right: Deformed contour, along which the magnitudes of the integrands of $\mathrm{E}_+(\gamma)$ and $\mathrm{E}_-(\delta)$ are approximately constant. The dash-dot curve in both diagrams shows a possible image of the negative real axis of the z-plane, where singularities might be located.

where $F(z)$ is the transform to be inverted. This is not a real-valued function, which is why the rectangle of integration in Figure 2 is not shown symmetric with respect to the real axis. As we shall see (and similar observations were made in [Weideman 2006; Weideman and Trefethen 2007]), in the upper half-plane the width of the strip is restricted by the singularities of $F(z(\theta))$, while in the lower half-plane the restriction is the size of $e^{z(\theta)t}$.

To keep the analysis tractable, we limit our discussion primarily to the model transform

$$F(z) = (z+\lambda)^{-1}, \qquad \lambda > 0, \tag{7}$$

i.e., singularities on the negative real axis only. In Section 4 we shall consider a matrix version of this transform when considering applications to PDEs.

For (7) and similar transforms, we can estimate the error using Theorem 2.1 and the rectangular contour shown on the left in Figure 2. For a better estimate, one can try and deform the contour as shown on the right in Figure 2.

To explain how this can be done, consider the error integrals in Theorem 2.1. Ignoring the constant factors $1/2$ in those integrals, as well as the factor $1/(2\pi i)$ in (6), the error formulas can be expressed as

$$E_\pm = \int_{C_\pm} e^{Nh_\pm(\theta)}\,d\theta, \quad h_\pm(\theta) = \zeta(\theta) + N^{-1}\Big(\log\big(1 \pm i\tan(\frac{N\theta}{2})\big) + \log F(z(\theta)) + \log z'(\theta)\Big),$$

where C_+ and C_- refer to contours in the upper and lower half-planes, respectively. Our strategy is now as follows: suppose it is possible to deform the rectangle on the left in Figure 2 to a contour along which $\mathrm{Re}\, h_\pm(\theta) = -c$ for some positive constant c. Then

$$C_\pm:\ \mathrm{Re}\, h_\pm(\theta) = -c \qquad \Longrightarrow \qquad |E_\pm| = O(e^{-cN}), \tag{8}$$

and the value of c can be maximized by picking the best among all such contours $C_\pm$.

The function $h_\pm(\theta)$ defined above is too complicated to analyze in this manner, and we make a few approximations in the case $N \gg 1$. First, we drop the terms involving $F(z(\theta))$ and $z'(\theta)$. This means $F(z(\theta))$ should not be zero nor exponentially small on any part of the contours $C_\pm$, and likewise $z'(\theta) \neq 0$. Regarding the term involving the tangent function in $h_\pm(\theta)$, we shall approximate it for $N \gg 1$ as follows. By converting the tangent function to exponentials one obtains, with $\theta = x + iy$,

$$1 \pm i\tan\big(\frac{N}{2}(x+iy)\big) = \frac{2\,e^{\mp\frac{1}{2}Ny}e^{\pm\frac{1}{2}iNx}}{e^{-\frac{1}{2}Ny}e^{\frac{1}{2}iNx} + e^{\frac{1}{2}Ny}e^{-\frac{1}{2}iNx}}.$$

Taking the logarithm yields

$$\log\left(1 \pm i\tan(\frac{N\theta}{2})\right) \sim \mp Ny, \qquad N \to \infty, \tag{9}$$

valid for fixed, nonzero y. The upper and lower sign choices correspond to $y > 0$ and $y < 0$, respectively. Accordingly, we approximate the contour $\mathrm{Re}\, h_\pm(\theta) = -c$ in (8) by

$$C_\pm : \quad \mathrm{Re}\, \zeta(x+iy) \mp y = -c. \tag{10}$$

First, consider the upper half of the θ-plane. As mentioned above, the error contour C_+ is restricted by the location of the singularities of $F(z(\theta))$ in the upper half of the θ-plane. For now, consider the model problem (7). Its pole is located at $z = -\lambda$, or $\zeta = -\lambda t/N$. In keeping with the limit $N \to \infty$, t fixed, we consider therefore $\zeta = 0$. The same argument will apply if there is a branch-cut on the negative real z-axis, for in practice we need to consider only a finite section of the branch-cut (e.g., the section between the origin and the dashed line in Figure 1).

We therefore need to examine the zeros of $\zeta(\theta)$, i.e.,

$$-\sigma + \mu\,\theta\cot(\alpha\theta) + \nu\, i\,\theta = 0.$$

It is no easy task to examine the complex roots of this equation for all possible parameter values, so we resorted to numerical experimentation, restricting searches to the strip $-\pi \le \mathrm{Re}\,\theta \le \pi$, which is of interest here. For certain parameter choices there are roots in the lower half of the θ-plane. For other parameter choices the roots are in the upper half-plane, located symmetrically with respect to the imaginary θ-axis. For yet other parameter values the roots are located precisely on the imaginary θ-axis, with the one closest to the real axis the critical one that limits the contour C_+.

Another limiting case is the fact that $\zeta'(\theta)$ may not be 0, as remarked below (8). By numerical experiments, we came to the same conclusion as [Weideman 2006], namely the configuration that gives the widest domain of analyticity in the upper half-plane occurs when these two critical points coincide precisely on the positive imaginary θ-axis. (This is the point marked P in Figure 2.) This leads to the two equations $\zeta(iy) = 0, \zeta'(iy) = 0$ for some $y > 0$. By considering (10), we conclude that $y = c$. If we further let $\theta \to \pm\pi$ on the contour C_+, we get $\zeta(\pm\pi) = -c$, which is by symmetry just one equation, not two. In summary, in the upper half-plane we require that

$$C_+ : \quad \zeta(ic) = 0, \qquad \zeta'(ic) = 0, \qquad \zeta(\pi) = -c. \tag{11}$$

By using these three equations, it is possible to solve for (σ, μ, ν) in terms of (α, c), namely

$$\sigma = 2\alpha c^2 B, \qquad \mu = 2\sinh^2(\alpha c)B, \qquad \nu = (\sinh(2\alpha c) - 2\alpha c)B, \tag{12}$$

where

$$B = \frac{c\sin^2(\alpha\pi)}{2\alpha c^2\sin^2(\alpha\pi) - \pi\sin(2\alpha\pi)\sinh^2(\alpha c)}.$$

It remains to fix the parameters α and c. For this, we consider the lower half of the θ-plane. Here the contour C_- is not restricted by analyticity but by the size of the exponential factor in (3). Again numerical experimentation confirmed the findings of [Weideman 2006], namely that there are two saddle points in the lower half of the θ-plane that are critical. (These are the points marked S and S' in Figure 2.) By making C_- pass through these saddle points, we can enclose a region as big as possible.

Denote the saddle point locations by $\theta = x_s + iy_s$. By inserting this into (10), and taking partial derivatives, yields

$$C_- : \quad \mathrm{Re}\,\zeta(x_s+iy_s) + y_s = -c, \quad \mathrm{Re}\,\zeta'(x_s+iy_s) = 0, \quad \mathrm{Im}\,\zeta'(x_s+iy_s) = 1. \tag{13}$$

We insert the expressions for σ, μ and ν from (12) and fix a value of α. Assuming a solution exists, the three equations (13) can then be solved numerically for the unknowns x_s, y_s, and c. By varying α, the value of c can thus be maximized.

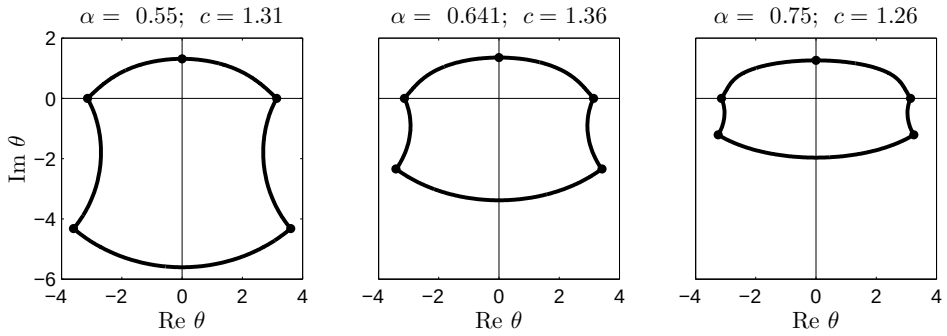

Fig. 3. Three possible deformations of the rectangle of Figure 2. The middle one maximizes c, which is both the value of the intersection on the positive imaginary axis and the decay rate in the error estimate $E = O(e^{-cN})$.

We made no attempt to establish theoretical existence of a solution of the equations (13), but direct numerical computation indicates that such a solution exists for α in the interval $[0.51, 0.82]$, roughly. Some of these configurations are shown in Figure 3, with the middle one corresponding to the maximum value of c. We computed this value of c using a univariate optimization code, and found

$$\alpha = 0.6407; \qquad c = 1.3580. \tag{14}$$

Substitution of these two values into (12) yields the modified Talbot contour

$$z(\theta) = \frac{N}{t}\Big(-0.6122 + 0.5017\,\theta\cot(0.6407\,\theta) + 0.2645\,i\,\theta\Big), \quad -\pi \le \theta \le \pi, \tag{15}$$

as reported in [Trefethen et al. 2006]. For completeness, we record that the corresponding saddle points are located at $x_s + iy_s = \pm 3.4208 - 2.3438\,i$, correct to all digits shown.

An alternative contour in which the cotangent is replaced by a rational function was respectively suggested and analyzed in [Talbot 1979] and [Weideman 2006]. Applying the same strategy as outlined above, we derived

$$z(\theta) = \frac{N}{t}\Big(0.1446 + \frac{3.0232\,\theta^2}{\theta^2 - 3.0767\,\pi^2} + 0.2339\,i\,\theta\Big), \quad -\pi \le \theta \le \pi. \tag{16}$$

This rational contour is virtually as good as the cotangent contour, as it gives a convergence rate of $O(e^{-1.311N})$, compared to the $O(e^{-1.358N})$ of (15). With either contour, an accuracy of close to machine precision $\epsilon \approx 2.2 \times 10^{-16}$ can be reached for values of N around 26 or 28, i.e., about 13 to 14 transform evaluations.

We emphasize that we have assumed here that all singularities of $F(z)$ are located on the negative real axis, and $F(z)$ is not exponentially small as discussed below (8). In Section 4 below, we test how well the contour (15) works when these assumptions are violated. But first, we address the issue of roundoff error that can contaminate the solution when N is too large.

3. ROUNDOFF ERROR CONTROL

When N gets large, the contours (15) and (16) move too far into the right half-plane. The exponential factor in (4) gets too big, and there is loss of precision in the summation. Using the same arguments as presented in [Weideman 2010] for the parabolic contour, we model this error by

$$R = O\Big(\epsilon\, e^{z(0)t}\Big) = O\Big(\epsilon\, e^{N\zeta(0)}\Big), \tag{17}$$

where ϵ is the unit roundoff.

Let the implied constants in (8) and (17) be denoted by k_1 and k_2, respectively. Then the roundoff error becomes significant when

$$k_1 e^{-cN} = k_2 \epsilon\, e^{N\zeta(0)} \quad \Longrightarrow \quad c + \zeta(0) + N^{-1}\log(\epsilon/k_0) = 0, \tag{18}$$

where we defined $k_0 = k_1/k_2$.

To check the accuracy of (18), consider $\epsilon \approx 2.2 \times 10^{-16}$ and assume for the moment that $k_0 \approx 1$. With the values of the contour parameters as defined by (12) and (14), a numerical solution of (18) yields $N \approx 24$ (25 in the case of the rational contour (16)). An examination of the convergence curves shown in Section 4 will confirm that $N = 24$ is often close to the critical value where roundoff errors start to dominate.

When k_0 is not $O(1)$, this critical value of N may be larger, say $N = N_*$. The value of N_* can be estimated by analyzing the convergence behavior. For example, we assume that for $N \leq N_*$ the error behaves like $f(t) - f_N(t) = k_1 e^{-cN}$, where $f_N(t)$ denotes the approximation (4). By computing $f_N(t)$ for a range of values of N, it is straightforward to detect at what value of N the convergence turns from exponential decay to exponential growth. By inserting this value of N_* into (18), and using the values of the contour parameters defined by (12) and (14), it is possible to compute an approximate value of k_0.

We use this value of k_0 to compute new contour parameters for values of $N > N_*$. Note that $\zeta(0) = -\sigma + \mu/\alpha$, which depends on c and α via (12). This means the equation in (18) has two unknowns, so we fix α at the value (14). Hence c can be solved by applying a univaritate rootfinding routine to equation on the right in (18), and following that the new contour parameters are computed from (12).

We can justify this approach by looking at the asymptotic behavior of the parameters. Letting $N \to \infty$ in (18) yields

$$c = -\log(\epsilon/k_0)N^{-1} + O(N^{-3}), \qquad \zeta(0) = O(N^{-3}).$$

Substitution of these two expressions into the error estimates (8) and (17), respectively, shows that both reduce to $O(\epsilon)$. Hence errors can be expected to remain at the roundoff level for large N.

Because $\zeta(0) = O(N^{-3})$, the contour is not only prevented from moving too far into the right half-plane, it actually moves back to the left (and becomes narrower) when $N > N_*$. For this reason this roundoff control strategy only works when the singularities of $F(z)$ are strictly on the negative real axis. In the next section, we investigate how well it works in practice.

We remark that roundoff control for the parabola and hyperbola was proposed in [Weideman 2010] and [López-Fernández et al. 2006], respectively. An entirely different strategy was developed by one of the authors of the present paper, based on methods that automatically compute a contour of optimal stability for (1). Details can be found in [Dingfelder 2012] and a forthcoming publication on this topic.

4. APPLICATIONS AND COMPARISONS

In this section we verify the theoretical results of the previous sections by applying the method to various model problems. These include a matrix transform that can be used for the time integration of parabolic PDEs, as well as some scalar transforms. The latter include four of the transforms considered in [Duffy 1993], all of which were taken from actual applications. In all these cases the inverses of the transforms are explicitly known as integrals or series, which enabled us to compute errors. Some of these transforms have singularities exclusively located on the negative real axis. The other problems deal with singularities on the imaginary axis and on the positive real axis as well as with various branch cuts in the complex plane. We shall see that good convergence is still possible for these cases, despite the fact that the parameter analysis of Section 2 assumed singularities strictly on the negative real axis.

In all figures, the relative error is plotted against the number of nodes, N, in the midpoint approximation (4). For reference, we also plot the theoretically predicted convergence rate $O(e^{-1.358N})$,

as a dash-dot curve. We also remind the reader about the N vs. $2N$ convention mentioned below (4). When compared to the error curves in [Duffy 1993; Weideman and Trefethen 2007; Weideman 2010], for example, the values of N on the horizontal axes of the graphs below should therefore be halved.

4.1. Singularities on the negative real axis

We start with the model problem (7) and its matrix analogue

$$F_1(z;\lambda) = (z+\lambda)^{-1}, \qquad F_1(z;A) = (zI+A)^{-1}. \tag{19}$$

Here $\lambda > 0$ and A is a symmetric positive definite matrix, with I the identity matrix of the same size.

The matrix version arises, for example, when semi-discretizations of parabolic PDEs are solved by Laplace transform techniques [Sheen et al. 2003; López-Fernández and Palencia 2004; Gavrilyuk and Makarov 2005]. Consider

$$\mathbf{u}_t + A\,\mathbf{u} = \mathbf{0} \quad \Longrightarrow \quad z\,\mathbf{U}(z) - \mathbf{u}_0 + A\,\mathbf{U}(z) = \mathbf{0} \quad \Longrightarrow \quad \mathbf{U}(z) = F_1(z;A)\,\mathbf{u}_0,$$

where A is an $M \times M$ matrix representation of the negative Laplacian, $\mathbf{u}$(t) an $M \times 1$ vector of unknowns with $\mathbf{U}(z)$ its Laplace transform, and $\mathbf{u}_0$ the initial condition. The inverse formula (1) yields

$$\mathbf{u}(t) = \exp(-At)\mathbf{u}_0 = \frac{1}{2\pi i}\int_C e^{zt}\Big(F_1(z;A)\,\mathbf{u}_0\Big)\,dz. \tag{20}$$

In this case the singularities of the transform are the negatives of the eigenvalues λ of A, which are located on the negative real axis because of the assumptions on A. The right-hand side is approximated by the midpoint sum (4), where each quadrature node requires the solution of a linear system with coefficient matrix $\big(z(\theta_k)I + A\big)$ and right-hand side $\mathbf{u}_0$.

Our test example is the heat equation $u_t - 0.01\nabla^2 u = 0$ on $[0,1]\times[0,1]$, supplemented with homogeneous Dirichlet boundary conditions. We take A as the familiar block-tridiagonal matrix based on the 5-point finite difference approximation to the negative of the Laplacian, which is known to be positive definite. The right-hand side $\mathbf{u}_0$ is taken to be random, and the reference solution $\mathbf{u}(t)$ was computed using the spectral decomposition of A, which is explicitly known.

The results for this example are shown in Figure 4. Note that we have plotted the error here in the computed value of $\exp(-At)\,\mathbf{u}_0$, i.e., only the temporal error. The error in the actual solution of the PDE is not shown but it would not reach such small values unless a super accurate space discretization is used.

Figure 4 confirms that (a) the modified Talbot contour converges as predicted by the error analysis of Section 2, namely $O(e^{-1.358N})$, and (b) the roundoff control strategy described in the previous section works well in practice. Ten digit accuracy is reached with $N = 16$ or fewer nodes, i.e., no more than 8 transform evaluations are required (each of which involves the solution of a sparse linear system). Almost full accuracy is reached for all $N > 24$.

The remaining transforms in this section are scalar problems from actual applications, as collected in [Duffy 1993]. As an example of a transform with a series of poles on the negative real axis we consider

$$F_2(z) = \frac{(100z-1)\sinh(\frac{1}{2}\sqrt{z})}{z(z\sinh(\sqrt{z}) + \sqrt{z}\cosh(\sqrt{z}))}.$$

This is Test 2 in [Duffy 1993], where it was noted that the apparent square root singularity can be removed by expanding the hyperbolic functions. We also consider Test 3 in [Duffy 1993], namely

$$F_3(z) = \frac{1}{z}\exp\left(-r\sqrt{\frac{z(1+z)}{1+cz}}\right), \qquad r>0,\ c>0.$$

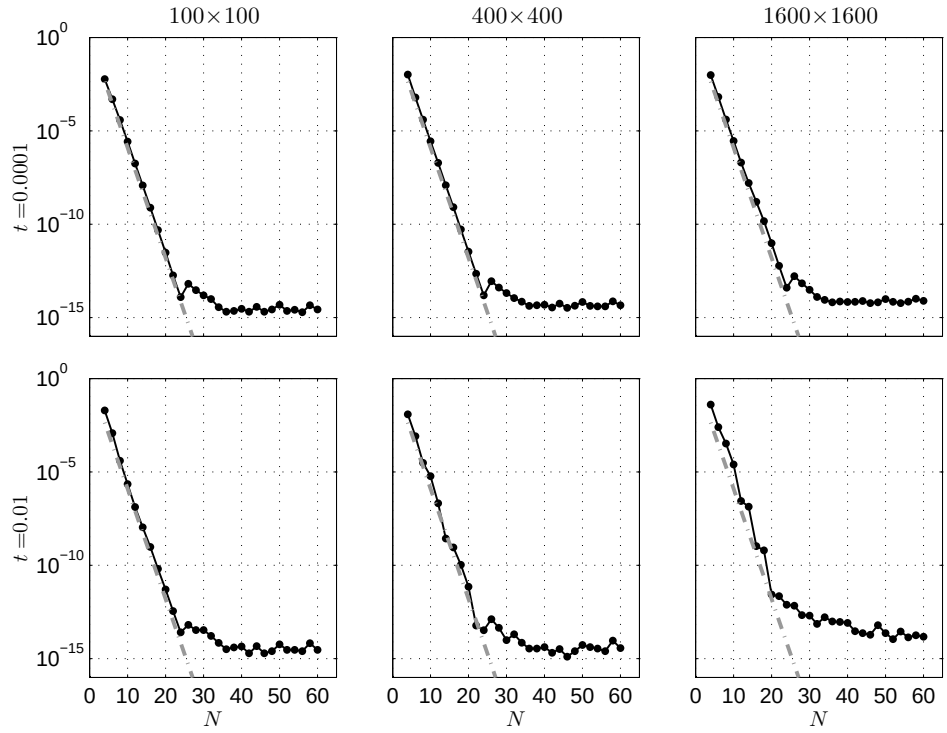

Fig. 4. Relative errors in the ∞-norm when $\exp(-At)\mathbf{u}_0$ is approximated by applying the midpoint sum (4) to the contour integral (20). For smaller values of N the contour (15) was used, while for larger N the roundoff control procedure of Section 3 was used. The three columns correspond to three sizes of A, while the two rows correspond to two values of t. The dashed line represents the theoretical predicted convergence rate $O(e^{-1.358N})$.

This transform has a pole at $z = 0$, an essential singularity $z = -1/c$ as well as branch point singularities at $z = 0$ and $z = -1$. With appropriate definition of the branch cuts, this transform is analytic everywhere off the negative real axis.

Numerical results for transforms $F_2(z)$ and $F_3(z)$ are presented in Figures 5 and 6, respectively. Again we see empirical confirmation of the theoretical $O(e^{-1.358N})$ convergence rate and the effectiveness of the roundoff control strategy. A comparison with the results presented in [Duffy 1993] shows a faster rate of convergence here, which reduces the total number of nodes in the midpoint sum (4) required for, say, ten-digit accuracy at $t = 1$, from about 26 and 36 for F_2 and F_3, to 18 for both transforms.

Note that $F_3(z)$ is an example for which the error analysis of Section 2 can fail. The reason is its rapid decay as $z \to +\infty$, particularly if $r \gg 1$; recall the discussion below (8). Nevertheless, for $0 < r < 1$ we get error curves of similar accuracy to those shown in Figure 6 for all values of t considered.

4.2. Singularities on the positive real axis

When the singularities are located on the positive real axis, the frequency shift formula

$$e^{-st} f(t) = \frac{1}{2\pi i} \int_C e^{zt} F(z+s) dz, \tag{21}$$

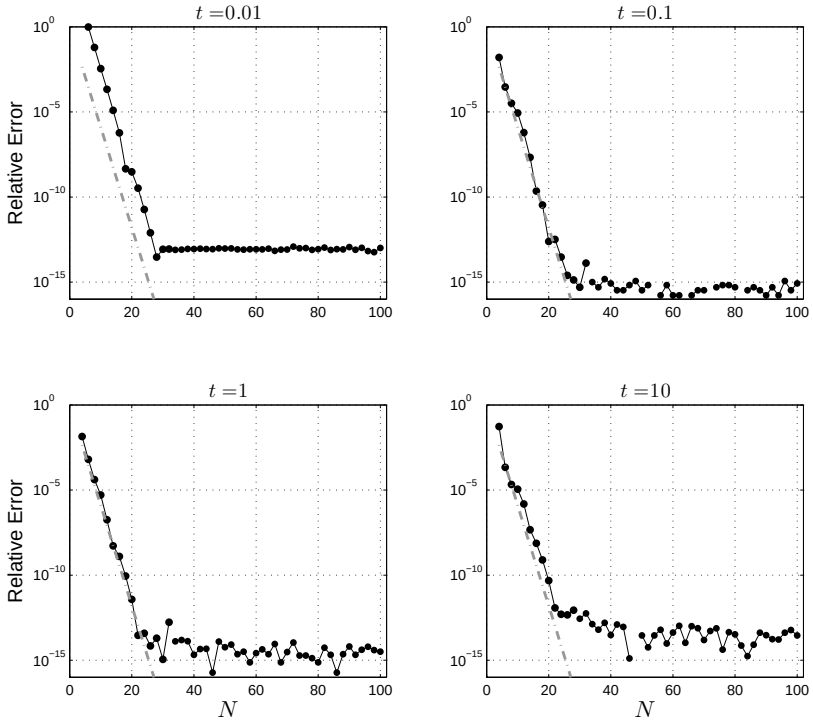

Fig. 5. Relative error in the inversion of transform $F_2(z)$. For smaller values of N the contour (15) was used, while for larger N the roundoff control procedure of Section 3 was implemented.

can be considered. The real shift s should be sufficiently large such that all the singularities are on the negative real axis. One may lose accuracy if s and t are large because roundoff errors are amplified by the factor e^{st}, but in our experiments this did not appear to be a major factor. In addition, for sufficiently large values of st, overflow can occur in the evaluation of the exponential factor.

To demonstrate this shifting formula, we examine the transform $F_1(z;-5)$ as defined in (19) with pole at $z = 5$. Applying the method to the unshifted transform fails, but by shifting the singularity to 0 we obtain the expected high rate of convergence, as shown in Figure 7.

4.3. Singularities off the real axis

The theoretical predicted convergence rate of $O(e^{-1.358N})$ is often observed even for moderate values of N when the transform has singularities off the real axis. For these calculations, however, the roundoff error control discussed in Section 3 does not apply and we have omitted it in the results presented in this section.

The first transform of this type is Test 4 given in [Duffy 1993],

$$F_4(z) = \frac{1}{z}\exp\left(-2\,\text{arcosh}(\sqrt{1+z^2+z^4/16})\right),$$

which we have simplified to

$$F_4(z) = \frac{1}{z(\sqrt{z^2+z^4/16}+\sqrt{1+z^2+z^4/16})^2}.$$

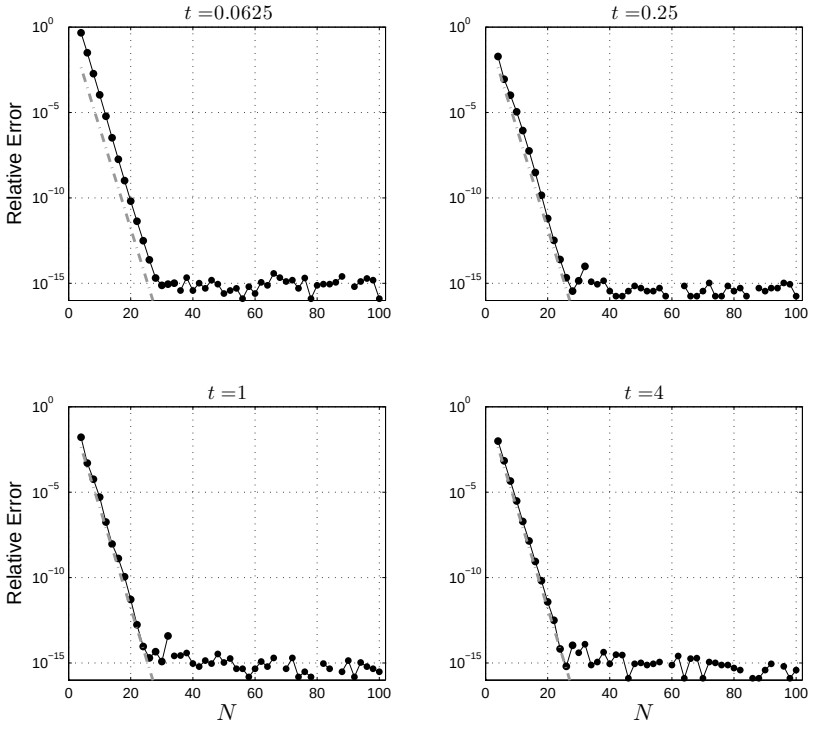

Fig. 6. Same as Figure 5 but the transform is $F_3(z)$, with $c = 0.4$ and $r = 0.5$.

There is a pole at $z = 0$ and six branch points on the imaginary axis, the outermost two of which are $z = \pm 4i$. The results for this transform are shown in Figure 8.

We also consider Test 5 given in [Duffy 1993], namely

$$F_5(z) = \frac{z - \sqrt{z^2 - c^2}}{\sqrt{z}\,\sqrt{z^2 - c^2}\,\sqrt{z - X\sqrt{z^2 - c^2}}}$$

with $c = (1 - X)/X$ and $X < 1$. This transform has branch point singularities on the positive real axis at c as well as on the imaginary axis at $\pm cX/\sqrt{X^2 - 1}$. To deal with the singularity on the positive real axis we shifted the transform by a distance c. If $X = \frac{1}{2}$, then $c = 1$ and the singularities are at 1 and $\pm i/\sqrt{3}$ before the shift, and at 0 and $-1 \pm i/\sqrt{3}$ thereafter. With proper attention to the branch cuts in the coding, this leads to the results shown in Figure 9.

Despite both transforms $F_4(z)$ and $F_5(z)$ having singularities off the real axis, we see that the modified Talbot method still shows a high order of convergence, asymptotically similar to the theoretical $O(e^{-1.358N})$. The implied constant is large, however, and increases with t. In the case of $F_4(z)$ the contour used here was not successful for large t. For this example the contour parameters of [Duffy 1993; Murli and Rizzardi 1990] are better, because they adjust for singularities off the negative real axis.

For $F_5(z)$ we observe good convergence in Figure 9, except for the roundoff error issue. Note that in [Duffy 1993] it is mentioned that Talbot's method "does rather poorly" for this example.

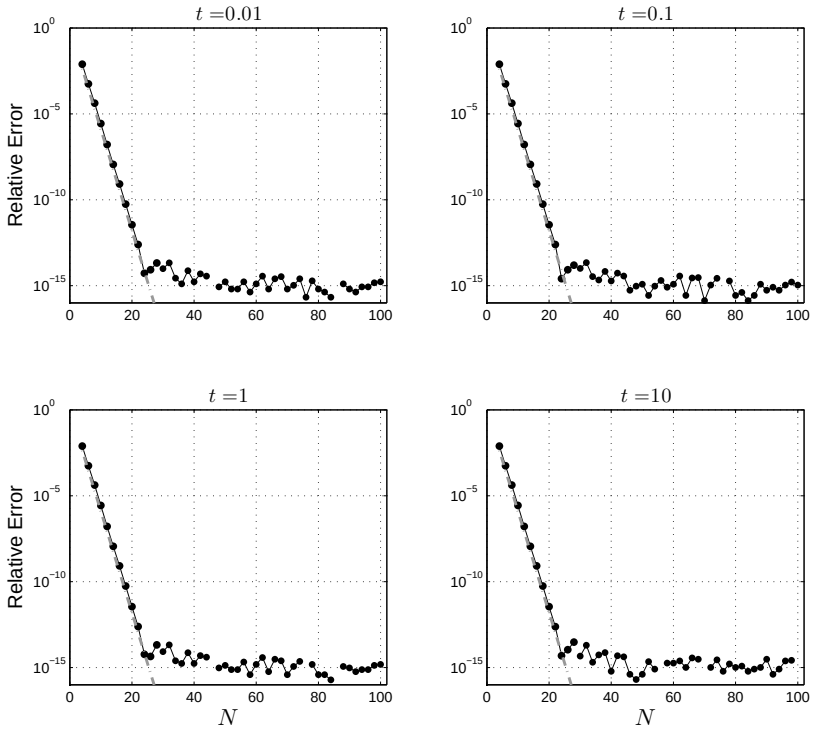

Fig. 7. Same as Figure 5 but the transform is $F_1(z;\lambda)$ with $\lambda = -5$. The frequency shift formula (21) is used, with $s = -\lambda$.

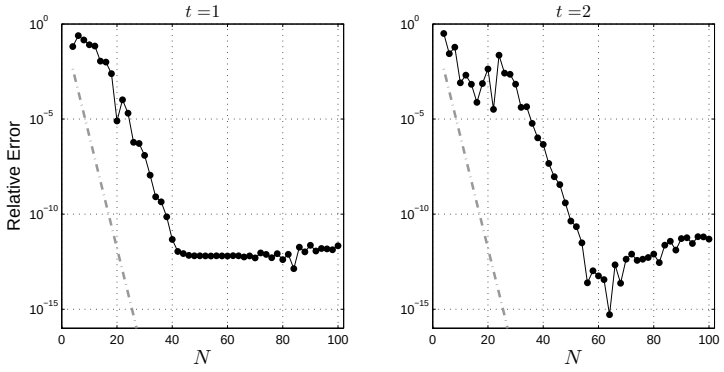

Fig. 8. Relative error in the inversion of transform $F_4(z)$. The contour (15) was used for all N shown (i.e., no roundoff control).

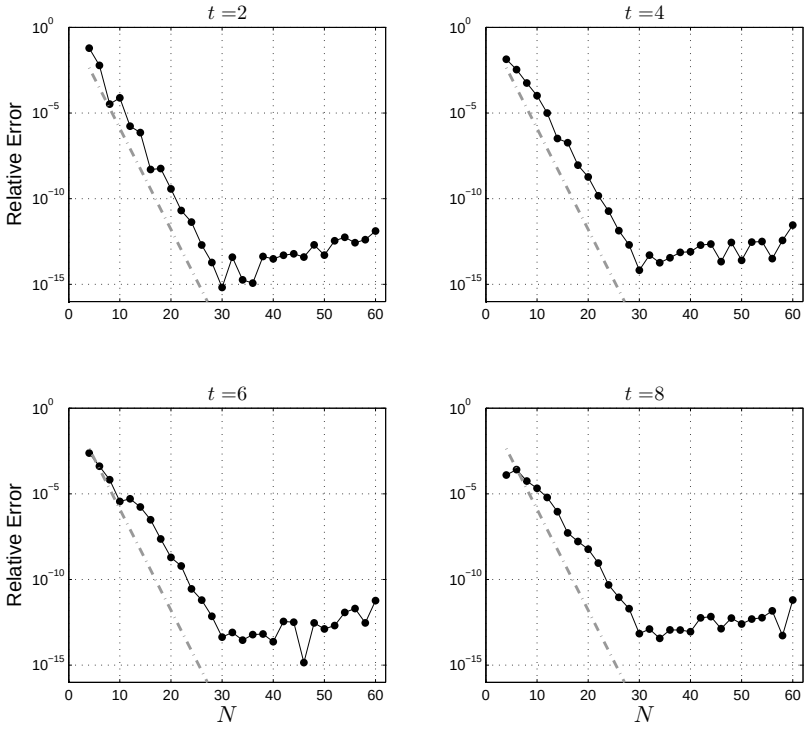

Fig. 9. Same as Figure 8, but for the transform $F_5(z)$, $c = 1$. The frequency shift formula (21) was used, with $s = c$.

This might be due to not shifting the transform or perhaps a branch-cut transgression. The results shown Figure 9 reach ten-digit precision with 24 nodes (i.e., 12 transform evaluations).

Finally, we remark that in the case of singularities in the right half-plane off the real axis, the frequency shift formula (21) can also be used to shift singularities to the left half-plane.

5. SOFTWARE

We have created a MATLAB function, `ModifiedTalbot`, which can be downloaded from the web site [Dingfelder and Weideman 2013]. It implements the method (4) on the contour (15). In its basic form the function takes as input the scalar transform $F(z)$ and a single value of t. It outputs an approximation to $f(t)$, and an estimated relative error. In addition, the user can

— specify a relative accuracy (ten-digit accuracy is the default),
— increase the maximum value of N (100 is the default),
— specify whether the roundoff error control strategy of Section 3 should be employed (recommended only when the singularities of the transform are on the negative real axis),
— include a frequency shift (based on formula (21)),
— turn on or off a graphical output of the convergence history, and
— specify an exact value for $f(t)$ (if known), for testing purposes.

More details on these options are given in the help lines of `ModifiedTalbot`.

The algorithm starts with a small value of N, and then increases its value in steps of 2 until an estimated error drops below the specified accuracy. The default stopping test is

$$\left| \frac{f_N(t) - f_{N-2}(t)}{f_N(t)} \right| \leq 10^{-10},$$

where $f_N(t)$ represents the approximation computed with N nodes in the midpoint sum (4). The left hand side is a reasonable estimate for the actual relative error when the convergence is exponential, as in $f(t) - f_N(t) = O(e^{-cN})$. In the ideal case (singularities on the negative real axis, $F(z)$ not exponentially small) the largest value of N required for this accuracy is typically around 20.

The strategy of increasing N until the desired accuracy is reached is somewhat inefficient, as transform values computed for $f_{N-2}(t)$ cannot be re-used in the computation of $f_N(t)$. If $F(z)$ is not too expensive to evaluate this is not a major drawback, however. Relatedly, keep in mind that this code does not require any information on the singularities, unless the frequency shift formula (21) is used. By contrast, software like [Murli and Rizzardi 1990] requires the user to specify the locations and type of some of the singularities of the transform. We consider the additional transform evaluations a relatively small price to pay for the convenience of not having to deal with a singularity analysis of the given transform.

The following MATLAB code snippet shows the usage of `ModifiedTalbot`. We chose test $F_3(z)$, in order to demonstrate the correct coding so that the branch-cut is located on the negative real axis. Also note the dot-multiplication and -division in the definition of the transform because the code in `ModifiedTalbot` is vectorized.

```
>> t = 1; F = @(z) (1./z).*exp(-0.5*sqrt(z).*sqrt(1+z)./sqrt(1+0.4*z));
>> [f,err] = ModifiedTalbot(F,t)
>> f =  0.722835907109632
>> err = 5.043215570091767e-012
```

For this example the exact value is $f(1) = 0.7228359071$, rounded to ten digits, so the default accuracy has been achieved. The actual relative error is about 1.8×10^{-13}, so `ModifiedTalbot` overestimates it by less than an order of magnitude.

In Table I we present a comparison of `ModifiedTalbot` with the results from [Murli and Rizzardi 1990]. The table shows the total number of nodes, N, in the quadrature sum (4), to reach an error of 10^{-6}. In [Murli and Rizzardi 1990] this error was defined as the relative error if the reference value is greater than one, and absolute error otherwise. Note that the numbers in parentheses shown in Table I are double the numbers shown in Tables I and II of [Murli and Rizzardi 1990], because of the N vs. $2N$ convention mentioned below (4).

The first three transforms in Table I have singularities on the negative real axis. For each of these transforms the value of N required by `ModifiedTalbot` is significantly less than the number required by the algorithm of [Murli and Rizzardi 1990]; in one instance the value of N is more than halved. The last three transforms in Table I have singularities in the complex plane. The last one of these has a singularity in the right half-plane, which we have shifted to the imaginary axis by choosing $s = 1$ in (21). For these three cases `ModifiedTalbot` yields reasonable results for small values of t. As t increases, however, the contour (15) moves closer to these singularities and the number of nodes has to be increased accordingly. For the cases marked (a) in Table I no convergence to the desired accuracy was achieved for N less than the default value of 100. In many of these cases, no convergence could be obtained even when this default value was increased. In these situations the algorithm of [Murli and Rizzardi 1990] is superior, but it too needs large numbers of transform evaluations to reach the accuracy goal. For these cases it is probably wise to consider alternatives to Talbot's method.

For additional performance tests, this time with a smaller error tolerance 10^{-10}, we offer six more transforms in Table II. The inverses can be looked up in tables such as [Abramowitz and Stegun 1970, Sect. 29]. The first five of these transforms have a combination of poles, branch-cuts

Table I. Comparison of the code `ModifiedTalbot` with the numerical results of [Murli and Rizzardi 1990]. The numbers without (resp. inside) the parentheses are the values of N used by `ModifiedTalbot` (resp. the algorithm of [Murli and Rizzardi 1990]) to reach a specified error of 10^{-6}. Cases where `ModifiedTalbot` could not reach this accuracy for N less than the default value of 100 are marked by (a).

Transform	$t = 10^{-1}$	$t = 10^{0}$	$t = 10^{1}$	$t = 10^{2}$	$t = 10^{3}$
z^{-2}	18 (22)	18 (22)	16 (22)	16 (22)	16 (22)
$\log(z)/z$	14 (22)	18 (22)	14 (22)	14 (22)	14 (22)
$\exp(-4\sqrt{z})$	16 (22)	14 (22)	12 (22)	12 (22)	10 (22)
$\arctan(1/z)$	16 (22)	20 (22)	46 (38)	(a) (140)	(a) (1050)
$\log\left(\frac{z^2+1}{z^2+4}\right)$	18 (22)	22 (22)	78 (46)	(a) (224)	(a) (5284)
$z^2(z^3+8)^{-1}$	18 (22)	26 (22)	80 (46)	(a) (304)	(a) (2224)

or essential singularities on the real axis, while the last one has a branch-cut on the imaginary axis. In those cases where the singularities are located on the positive real axis, a shift equal to the largest real singularity was implemented.

In most of these cases `ModifiedTalbot` performs reliably, providing the default ten-digit accuracy with around 18 to 24 nodes in the quadrature sum (4). The cases of non-convergence for N less than the default value 100 are marked either (a) the singularities are too far off the negative real axis and t is too large, or (b) the transform decays too rapidly for z in the right half-plane and t is too small. (Note that (a) was already encountered in Table I.) Whenever one of these exceptional situations occur, the graphical display of the convergence history offered by `ModifiedTalbot` is useful for identifying that something is wrong and that other methods should be used instead.

Considering the results of Tables I and II, we conclude that in the ideal case ($F(z)$ with singularities on the negative real-axis and not exponentially small) the code `ModifiedTalbot` performs reliably. In other cases it can still perform well, unless one encounters one or more of the exceptions (a) and (b) listed above. We emphasize that the main advantages of `ModifiedTalbot` are rapid convergence and the fact that minimal singularity information has to be specified by the user.

Table II. Performance of `ModifiedTalbot` on six other transforms. The value of N to reach a specified relative accuracy of 10^{-10} is tabulated. Unsuccessful cases are denoted by (a) or (b), as discussed in the penultimate paragraph of Section 5.

Transform	Parameters	$t = 10^{-2}$	$t = 10^{-1}$	$t = 1$	$t = 10^{1}$	$t = 10^{2}$
$\exp(-k\sqrt{z})/z$	$k = 1$	40	24	22	20	20
	$k = 5$	(b)	(b)	26	22	20
$\exp(-k/z)/z$	$k = 1$	20	22	24	28	44
	$k = 5$	22	22	28	38	70
$\sqrt{z}/(z-a^2)$	$a = 1$	20	20	20	20	20
	$a = 5$	20	20	20	20	(c)
$(\sqrt{z+a}\sqrt{z+b})^{-1}$	$a = -3,\ b = 4$	20	18	20	20	20
$\sqrt{z+a} - \sqrt{z+b}$	$a = -5,\ b = 1$	18	20	22	22	24
$(\sqrt{z^2+a^2})^{-1}$	$a = 1$	20	22	28	64	(a)
	$a = 2$	20	22	34	(a)	(a)
	$a = 10$	22	28	64	(a)	(a)

ACKNOWLEDGMENTS

The research of BD was supported by the DFG Collaborative Research Center TRR 109, "Discretization in Geometry and Dynamics." He further acknowledges support from the graduate program TopMath of the Elite Network of Bavaria and the TopMath Graduate School at Technische Universität München. The research of JACW was supported the National Research Foundation of South Africa. Folkmar Bornemann and Nick Trefethen both offered valuable suggestions.

REFERENCES

Abramowitz, M. and Stegun, I. A. 1970. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover Publications, Inc., New York.

Butcher, J. C. 1957. On the numerical inversion of Laplace and Mellin transforms. Conference on Data Processing and Automatic Computing Machines. Salisbury, Australia.

Dingfelder, B. 2012. Raten- und stabilitätsoptimierte Algorithmen zur Berechnung eines Integrals mit Hankelintegrationsweg, Bachelor's thesis, Technische Universität München, Germany.

Dingfelder, B. and Weideman, J. A. C. 2013. `ModifiedTalbot`. *`http://dip.sun.ac.za/~weideman/research/int.html`*. (temporary location).

Duffy, D. G. 1993. On the numerical inversion of Laplace transforms: comparison of three new methods on characteristic problems from applications. *ACM Trans. Math. Software 19,* 3, 333–359.

Gavrilyuk, I. P. and Makarov, V. L. 2005. Exponentially convergent algorithms for the operator exponential with applications to inhomogeneous problems in Banach spaces. *SIAM J. Numer. Anal. 43,* 5, 2144–2171.

Kress, R. 1998. *Numerical analysis*. Springer-Verlag, New York.

López-Fernández, M. and Palencia, C. 2004. On the numerical inversion of the Laplace transform of certain holomorphic mappings. *Appl. Numer. Math. 51,* 2-3, 289–303.

López-Fernández, M., Palencia, C., and Schädle, A. 2006. A spectral order method for inverting sectorial Laplace transforms. *SIAM J. Numer. Anal. 44,* 3, 1332–1350.

MATLAB. 2012. *version 7.14 (R2012a)*. The MathWorks Inc., Natick, Massachusetts.

Murli, A. and Rizzardi, M. 1990. Algorithm 682: Talbot's method for the Laplace inversion problem. *ACM Trans. Math. Software 16,* 2, 158–168.

Rizzardi, M. 1995. A modification of Talbot's method for the simultaneous approximation of several values of the inverse Laplace transform. *ACM Trans. Math. Software 21,* 4, 347–371.

Sheen, D., Sloan, I. H., and Thomée, V. 2003. A parallel method for time discretization of parabolic equations based on Laplace transformation and quadrature. *IMA J. Numer. Anal. 23,* 2, 269–299.

Strang, G. 2007. *Computational science and engineering*. Wellesley-Cambridge Press, Wellesley, MA.

Talbot, A. 1979. The accurate numerical inversion of Laplace transforms. *J. Inst. Math. Appl. 23,* 1, 97–120.

Trefethen, L. N. 2006. Private communication.

Trefethen, L. N., Weideman, J. A. C., and Schmelzer, T. 2006. Talbot quadratures and rational approximations. *BIT 46,* 3, 653–670.

Weideman, J. A. C. 2006. Optimizing Talbot's contours for the inversion of the Laplace transform. *SIAM J. Numer. Anal. 44,* 6, 2342–2362.

Weideman, J. A. C. 2010. Improved contour integral methods for parabolic PDEs. *IMA J. Numer. Anal. 30,* 1, 334–350.

Weideman, J. A. C. and Trefethen, L. N. 2007. Parabolic and hyperbolic contours for computing the Bromwich integral. *Math. Comp. 76,* 259, 1341–1356.