Computing Fresnel integrals via modified trapezium rules

Article

Accepted Version

Noname manuscript No.
(will be inserted by the editor)

Computing Fresnel Integrals via Modified Trapezium Rules

Mohammad Alazah · Simon N. Chandler-Wilde · Scott La Porte

Dedicated to David Hunter on the occasion of his 80th birthday

October 23, 2013

Abstract In this paper we propose methods for computing Fresnel integrals based on truncated trapezium rule approximations to integrals on the real line, these trapezium rules modified to take into account poles of the integrand near the real axis. Our starting point is a method for computation of the error function of complex argument due to Matta and Reichel (*J. Math. Phys.* **34** (1956), 298–307) and Hunter and Regan (*Math. Comp.* **26** (1972), 539–541). We construct approximations which we prove are exponentially convergent as a function of N, the number of quadrature points, obtaining explicit error bounds which show that accuracies of 10^{-15} uniformly on the real line are achieved with $N = 12$, this confirmed by computations. The approximations we obtain are attractive, additionally, in that they maintain small relative errors for small and large argument, are analytic on the real axis (echoing the analyticity of the Fresnel integrals), and are straightforward to implement.

Mathematics Subject Classification (2000) 65D30 · 33B32

1 Introduction

Let $C(x)$, $S(x)$, and $F(x)$ be the Fresnel integrals defined by

$$C(x) := \int_0^x \cos\left(\tfrac{1}{2}\pi t^2\right)\,\mathrm{d}t, \quad S(x) := \int_0^x \sin\left(\tfrac{1}{2}\pi t^2\right)\,\mathrm{d}t, \tag{1}$$

Mohammad Alazah · Simon N. Chandler-Wilde
Department of Mathematics and Statistics, University of Reading, Whiteknights, PO Box 220, Reading RG6 6AX, UK
E-mail: m.a.m.alazah@pgr.reading.ac.uk
E-mail: S.N.Chandler-Wilde@reading.ac.uk

Scott La Porte
Department of Mathematical Sciences, John Crank Building, Brunel University, Uxbridge UB8 3PH, UK
E-mail: scottis@ntlworld.com

and

$$F(x) := \frac{\mathrm{e}^{-\mathrm{i}\pi/4}}{\sqrt{\pi}} \int_x^\infty \mathrm{e}^{\mathrm{i}t^2}\,\mathrm{d}t. \tag{2}$$

Our definitions in (1) are those of [3] and [1, §7.2(iii)], and F, C and S are related through

$$\sqrt{2}\,\mathrm{e}^{\mathrm{i}\pi/4}F(x) = \tfrac{1}{2} - C\left(\sqrt{2/\pi}\,x\right) + \mathrm{i}\left(\tfrac{1}{2} - S\left(\sqrt{2/\pi}\,x\right)\right). \tag{3}$$

In this paper we derive new methods for computing these Fresnel integrals $F(x)$, $C(x)$ and $S(x)$. The derivation of our approximations makes use of the relationship between the Fresnel integral and the error function, that

$$F(x) = \tfrac{1}{2}\mathrm{erfc}(\mathrm{e}^{-\mathrm{i}\pi/4}x) = \tfrac{1}{2}\,\mathrm{e}^{\mathrm{i}x^2}\,w\left(\mathrm{e}^{\mathrm{i}\pi/4}x\right) \tag{4}$$

where erfc is the complementary error function, defined by

$$\mathrm{erfc}(z) := \frac{2}{\sqrt{\pi}} \int_z^\infty \mathrm{e}^{-t^2}\mathrm{d}t,$$

and

$$w(z) := \mathrm{e}^{-z^2}\mathrm{erfc}(-\mathrm{i}z).$$

It also depends on the integral representation [3, (7.1.4)] that

$$w(z) = \frac{\mathrm{i}}{\pi}\int_{-\infty}^{\infty} \frac{\mathrm{e}^{-t^2}}{z-t}\,\mathrm{d}t = \frac{\mathrm{i}z}{\pi}\int_{-\infty}^{\infty} \frac{\mathrm{e}^{-t^2}}{z^2-t^2}\,\mathrm{d}t, \quad \text{for } \mathrm{Im}(z) > 0. \tag{5}$$

Combining (4) and (5) gives an integral representation for $F(x)$, that

$$F(x) = \frac{x}{2\pi}\,\mathrm{e}^{\mathrm{i}(x^2+\pi/4)}\int_{-\infty}^{\infty} \frac{\mathrm{e}^{-t^2}}{x^2+\mathrm{i}t^2}\,\mathrm{d}t, \quad \text{for } x > 0. \tag{6}$$

Fresnel integrals arise in applications throughout science and engineering, especially in problems of wave diffraction and scattering (*e.g.*, [5, §8.2], [6]), so that methods for the efficient and accurate computation of these functions are of wide application. The purpose of this paper is to present new approximations for the Fresnel integrals, based on N-point trapezium rule approximations to the integral representation (6) for $F(x)$, these trapezium rules modified to take into account the poles of the integrand. These poles lie near the path of integration when x is small.

The observation that the trapezium rule is exponentially convergent when applied to integrals of the form

$$\int_{-\infty}^{\infty} \mathrm{e}^{-t^2}\,f(t)\,\mathrm{d}t, \tag{7}$$

with $f(t)$ analytic in a strip surrounding the real axis, dates back at least to Turing [29] and Goodwin [14]. The derivation of this result uses contour integration and Cauchy's residue theorem; see §2 below. Applying the trapezium rule with step-length $h > 0$ to (6) leads to the approximation

$$F(x) \approx \frac{xh}{\pi}\, \mathrm{e}^{\mathrm{i}(x^2+\pi/4)} \sum_{k=1}^{\infty} \frac{\mathrm{e}^{-\tau_k^2}}{x^2 + \mathrm{i}\tau_k^2}, \quad \text{for } x > 0, \tag{8}$$

where

$$\tau_k := (k - 1/2)h. \tag{9}$$

When $x > 0$ is large this approximation is very accurate. Indeed, if we choose

$$h = \sqrt{\pi/(N + 1/2)} \tag{10}$$

for some large integer N, then this approximation is essentially identical to the approximation $F_N(x)$ for $F(x)$ that we propose in (14) below. However, the approximation (8) becomes increasingly poor as $x > 0$ approaches zero.

In the context of developing methods for evaluating the complementary error function of complex argument (by (4), evaluating $F(x)$ for x real is just a special case of this larger problem), Chiarella and Reichel [8], Matta and Reichel [20], and Hunter and Regan [16] proposed modifications of the trapezium rule that follow naturally from the contour integration argument used to prove that the trapezium rule is exponentially convergent. The most appropriate form of this modification is that in [16] where the modified trapezium rule approximation

$$F(x) \approx \frac{xh}{\pi}\, \mathrm{e}^{\mathrm{i}(x^2+\pi/4)} \sum_{k=1}^{\infty} \frac{\mathrm{e}^{-\tau_k^2}}{x^2 + \mathrm{i}\tau_k^2} + R(h, x), \quad \text{for } x > 0, \tag{11}$$

is proposed. Here the correction term $R(h, x)$ is defined by

$$R(h,x) := \begin{cases} 1/(\exp(2\pi \mathrm{e}^{-\mathrm{i}\pi/4} x/h) + 1), & \text{if } 0 < x < \sqrt{2}\,\pi/h, \\ 0.5/(\exp(2\pi \mathrm{e}^{-\mathrm{i}\pi/4} x/h) + 1), & \text{if } x = \sqrt{2}\,\pi/h, \\ 0, & \text{if } x > \sqrt{2}\,\pi/h. \end{cases}$$

The approximation (11) clearly coincides with $F_N(x)$, given by (14), for $0 < x < \sqrt{2}\,\pi/h$, if the range of summation in (11) is truncated to $1, ..., N$ and the choice (10) for h is made. Hunter and Regan prove that the magnitude of the error in (11) is

$$\leq \frac{x \mathrm{e}^{-\pi^2/h^2}}{\sqrt{\pi}\left(1 - \mathrm{e}^{-2\pi^2/h^2}\right)\, |x^2/2 - \pi^2/h^2|}, \tag{12}$$

for $x > 0$, provided $x \neq \sqrt{2}\pi/h$. Similar estimates, it appears arrived at independently, are derived by Mori [21], in which paper the emphasis is on computing $\mathrm{erfc}(x)$ for real x.

The approximation (11) is the starting point for the method we propose in this paper. Our main contributions (see §1.2 for detail) are: (i) to point

out that the approximation proposed in (11) for $0 < x < \sqrt{2}\,\pi/h$ in fact provides an accurate (and real-analytic) approximation to the entire function F on the whole real line; (ii) to provide an optimal formula for the choice of the step-size h as a function of N, the number of terms retained in the sum in (11); (iii) to prove that, with this choice of h, the resulting approximations are exponentially convergent as a function of N, uniformly on the real line (this in contrast to (12) which blows up at $x = \sqrt{2}\,\pi/h$).

1.1 Other methods for computing Fresnel integrals

Naturally, there exist already a number of effective schemes for computation of Fresnel integrals, and we briefly summarise now the best of these. An effective computational method for smaller values of $|x|$ is to make use of the power series for $C(x)$ and $S(x)$ (see (69) below). These converge for all x, and very rapidly for smaller x, and so are widely used for computation. For example, the algorithm in the standard reference [25] uses these power series for $|x| \leq 1.5$. For this range, after the first two terms, these series are alternating series of monotonically decreasing terms, and the error in truncation has magnitude smaller than the first neglected term. Thus, for $|x| \leq 1.5$, the errors in computing $C(x)$ and $S(x)$ by these power series truncated to N terms are $\leq 2 \times 10^{-16}$ and $\leq 2.3 \times 10^{-17}$, respectively, for $N = 14$.

For$|x| > 1.5$, [25] recommends computation using the representations in terms of erfc which follow from (3) and (4), and the continued fraction representation for $\mathrm{e}^{z^2}\mathrm{erfc}(z) = w(\mathrm{i}z)$ given as [1, (7.9.2)]. Methods for evaluation of $w(z)$ based on continued fractions for larger complex z (which can be used to evaluate $F(x)$ and hence $C(x)$ and $S(x)$) are also discussed in Gautschi [13] and are finely tuned to form TOMS "Algorithm 680" in Poppe and Wijers [23,24]. This algorithm achieves relative errors of 10^{-14} over "nearly all" the complex plane by Taylor expansions of degree up to 20 in an ellipse around the origin, convergents of up to order 20 of continued fractions outside a larger ellipse, and a more expensive mix of Taylor expansion and continued fraction calculations in between.

Weideman [30] presents an alternative method of computation (the derivation starts from the integral representation (5)) which approximates $w(z)$, for $\mathrm{Im}(z) > 0$, by the polynomial

$$w_M(z) = \frac{2}{L^2 + z^2} \sum_{n=0}^{M} a_n Z^n \tag{13}$$

in the transformed variable $Z = (L + \mathrm{i}z)/(L - \mathrm{i}z)$. Here $L = \sqrt{M/\sqrt{2}}$ and the coefficients a_n can be viewed as Fourier coefficients and efficiently computed by the FFT. We will see in §4 that a polynomial degree $M = 36$ in (13) suffices to compute $F(x) = \mathrm{e}^{\mathrm{i}x^2} w(\mathrm{e}^{\mathrm{i}\pi/4}x)/2$ with relative error $\leq 10^{-15}$ uniformly on the positive real axis. Weideman [30] argues carefully and persuasively that,

for intermediate values of $|z|$ (values in approximately the range $1.5 \le |z| \le 5$ for the case $\arg(z) = \pi/4$ which we require), and as measured by operation counts, the work required to compute $w(z)$ to 10^{-14} relative accuracy is much smaller for the approximation (13) than for Algorithm 680 [24].

All the approximations described above are polynomial or rational approximations (or piecewise polynomial/rational approximations, proposing different approximations on different regions). Many other authors describe approximations of these types for computing the Fresnel integrals specifically with real arguments. The best of these in terms of accuracy is Cody [9], where numerical coefficient values are given for piecewise rational approximations to $C(x)$ and $S(x)$ for $0 \le x \le 1.6$, and for piecewise rational approximations to the related functions $f(x)$ and $g(x)$ (see (64) and (65) below), for $x \ge 1.6$. These approximations, in their respective regions of validity, achieve relative errors $\le 10^{-15.58} \approx 2.7 \times 10^{-16}$, this using rational approximations which are ratios of polynomials of degree ≤ 6; in total five different approximations are used on different subintervals of the real axis. Single rational approximations, based on a "polar" version of (64) and (65), are computed in [15], but these are of limited accuracy (absolute errors $\le 4 \times 10^{-8}$).

1.2 Summary of the main results

The main result of this paper is to derive, with rigorous error bounds, a new family of approximations to $F(x)$ based on modified trapezium rules, given by

$$F_N(x) := \frac{1}{2} + \frac{\mathrm{i}}{2} \tan\left(A_N x \mathrm{e}^{\mathrm{i}\pi/4}\right) + \frac{x}{A_N}\, \mathrm{e}^{\mathrm{i}(x^2+\pi/4)} \sum_{k=1}^{N} \frac{\mathrm{e}^{-t_k^2}}{x^2 + \mathrm{i}t_k^2} \tag{14}$$

$$= \frac{1}{\exp\left(2A_N x \mathrm{e}^{-\mathrm{i}\pi/4}\right) + 1} + \frac{x}{A_N}\, \mathrm{e}^{\mathrm{i}(x^2+\pi/4)} \sum_{k=1}^{N} \frac{\mathrm{e}^{-t_k^2}}{x^2 + \mathrm{i}t_k^2}, \tag{15}$$

where

$$t_k := \frac{(k-1/2)\,\pi}{\sqrt{(N+1/2)\,\pi}}, \quad A_N := t_{N+1} = \sqrt{(N+1/2)\pi}. \tag{16}$$

The corresponding approximations to $C(x)$ and $S(x)$ that we propose (obtained by substituting in (3) and separating real and imaginary parts) are

$$C_N(x) := \frac{1}{2}\, \frac{\sinh\left(\sqrt{\pi}\, A_N\, x\right) + \sin\left(\sqrt{\pi}\, A_N\, x\right)}{\cos(\sqrt{\pi}\, A_N\, x) + \cosh(\sqrt{\pi}\, A_N\, x)} + \frac{\sqrt{\pi}\, x}{A_N} \left(a_N\left(\frac{\pi}{2}x^2\right) \sin\left(\frac{\pi}{2}x^2\right) - b_N\left(\frac{\pi}{2}x^2\right) \cos\left(\frac{\pi}{2}x^2\right)\right) \tag{17}$$

and

$$S_N(x) := \frac{1}{2}\, \frac{\sinh\left(\sqrt{\pi}\, A_N\, x\right) - \sin\left(\sqrt{\pi}\, A_N\, x\right)}{\cos(\sqrt{\pi}\, A_N\, x) + \cosh(\sqrt{\pi}\, A_N\, x)} - \frac{\sqrt{\pi}\, x}{A_N} \left(a_N\left(\frac{\pi}{2}x^2\right) \cos\left(\frac{\pi}{2}x^2\right) + b_N\left(\frac{\pi}{2}x^2\right) \sin\left(\frac{\pi}{2}x^2\right)\right), \tag{18}$$

where

$$a_N(s) := s\sum_{k=1}^{N}\frac{\mathrm{e}^{-t_k^2}}{s^2+t_k^4}, \quad b_N(s) := \sum_{k=1}^{N}\frac{t_k^2\,\mathrm{e}^{-t_k^2}}{s^2+t_k^4}. \tag{19}$$

These approximations, designed for computation of $F(x)$, $C(x)$ and $S(x)$ for all $x \in \mathbb{R}$, are attractive in several respects.

- The approximation F_N is proven in Theorems 3 and 5 to converge to F approximately in proportion to $\exp(-\pi N)$, uniformly on the real line with respect to both absolute and relative error, and this predicted rate of exponential convergence is observed in numerical experiments (see §4).
- The approximations $F_N(z)$, $C_N(z)$ and $S_N(z)$ to the entire functions F, C, and S, are analytic in the strip $|\mathrm{Im}(z)| < \sqrt{(N+1/2)\pi/2}$ and the error bounds we prove extend in modified form into this strip. This implies exponentially convergent error estimates, presented in §2.1 and §3, for the difference between the coefficients in the Maclaurin series of F, C, and S and those in the corresponding series for F_N, C_N and S_N. In turn (see §3), this implies that the approximations all retain small relative error for $|x|$ small, and the computations in §4 demonstrate this.
- These approximations inherit symmetries of the Fresnel integrals. In particular, our normalisation of $F(x)$ is such that

$$F(-x) = 1 - F(x), \tag{20}$$

so that, in particular, $F(0) = 1/2$. It is clear from (14) that the same holds for $F_N(x)$, *i.e.*,

$$F_N(-x) = 1 - F_N(x). \tag{21}$$

Similarly, where an overline denotes a complex conjugate,

$$\overline{F(z)} = F(\mathrm{i}\bar{z}) \text{ and } \overline{F_N(z)} = F_N(\mathrm{i}\bar{z}). \tag{22}$$

Both these symmetries can be deduced from the structure of C and S and their approximations: by inspection of (17) and (18) we see that

$$C_N(x) = xf_C(x^4), \quad S_N(x) = x^3 f_S(x^4), \tag{23}$$

where f_C and f_S are analytic in a neighbourhood of the real line and are real-valued for real arguments. This is the same structure as C and S (see (69)). In particular, (23) implies that C_N and S_N, like C and S, are odd functions.

- These approximations are straightforward to code. Tables 1 and 2 show the short Matlab codes used to evaluate F_N, C_N and S_N for all the computations in this paper.

We end this introduction by outlining the remainder of the paper. In §2 we derive the approximation (14) to $F(x)$ and prove rigorous bounds on $|F(x) - F_N(x)|$. In §3 we deduce from this the approximations (17) and (18) and bounds on the errors $C(x) - C_N(x)$ and $S(x) - S_N(x)$, especially bounds for x small. In §4 we show numerical results, comparing our new approximations

```
function f = fresnel(x,N)
% Evaluates the approximation F_N(x) to the Fresnel integral F(x).
% x is a real scalar or matrix,
% N is the positive integer controlling accuracy (suggest N=12),
% f is the corresponding scalar or matrix of values of F_N(x).
select = x>=0;
f = zeros(size(x));
if any(select), f(select) = F(x(select),N); end
if any(~select), f(~select) = 1-F(-x(~select),N); end

function f = F(x,N)
h = sqrt(pi/(N+0.5));
t = h*((N:-1:1)-0.5);  AN = pi/h;
t2 = t.*t; t4 = t2.*t2; et2 = exp(-t2);
rooti = exp(i*pi/4);
z = rooti*x; x2 = x.*x; x4 = x2.*x2; z2 = i*x2;
S = (-et2(1)./(x4+t4(1))).*(z2+t2(1));
for n = 2:N
    S = S + (-et2(n)./(x4+t4(n))).*(z2+t2(n));
end
ez = exp((2*AN*i*rooti)*x);
f =  (i/AN)*z.*exp(z2).*S + ez./(ez+1);
```

Table 1 Matlab code to evaluate $F_N(x)$ given by (15), making use of (21) for $x < 0$.

with the error bounds derived in the earlier sections and with certain rival methods for computing Fresnel integrals. The appendix proves what appears to be a new, sharp lower bound on $|\mathrm{erfc}(z)|$, for $\mathrm{Re}(z) \geq 0$, of some independent interest, potentially useful for deriving rigorous upper bounds on the relative error in approximate methods for computing erfc (e.g., the methods of [16, 23, 30]). The relevance of this lower bound to the rest of the paper is that it implies, via (4), a new lower bound on $|F(x)|$ for $x > 0$, of independent interest and a key component in our theoretical bounds on relative errors in §2.

2 The Approximation for $F(x)$ and its Error Bounds

In this section we derive the approximation $F_N(x)$ to $F(x)$ and derive error bounds for this approximation demonstrating that both absolute and relative errors converge exponentially to zero as N increases, uniformly on the real line, and that $N = 12$ is enough to achieve errors $< 10^{-15}$. The first part of our derivation follows in large part Matta and Reichel [20] and Hunter and Regan [16]. From (6) we have that, for $x > 0$,

$$I := \int_{-\infty}^{\infty} f(t)\,\mathrm{d}t = F(x), \text{ where } f(t) := \mathrm{e}^{\mathrm{i}(x^2+\pi/4)}\,\frac{x}{2\pi}\,\frac{\mathrm{e}^{-t^2}}{x^2+\mathrm{i}t^2}, \tag{24}$$

and we have suppressed in our notation the dependence of $f(t)$ on x.

Given $h > 0$ let

$$g(z) = \mathrm{i}\tan(\pi z/h),$$

which is an odd meromorphic function with simple poles at the points τ_k, defined by (9), which has the property that, for $z = X + \mathrm{i}H$ with $X \in \mathbb{R}$, $H > 0$,

$$|1 + g(z)| \leq \frac{2\mathrm{e}^{-2\pi H/h}}{1 - \mathrm{e}^{-2\pi H/h}}. \tag{25}$$

The approximation (11) is obtained by considering the integral in the complex plane

$$J = \int_\Gamma f(z)(1 + g(z))\,\mathrm{d}z, \tag{26}$$

where the path of integration is from $-\infty$ to ∞ along the real axis, except that the path makes small semicircular deformations to pass above each of the simple poles at the points τ_k, $k \in \mathbb{Z}$. Explicitly, the kth deformation is the semicircle $\gamma_k = \{\tau_k + \epsilon \mathrm{e}^{-\mathrm{i}\theta} : \pi \leq \theta \leq 2\pi\}$, with ϵ in the range $(0, h/2)$ small enough so that the simple pole singularity in $f(z)$ at $z = z_0 := \mathrm{e}^{\mathrm{i}\pi/4}x$ lies above Γ. Then, since $f(z)g(z)$ is an odd function, we see that

$$J = \int_\Gamma f(z)\,\mathrm{d}z + \int_\Gamma f(z)g(z)\,\mathrm{d}z = I + \sum_{k\in\mathbb{Z}} \int_{\gamma_k} f(z)g(z)\,\mathrm{d}z.$$

In the limit $\epsilon \to 0$, $\int_{\gamma_k} f(z)g(z)\,\mathrm{d}z \to -\pi\mathrm{i}\,\mathrm{Res}(fg, \tau_k) = -hf(\tau_k)$, where $\mathrm{Res}(fg, \tau_k)$ denotes the residue of fg at τ_k. Thus $J = I - I_h$, where

$$I_h = h\sum_{k\in\mathbb{Z}} f(\tau_k) = 2h\sum_{k=1}^{\infty} f((k - 1/2)h) \tag{27}$$

is a trapezium/midpoint rule approximation to I.

For $H > 0$ let

$$J_H = \int_{\Gamma_H} f(z)(1 + g(z))\,\mathrm{d}z,$$

where the path of integration Γ_H is the line $\mathrm{Im}(z) = H$, traversed in the direction of increasing $\mathrm{Re}(z)$. It follows from Cauchy's residue theorem that

$$J - J_H = \mathbf{H}\left(\sqrt{2}\,H - x\right)\, PC_h, \tag{28}$$

where $\mathbf{H}$ is the Heaviside step function (defined by $\mathbf{H}(t) = 1$, for $t > 0$, $\mathbf{H}(0) = 1/2$, and $\mathbf{H}(t) = 0$, for $t < 0$), and

$$PC_h = 2\pi\mathrm{i}\,\mathrm{Res}(f(1 + g), z_0) = \frac{1}{2}\,(1 + g(z_0)) = \frac{1}{2}\,\left(1 + \mathrm{i}\tan\left(\mathrm{e}^{\mathrm{i}\pi/4}x\pi/h\right)\right).$$

Thus

$$I = I_h + \mathbf{H}\left(\sqrt{2}\,H - x\right)\, PC_h + J_H. \tag{29}$$

The point of this formula is that $I_h + \mathbf{H}\left(\sqrt{2}\,H - x\right)\, PC_h$ is a computable approximation to I and the integral J_H is small, as quantified in the following proposition.

Proposition 1 *Let e_h denote the value of the integral J_H when we choose $H = \pi/h$. Then, for $x > 0$,*

$$|e_h| \le \delta_1(x) := \frac{x\ \mathrm{e}^{-\pi^2/h^2}}{\sqrt{\pi}\,|\pi^2/h^2 - x^2/2|\ \left(1 - \mathrm{e}^{-2\pi^2/h^2}\right)}. \tag{30}$$

Proof For $z = X + \mathrm{i}H$,

$$|x^2 + \mathrm{i}z^2| = |z_0 - z|\,|z_0 + z| \ge |x/\sqrt{2} - H|\,|x/\sqrt{2} + H| = |x^2/2 - H^2|$$

so, using (25) and recalling that $\int_{-\infty}^{\infty} \mathrm{e}^{-t^2}\,dt = \sqrt{\pi}$, we see that

$$|J_H| \le \frac{x\,\mathrm{e}^{H^2 - 2\pi H/h}}{\sqrt{\pi}\,|H^2 - x^2/2|\ \left(1 - \mathrm{e}^{-2\pi H/h}\right)}.$$

Choosing $H = \pi/h$, to minimise the exponent $H^2 - 2\pi H/h$, the result (30) follows. □

Note that $I_h + \mathbf{H}\left(\sqrt{2}\,\pi/h - x\right)\ PC_h = I_h + R(h,x)$ is precisely the approximation (11), and that the above bound on e_h is precisely the bound (12) from [16].

Theorem 1 *Let $I_h^* := I_h + PC_h$ and $e_h^* := I - I_h^*$. Then, for $x > 0$,*

$$|e_h^*| \le \Delta_h(x), \tag{31}$$

where

$$\Delta_h(x) := \begin{cases} \delta_1(x), & 0 \le \frac{x}{\sqrt{2}} \le \frac{3}{4}\frac{\pi}{h}, \\ \delta_2(x), & \frac{3}{4}\frac{\pi}{h} < \frac{x}{\sqrt{2}} < \frac{5}{4}\frac{\pi}{h}, \\ \delta_3(x), & \frac{x}{\sqrt{2}} \ge \frac{5}{4}\frac{\pi}{h}. \end{cases} \tag{32}$$

Here δ_1 is defined by (30)*,*

$$\delta_2(x) := \frac{4hx\ \mathrm{e}^{-\pi^2/h^2}}{\sqrt{\pi}\,\pi(\pi/h + x/\sqrt{2})\ \left(1 - \mathrm{e}^{-2\pi^2/h^2}\right)} \left(1 + 2\sqrt{\pi}\,\mathrm{e}^{-\beta\pi^2/h^2}\right), \tag{33}$$

with $\beta = 1 - \sqrt{2}/2 - (2\sqrt{2}+1)/16 \approx 0.0536$, and

$$\delta_3(x) := \delta_1(x) + \frac{\mathrm{e}^{-\sqrt{2}\,\pi x/h}}{1 - \mathrm{e}^{-\sqrt{2}\,\pi x/h}}. \tag{34}$$

Proof The bound (30) implies that $|e_h^*| \le \delta_1(x)$, for $0 < x < \sqrt{2}\,\pi/h$. Since, applying (25),

$$|PC_h| \le \frac{\mathrm{e}^{-\sqrt{2}\,\pi x/h}}{1 - \mathrm{e}^{-\sqrt{2}\,\pi x/h}},$$

the bound (30) also implies that $|e_h^*| \le \delta_3(x)$, for $x > \sqrt{2}\,\pi/h$.

Setting $H = \pi/h$, select ϵ in the range $(0, H)$ and consider the case that $\left|x/\sqrt{2} - H\right| < \epsilon$. In this case we observe that the derivation of (29) can be modified to show that

$$e_h^* = \int_{\Gamma_H^*} f(z)(1 + g(z))\,\mathrm{d}z, \tag{35}$$

where the contour Γ_H^* passes above the pole in f at z_0; precisely, Γ_H^* is the union of Γ' and γ, where $\Gamma' = \{z \in \Gamma_H : |z - z_0| > \epsilon\}$ and γ is the circular arc $\gamma = \{z_0 + \epsilon \mathrm{e}^{\mathrm{i}\theta} : \theta_0 \le \theta \le \pi - \theta_0\}$, where $\theta_0 = \sin^{-1}((H - x/\sqrt{2})/\epsilon) \in (-\pi/2, \pi/2)$. For $z \in \Gamma'$ it holds that

$$|x^2 + \mathrm{i}z^2| = |z_0 - z|\,|z_0 + z| \ge \epsilon\,|x/\sqrt{2} + H|. \tag{36}$$

Thus, and applying (25), similarly to (30) we deduce that

$$\left|\int_{\Gamma'} f(z)(1 + g(z))\,\mathrm{d}z\right| \le \frac{x\ \mathrm{e}^{-\pi^2/h^2}}{\sqrt{\pi}\,\epsilon|\pi/h + x/\sqrt{2}|\ \left(1 - \mathrm{e}^{-2\pi^2/h^2}\right)}. \tag{37}$$

To bound the integral over γ we note that, for $z = X + \mathrm{i}Y = z_0 + \epsilon \mathrm{e}^{\mathrm{i}\theta} \in \gamma$, (36) is true and $Y \ge H$. Further, $|\mathrm{e}^{-z^2}| = \mathrm{e}^P$, where

$$P = Y^2 - X^2 = 2x\epsilon \sin(\theta - \pi/4) - \epsilon^2 \cos(2\theta) < 2x\epsilon + \epsilon^2 \le 2\sqrt{2}H\epsilon + (2\sqrt{2}+1)\epsilon^2,$$

since $\left|x/\sqrt{2} - H\right| < \epsilon$. From these bounds and (25), defining $\alpha = \epsilon/H \in (0, 1)$, we deduce that

$$\left|\int_{\gamma} f(z)(1 + g(z))\,\mathrm{d}z\right| \le \frac{2x\ \exp((2\sqrt{2}\alpha + (2\sqrt{2}+1)\alpha^2 - 2)\pi^2/h^2)}{\epsilon|\pi/h + x/\sqrt{2}|\ \left(1 - \mathrm{e}^{-2\pi^2/h^2}\right)}. \tag{38}$$

For x in the range $\left|x/\sqrt{2} - H\right| < \epsilon$ we can bound e_h^* using (35), (37), (38), and the triangle inequality, to get that

$$|e_h^*| \le \frac{hx\ \mathrm{e}^{-\pi^2/h^2}}{\alpha\sqrt{\pi}\,\pi|\pi/h + x/\sqrt{2}|\ \left(1 - \mathrm{e}^{-2\pi^2/h^2}\right)} \left(1 + 2\sqrt{\pi}\,\mathrm{e}^{-\beta\pi^2/h^2}\right), \tag{39}$$

where $\beta = 1 - 2\sqrt{2}\alpha - (2\sqrt{2}+1)\alpha^2$. Noting that $\beta > 0$ if and only if $0 < \alpha < \alpha_0$, where $\alpha_0 = (1 + 2\sqrt{2})^{-1} \approx 0.2612$, we choose $\alpha < \alpha_0$ to be $\alpha = 1/4$. With this choice it follows from (39) that $|e_h^*| \le \delta_2(x)$ for $\frac{3}{4}\frac{\pi}{h} < \frac{x}{\sqrt{2}} < \frac{5}{4}\frac{\pi}{h}$, and the proof is complete. □

The approximation $F_N(x)$, given by (14), that we propose for $I = F(x)$ is just $I_h^* = I_h + PC_h$ with a particular choice of h and with the range of summation in (27) reduced to the finite range $1, ..., N$. This induces an additional error,

$$T_N := 2h \sum_{m=N+1}^{\infty} f(\tau_m), \tag{40}$$

that we bound in the next proposition.

Proposition 2 *For $x > 0$,*

$$|T_N| \le \frac{(2h\tau_{N+1}+1)x}{2\pi\tau_{N+1}\sqrt{x^4+\tau_{N+1}^4}}\,\mathrm{e}^{-\tau_{N+1}^2}.$$

Proof

$$\begin{aligned}
|T_N| &\le \frac{hx}{\pi}\sum_{m=N+1}^{\infty}\frac{\mathrm{e}^{-\tau_m^2}}{\sqrt{x^4+\tau_m^4}}\\
&\le \frac{x}{2\pi\sqrt{x^4+\tau_{N+1}^4}}\left(2h\mathrm{e}^{-\tau_{N+1}^2}+2h\sum_{m=N+2}^{\infty}\mathrm{e}^{-\tau_m^2}\right)\\
&\le \frac{x}{2\pi\sqrt{x^4+\tau_{N+1}^4}}\left(2h\mathrm{e}^{-\tau_{N+1}^2}+2\int_{\tau_{N+1}}^{\infty}\mathrm{e}^{-t^2}\mathrm{d}t\right)\\
&\le \frac{x}{2\pi\sqrt{x^4+\tau_{N+1}^4}}\left(2h\mathrm{e}^{-\tau_{N+1}^2}+\frac{\mathrm{e}^{-\tau_{N+1}^2}}{\tau_{N+1}}\right)=\frac{(2h\tau_{N+1}+1)x}{2\pi\tau_{N+1}\sqrt{x^4+\tau_{N+1}^4}}\,\mathrm{e}^{-\tau_{N+1}^2}.
\end{aligned}$$

To arrive at the last line we have used that, for $x > 0$,

$$2\int_x^{\infty}\mathrm{e}^{-t^2}\mathrm{d}t=\frac{\mathrm{e}^{-x^2}}{x}-\int_x^{\infty}\frac{\mathrm{e}^{-t^2}}{t^2}\mathrm{d}t<\frac{\mathrm{e}^{-x^2}}{x}. \tag{41}$$

□

At this point we make a choice of h to approximately equalise $\Delta_h(x)$ in Theorem 1 (which is approximately proportional to $\exp(-\pi^2/h^2)$) and the bound on T_N in Proposition 2, choosing h so that $\pi/h = \tau_{N+1} = (N+1/2)h$. In other words, we make the choice $h = \sqrt{\pi/(N+1/2)}$ given by (10), in which case $\tau_{N+1} = A_N = \sqrt{(N+1/2)\pi}$, and $\tau_k = t_k$, where t_k is defined by (16). Making this choice of h we see that

$$E_N(x) := F(x) - F_N(x) = e_h^* + T_N \tag{42}$$

and that

$$|T_N| \le \frac{(2\pi+1)x}{2\pi A_N\sqrt{x^4+A_N^4}}\,\mathrm{e}^{-A_N^2}. \tag{43}$$

Combining (42) and (43) with Theorem 1, we arrive at the following theorem which is our main pointwise error bound. Theorem 1, (42), and (43) prove this theorem only for $x > 0$, but the symmetries (20) and (21) imply that $E_N(-x) = -E_N(x)$, so that (44) holds also for $x < 0$, and, by continuity, also for $x = 0$ (and in fact $E_N(0) = \eta_N(0) = 0$).

Theorem 2 *For $x \in \mathbb{R}$,*

$$|E_N(x)| \le \eta_N(x) := \Delta_h(|x|) + \frac{(2\pi+1)|x|}{2\pi A_N\sqrt{x^4+A_N^4}}\,\mathrm{e}^{-A_N^2}, \tag{44}$$

where

$$\Delta_h(x) = \begin{cases} \dfrac{x\,\mathrm{e}^{-A_N^2}}{\sqrt{\pi}\,(A_N^2 - x^2/2)\,\left(1-\mathrm{e}^{-2A_N^2}\right)}, & 0 \le \dfrac{x}{\sqrt{2}} \le \frac{3}{4}A_N, \\ \dfrac{4x\,\mathrm{e}^{-A_N^2}\left(1+2\sqrt{\pi}\,\mathrm{e}^{-\beta A_N^2}\right)}{\sqrt{\pi}\,A_N(A_N+x/\sqrt{2})\,\left(1-\mathrm{e}^{-2A_N^2}\right)}, & \frac{3}{4}A_N < \dfrac{x}{\sqrt{2}} < \frac{5}{4}A_N, \\ \dfrac{x\;\mathrm{e}^{-A_N^2}}{\sqrt{\pi}\,(x^2/2 - A_N^2)\,\left(1-\mathrm{e}^{-2A_N^2}\right)} + \dfrac{\mathrm{e}^{-\sqrt{2}\,A_N x}}{1-\mathrm{e}^{-\sqrt{2}A_N x}}, & \dfrac{x}{\sqrt{2}} \ge \frac{5}{4}A_N. \end{cases} \tag{45}$$

We will compare $|E_N(x)|$ to the upper bound $\eta_N(x)$ for $N = 9$ in Figure 3 below. The following theorem estimates the maximum value of $\eta_N(x)$ on the real line.

Theorem 3 *For $x \in \mathbb{R}$,*

$$|F(x) - F_N(x)| \le \eta_N(x) \le c_N \frac{\mathrm{e}^{-\pi N}}{\sqrt{N+1/2}}, \tag{46}$$

where

$$c_N = \frac{20\sqrt{2}\mathrm{e}^{-\pi/2}}{9\pi\left(1-\mathrm{e}^{-2A_N^2}\right)}\left(1+2\sqrt{\pi}\,\mathrm{e}^{-\beta A_N^2}\right) + \frac{(2\pi+1)\mathrm{e}^{-\pi/2}}{2\sqrt{2}\,\pi^{3/2}A_N}, \tag{47}$$

which decreases as N increases, with

$$c_1 \approx 0.825 \quad \textit{and} \quad \lim_{N\to\infty} c_N = \frac{20\sqrt{2}\mathrm{e}^{-\pi/2}}{9\pi} \approx 0.208. \tag{48}$$

Proof It is easy to see that $\Delta_h(x)$ is increasing on $[0, \frac{5}{4}\sqrt{2}\,A_N)$ and decreasing on $[\frac{5}{4}\sqrt{2}\,A_N, \infty)$. Further, where $\Delta_h(\frac{5}{4}\sqrt{2}\,A_N^-)$ denotes the limiting value of $\Delta_h(x)$ as $x \to \frac{5}{4}\sqrt{2}\,A_N$ from below, since $2A_N^{-1} > \mathrm{e}^{-A_N^2}$,

$$\begin{aligned} \Delta_h\left(\tfrac{5}{4}\sqrt{2}A_N^-\right) &= \frac{20\sqrt{2}\,\mathrm{e}^{-A_N^2}}{9\sqrt{\pi}\,A_N\,\left(1-\mathrm{e}^{-2A_N^2}\right)}\left(1+2\sqrt{\pi}\,\mathrm{e}^{-\beta A_N^2}\right) \\ &> \frac{20\sqrt{2}\;\mathrm{e}^{-A_N^2}}{9\sqrt{\pi}\,A_N\,\left(1-\mathrm{e}^{-2A_N^2}\right)} + \frac{\mathrm{e}^{-5A_N^2/2}}{1-\mathrm{e}^{-5A_N^2/2}} = \Delta_h\left(\tfrac{5}{4}\sqrt{2}\,A_N\right). \end{aligned}$$

Similarly, $x\Delta_h(x)$ is increasing on $[0, \frac{5}{4}\sqrt{2}\,A_N)$ and decreasing on $[\frac{5}{4}\sqrt{2}\,A_N, \infty)$. Thus, for $x \ge 0$,

$$\Delta_h(x) \le \Delta_h\left(\tfrac{5}{4}\sqrt{2}\,A_N^-\right) \quad \text{and} \quad x\Delta_h(x) \le \tfrac{5}{4}\sqrt{2}\,A_N\Delta_h\left(\tfrac{5}{4}\sqrt{2}\,A_N^-\right). \tag{49}$$

Moreover,

$$\frac{|x|}{\sqrt{x^4+A_N^4}} \le \frac{1}{\sqrt{2}\,A_N} \quad \text{and} \quad \frac{x^2}{\sqrt{x^4+A_N^4}} < 1, \quad \text{for } x \in \mathbb{R}. \tag{50}$$

Combining (44), (49) and (50) we reach the result. □

We can also bound the relative error in our approximation $F_N(x)$. The proof of Theorem 4 is postponed to the appendix.

Theorem 4

$$|F(x)| \ge \frac{1}{2+2\sqrt{\pi}\,x}, \quad \textit{for } x \ge 0, \tag{51}$$

and

$$|F(x)| \ge \frac{1}{2}, \quad \textit{for } x \le 0. \tag{52}$$

Theorem 5

$$\frac{|F(x)-F_N(x)|}{|F(x)|} \le \frac{\eta_N(x)}{|F(x)|} \le \begin{cases} c_N^* \mathrm{e}^{-\pi N}, & \textit{for } x \ge 0, \\ 2c_N \dfrac{\mathrm{e}^{-\pi N}}{\sqrt{N+1/2}}, & \textit{for } x \le 0, \end{cases} \tag{53}$$

where

$$c_N^* = \frac{10\sqrt{2}\left(4+5\sqrt{2\pi}A_N\right)\left(1+2\sqrt{\pi}\mathrm{e}^{-\beta A_N^2}\right)}{9\sqrt{\pi}\,\mathrm{e}^{\pi/2}\,A_N\left(1-\mathrm{e}^{-2A_N^2}\right)} + \frac{(2\pi+1)}{\pi \mathrm{e}^{\pi/2} A_N}\left(\frac{1}{\sqrt{2}\,A_N}+\sqrt{\pi}\right),$$

which decreases as N increases, with $c_1^* \approx 10.4$ and $\lim_{N\to\infty} c_N^* = 100\mathrm{e}^{-\pi/2}/9 \approx 2.3$.

Proof Combining (51), (44), (49), and (50), we see that, for $x \ge 0$,

$$\frac{\eta_N(x)}{|F(x)|} \le \left(2+\tfrac{5}{2}\sqrt{2\pi}A_N\right)\Delta_h\left(\tfrac{5}{4}\sqrt{2}\,A_N^-\right) + \frac{(2\pi+1)}{\pi}\frac{\mathrm{e}^{-A_N^2}}{A_N}\left(\frac{1}{\sqrt{2}\,A_N}+\sqrt{\pi}\right).$$

This implies the bound (53) for $x \ge 0$. The bound (53) for $x \le 0$ follows immediately from (52) and (46). □

In the above theorems we use (44) and (45) to bound the maximum absolute and relative errors in the approximation $F_N(x)$. These inequalities, additionally, imply that $F_N(x)$ is particularly accurate for $|x|$ small. For $|x| \le A_N/\sqrt{2} = \sqrt{(N+1/2)\pi/2}$, it follows from (44) and (45) that

$$|F(x)-F_N(x)| \le \eta(x) \le \tilde{c}_N|x|\frac{\mathrm{e}^{-\pi N}}{2N+1} \tag{54}$$

where

$$\tilde{c}_N = \frac{8}{3\pi^{3/2}\mathrm{e}^{\pi/2}\left(1-\mathrm{e}^{-2A_N^2}\right)} + \frac{(2\pi+1)}{\pi^2\mathrm{e}^{\pi/2}A_N}, \tag{55}$$

which decreases as N increases, with $\tilde{c}_1 \approx 0.17$ and $\lim_{N\to\infty}\tilde{c}_N = 8/(3\pi^{3/2}\mathrm{e}^{\pi/2}) \approx 0.10$.

2.1 Extensions of the error bounds into the complex plane

In §1 we have made claims regarding the analyticity of the approximation $F_N(x)$, considered as a function of x in the complex plane. We justify these claims now. One attractive feature of the modified trapezium rule approximation I_h^* is that, in contrast to I_h, it is entire as a function of x. This is not immediately obvious: $I_h^* = I_h + PC_h$, and PC_h has simple pole singularities at $x = \mathrm{e}^{-\mathrm{i}\pi/4}\tau_k$, $k \in \mathbb{Z}$. But I_h also has simple poles at the same points and it is an easy calculation to see that the residues add to zero, so that the singularities cancel out. Since $F_N(x) = I_h^* - T_N$, with h given by (10), it follows that the singularities of $F_N(x)$ are those of T_N, *i.e.*, simple poles at $\pm\mathrm{e}^{-\mathrm{i}\pi/4}t_k$, for $k = N+1, N+2, \ldots$. Thus $F_N(x)$ is a meromorphic function and, in particular, is analytic in the strip $|\mathrm{Im}(x)| < A_N/\sqrt{2}$ and in the first and third quadrants of the complex plane.

We will note two consequences of this analyticity and the bounds that we have already proved. In these arguments we will use an extension of the maximum principle for analytic functions to unbounded domains, that if $w(z)$ is analytic in an open quadrant in the complex plane, let us say $Q = \{z \in \mathbb{C} : 0 < |\arg(z)| < \pi/2\}$, and is continuous and bounded in its closure, then

$$\sup_{z\in Q} |w(z)| \le \sup_{z\in \partial Q} |w(z)|, \tag{56}$$

where ∂Q denotes the boundary of the quadrant. (This sort of extension of the maximum principle to unbounded domains is due to Phragmen and Lindelöf; see, *e.g.*, [26].)

The first consequence is that, from (42), (46), and (22), it follows that the bound (46) holds on both the real and imaginary axes. Further, from (4) and the asymptotics of erfc(z) in the complex plane [3, (7.1.23)], it follows that $F(z) \to 0$, uniformly in $\arg(z)$, for $0 \le \arg(z) \le \pi/2$; moreover, it is clear from (15) that the same holds for $F_N(z)$ and hence for $E_N(z)$. Thus (56) implies that (46) holds for $0 \le \arg(z) \le \pi/2$, and (20) and (21) then imply that (46) holds also for $\pi \le \arg(z) \le 3\pi/4$.

It is clear from the derivations above that, if h is given by (10), then I_h^* also satisfies the bound (46), *i.e.*,

$$|F(z) - I_h^*| \le c_N \frac{\mathrm{e}^{-\pi N}}{\sqrt{N+1/2}}, \tag{57}$$

this holding in the first instance for real z, then for imaginary z, and finally for all z in the first and third quadrants. The bound (46) cannot hold in the second or fourth quadrant because $E_N(z) = F(z) - F_N(z)$ has poles there. This issue does not hold for $F(z) - I_h^*$, which is an entire function, but (57) cannot hold in the whole complex plane because this, by Liouville's theorem [26], would imply that $F(z) - I_h^*$ is a constant. What does hold is that $\mathrm{e}^{-\mathrm{i}z^2}(F(z) - I_h^*)$ is bounded in the second and fourth quadrants, this a consequence of the

definition of I_h^* and the asymptotics of $\mathrm{e}^{z^2}\mathrm{erfc}(z)$ at infinity. Thus it follows from (56), and since $|\mathrm{e}^{-\mathrm{i}z^2}| = 1$ if z is real or pure imaginary, that

$$|F(z) - I_h^*| \le c_N \mathrm{e}^{-xy} \frac{\mathrm{e}^{-\pi N}}{\sqrt{N+1/2}}, \tag{58}$$

for $z = x + \mathrm{i}y$ in the second and fourth quadrants.

We can use the bound (58) to obtain a bound on $E_N(x)$ in the second and fourth quadrants. Clearly, where T_N is defined by (40), with h given by (10), for $z = x + \mathrm{i}y$ in the second and fourth quadrants,

$$|F(z) - F_N(z)| \le c_N \mathrm{e}^{-xy} \frac{\mathrm{e}^{-\pi N}}{\sqrt{N+1/2}} + |T_N|.$$

Further, arguing as below (40), if $|y| \le A_N/(2\sqrt{2})$ so that

$$|z^2 + \mathrm{i}t_k^2| \ge \left(\frac{A_N}{\sqrt{2}} - |y|\right)\left(\left(\frac{A_N}{\sqrt{2}} - |y|\right)^2 + \left(\frac{A_N}{\sqrt{2}} + |x|\right)^2\right) \ge \frac{A_N}{2\sqrt{2}}\left(A_N^2/8 + |x|^2\right),$$

which implies that $|z^2 + \mathrm{i}t_k^2| \ge |z|A_N/(2\sqrt{2})$, then

$$|T_N| \le \mathrm{e}^{-xy} \frac{(2\pi+1)\sqrt{2}}{\pi A_N^2} \mathrm{e}^{-A_N^2} = \mathrm{e}^{-xy} \frac{\sqrt{2}(2\pi+1)}{\pi^{3/2}\exp(\pi/2)(N+1/2)} \mathrm{e}^{-\pi N}.$$

Thus, for $z = x + \mathrm{i}y$ in the second and fourth quadrants with $|y| \le A_N/(2\sqrt{2})$,

$$|F(z) - F_N(z)| \le \hat{c}_N \mathrm{e}^{-xy} \frac{\mathrm{e}^{-\pi N}}{\sqrt{N+1/2}} \tag{59}$$

where

$$\hat{c}_N := c_N + \frac{\sqrt{2}(2\pi+1)}{\pi^{3/2}\exp(\pi/2)\sqrt{N+1/2}}, \tag{60}$$

which is decreasing with $\hat{c}_1 \approx 1.14$ and $\lim_{N\to\infty} \hat{c}_N = \lim_{N\to\infty} c_N \approx 0.208$.

We observe above that the bound (46) on $E_N(z) = F(z) - F_N(z)$ holds for all complex z in the first and third quadrants of the complex plane, and on the boundaries of those quadrants, the real and imaginary axes, while the bound (59) holds in the second and fourth quadrants for $|\mathrm{Im}(z)| \le A_N/(2\sqrt{2})$. These bounds imply that the coefficients in the Maclaurin series of $F_N(z)$ are close to those of $F(z)$. Precisely, at least for $|z| < A_N/\sqrt{2}$,

$$F(z) = \sum_{n=0}^{\infty} a_n z^n \quad \text{and} \quad F_N(z) = \sum_{n=0}^{\infty} b_n z^n,$$

with $a_n = F^{(n)}(0)/n!$, $b_n = F_N^{(n)}(0)/n!$. Thus, where $M_N = \sup_{|z|<\sqrt{\pi/2}} |E_N(z)|$, it follows from Cauchy's estimate [26, Theorem 10.26] and the bounds (46) and (59) that, for $N \ge 4$ so that $A_N/(2\sqrt{2}) \ge \sqrt{\pi/2}$,

$$|a_n - b_n| = \frac{|E_N^{(n)}(0)|}{n!} \le M_N \left(\frac{2}{\pi}\right)^{n/2} \le \hat{c}_N \left(\frac{2}{\pi}\right)^{n/2} \frac{\mathrm{e}^{-\pi(N-1/4)}}{\sqrt{N+1/2}}. \tag{61}$$

3 Approximating $C(x)$ and $S(x)$

From (3) we see that, for x real,

$$C(x) = \mathrm{Re}\left(\sqrt{2}\,\mathrm{e}^{\mathrm{i}\pi/4}(\tfrac{1}{2} - F(\sqrt{\pi/2}\,x))\right),\ \ S(x) = \mathrm{Im}\left(\sqrt{2}\,\mathrm{e}^{\mathrm{i}\pi/4}(\tfrac{1}{2} - F(\sqrt{\pi/2}\,x))\right). \tag{62}$$

Clearly, given the approximation $F_N(x)$ to $F(x)$, these relationships can be used to generate approximations for the Fresnels integrals $C(x)$ and $S(x)$. These approximations are defined, for $x \in \mathbb{R}$, by

$$\begin{aligned} C_N(x) &= \mathrm{Re}\left(\sqrt{2}\,\mathrm{e}^{\mathrm{i}\pi/4}(\tfrac{1}{2} - F_N(\sqrt{\pi/2}\,x))\right), \\ S_N(x) &= \mathrm{Im}\left(\sqrt{2}\,\mathrm{e}^{\mathrm{i}\pi/4}(\tfrac{1}{2} - F_N(\sqrt{\pi/2}\,x))\right), \end{aligned} \tag{63}$$

and are given explicitly in (17) and (18). We note the similarity between (17) and (18) and the formulae [1, (7.5.3)-(7.5.4)]

$$C(x) = \tfrac{1}{2} + f(x)\sin\left(\tfrac{1}{2}\pi x^2\right) - g(x)\cos\left(\tfrac{1}{2}\pi x^2\right), \tag{64}$$

$$S(x) = \tfrac{1}{2} - f(x)\cos\left(\tfrac{1}{2}\pi x^2\right) - g(x)\sin\left(\tfrac{1}{2}\pi x^2\right), \tag{65}$$

which express $C(x)$ and $S(x)$ in terms of the auxiliary functions, $f(x)$ and $g(x)$, for the Fresnel integrals [1, §7.2(iv)]. Indeed, it follows from [1, (7.7.10)-(7.7.11)] that, for $x > 0$, $f(x)$ and $g(x)$ have the integral representations

$$f(x) = \frac{\sqrt{\pi}\,x^3}{2}\int_0^\infty \frac{\mathrm{e}^{-t^2}}{\left(\frac{\pi}{2}x^2\right)^2 + t^4}\,\mathrm{d}t \ \text{ and } \ g(x) = \frac{x}{\sqrt{\pi}}\int_0^\infty \frac{t^2\mathrm{e}^{-t^2}}{\left(\frac{\pi}{2}x^2\right)^2 + t^4}\,\mathrm{d}t,$$

and, recalling that A_N is linked to the quadrature step-size through (10), it is clear that, for $x > 0$, $\sqrt{\pi}\,x a_N\left(\frac{\pi}{2}x^2\right)/A_N$ and $\sqrt{\pi}\,x b_N\left(\frac{\pi}{2}x^2\right)/A_N$ can be viewed as quadrature approximations to these integrals.

The approximations (17) and (18) inherit the accuracy of $F_N(x)$ on the real line: from (62) and (63) we see, for $x \in \mathbb{R}$, that

$$|C(x) - C_N(x)| \le \sqrt{2}\,|E_N(\sqrt{\pi/2}\,x)| \ \text{ and } |S(x) - S_N(x)| \le \sqrt{2}\,|E_N(\sqrt{\pi/2}\,x)|. \tag{66}$$

where $E_N(x) = F(x) - F_N(x)$. Thus the error bounds of the previous section can be applied. In particular, from (46) and (54) it follows that both $|C(x) - C_N(x)|$ and $|S(x) - S_N(x)|$ are

$$\le 2c_N\frac{\mathrm{e}^{-\pi N}}{\sqrt{2N+1}}, \quad \text{for } x \in \mathbb{R}, \tag{67}$$

and

$$\le \sqrt{\pi}\,\tilde{c}_N|x|\,\frac{\mathrm{e}^{-\pi N}}{2N+1}, \quad \text{for } |x| \le \sqrt{N+1/2}\,. \tag{68}$$

Here $c_N < 0.83$ and $\tilde{c}_N < 0.18$ are the decreasing sequences of positive numbers defined by (47) and (55), respectively.

These bounds show that $C_N(x)$ and $S_N(x)$ are exponentially convergent as $N \to \infty$, uniformly on the real line, so that very accurate approximations can be obtained with very small values of N ((67) shows that both $|C_N(x)-C(x)|$ and $|S_N(x)-S(x)|$ are $\le 1.4\times10^{-16}$ on the real line for $N \ge 11$). In §4 we will confirm the effectiveness of these approximations by numerical experiments, checking the accuracy of (17) and (18) by comparison with the power series [1, §7.6(i)]

$$C(x) = \sum_{n=0}^{\infty} \frac{(-1)^n \left(\frac{1}{2}\pi\right)^{2n} x^{4n+1}}{(2n)!(4n+1)}, \quad S(x) = \sum_{n=0}^{\infty} \frac{(-1)^n \left(\frac{1}{2}\pi\right)^{2n+1} x^{4n+3}}{(2n+1)!(4n+3)}. \tag{69}$$

It follows from the analyticity of $F_N(x)$ in the complex plane, discussed in §2.1, that $F_N(x)$ has a power series convergent in $|x| < A_N/\sqrt{2}$, and from (63) that $C_N(x)$ and $S_N(x)$ have convergent power series representations in $|x| < A_N/\sqrt{\pi}$. From the observations below (23) it is clear that, echoing (69), these take the form

$$C_N(x) = \sum_{n=0}^{\infty} \mathfrak{c}_n x^{4n+1}, \quad S_N(x) = \sum_{n=0}^{\infty} \mathfrak{s}_n x^{4n+3}. \tag{70}$$

Further, it follows from (63) and (61) that the coefficients $\mathfrak{c}_n$ and $\mathfrak{s}_n$ are close to the corresponding coefficients of $C(x)$ and $S(x)$, with the difference having absolute value

$$\le \sqrt{2}\,\hat{c}_N \frac{\mathrm{e}^{-\pi(N-1/4)}}{\sqrt{N+1/2}}, \tag{71}$$

for $N \ge 4$, where $\hat{c}_N \le \hat{c}_4 < 0.77$ is the decreasing sequence of positive numbers given by (60). This implies that, near zero, where $C(x)$ has a simple zero and $S(x)$ a zero of order three, the approximations $C_N(x)$ and $S_N(x)$ retain small relative error. For $C_N(x)$ this follows already from (68) but to see this for $S_N(x)$ we need the stronger bound implied by (71) that, for $|x| < 1$,

$$|S(x) - S_N(x)| \le \sqrt{2}\,\hat{c}_N \frac{\mathrm{e}^{-\pi(N-1/4)}}{\sqrt{N+1/2}} \sum_{n=0}^{\infty} |x|^{4n+3} = \frac{|x|^3}{1-|x|^4} \frac{\sqrt{2}\,\hat{c}_N\,\mathrm{e}^{-\pi(N-1/4)}}{\sqrt{N+1/2}}. \tag{72}$$

Table 2 shows the Matlab implementing (17) and (18) that we use in the next section. To evaluate $(\sinh t \pm \sin t)/(\cosh t + \cos t)$, with $t = \sqrt{\pi}\,A_N x$, in (17) and (18), we note that, for $|t| \ge 39$, $\cosh(t) + \cos(t)$ and $\exp(t)/2$ have the same value in double precision arithmetic, as do $\sinh t \pm \sin t$ and $\mathrm{sign}(t)\exp(t)/2$. Thus this expression evaluates as $\mathrm{sign}(t)$ in double precision arithmetic for $39 \le |t| \lessapprox 710$. To avoid underflow and reduce computation time, we evaluate it as $\mathrm{sign}(t)$ for $|t| \ge 39$. For small t there is an additional issue of loss of precision in evaluating $\sinh t - \sin t$ for $|t|$ small. This is avoided in Table 2 by using $\sinh t - \sin t = 2t^3/3! + 2t^7/7! + \ldots$ for $|t| < 1$, truncating after four terms as the 5th term is negligible in double precision.

```
function [C,S] = fresnelCS(x,N)
% Evaluates approximations to the Fresnel integrals C(x) and S(x).
% x is a real scalar or matrix,
% N is a positive integer controlling accuracy (suggest N=12),
% C and S are the scalars/matrices of the same size as x approximating C(x) and S(x).
h = sqrt(pi/(N+0.5));
t = h*((N:-1:1)-0.5);  AN = pi/h; rootpi = sqrt(pi);
t2 = t.*t; t4 = t2.*t2; et2 = exp(-t2);
x2pi2 = (pi/2)*x.*x; x4 = x2pi2.*x2pi2;
a = et2(1)./(x4+t4(1)); b = t2(1)*a;
for n = 2:N
    term = et2(n)./(x4+t4(n));
    a = a + term;  b = b + t2(n)*term;
end
a = a.*x2pi2;
mx = (rootpi*AN)*x; Mx = (rootpi/AN)*x;
Chalf = 0.5*sign(mx); Shalf = Chalf;
select = abs(mx)<39;
if any(select)
    mxs = mx(select); shx = sinh(mxs); sx = sin(mxs);
    den = 0.5./(cos(mxs)+cosh(mxs));
    Chalf(select) = (shx+sx).*den;
    ssdiff = shx-sx;
    select2 = abs(mxs)<1;
    if any(select2)
        mxs = mxs(select2); mxs3 = mxs.*mxs.*mxs; mxs4 = mxs3.*mxs;
        ssdiff(select2) = mxs3.*(1/3 + mxs4.*(1/2520 ...
            + mxs4.*((1/19958400)+(0.001/653837184)*mxs4)));
    end
    Shalf(select) = ssdiff.*den;
end
cx2 = cos(x2pi2); sx2 = sin(x2pi2);
C = Chalf + Mx.*(a.*sx2-b.*cx2); S = Shalf - Mx.*(a.*cx2+b.*sx2);
```

Table 2 Matlab to evaluate $C_N(x)$ and $S_N(x)$ given by (17) and (18). See §3 for details.

4 Numerical Results and Comparison of Methods

In this section we show numerical computations that confirm and illustrate the theoretical error bounds in §2 and §3, and that explore the accuracy and efficiency of our new methods, through qualitative and quantitative comparisons with certain of the other computational methods described in §1.1.

In Figure 1 it can be seen that the exponential convergence predicted by the bounds (46) and (53) is achieved, indeed these bounds overestimate their respective maximum errors by at most a factor of 10. Further, with N as small as 12 it appears that we achieve maximum absolute and relative errors in $F_N(x)$ which are $< 2.9\times10^{-16}$ and $< 9.3\times10^{-16}$, respectively; these values are upper bounds whichever of the two methods for approximating $F(x)$ accurately is used. (We should add a note of caution here: the different approximations agree to high accuracy, but the accuracy of each approximation is limited, for large x, by the accuracy with which $\mathrm{e}^{\mathrm{i}x^2}$ is computed.) These plots also

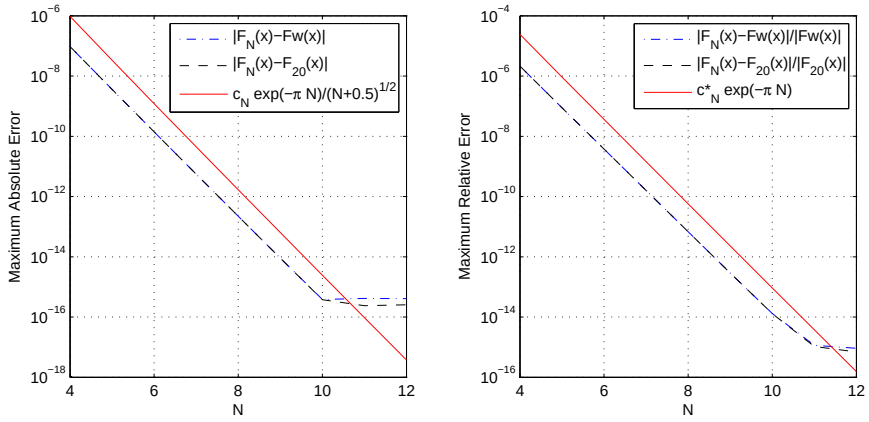

Fig. 1 Left hand side: maximum error, $\max_{x\geq 0}|F(x)-F_N(x)|$, and its upper bound (46) (−), plotted against N, in one case where $F(x)$ is approximated by $Fw(x) := \mathrm{e}^{\mathrm{i}x^2} w_{36}(\mathrm{e}^{\mathrm{i}\pi/4}x)/2$ (− · −) with $w_{36}(z)$ defined by (13) and computed by the function in Table 1 of [30], and in the other case where $F(x)$ is approximated by $F_{20}(x)$ (−−). Right hand side: maximum relative error, $\max_{x\geq 0}|(F(x)-F_N(x))/F(x)|$, and its upper bound (53) (−), plotted against N, where $F(x)$ is approximated in the two curves as on the left hand side. (All maximums are taken over 40,000 equally spaced points between 0 and 1,000, and all values of $F_N(x)$ are computed using the code in Table 1.)

verify the high accuracy of the approximation (13) for $w(z)$ from [30], at least for $\arg(z) = \pi/4$ and if M is large enough in (13). Figure 2 explores this in more detail: in each plot the trend is one of exponential convergence, but the convergence is not monotonic and is slower than that in Figure 1.

In Figure 3 we see that our pointwise theoretical error bounds are upper bounds as claimed, and that these bounds appear to capture the x-dependence of the errors fairly well, for example that $E_N(x) = O(x)$ as $x \to 0$, $= O(x^{-1})$ as $x \to \infty$, and that $E_N(x)$ reaches a maximum at about $x = \sqrt{2}\,A_N = \sqrt{\pi(2N+1)}$ (≈ 7.7 when $N = 9$).

The above figures explore the accuracy of the approximation $F_N(x)$. Let us comment on efficiency. Most straightforward is a comparison of the Matlab function `F(x,N)` in Table 1 with computation of $F(x)$ via the Matlab code `Fw(x,M)=exp(i*x.^2).*cef(exp(i*pi/4)*x,M)/2` that uses `cef.m` from [30] implementing (13). Both `F(x,N)` and `cef(x,M)` are optimised for efficiency when `x` is a large vector. The main cost in computation of $F(x)$ via `cef` when x is a large vector is a complex vector exponential (for $\mathrm{e}^{\mathrm{i}x^2}$), and the M complex vector multiplications and M additions required to evaluate the polynomial (13) using Horner's algorithm. In comparison, evaluation of $F(x)$ using `F(x,N)` in Table 1 requires 2 complex vector exponentials, and slightly more than N real vector multiplications/divisions, real vector additions, complex vector multiplications, and complex vector additions. From Figures 1 and 2 we read off that to achieve absolute and relative errors below 10^{-8} requires $N = 6$ and

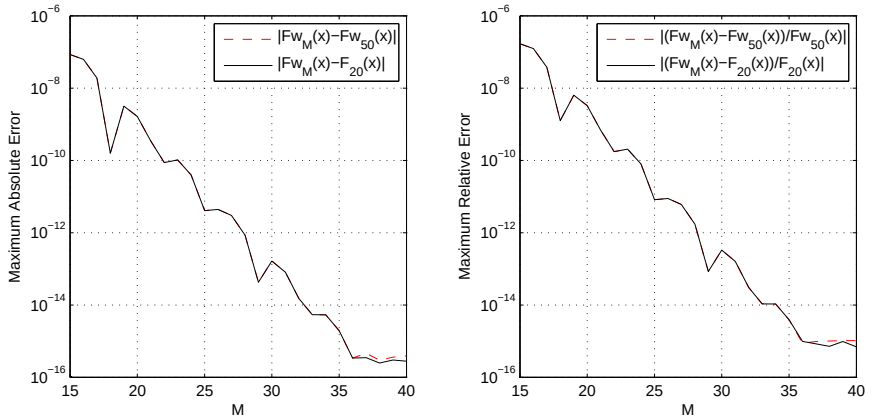

Fig. 2 Left hand side: maximum error, $\max_{x\geq 0}|F(x) - Fw(x)|$, where $Fw(x) := \mathrm{e}^{\mathrm{i}x^2} w_M(\mathrm{e}^{\mathrm{i}\pi/4}x)/2$ with $w_M(z)$ defined in (13). Right hand side: same, but maximum relative error, $\max_{x\geq 0}|(F(x) - Fw(x))/F(x)|$, is plotted against M. In each plot the two curves correspond to different methods for approximating the exact value of $F(x)$, either $F(x) \approx F_{20}(x)$ given by (14) (−), or $F(x) \approx Fw(x)$ with $M = 50$ (−−). (The maximums, as in Figure 1, are taken over 40,000 equally spaced points between 0 and 1,000.)

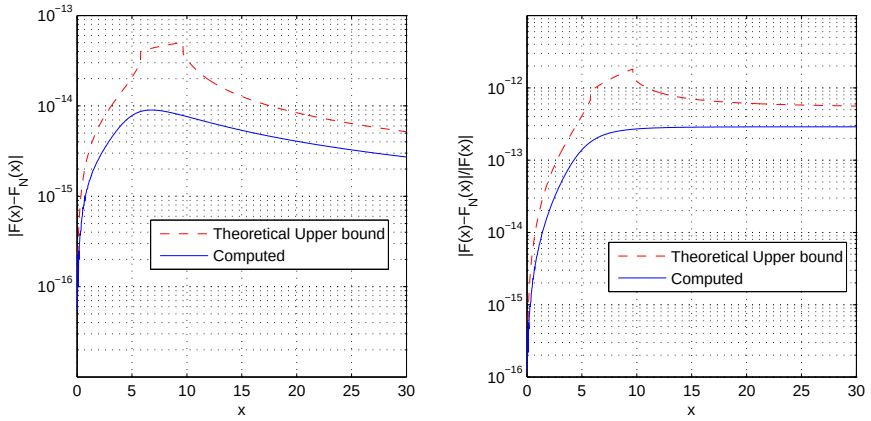

Fig. 3 Left hand side: absolute error, $|F(x) - F_N(x)|$ (−), and its upper bound $\eta_N(x)$ given by (44) (−−), plotted against x. Right hand side: relative error, $|F(x) - F_N(x)|/|F(x)|$ (−), and its upper bound $2(1 + \sqrt{\pi}\,x)\eta_N(x)$ (−−), plotted against x. In both plots $N = 9$ and $F(x)$ is approximated by $F_{20}(x)$.

$M = 18$; to achieve errors below 10^{-15} requires $N = 12$ and $M = 36$. Thus computing $F(x)$ via `F(x,N)` requires a substantially lower operation count than computing via `cef`. (We note, moreover, as discussed in §1.1 and in §7 of [30], that, at least for intermediate values of x ($1.5 \leq x \leq 5$), the operation counts for `cef` are lower than those of the method for $w(z)$ of [23,24].)

To test whether `F(x,N)` is faster we have compared computation times in Matlab (version 7.8.0.347 (R2009a) on a laptop with dual 2.4GHz P8600 Intel processors) between `Fw(x,36)` and `F(x,12)` when `x` is a length 10^7 vector of equally spaced numbers between 0 and 1,000. The average elapsed times were 11.1 and 15.6 seconds, respectively, so that `F(x,12)` is almost 50% faster.

Turning to $C(x)$ and $S(x)$, in Figure 4 we have plotted the maximum values of the absolute and relative errors in $S_N(x)$ and $C_N(x)$, computed using `fresnelCS` in Table 2. As accurate values for $C(x)$ and $S(x)$ we use $C_{20}(x)$ and $S_{20}(x)$ for $x > 1.5$ while, for $0 < x < 1.5$ (following [25]) we approximate by the series (69) truncated after 15 terms, evaluated by the Horner algorithm. Exponential convergence is seen in Figure 4: the absolute errors are $\leq 4.5 \times 10^{-16}$ for $N \geq 11$, the maximum relative error in $C_N(x)$ is $\approx 3.6 \times 10^{-15}$ for $N = 11$ but that in $S_N(x)$ as large as 2.7×10^{-13}. These errors may be entirely acceptable, but the truncated power series (69) must achieve smaller errors for small x and is cheaper to evaluate. (Evaluating at 10^7 equally spaced points between 0 and 1.5 takes 2.9 times longer in Matlab with `fresnelCS` than evaluating 15 terms of both the series (69) via Horner's algorithm.)

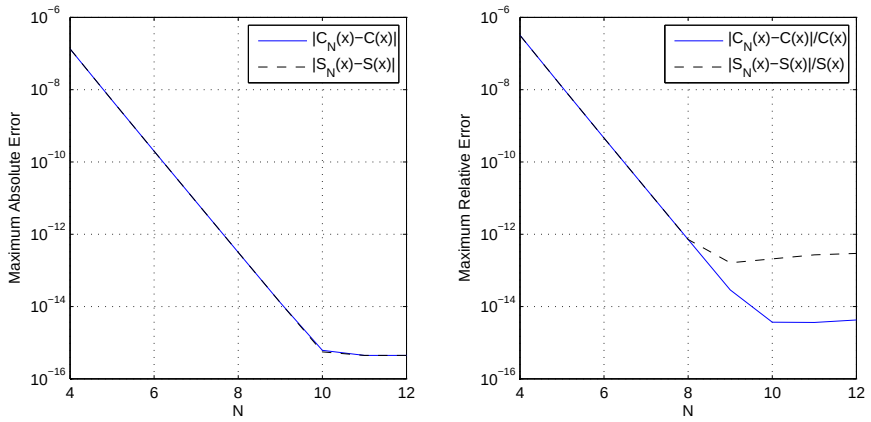

Fig. 4 Left hand side: maximum values of $|C_N(x)-C(x)|$ and $|S_N(x)-S(x)|$ on $0 \leq x \leq 20$. Right hand side: maximum values of $|C_N(x) - C(x)|/C(x)$ and $|S_N(x) - S(x)|/S(x)$ on $0 \leq x \leq 20$.

5 Concluding Remarks

To conclude, we have presented in this paper new approximations for the Fresnel integrals, derived from and inspired by modified trapezium rule approximations previously suggested for the complementary error function of

complex argument in [20,16]. These approximations are simple to implement (Matlab codes are included in Tables 1 and 2): the computation of $F_N(x)$ requires a couple of complex exponentiations and a short summation to compute a quadrature sum, and that of $C_N(x)$ and $S_N(x)$ evaluation of trigonometric and hyperbolic functions and a similar short summation.

Operation counts and timings suggest that $F_N(x)$ with $N = 12$ may be faster than previous methods, at least for intermediate values of $|x|$. In particular, the Matlab function in Table 1 outperforms that in Table 1 of [30] for this application. The code for $S_N(x)$ and $C_N(x)$ is faster still, but the power series (69), truncated after 15 terms, are more accurate and efficient on the interval $[0, 1.5]$, this conclusion endorsing recommendations in [25].

Part of the motivation for this paper was a remark in Weideman [30] regarding the modified trapezium rule methods of [20,16] for computing $\mathrm{erfc}(z)$, that they are "very accurate, provided for given z and N [the finite number of quadrature points retained] the optimal stepsize h is selected. It is not easy, however, to determine this optimal h a priori." At least as far as computing $\mathrm{erfc}(z)$ for $\arg(z) = -\pi/4$ is concerned (which, by (4), is the same as computing $F(x)$) this problem is solved in this paper, so that the effectiveness of the modified trapezium rule methods of [20,16,30] is clearly demonstrated. We hope that the methodology and positive results of this paper will inspire further applications of this truncated, modified trapezium rule method.

We finish by flagging that the modified trapezium rule method that we have used in this paper is applicable widely to the evaluation of integrals on the real line of functions that are analytic but with poles near the real axis. Indeed, general theories of the method are presented in Bialecki [4], Hunter [17] (and see [10], [18, §5.1.4]), and in the thesis of one of the authors [19], where the emphasis is on the particular case (7), where the analytic function $f(t) = O(1)$ as $t \to \pm\infty$. Integrals of the form (7) arise in probabilistic applications [10] and as representations in integral form of solutions to linear PDEs with constant coefficients, after solution by Fourier transform methods and deformation of the path of integration to a steepest descent path. One example which continues to be the subject of computational studies [7,11,22] is the Green's function for the Helmholtz equation $\Delta u + k^2 u = 0$ in a half-space with an impedance boundary condition, $\partial u/\partial n = \mathrm{i}k\beta u$. Representations for this Green's function in terms of a steepest descent path integral of the form (7), in both the 2D and 3D cases, are given in [7], and the application of the truncated modified trapezium rule method is discussed in [19].

Acknowledgements This paper is dedicated to David Hunter, formerly of the University of Bradford, UK, who celebrated his 80th birthday in April 2013. Sadly David passed away on 15 August 2013. David was a kind and gentle man and a fine mathematician and teacher and the second author acknowledges his gratitude for David's contribution to his education as a numerical analyst at Bradford in the 80s. We also acknowledge the very helpful and thorough comments of the two anonymous referees.

References

1. Digital Library of Mathematical Functions. National Institute of Standards and Technology, from http://dlmf.nist.gov/, release date: 2010-05-07 (2010)
2. Arens, T., Sandfort, K., Schmitt, S., Lechleiter, A.: Analysing Ewald's method for the evaluation of Green's functions for periodic media, IMA J. Appl. Math. **78**, 405–431 (2013)
3. Abramowitz, M., Stegun, I. A.: Handbook of Mathematical Functions, Dover, New York (1964)
4. Bialecki, B.: A modified sinc quadrature rule for functions with poles near the arc of integration, BIT **29**, 464–476 (1989)
5. Bowman, J. J., Senior, T. B. A., Uslenghi, P. L. E.: Electromagnetic and Acoustic Scattering by Simple Shapes, North-Holland, Amsterdam (1969)
6. Chandler-Wilde, S. N., Hewett, D. P., Langdon, S., Twigger, A.: A high frequency boundary element method for scattering by a class of nonconvex obstacles. University of Reading, Department of Mathematics and Statistics Preprint MPS-2012-04 (2012)
7. Chandler-Wilde, S.N., Hothersall, D. C.: Efficient calculation of the Green function for acoustic propagation above a homogeneous impedance plane, J. Sound Vib. **180**, 705-724 (1995)
8. Chiarella, C., Reichel, A.: On the evaluation of integrals related to the error function, Math. Comp. **22**, 137–143 (1968)
9. Cody, W. J.: Chebyshev approximations for the Fresnel integrals, Math. Comp. **22**, 450–453 + s1–s18 (1968)
10. Crouch, E. A. C., Spiegelman, D.: The evaluation of integrals of the form $\int_{-\infty}^{+\infty} f(t)\exp(-t^2)dt$: application to logistic-normal models, J. Amer. Stat. Assoc. **85**, 464–469 (1990)
11. Durán, M., Hein, R., Nédélec, J.-C.: Computing numerically the Green's function of the half-plane Helmholtz operator with impedance boundary conditions, Numer. Math. **107**, 295-314 (2007)
12. Fettis, H. E., Caslin, J. C., Cramer, K. R.: Complex zeros of the error function and of the complementary error function, Math. Comp. **27**, 401–407 (1973)
13. Gautschi, W.: Efficient computation of the complex error function, SIAM J. Numer. Anal. **7**, 187–198 (1970)
14. Goodwin, E. T.: The evaluation of integrals of the form $\int_{-\infty}^{\infty} f(x)\mathrm{e}^{-x^2}dx$, Proc. Camb. Phil. Soc. **45**, 241–245 (1949)
15. Heald, M. A.: Rational approximations for the Fresnel integrals, Math. Comp. **44**, 459–461 (1985)
16. Hunter, D. B., Regan, T.: A note on evaluation of the complementary error function, Math. Comp. **26**, 539–541 (1972)
17. Hunter, D. B.: The numerical evaluation of definite integrals affected by singularities near the interval of integration. In: Numerical Integration, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., **357**, pp. 111-120. Kluwer Acad. Publ., Dordrecht (1992)
18. Kythe, P. M., Schäferkotter, M. R.: Handbook of Computational Methods for Integration, Chapman and Hall/CRC, Boca Raton, FL (2005)
19. La Porte, S.: Modified Trapezium Rule Methods for the Effcient Evaluation of Green's Functions in Acoustics. PhD Thesis, Brunel University, UK (2007)
20. Matta, F., Reichel, A.: Uniform computation of the error function and other related functions, J. Math. Phys. **34**, 298–307 (1956)
21. Mori, M.: A method for evaluation of the error function of real and complex variable with high relative accuracy, Publ. RIMS, Kyoto Univ. **19**, 1081–1094 (1983)
22. O'Neil, M., Greengard, L., Pataki, A.: On the efficient representation of the half-space impedance Green's function for the Helmholtz equation, Wave Motion, in press.
23. Poppe, G. P., Wijers, C. M.: More efficient computation of the complex error function, ACM Trans. Math. Software **16**, 38–46 (1990)
24. Poppe, G. P., Wijers, C. M.: Algorithm 680 – Evaluation of the complex error functon, ACM Trans. Math. Software **16**, 47–47 (1990).
25. Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P.: Numerical Recipes 3rd Edition: The Art of Scientific Computing, Cambridge University Press (2007)

26. Rudin, W.: Real and Complex Analysis, 3rd Edition, Mc-Graw Hill (1987)
27. Salzer, H.: Formulas for computing the error function of a complex variable, MTAC **5**, 67–70 (1951)
28. Strand, O.: A method for the computation of the error function of a complex variable, Math. Comp. **19**, 127–129 (1965)
29. Turing, A. M.: A method for the calculation of the zeta-function, Proc. London Math. Soc. **s2-48**, 180–197 (1945)
30. Weideman, J. A. C.: Computation of the complex error function, SIAM J. Numer. Anal. **5**, 1497–1518 (1994)

A Appendix: Bounds on erfc

In this appendix we prove Theorem 4 as a corollary of bounds on erfc in the right hand complex plane contained in Theorem 6 below. In particular (51) follows immediately from (4) and the first bound in (73), while (52) follows from (20), (4), and the second of the bounds (73). The bounds in Theorem 6 are well-known in the case $z \geq 0$ [1, (7.8.2)-(7.8.3)], and the second bound (equivalent by (4) to the bound $|w(z)| \leq 1$ for $\mathrm{Im}(z) \geq 0$) is recently proved by an alternative argument on p. 413 of [2].

Theorem 6 *For $z = x + \mathrm{i}y$ with $x \geq 0$, $y \in \mathbb{R}$, we have that*

$$|\mathrm{erfc}(z)| \geq \frac{\mathrm{e}^{y^2-x^2}}{\sqrt{(1+\sqrt{\pi}\,x)^2+\pi y^2}} \geq \frac{\mathrm{e}^{y^2-x^2}}{1+\sqrt{\pi}\,|z|} \quad \textit{and} \quad |\mathrm{erfc}(z)| \leq \mathrm{e}^{y^2-x^2}. \tag{73}$$

Proof The first of the bounds (73) is equivalent to the bound

$$|\mathcal{G}(z)| \geq 1, \quad \text{for } \mathrm{Re}(z) \geq 0, \tag{74}$$

where $\mathcal{G}(z) = (1+\sqrt{\pi}\,z)\mathrm{e}^{z^2}\mathrm{erfc}(z)$ is an entire function which has the properties that $\mathcal{G}(0) = 1$ and $\mathcal{G}(z) \to 1$ as $|z| \to \infty$ in the right hand plane, uniformly in $\arg(z)$ [3, (7.1.23)]. (These properties imply that the first of the bounds (73) is sharp for $z = 0$ and in the limit $|z| \to \infty$.) We will show (74) by showing that (74) holds for all z in the right hand plane if it holds on the imaginary axis, and then showing that (74) holds on the imaginary axis.

To see that it is enough to prove that (74) holds for imaginary z, observe that, since $\mathrm{erfc}(z)$ has no zeros in the right hand complex plane [28,12] (or on the imaginary axis where $\mathrm{Re}(\mathrm{erfc}(z)) = 1$, see (76)), the function $\mathcal{H}(z) := 1/\mathcal{G}(z)$ is also analytic in the right hand complex plane and is continuous up to the imaginary axis. Moreover, $\mathcal{H}(z)$ is bounded in the right hand plane since, as observed above, $\mathcal{G}(z) \to 1$ as $|z| \to \infty$ in the right hand plane (uniformly in $\arg(z)$). Since $\mathcal{H}(z)$ is bounded in the right hand plane, it follows from the maximum principle that

$$\sup_{\mathrm{Re}(z)\geq 0} |\mathcal{H}(z)| = \sup_{\mathrm{Re}(z)=0} |\mathcal{H}(z)|. \tag{75}$$

To see this, note that this equality holds for $\mathcal{H}_\alpha(z) := 1/\mathcal{G}_\alpha(z)$, with $\alpha > 1$, where $\mathcal{G}_\alpha(z) := (1+\sqrt{\pi}\,z)^\alpha \mathrm{e}^{z^2}\mathrm{erfc}(z)$ with the branch cut taken as the

negative real axis. This is clear since $\mathcal{H}_\alpha(z)$ is analytic in the right half-plane, continuous up to the imaginary axis, and vanishes at infinity, so that the standard maximum principle implies that $\mathcal{H}_\alpha(z)$ takes its maximum value on the imaginary axis. But then (75) follows by taking the limit $\alpha \to 1^+$.

In view of (75), to establish (74) we need only show that it holds for $z = \mathrm{i}y$ with $y \in \mathbb{R}$; indeed, establishing this bound for $y \geq 0$ is sufficient since $\operatorname{erfc}(-\mathrm{i}y) = \overline{\operatorname{erfc}(\mathrm{i}y)}$. Now, for $z = \mathrm{i}y$ with $y \geq 0$, using [1, (7.5.1)], which implies

$$\mathrm{e}^{z^2}\operatorname{erfc}(z) = \mathrm{e}^{-y^2}\left(1 - \frac{2\mathrm{i}}{\sqrt{\pi}}\int_0^y \mathrm{e}^{t^2}\mathrm{d}t\right) \tag{76}$$

we see that

$$\begin{aligned} |\mathcal{G}(\mathrm{i}y)|^2 &= (1+\pi y^2)\mathrm{e}^{-2y^2}\left(1 + \frac{4}{\pi}\left(\int_0^y \mathrm{e}^{t^2}\mathrm{d}t\right)^2\right) \\ &\geq (1+\pi y^2)\mathrm{e}^{-2y^2}\left(1 + \frac{4}{\pi}y^2\right) \\ &= \left(1 + \left(\pi + \frac{4}{\pi}\right)y^2 + 4y^4\right)\mathrm{e}^{-2y^2}. \end{aligned} \tag{77}$$

It is an easy calculus exercise to show the right hand side takes its minimum value on $[0,1]$ at either 0 or 1, and hence that $|\mathcal{G}(\mathrm{i}y)| \geq 1$, for $0 \leq y \leq 1$, since $|\mathcal{G}(\mathrm{i})|^2 > (5+\pi)/\mathrm{e}^2 > 8/2.8^2 > 1$. Further, (77) implies that

$$|\mathcal{G}(\mathrm{i}y)| \geq 2y\mathrm{e}^{-y^2}\int_0^y \mathrm{e}^{t^2}\mathrm{d}t$$

and, for $y \geq 1$, it follows on integrating by parts that

$$\begin{aligned} \int_0^y \mathrm{e}^{t^2}\mathrm{d}t = \int_0^1 \mathrm{e}^{t^2}\mathrm{d}t + \int_1^y e^{t^2}\mathrm{d}t &= \int_0^1 \mathrm{e}^{t^2}\mathrm{d}t + \frac{\mathrm{e}^{y^2}}{2y} - \frac{\mathrm{e}}{2} + \int_1^y \frac{\mathrm{e}^{t^2}}{2t^2}\mathrm{d}t \\ &> \int_0^1 (1 + t^2 + \tfrac{1}{2}t^4)\mathrm{d}t + \frac{\mathrm{e}^{y^2}}{2y} - \frac{\mathrm{e}}{2} > \frac{\mathrm{e}^{y^2}}{2y}, \end{aligned}$$

since $\mathrm{e} < 2.8 < 2(1 + 1/3 + 1/10)$. Thus $|\mathcal{G}(\mathrm{i}y)| \geq 1$ on $[1, \infty)$ and the bound (74) is proved.

Similarly,

$$\sup_{\operatorname{Re}(z)\geq 0} |\mathrm{e}^{-z^2}\operatorname{erfc}(z)| = \sup_{\operatorname{Re}(z)=0} |\mathrm{e}^{-z^2}\operatorname{erfc}(z)| = \sup_{y\geq 0} |\mathrm{e}^{-y^2}\operatorname{erfc}(\mathrm{i}y)|. \tag{78}$$

Further, (76) implies that, for $y \geq 0$,

$$\begin{aligned} |\operatorname{erfc}(\mathrm{i}y)|^2 - 1 = \frac{4}{\pi}\left(\int_0^y \mathrm{e}^{t^2}\mathrm{d}t\right)^2 &= \frac{4y^2}{\pi}\left(\sum_{n=0}^{\infty}\frac{y^{2n}}{n!(2n+1)}\right)^2 \\ &= \frac{2y^2}{\pi}\sum_{n=0}^{\infty} a_n y^{2n} \leq \frac{2}{\pi}\left(\mathrm{e}^{2y^2} - 1\right) \end{aligned}$$

where

$$a_n = \sum_{m=0}^{n} \frac{2}{m!(n-m)!(2m+1)(2(n-m)+1)} \leq \frac{2}{n+1} \sum_{m=0}^{n} \frac{1}{m!(n-m)!} = \frac{2^{n+1}}{(n+1)!}.$$

Thus, for $y \geq 0$,

$$|\mathrm{e}^{-y^2}\mathrm{erfc}(\mathrm{i}y)|^2 \leq \frac{2}{\pi} + \left(1 - \frac{2}{\pi}\right)\mathrm{e}^{-2y^2} \leq 1.$$

Combining this with (78) we see that the second of the bounds (73) holds.