IMA Journal of Applied Mathematics (2025) Page 1 of 31
doi:10.1093/imamat/xxx000

Solution of Wiener-Hopf and Fredholm integral equations by fast Hilbert and Fourier transforms

GUIDO GERMANO
Department of Computer Science, University College London
Systemic Risk Centre, London School of Economics and Political Science
g.germano@ucl.ac.uk

CAROLYN E. PHELAN
Department of Computer Science, University College London
carolyn.phelan.14@ucl.ac.uk (corresponding author)

DANIELE MARAZZINA
Dipartimento di Matematica, Politecnico di Milano
daniele.marazzina@polimi.it

GIANLUCA FUSAI
Bayes Business School, City St Georges, University of London
Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale Amedeo Avogadro, Novara
gianluca.fusai.1@citystgeorges.ac.uk, gianluca.fusai@unipmn.it

[Received on 25 September 2025]

We present numerical methods based on the fast Fourier transform (FFT) to solve convolution integral equations on a semi-infinite interval (Wiener-Hopf equation) or on a finite interval (Fredholm equation). We improve a FFT-based method for the Wiener-Hopf equation due to Henery by expressing it in terms of the Hilbert transform and computing the latter in a more sophisticated way with a sinc function expansion. We further enhance the error convergence using a spectral filter. We then generalise our method to the Fredholm equation by reformulating it as two coupled Wiener-Hopf equations and solving them iteratively. We provide numerical tests and open-source code.

Keywords: Wiener-Hopf, Fredholm, integral equation, fast Fourier transform, fast Hilbert transform.

1. Introduction

We consider the linear integral equation of convolution type with constant limits of integration

$$\lambda f(x) - \int_a^b k(x-x')f(x')dx' = g(x), \quad x \in [a,b], \tag{1.1}$$

where $f(x)$ is the unknown function, $k(x)$ is a given kernel, and $g(x)$ is a given so-called forcing function. The domain of $f(x)$ and $g(x)$ is $[a,b]$, the domain of $k(x)$ is $[a-b,b-a]$; an endpoint is excluded if it is infinite or a function is undefined there. If $a=-\infty$ or $b=+\infty$ Eq. (1.1) is called a Wiener-Hopf equation (Wiener & Hopf, 1931; Noble, 1958; Krein, 1962; Polyanin & Manzhirov, 1998; Lawrie & Abrahams, 2007); if both integration limits are finite, it is called a Fredholm equation (Fredholm, 1903; Whittaker & Watson, 1927; Polyanin & Manzhirov, 1998). The latter case is also called a Wiener-Hopf equation on a

finite interval (Voronin, 2004) or, because of an application in electrotechnics, a longitudinally modified Wiener-Hopf equation (LMWHE), while the former case is also called a classical Wiener-Hopf equation (CWHE) (Daniele & Zich, 2014). If $\lambda = 0$ it is an equation of the first kind; if $\lambda \neq 0$ it is an equation of the second kind. In the latter case it can be assumed that $\lambda = 1$, dividing the kernel and the forcing function by values of this parameter different from 1. Historically these equations arose in physics, e.g. to describe diffraction in the presence of an impenetrable wedge or of planar waveguides (Daniele & Lombardi, 2007), but also for problems in crystal growth, fracture mechanics, flow mechanics (Choi et al., 2005), geophysics, and diffusion (Lawrie & Abrahams, 2007). The connection of the Wiener-Hopf equation with probabilistic problems was noticed by Spitzer (1957) and is discussed by Feller (1971) together with the application of Fourier transform methods to stochastic processes. More recently these equations have become of interest in finance for the pricing of discretely monitored path-dependent options like barrier, first-touch, lookback (or hindsight), quantile and Bermudan options (Fusai et al., 2006; Green et al., 2010; Fusai et al., 2012; Marazzina et al., 2012; Fusai et al., 2016; Phelan et al., 2018, 2019, 2020). The Wiener-Hopf method is employed also to solve a large collection of mixed boundary value problems (Duffy, 2008).

2. Mathematical tools

2.1 *Fourier transform and projection operators*

We define the Fourier transform of a function $f(x)$,

$$\widehat{f}(\xi) = \mathscr{F}_{x\to\xi}[f(x)] := \int_{-\infty}^{+\infty} e^{i\xi x} f(x)dx, \tag{2.1}$$

where i is the imaginary unit, and correspondingly the inverse transform of $\widehat{f}(\xi)$,

$$f(x) = \mathscr{F}^{-1}_{\xi\to x}[\widehat{f}(\xi)] := \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ix\xi}\widehat{f}(\xi)d\xi. \tag{2.2}$$

We choose this definition because it is the one normally used in major application fields of Eq. (1.1), i.e. probability, physics and finance, so that the Fourier transform of the probability density function (PDF) $f_X(x)$ of a random variable X coincides with its characteristic function $\varphi_X(\xi) := E(e^{i\xi X}) = \widehat{f}_X(\xi)$, where E is the expectation. However, it would be better to define the Fourier transform in terms of frequency ν rather than angular frequency or pulsation $\xi = \omega = 2\pi\nu$ (in physics terminology if x is interpreted as time): "We were raised on the ω-convention, but we changed!" (Press et al., 2007, Section 12.0). With ν, the transform is unitary, i.e. the inverse transform is the adjoint of the forward transform, and norm-preserving (Plancherel, 1910), the inverse transform lacks the factor $1/(2\pi)$ making it symmetric with respect to the forward transform, and the Nyquist relation between grids in the normal and Fourier spaces simplifies to $\Delta x\Delta\nu = 1/N$, where N is the number of grid steps, without a factor 2π on the right-hand side. Moreover, a minus sign in the exponent of the forward Fourier transform is consistent with the definition of the Laplace transform. Indeed the Fourier kernel $e^{-i2\pi\nu x}$ is the more common choice in fast Fourier transform (FFT) libraries, including the FFTW Frigo & Johnson (2005) used in MATLAB. Thus an inconvenience is that the Fourier transform $\mathscr{F}$ of Eq. (2.1) with the Fourier kernel $e^{i\xi x}$ translates into `ifft()*N*Dx` in the MATLAB code that we give in the supplementary material, and the inverse transform $\mathscr{F}^{-1}$ into `fft()*Dxi/(2*pi)`.

The function space $L_p(\mathbb{R})$ is the set of functions $f : \mathbb{R} \to \mathbb{C}$ where $|f|^p$ has a finite Lebesgue integral over $\mathbb{R}$. The Fourier transform $\widehat{f}$ naturally exists if $f \in L_1(\mathbb{R})$, i.e. if f is absolutely integrable. Necessary

conditions for $f \in L_1(\mathbb{R})$ are that f vanishes at infinity faster than $1/|x|$ and tends to infinity slower than $1/(x-x_0)$ at any pole x_0. The PDF f_X of a random variable X is in $L_1(\mathbb{R})$, and its non-unitary Fourier transform given by Eq. (2.1) is the characteristic function of X, $\varphi_X(\xi) := E(e^{i\xi X}) = \widehat{f_X}(\xi)$. It does not necessarily follow from $f \in L_1(\mathbb{R})$ that $\widehat{f} \in L_1(\mathbb{R})$ too; the Riemann-Lebesgue lemma just states that $\widehat{f} \in C_0(\mathbb{R})$, the space of continuous functions that vanish at infinity, which is a subset of $L_\infty(\mathbb{R})$. However, if $f \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$, i.e. f is also square-integrable, the Plancherel (1910) theorem states that $\widehat{f} \in L_2(\mathbb{R}) \cap L_\infty(\mathbb{R})$ and the unitary Fourier transform is an isometry. The Fourier transform is then extended to the closure of $L_1(\mathbb{R}) \cap L_2(\mathbb{R})$, which is the whole $L_2(\mathbb{R})$. Because $\mathscr{F} : L_1(\mathbb{R}) \to L_\infty(\mathbb{R})$ and $\mathscr{F} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ are continuous linear maps, the Riesz-Thorin interpolation theorem further extends the Fourier transform to $\mathscr{F} : L_p(\mathbb{R}) \to L_q(\mathbb{R})$ with $1 \leqslant p \leqslant 2$ (Hausdorff-Young inequality) and $1/p + 1/q = 1$ (p, q are Hölder conjugates). Thus, $f \in L_2(\mathbb{R})$ is a sufficient condition for both the forward and the inverse Fourier transforms of f to exist.

Eq. (1.1) is solved in Fourier space, requiring that the Fourier transform $\widehat{k}$ of the kernel exists. In his fundamental work on the Wiener-Hopf equation, i.e. Eq. (1.1) with $a = 0$ and $b = +\infty$, Krein (1962) proved that a sufficient condition for a solution f to exist is that for all $\xi \in \mathbb{R}$, $\widehat{l}(\xi) := \lambda - \widehat{k}(\xi) \neq 0$, and that if and only if the winding number of $\widehat{l}$, called the index of the equation, is 0, for any forcing function $g \in L_p(\mathbb{R}_+)$, $1 \leqslant p \leqslant \infty$, there is a unique solution $f \in L_p(\mathbb{R}_+)$. Similar results exist for other values of a and b. However, to operatively find $\widehat{f}$ requires $1 \leqslant p \leqslant 2$, and to be sure that f can be retrieved, a sufficient condition is $p = 2$. Actually, most solution methods assume $k \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$, and several methods require additional assumptions, e.g. that $k(x)$ decays exponentially for $|x| \to \infty$. Like the seminal paper by Wiener and Hopf Wiener & Hopf (1931), our method assumes only that all functions are in L_2, although our numerical examples are in $L_1 \cap L_2$ like many practically relevant cases, where often the unknown f is a PDF and the kernel k is a transition density, i.e. a conditional PDF. A notable case in $L_2 \setminus L_1$ is the sinc function.

We define the projection of a function $f(x)$ on the positive or negative real half-axis through the multiplication with the indicator function of that set,

$$f_+(x) = \mathscr{P}_{+,x}[f(x)] := 1_{\mathbb{R}_+}(x) f(x) \tag{2.3}$$

$$f_-(x) = \mathscr{P}_{-,x}[f(x)] := 1_{\mathbb{R}_-}(x) f(x). \tag{2.4}$$

A function that, like $f_+(x)$, is 0 for $x < 0$ and nonzero for $x > 0$ is called "causal" because it can be used to describe the effect of something that happens at $x = 0$ and causes the function to become nonzero. The two half-range Fourier transforms are

$$\widehat{f_+}(\xi) = \mathscr{F}_{x\to\xi}[f_+(x)] = \int_0^{+\infty} e^{i\xi x} f(x)\,dx \tag{2.5}$$

$$\widehat{f_-}(\xi) = \mathscr{F}_{x\to\xi}[f_-(x)] = \int_{-\infty}^{0} e^{i\xi x} f(x)\,dx. \tag{2.6}$$

Notice that $\widehat{f_+}(\xi)$ is the Fourier transform of a projected function, while $\widehat{f}_+(\xi)$ is the projection of a Fourier-transformed function,

$$\widehat{f_+}(\xi) = \mathscr{F}_{x\to\xi}[\mathscr{P}_{+,x}[f(x)]] \neq \widehat{f}_+(\xi) = \mathscr{P}_{+,\xi}[\mathscr{F}_{x\to\xi}[f(x)]]. \tag{2.7}$$

In other words, $\widehat{f_+}(\xi)$ is the Fourier transform of a function $f(x)$ that vanishes for negative arguments x, but $\widehat{f_+}(\xi)$ does not vanish itself for negative arguments ξ, which instead happens with $\widehat{f}_+(\xi)$; similarly

for the $-$ case. The function $\widehat{f_+}(\xi)$ is analytic (or holomorphic), i.e. locally given by a convergent power series, in an upper complex half-plane that includes the real line; the function $\widehat{f_-}(\xi)$ is analytic in a lower complex half-plane that includes the real line. The half-range Fourier transforms can be considered special cases of the Laplace transform,

$$\widetilde{f}(s) = \mathscr{L}_{x\to s}[f(x)] = \int_0^{+\infty} e^{-sx} f(x)dx, \quad s \in \mathbb{C}, \tag{2.8}$$

where $s = \pm i\xi$, while the Fourier transform can be considered a special case of the bilateral or two-sided Laplace transform. Except possibly for $x = 0$, the indicator function $1_{\mathbb{R}_+}(x)$ coincides with the Heaviside step function $H(x)$, and $1_{\mathbb{R}_-}(x)$ with $1-H(x)$; $H(x) = 1$ if $x > 0$ and 0 if $x < 0$, while for $x = 0$ it can be assigned the value 0 (left-continuous choice), 1 (right-continuous choice), or 1/2 (symmetric choice). When integrating as in Eqs. (2.5) and (2.6), the value for $x = 0$ matters only numerically and only if $x = 0$ is a grid point, as analytically the measure of a point is zero. Clearly the sum of the two projections, Eqs. (2.3) and (2.4), is the full function,

$$f_+(x) + f_-(x) = f(x), \tag{2.9}$$

and the sum of the two half-range Fourier transforms, Eqs. (2.5) and (2.6), is the full Fourier transform,

$$\widehat{f_+}(\xi) + \widehat{f_-}(\xi) = \widehat{f}(\xi). \tag{2.10}$$

2.2 *Gibbs phenomenon*

As explained in the previous subsection, we numerically implement the forward and inverse Fourier transform using the FFTW library in MATLAB. The ranges of x and ξ cease to be infinite and continuous, and are approximated with grids of size N. The other parameter which defines both grids, which we centre around zero, is the truncation in the x domain $x_{\max}$. The step is $\Delta x = 2x_{\max}/N$ and the x grid is

$$x_n = n\Delta x, \quad n = -\frac{N}{2}, -\frac{N}{2}+1, \ldots, \frac{N}{2}-1. \tag{2.11}$$

The step of the ξ grid is given by the Nyquist relation, $\Delta\xi = 2\pi/(N\Delta x) = \pi/x_{\max}$; the truncation in the ξ domain is $\xi_{\max} = \pi/\Delta x$ and the ξ grid is

$$\xi_m = m\Delta\xi, \quad m = -\frac{N}{2}, -\frac{N}{2}+1, \ldots, \frac{N}{2}-1. \tag{2.12}$$

The discrete forward and inverse Fourier transforms are

$$\widehat{f}(\xi_m, \Delta x, N) = \Delta x \sum_{n=-N/2}^{N/2-1} e^{i\xi_m x_n} f(x_n) \tag{2.13}$$

$$f(x_n, \Delta\xi, N) = \frac{\Delta\xi}{2\pi} \sum_{m=-N/2}^{N/2-1} e^{-ix_n\xi_m} \widehat{f}(\xi_m). \tag{2.14}$$

The truncation of the sums in Eqs. (2.13) and (2.14) causes the Gibbs phenomenon. For a detailed explanation of its effect on the solution to Wiener-Hopf type equations see Phelan et al. (2019). In this case we must consider two main issues: firstly, if the function $f(x)$ has a discontinuity, the truncation

of $\widehat{f}(\xi_m, \Delta x, N)$ causes oscillations in $f(x_n, \Delta\xi, N)$ close to the discontinuity; secondly, the error away from that discontinuity will decay with the grid size N as $|f(x_n) - f(x_n, \Delta\xi, N)| = O(1/N)$.

There have been many different approaches to solve or mitigate the Gibbs phenomenon (Vandeven, 1991; Gottlieb & Shu, 1997; Tadmor & Tanner, 2005; Tadmor, 2007; Ruijter et al., 2015). As in Phelan et al. (2019), we apply a spectral filter in the Fourier domain, specifically the exponential filter of Gottlieb & Shu (1997)

$$\sigma(\eta) = e^{-\vartheta\eta^p}, \tag{2.15}$$

where $p \in \mathbb{N}$ is even and $\eta = \xi/\xi_{\max}$. This function does not strictly meet the usual filter requirements described for example by Vandeven (1991), as it does not go exactly to zero when $|\eta| = 1$, nor do so its derivatives. However, if we select $\vartheta > -\log\varepsilon_{\mathrm{m}}$, where ε_{m} is the machine precision, then the filter coefficients are within computational accuracy of the requirements. Advantages of the exponential filter are its good performance, its simple form, and the order of the filter being equal to the parameter p which is directly input into the filter equation.

We also investigated the use of the Planck taper described in McKechan et al. (2010), which is defined piecewise as

$$\sigma(\eta) := \begin{cases} 0, & \eta \leqslant \eta_1, \quad \eta_1 = -1 \\ \frac{1}{e^{z(\eta)}+1},\ z(\eta) = \frac{\eta_2-\eta_1}{\eta-\eta_1} + \frac{\eta_2-\eta_1}{\eta-\eta_2}, & \eta_1 < \eta < \eta_2,\ \eta_2 = \varepsilon - 1 \\ 1, & \eta_2 \leqslant \eta \leqslant \eta_3,\ \eta_3 = 1-\varepsilon \\ \frac{1}{e^{z(\eta)}+1},\ z(\eta) = \frac{\eta_3-\eta_4}{\eta-\eta_3} + \frac{\eta_3-\eta_4}{\eta-\eta_4}, & \eta_3 < \eta < \eta_4,\ \eta_4 = 1 \\ 0, & \eta \geqslant \eta_4. \end{cases} \tag{2.16}$$

Here, the value of ε gives the proportion of the range of η which is used for the slope regions. In common with the findings by Phelan et al. (2019), the Planck taper, whilst having interesting characteristics such as a flat central section and a filter order of ∞, when tested did not offer any advantage over the exponential filter, so we did not pursue its use any further.

2.3 *Hilbert transform and Wiener-Hopf factorisation*

The Hilbert transform (Pandey, 1996; Vergara Caffarelli & Loreti, 1999; King, 2009) of $\widehat{f}(\xi)$ is the Cauchy principal value of the convolution of $\widehat{f}(\xi)$ with $1/(\pi\xi)$,

$$\begin{aligned} \mathscr{H}_\xi[\widehat{f}(\xi)] &:= \mathrm{p.v.}\,\frac{1}{\pi\xi} * \widehat{f}(\xi) = \mathrm{p.v.}\,\frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{\widehat{f}(\xi')}{\xi-\xi'}d\xi' \\ &= \lim_{\varepsilon\to 0^+} \frac{1}{\pi}\left(\int_{-\infty}^{\xi-\varepsilon} \frac{\widehat{f}(\xi')}{\xi-\xi'}d\xi' + \int_{\xi+\varepsilon}^{+\infty} \frac{\widehat{f}(\xi')}{\xi-\xi'}d\xi'\right). \end{aligned} \tag{2.17}$$

The principal value avoids that the improper integral evaluates to the indefinite form $+\infty - \infty$. The Hilbert transform is well defined for $\widehat{f} \in L_p(\mathbb{R})$, $1 < p < \infty$, and maps to the same space, $\mathscr{H} : L_p(\mathbb{R}) \to L_p(\mathbb{R})$. Because with the above definition the Hilbert transform often appears multiplied by the imaginary unit (see the following equations), some authors such as Stenger (1973) define the Hilbert transform as the principal value of the convolution with $i/(\pi\xi)$. The Hilbert transform is a functional like the Fourier and Laplace transforms; as it maps to the same space, we just write $\mathscr{H}_\xi[\widehat{f}(\xi)]$ instead of

the more cumbersome $\mathscr{H}_{\xi'\to\xi}[f(\xi')]$. For clarity we will keep the subscripts $x\to\xi$, $\xi\to x$ and ξ although they could be omitted when there is no misunderstanding about which variable the operators $\mathscr{F},\mathscr{F}^{-1},\mathscr{P}_+,\mathscr{P}_-$ and $\mathscr{H}$ act on, notably when the argument function depends on a single variable which is always the case here, whereas in applications the argument functions often depend also on time. The operator $i\mathscr{H}_\xi$ is its own inverse,

$$(i\mathscr{H}_\xi)^2[\widehat{f}(\xi)]=\widehat{f}(\xi); \tag{2.18}$$

equivalently, $\mathscr{H}_\xi^{-1}=-\mathscr{H}_\xi$. The convolution theorem

$$\mathscr{F}^{-1}_{\xi\to x}[(\widehat{f}*\widehat{g})(\xi)]=f(x)g(x), \tag{2.19}$$

which maps the convolution to a product via a Fourier transform, together with the inverse Fourier transform (Weisstein, 2025)

$$\text{p.v.}\,\mathscr{F}^{-1}_{\xi\to x}\left[\frac{1}{\pi\xi}\right]=-i\,\text{sgn}\,x \tag{2.20}$$

enables to express the Hilbert transform through an inverse and a forward Fourier transform,

$$i\mathscr{H}_\xi\left[\widehat{f}(\xi)\right]=\mathscr{F}_{x\to\xi}[\text{sgn}(x)f(x)]. \tag{2.21}$$

Thus a fast method to numerically compute the Hilbert transform consists simply in evaluating Eq. (2.21) through an inverse and a forward FFT. In the next subsection, we shall see more sophisticated numerical methods.

Substituting $\text{sgn}\,x=1_{\mathbb{R}_+}(x)-1_{\mathbb{R}_-}(x)$ (this is true also for $x=0$, while $\text{sgn}\,x=2H(x)-1$ is fulfilled for $x=0$ only if $H(0)=1/2$), and applying the definitions of the half-range Fourier transforms, Eqs. (2.5) and (2.6), yields

$$\widehat{f_+}(\xi)-\widehat{f_-}(\xi)=i\mathscr{H}_\xi[\widehat{f}(\xi)]. \tag{2.22}$$

This can be shown also by evaluating the integral in Eq. (2.17) with contour integration methods in the complex plane. Together, Eqs. (2.10) and (2.22) are known as Plemelj-Sokhotsky relations (Pandey, 1996; Vergara Caffarelli & Loreti, 1999; King, 2009). They can be rearranged as

$$\widehat{f_+}(\xi)=\frac{1}{2}\left\{\widehat{f}(\xi)+i\mathscr{H}_\xi[\widehat{f}(\xi)]\right\} \tag{2.23}$$

$$\widehat{f_-}(\xi)=\frac{1}{2}\left\{\widehat{f}(\xi)-i\mathscr{H}_\xi[\widehat{f}(\xi)]\right\} \tag{2.24}$$

or, with a different notation involving the Fourier-transform and projection operators, as

$$\mathscr{F}_{x\to\xi}[\mathscr{P}_{+,x}[f(x)]]=\frac{1}{2}\left\{\mathscr{F}_{x\to\xi}[f(x)]+i\mathscr{H}_\xi[\mathscr{F}_{x\to\xi}[f(x)]]\right\} \tag{2.25}$$

$$\mathscr{F}_{x\to\xi}[\mathscr{P}_{-,x}[f(x)]]=\frac{1}{2}\left\{\mathscr{F}_{x\to\xi}[f(x)]-i\mathscr{H}_\xi[\mathscr{F}_{x\to\xi}[f(x)]]\right\}. \tag{2.26}$$

Substituting $f(x)$ with $\mathscr{P}_+f(x)$ in Eq. (2.25) and $f(x)$ with $\mathscr{P}_-f(x)$ in Eq. (2.26), and taking into account that projection operators are idempotent, i.e., $\mathscr{P}\mathscr{P}f(x)=\mathscr{P}f(x)$, shows that the half-range Fourier transforms are eigenfunctions of the Hilbert transform operator,

$$i\mathscr{H}_\xi[\widehat{f_+}(\xi)]=\quad\widehat{f_+}(\xi) \tag{2.27}$$

$$i\mathscr{H}_\xi[\widehat{f_-}(\xi)]=-\widehat{f_-}(\xi). \tag{2.28}$$

This is evident also by substituting $f(x)$ with $f_+(x)$ or with $f_-(x)$ in Eq. (2.21), or applying the operator $i\mathscr{H}_\xi$ to both sides of Eqs. (2.23)–(2.24) and simplifying with Eq. (2.18). Eqs. (2.27) and (2.28) allow us to obtain Eq. (2.22) by applying the operator $i\mathscr{H}_\xi$ to both sides of Eq. (2.10); conversely, Eq. (2.10) can be reobtained by applying $i\mathscr{H}_\xi$ to both sides of Eq. (2.22). Eqs. (2.23) and (2.24) are invariant with respect to an application of $i\mathscr{H}_\xi$ to both sides.

The key step in the Wiener-Hopf solution of Eq. (1.1) described in the following section is the decomposition of a function $\widehat{f}$, i.e., the reverse of Eq. (2.10),

$$\widehat{f}(\xi) = \widehat{f_+}(\xi) + \widehat{f_-}(\xi). \tag{2.29}$$

The factorisation of a function $\widehat{f}(\xi)$

$$\widehat{f}(\xi) = \widehat{f_+}(\xi)\widehat{f_-}(\xi), \tag{2.30}$$

which is required too, can be reduced to a decomposition by taking logarithms,

$$\log\widehat{f}(\xi) = \log\widehat{f_+}(\xi) + \log\widehat{f_-}(\xi). \tag{2.31}$$

This procedure is called logarithmic decomposition. The decomposition can be achieved by (Rino, 1970; Henery, 1974; Bart et al., 2004)

$$\widehat{f_+}(\xi) = \mathscr{F}_{x\to\xi}[\mathscr{P}_{+,x}[\mathscr{F}^{-1}_{\xi\to x}[\widehat{f}(\xi)]]] \tag{2.32}$$

$$\widehat{f_-}(\xi) = \mathscr{F}_{x\to\xi}[\mathscr{P}_{-,x}[\mathscr{F}^{-1}_{\xi\to x}[\widehat{f}(\xi)]]], \tag{2.33}$$

as can also be seen from the definitions of the half-range Fourier transforms, Eqs. (2.5) and (2.6). For the state of the art of the solution of convolution equations with projection methods in the early 1970s, see also Gohberg & Fel'dman (1974). More in general, the Plemelj-Sokhotsky relations, Eqs. (2.23) and (2.24) can be used (Stenger, 1973): Eqs. (2.32) and (2.33) are a special case of the latter if the Hilbert transform is computed through Eq. (2.21).

The definition of the two half-range Fourier transforms, Eqs. (2.5) and (2.6), can be generalised by splitting the x axis around a constant $a \neq 0$. Feng & Linetsky (2008) showed how the shift theorem,

$$\mathscr{F}_{x\to\xi}[f(x+a)] = e^{-ia\xi}\widehat{f}(\xi), \tag{2.34}$$

can be exploited to generalise the Plemelj-Sokhotsky relations to

$$\widehat{f_+}(\xi) == \frac{1}{2}\left\{\widehat{f}(\xi) + e^{ia\xi} i\mathscr{H}_\xi\left[e^{-ia\xi}\widehat{f}(\xi)\right]\right\} \tag{2.35}$$

$$\widehat{f_-}(\xi) == \frac{1}{2}\left\{\widehat{f}(\xi) - e^{ia\xi} i\mathscr{H}_\xi\left[e^{-ia\xi}\widehat{f}(\xi)\right]\right\}. \tag{2.36}$$

It might be a good idea to write $\widehat{f_{+,a}}$ and $\widehat{f_{-,a}}$ on the left-hand side, but we will avoid it to not overburden the notation, as it will be clear from the context with respect to which parameter a function is decomposed. In the above formulas Eq. (2.21) generalises to

$$e^{ia\xi} i\mathscr{H}_\xi\left[e^{-ia\xi}\widehat{f}(\xi)\right] = e^{ia\xi}\mathscr{F}_{x\to\xi}\left[\operatorname{sgn}(x)f(x+a)\right] \tag{2.37}$$

$$= \mathscr{F}_{x\to\xi}\left[\operatorname{sgn}(x-a)f(x)\right] \tag{2.38}$$

$$= \mathscr{F}_{x\to\xi}\left[(1_{(a,+\infty)}(x) - 1_{(-\infty,a)}(x))f(x)\right]. \tag{2.39}$$

Thus it is easy to show that

$$\lim_{a\to-\infty} e^{ia\xi} i\mathscr{H}_\xi \left[e^{-ia\xi}\widehat{f}(\xi)\right] = \widehat{f}(\xi) \tag{2.40}$$

$$\lim_{a\to+\infty} e^{ia\xi} i\mathscr{H}_\xi \left[e^{-ia\xi}\widehat{f}(\xi)\right] = -\widehat{f}(\xi) \tag{2.41}$$

and that $\lim_{a\to-\infty}\widehat{f_+}(\xi) = \widehat{f}(\xi)$, $\lim_{a\to+\infty}\widehat{f_+}(\xi) = 0$, $\lim_{a\to-\infty}\widehat{f_-}(\xi) = 0$, $\lim_{a\to+\infty}\widehat{f_-}(\xi) = \widehat{f}(\xi)$. These limits are useful to retrieve the results for the classical Wiener-Hopf equation from those for the Fredholm equation.

2.4 *Fast Hilbert transform with sinc functions*

Eq. (2.21) provides a straightforward method to evaluate numerically the Hilbert transform. As it is based on two FFTs, its computational cost is $O(N\log N)$, where N is the number of grid points, and thus is called fast. We compared this method with the quadrature method described in Eqs. (4.19)–(4.20) of King (2009), where the summation is taken over every second point in order to avoid the singularity which results when $x_i - x_j = 0$. We tried various quadrature weights, including Simpson's rule and 3rd and 4th order quadrature (Press et al., 2007, Chapter 4). For our implementation, see the MATLAB functions htq.m and weights.m in the supplementary material. All weights give the same result and have polynomial convergence with N. Therefore, as with quadrature the computation speed is $O(N^2)$, the FFT-based method is preferable because of its higher speed.

An alternative, but equally fast $O(N\log N)$ approach to compute numerically the Hilbert transform is based on the sinc expansion approximation of analytical functions. Sinc functions were deeply studied by Stenger (1993, 2011), who proved that a function $f(z)\in L_2(\mathbb{C})$ analytical in the whole complex plane and of exponential type with parameter π/h, i.e.,

$$|f(z)| \leqslant Ce^{\pi|z|/h}, \quad z\in\mathbb{C}, \tag{2.42}$$

can be reconstructed exactly from the knowledge of its values on an equispaced grid of step h. We consider the latter constraint further down in this section. Defining the sinc functions

$$S_n(z,h) = \frac{\sin(\pi(z-nh)/h)}{\pi(z-nh)/h}, \quad n\in\mathbb{Z}, \tag{2.43}$$

the function f admits the sinc expansion (Stenger, 1993, Theorem 1.10.1)

$$f(z) = \sum_{n=-\infty}^{+\infty} f(nh)S_n(z,h). \tag{2.44}$$

Also its Fourier transform admits the sinc expansion

$$\widehat{f}(\xi) = h\sum_{n=-\infty}^{+\infty} f(nh)e^{i\xi nh} \quad \text{if } |\xi|<\pi/h, \tag{2.45}$$

while it is zero if $|\xi|\geqslant\pi/h$. Moreover, the integrals of f and $|f|^2$ over $\mathbb{R}$ can be written as sums of the coefficients of the sinc expansion of f (Stenger, 1993, Corollary 1.10.2),

$$\int_{-\infty}^{+\infty} f(x)dx = h\sum_{n=-\infty}^{+\infty} f(nh), \quad \int_{-\infty}^{+\infty}|f(x)|^2dx = h\sum_{n=-\infty}^{+\infty}|f(nh)|^2. \tag{2.46}$$

The above results show that the trapezoidal quadrature rule with step size h is exact. Using the following result on the Hilbert transform of the sinc functions (Feng & Linetsky, 2008, Corollary 6.1)

$$\mathscr{H}_\xi[S_n(\xi,h)] = \frac{1-\cos(\pi(\xi-nh)/h)}{\pi(\xi-nh)/h}, \tag{2.47}$$

also the Hilbert transform can be evaluated exactly,

$$\mathscr{H}_\xi[\widehat{f}(\xi)] = \sum_{n=-\infty}^{+\infty} f(nh)\frac{1-\cos(\pi(\xi-nh)/h)}{\pi(\xi-nh)/h}. \tag{2.48}$$

The equality holds for a function f analytic in the whole complex plane. It becomes an approximation for a function analytic only in a strip that includes the real axis; the proof and error bounds were given by Stenger (1993), Chapter 3. Feng & Linetsky (2008) described this approximation for its application in option pricing and gave details of the error for various classes of functions. The following convergence result was proven: if a function is analytic in a suitable strip around the real axis, then the discretisation error of its numerical decomposition or factorisation decays exponentially with respect to the discretisation step h, see Press et al. (2007), Section 4.5.5, Feng & Linetsky (2008), Section 6.3 and Stenger (1993), Theorems 3.1.3, 3.2.1 and 3.1.4.

Now this approximation can be exploited to compute the Hilbert transform with an exponentially decaying discretisation error by combining a FFT with a sinc expansion (Feng & Linetsky, 2008, Section 6.5). The idea is that to compute a discrete Hilbert transform it is necessary to do matrix-vector multiplications involving Toeplitz matrices. These multiplications can be performed exploiting the FFT once those matrices are embedded in a circulant matrix (Fusai et al., 2012; Feng & Linetsky, 2008, Appendix B). Feng & Linetsky (2009), Theorem 3.3, concerns the computation of the Hilbert transform; Feng & Linetsky (2008), Theorem 6.5, and Feng & Linetsky (2009), Theorem 3.4, consider in particular the calculation of the Plemelj-Sokhotsky relations Eqs. (2.23) and (2.24). An implementation of both the $O(N\log N)$ methods presented above is available in the function ifht.m in the supplementary material.

In addition to the discretisation error, an error is caused also by the truncation of the infinite sum in Eq. (2.48) to the number of FFT grid points (Stenger, 1993). This truncation error depends on the shape of the function under transform; its bounds have been explored further by Feng & Linetsky (2008), Section 6.4.2, and Phelan et al. (2019), Section 3. For a function that decays exponentially for $|\xi|\to\infty$, the truncation error converges exponentially. For a function with a polynomial decay, the convergence of the truncation error is only polynomial. In their paper on lookback options, Feng & Linetsky (2009) report a result by Stenger which proves the exponential convergence of the discrete sinc-based fast Hilbert transform to the continuous Hilbert transform. They also examine the truncation error, specifically observing in a footnote that this converges exponentially only under certain conditions, notably $f(x)\leqslant a\exp(-b|x|^c)$ for some $a,b,c>0$. This algorithm can be obtained with an eigenfunction expansion of $\mathscr{H}_\xi$ and is identical to the Kress and Martensen method, which was introduced with a proof that its error converges exponentially (Kress & Martensen, 1970; Weideman, 1995).

We can also revisit the requirement that $\widehat{f}\in L_2(R)$ to show that the approximation is valid for the functions and methodology in this paper. The Fourier transform of a piecewise continuous function is bounded and for $|\xi|\to\infty$ decays at least as $O(1/|\xi|)$ (Boyd, 2001) and so the functions used in our experiments meet this requirement.

3. The classical Wiener-Hopf method for a convolution equation on a semi-infinite interval

Consider Eq. (1.1) with $b = +\infty$. The lower integration limit a can be set to 0 shifting the x scale horizontally by the constant a to a new scale $x' = x - a$; the prime is dropped hereafter. The functions $f(x)$ and $g(x)$, whose domain is $[0, +\infty)$, are extended to the whole real axis defining $f_0(x) = 0$ for $x < 0$, $f_0(x) = f(x)$ for $x \geqslant 0$ and $g_0(x) = 0$ for $x < 0$, $g_0(x) = g(x)$ for $x \geqslant 0$. Define moreover the auxiliary function

$$f_1(x) := \int_0^{\infty} k(x-x')f(x')dx' = \int_{-\infty}^{+\infty} k(x-x')f_0(x')dx', \quad x < 0, \tag{3.1}$$

and $f_1(x) := 0$ for $x \geqslant 0$, i.e., $\widehat{f_1} = \left(\widehat{k}\widehat{f_0}\right)_-$. As f_0 and g_0 are $+$ functions and f_1 is a $-$ function, it is customary to denote these functions f_+, g_+, and f_- respectively. With them Eq. (1.1) is extended to

$$\lambda f_+(x) - \int_{-\infty}^{+\infty} k(x-x')f_+(x')dx' + f_-(x) = g_+(x), \quad x \in \mathbb{R}, \tag{3.2}$$

or, with a more compact notation for the convolution,

$$\lambda f_+(x) - (k * f_+)(x) + f_-(x) = g_+(x), \quad x \in \mathbb{R}. \tag{3.3}$$

The extension of the integration domain to the whole real axis does not affect the equation and its solution on the positive half-axis. Assuming that $\widehat{k}, \widehat{f_+}, \widehat{f_-}, \widehat{g_+}$ exist, we can now apply the convolution theorem, Eq. (2.19), and obtain the equation in Fourier space,

$$\widehat{l}(\xi)\widehat{f_+}(\xi) + \widehat{f_-}(\xi) = \widehat{g_+}(\xi), \quad \xi \in \mathbb{R}, \tag{3.4}$$

where $\widehat{l}(\xi) := \lambda - \widehat{k}(\xi)$; l and $\widehat{l}$ are the functional derivatives of the equation with respect to the solution in normal and Fourier space. Dropping the argument ξ for brevity, factorising $\widehat{l} = \widehat{l_-}\widehat{l_+}$ and dividing the equation by $\widehat{l_-}$ gives

$$\widehat{l_+}\widehat{f_+} + \widehat{l_-}^{-1}\widehat{f_-} = \widehat{l_-}^{-1}\widehat{g_+}. \tag{3.5}$$

This is subject to the condition (Krein, 1962) $\widehat{l} = \lambda - \widehat{k} \neq 0 \Leftrightarrow \int_{-\infty}^{+\infty} k(x)dx < \lambda$. Defining

$$\widehat{c} = \widehat{l_-}^{-1}\widehat{g_+} \tag{3.6}$$

and decomposing it as $\widehat{c} = \widehat{c_+} + \widehat{c_-}$ yields finally

$$\widehat{l_+}\widehat{f_+} + \widehat{l_-}^{-1}\widehat{f_-} = \widehat{c_+} + \widehat{c_-}. \tag{3.7}$$

The $+$ and $-$ components are separated into

$$\widehat{f_+} = \widehat{l_+}^{-1}\widehat{c_+} \tag{3.8}$$

$$\widehat{f_-} = \widehat{l_-}\widehat{c_-}, \tag{3.9}$$

which allows to obtain the sought solution from

$$f_+(x) = \mathscr{F}^{-1}_{\xi \to x}\left[\widehat{l_+}^{-1}(\xi)\widehat{c_+}(\xi)\right], \tag{3.10}$$

while $f_-(x)$ was introduced as an auxiliary function and is not of further interest.

The case with $a=-\infty$ is treated in a similar fashion. The upper integration limit b is set to 0 shifting the x scale horizontally by the constant b to a new scale $x'=x-b$; the prime is dropped hereafter. The functions $f(x)$ and $g(x)$, whose domain is $(-\infty,0]$, are extended to the whole real axis defining $f_0(x)=f(x)$ for $x\leqslant 0$, $f_0(x)=0$ for $x>0$ and $g_0(x)=g(x)$ for $x\leqslant 0$, $g_0(x)=0$ for $x>0$. Define moreover the auxiliary function

$$f_2(x)=\int_{-\infty}^{0}k(x-x')f(x')dx'=\int_{-\infty}^{+\infty}k(x-x')f_0(x')dx',\quad x>0, \tag{3.11}$$

and $f_2(x)=0$ for $x\leqslant 0$, i.e., $\widehat{f_2}=\left(\widehat{k}\widehat{f_0}\right)_+$. Now f_0 and g_0 are $-$ functions and f_2 is a $+$ function, so it is customary to denote these functions f_-, g_-, and f_+ respectively. With them Eq. (1.1) is extended to equations identical to Eqs. (3.2)–(3.10), except that the $+$ and $-$ indices are swapped. In particular, the sought solution is obtained from

$$\widehat{c}(\xi)=\widehat{l_+}^{-1}(\xi)\widehat{g_-}(\xi) \tag{3.12}$$

$$f_-(\xi)=\mathscr{F}^{-1}_{\xi\to x}\left[\widehat{l_-}^{-1}(\xi)\widehat{c_-}(\xi)\right]. \tag{3.13}$$

A more elegant alternative to shifting the x scale forth and back by the constant a or b is to modulate the functions in Fourier space decomposing $\widehat{c}$ with respect to this constant by the generalised Plemelj-Sokhotsky relations, Eqs. (2.35) and (2.36). The function $\widehat{l}$ is always factorised with respect to 0, while $\widehat{c}$ is decomposed with respect to a when $b=+\infty$ or to b when $a=-\infty$. For details, see the function whf_gmf_filt4.m in the supplementary material.

4. Generalisation of the Wiener-Hopf method to a convolution equation on a finite interval: the Fredholm equation

4.1 *Theory*

In the Fredholm equation both integration limits a and b are finite; either a or, less usually, b can be set to 0 shifting the x scale, but, unlike with the classical Wiener-Hopf equation described in the previous section, we prefer not to modify any of the two integration limits; instead, we will use the generalised Plemelj-Sokhotsky relations. The functions $f(x)$ and $g(x)$, whose domain is $[a,b]$, are extended to the whole real axis defining $f_0(x)=f(x)$ for $x\in[a,b]$, $f_0(x)=0$ for $x\notin[a,b]$ and $g_0(x)=g(x)$ for $x\in[a,b]$, $g_0(x)=0$ for $x\notin[a,b]$. The kernel $k(x)$, whose domain is $[a-b,b-a]$, is extended to the whole real axis defining $k_0(x)=k(x)$ for $x\in[a-b,b-a]$ and $k_0(x)=0$ for $x\notin[a-b,b-a]$. Define moreover the two auxiliary functions

$$f_1(x)=\int_{a}^{b}k(x-x')f(x')dx'=\int_{-\infty}^{+\infty}k_0(x-x')f_0(x')dx',\quad x<a, \tag{4.1}$$

$f_1(x)=0$ for $x\geqslant a$, i.e., $\widehat{f_1}=e^{ia\xi}\left(e^{-ia\xi}\widehat{k_0}\widehat{f_0}\right)_-=\widehat{f_-}$,

$$f_2(x)=\int_{a}^{b}k(x-x')f(x')dx'=\int_{-\infty}^{+\infty}k_0(x-x')f_0(x')dx',\quad x>b, \tag{4.2}$$

and $f_2(x)=0$ for $x\leqslant b$, i.e., $\widehat{f_2}=e^{ib\xi}\left(e^{-ib\xi}\widehat{k}\widehat{f_0}\right)_+=\widehat{f_+}$. Because $k_0(x)=0$ for $x\notin[a-b,b-a]$, $f_-(x)=0$ also for $x<a-(b-a)=2a-b$ and $f_+(x)=0$ also for $x>b-(a-b)=2b-a$. Thus

Eq. (1.1) extends to

$$\lambda f_0(x) - \int_{-\infty}^{+\infty} k_0(x-x')f_0(x')dx' + f_-(x) + f_+(x) = g_0(x) \tag{4.3}$$

or, with a more compact notation for the convolution,

$$\lambda f_0(x) - (k_0 * f_0)(x) + f_-(x) + f_+(x) = g_0(x), \tag{4.4}$$

and upon Fourier transformation, setting $\widehat{l}(\xi) = \lambda - \widehat{k_0}(\xi)$,

$$\widehat{l}(\xi)\widehat{f_0}(\xi) + \widehat{f_-}(\xi) + \widehat{f_+}(\xi) = \widehat{g_0}(\xi). \tag{4.5}$$

Eqs. (4.3)–(4.5) look similar to Eqs. (3.2)–(3.4), but now we have two auxiliary functions, $\widehat{f_-}(\xi)$, which is $-$ with respect to any $c \geqslant a$, and $\widehat{f_+}(\xi)$, which is $+$ with respect to any $d \leqslant b$, while both the unknown function $\widehat{f_0}(\xi)$ and the forcing function $\widehat{g_0}(\xi)$ are $+$ with respect to a (or any number $\leqslant a$) and $-$ with respect to b (or any number $\geqslant b$). Therefore the usual approach is to split Eq. (4.5) into two coupled Wiener-Hopf equations, one with the origin shifted to a, the other with the origin shifted to b (Green et al., 2010). These functions typically involve the four redundant unknowns $e^{-ia\xi}\widehat{f_0}(\xi)$, $e^{-ib\xi}\widehat{f_+}(\xi)$ (which are $+$ functions), $e^{-ib\xi}\widehat{f_0}(\xi)$, $e^{-ia\xi}\widehat{f_-}(\xi)$ (which are $-$ functions). In the next subsection, the functions $\widehat{f_-}(\xi)$ and $\widehat{f_+}(\xi)$ correspond to J_- and J_+ from Green et al. (2010), Eq. (2.51), while $\widehat{c}_1$ and $\widehat{c}_2$ correspond to P and Q from Eqs. (2.12) and (2.24) in that paper.

4.2 *Iterative solution*

We solved the system of integral equations described in Eqs. (4.1)–(4.3) iteratively observing that, if we know $\widehat{f_+}(\xi)$ and subtract it from both sides of Eq. (4.5) with the origin of the x axis shifted to a, the result looks like Eq. (3.4), so that we can use the method described in Section 3 to obtain $\widehat{f_-}(\xi)$; similarly, if we know $\widehat{f_-}(\xi)$ and subtract it from both sides of Eq. (4.5) with the origin of the x axis shifted to b, we can use the method described in Section 3 to obtain $\widehat{f_+}(\xi)$. Thus, once again dropping the argument ξ for brevity of notation, our procedure is to write Eq. (4.5) divided once by $\widehat{l_-}$, as in Eq. (3.5), and once by $\widehat{l_+}$,

$$\widehat{l_+}\widehat{f_0} + \widehat{l_-}^{-1}\widehat{f_-} + \widehat{l_-}^{-1}\widehat{f_+} = \widehat{l_-}^{-1}\widehat{g_0} \tag{4.6}$$

$$\widehat{l_-}\widehat{f_0} + \widehat{l_+}^{-1}\widehat{f_-} + \widehat{l_+}^{-1}\widehat{f_+} = \widehat{l_+}^{-1}\widehat{g_0}, \tag{4.7}$$

start from the guess $\widehat{f_+} = 0$ in Eq. (4.6), set

$$\widehat{c_1} = \widehat{l_-}^{-1}(\widehat{g_0} - \widehat{f_+}), \tag{4.8}$$

decompose $\widehat{c_1} = \widehat{c_{1+}} + \widehat{c_{1-}}$ with respect to a, and compute the approximations

$$\widehat{f_0} = \widehat{l_+}^{-1}\widehat{c_{1+}} \tag{4.9}$$

$$\widehat{f_-} = \widehat{l_-}\widehat{c_{1-}} \tag{4.10}$$

as $+$ and $-$ functions with respect to a; then turn to Eq. (4.7), set

$$\widehat{c_2} = \widehat{l_+}^{-1}(\widehat{g_0} - \widehat{f_-}), \tag{4.11}$$

decompose $\widehat{c_2} = \widehat{c_{2+}} + \widehat{c_{2-}}$ with respect to b, and compute the new approximations

$$\widehat{f_0} = \widehat{l_-}^{-1}\widehat{c_{2-}} \tag{4.12}$$

$$\widehat{f_+} = \widehat{l_+}\widehat{c_{2+}} \tag{4.13}$$

as $+$ and $-$ functions with respect to b; and so on until the difference between the values of $\widehat{f_0}$ at an iteration and the previous falls below a threshold. An equivalent result is obtained starting from the guess $\widehat{f_-} = 0$ in Eq. (4.7) and the computation of $\widehat{c_2}$. Notice that the iterations are performed looking for a fixed point on the variables $\widehat{f_-}$ and $\widehat{f_+}$, while $\widehat{f_0}$ is a side product output at each step, but not used to compute the next step. For details, see the function whf_gmf_filt4.m in the supplementary material.

4.3 *Other iterative solutions*

Henery (1977) proposed an iterative solution of the Fredholm equation, but presented only the theory without a numerical validation. In our tests, its literal implementation does not work. The procedure can be mapped to ours including a missing projection and an untold detail: the y_n found in the residual correction scheme are corrections to the solution and thus must be added together. Besides these omissions, Henery (1977) did not express the algorithm in terms of the Hilbert transform and thus used only the simple implementation with the sign function, not the more sophisticated with a sinc function expansion as we did.

Margetis & Choi (2006) presented an iterative solution limited to algebraic kernel functions. Moreover, in the example they implemented which is based on a steady advection-diffusion problem first suggested by Choi et al. (2005), they noted that "this choice of source function and kernel causes fortuitous algebraic simplifications." Therefore this method, whilst interesting as an iterative procedure, cannot be considered to have a general validity.

4.4 *Noble's matrix factorisation approach*

To avoid the iterations, we tried to solve the two simultaneous Wiener-Hopf equations cast in matrix form according to the classic approach of Noble (1958) pp. 153–157; see also Daniele (1984) and Daniele & Zich (2014), Section 1.5.2. We write Eq. (4.5) multiplied once by $e^{-ia\xi}$ and once by $e^{-ib\xi}$ as

$$\begin{pmatrix} \widehat{l} & e^{i(d-a)\xi} \\ 0 & e^{i(d-b)\xi} \end{pmatrix} \begin{pmatrix} e^{-ia\xi}\widehat{f_0} \\ e^{-id\xi}\widehat{f_+} \end{pmatrix} + \begin{pmatrix} 0 & e^{i(c-a)\xi} \\ \widehat{l} & e^{i(c-b)\xi} \end{pmatrix} \begin{pmatrix} e^{-ib\xi}\widehat{f_0} \\ e^{-ic\xi}\widehat{f_-} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \widehat{l} & e^{i(a-b)\xi} \end{pmatrix} \begin{pmatrix} 0 \\ e^{-ia\xi}\widehat{g_0} \end{pmatrix}, \tag{4.14}$$

where $a \leqslant c$ and $d \leqslant b$, as described in Section 4.1. Convenient choices of the parameters c and d are $c = a,\ d = b$; $c = b,\ d = a$; $c = d = a$; $c = d = b$. We choose $c = a,\ d = b$ and write for short

$$\widehat{\mathbf{L}_1}\widehat{\mathbf{f}_+} + \widehat{\mathbf{L}_2}\widehat{\mathbf{f}_-} = \widehat{\mathbf{L}_2}\widehat{\mathbf{g}_+}. \tag{4.15}$$

Multiplying from the left with $\widehat{\mathbf{L}_2}^{-1}$ yields a matrix version of Eq. (3.4),

$$\widehat{\mathbf{L}}\widehat{\mathbf{f}_+} + \widehat{\mathbf{f}_-} = \widehat{\mathbf{g}_+}, \tag{4.16}$$

where

$$\widehat{\mathbf{L}} = \widehat{\mathbf{L}_2}^{-1}\widehat{\mathbf{L}_1} = \widehat{l}^{-1}\begin{pmatrix} -e^{i(a-b)\xi} & 1 \\ \widehat{l} & 0 \end{pmatrix}\begin{pmatrix} \widehat{l} & e^{i(b-a)\xi} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -e^{i(a-b)\xi} & 0 \\ \widehat{l} & e^{i(b-a)\xi} \end{pmatrix} \tag{4.17}$$

is a triangular matrix. Swapping the elements of $\widehat{\mathbf{f}_+}$ and $\widehat{\mathbf{f}_-}$ permutes the elements of $\widehat{\mathbf{L}}$. If we knew how to factorise $\widehat{\mathbf{L}} = \widehat{\mathbf{L}_-}\,\widehat{\mathbf{L}_+}$, multiplying Eq. (4.16) from the left with $\widehat{\mathbf{L}_-}^{-1}$ would lead finally to

$$\widehat{\mathbf{L}_+}\,\widehat{\mathbf{f}_+} + \widehat{\mathbf{L}_-}^{-1}\widehat{\mathbf{f}_-} = \widehat{\mathbf{L}_-}^{-1}\widehat{\mathbf{g}_+}, \tag{4.18}$$

which is a matrix version of Eq. (3.5) and is solved in a similar fashion decomposing its right-hand side. The same result is obtained multiplying Eq. (4.15) from the left with $\widehat{\mathbf{L}_1}^{-1}$ or Eq. (4.16) from the left with $\widehat{\mathbf{L}}^{-1}$, yielding

$$\widehat{\mathbf{f}_+} + \widehat{\mathbf{L}}^{-1}\widehat{\mathbf{f}_-} = \widehat{\mathbf{L}}^{-1}\widehat{\mathbf{g}_+}, \tag{4.19}$$

where

$$\widehat{\mathbf{L}}^{-1} = \widehat{\mathbf{L}_1}^{-1}\widehat{\mathbf{L}_2} = \widehat{l}^{-1}\begin{pmatrix} 1 & -e^{i(b-a)\xi} \\ 0 & \widehat{l} \end{pmatrix}\begin{pmatrix} 0 & 1 \\ \widehat{l} & e^{i(a-b)\xi} \end{pmatrix} = \begin{pmatrix} -e^{i(b-a)\xi} & 0 \\ \widehat{l} & e^{i(a-b)\xi} \end{pmatrix} \tag{4.20}$$

If we knew how to factorise $\widehat{\mathbf{L}}^{-1} = \widehat{\mathbf{L}_+}^{-1}\widehat{\mathbf{L}_-}^{-1}$, multiplying Eq. (4.19) from the left with $\widehat{\mathbf{L}_+}$ would lead again to Eq. (4.18).

The matrix $\widehat{\mathbf{L}}$ does not have a commutative factorisation because $\widehat{\mathbf{L}}(\xi)\widehat{\mathbf{L}}(\xi') \neq \widehat{\mathbf{L}}(\xi')\widehat{\mathbf{L}}(\xi)$: this condition is fulfilled by the elements $\widehat{L}_{11}$, $\widehat{L}_{12}$, $\widehat{L}_{22}$, but not by $\widehat{L}_{21}$. A formula to factorise triangular 2×2 matrices due to Jones (1984), Eq. (21), and Jones (1991), Eq. (6), cannot be applied because the oscillatory elements of $\widehat{\mathbf{L}}$ do not fulfil the required condition that $+$ or $-$ factors remain $+$ or $-$ when inverted: the inverse of the $+$ function $e^{i(b-a)\xi}$ is the $-$ function $e^{i(a-b)\xi}$; see also Daniele & Zich (2014), Section 4.3, Example 2. Feldman et al. (2000) proposed a factorisation of a matrix $\mathbf{G}$ which coincides with our matrix $\widehat{\mathbf{L}}$ except for the sign of the element $\widehat{L}_{11}$, but it requires the factorisation of another matrix $\mathbf{A}$ built from the elements of $\mathbf{G}$ which is not straightforward.

4.5 *Voronin's matrix factorisation approach*

Voronin (2004) proposed a different matrix form of the two simultaneous Wiener-Hopf equations. We present it with slight modifications. Start from Eq. (4.5) and decompose the kernel, $\widehat{k_0} = \widehat{k_-} + \widehat{k_+}$ (for simplicity, we drop the 0 subscript from $\widehat{k_{0-}}, \widehat{k_{0+}}$), obtaining

$$(\lambda - \widehat{k_-} - \widehat{k_+})\widehat{f_0} + \widehat{f_-} + \widehat{f_+} = \widehat{g}_0. \tag{4.21}$$

Multiply by $e^{-ia\xi}$, take the $+$ part, thus eliminating $\widehat{f_-}$, which is a $-$ function with respect to a, and multiply by $e^{ia\xi}$, yielding

$$(\lambda - \widehat{k_+})\widehat{f_0} - e^{ia\xi}(e^{-ia\xi}\widehat{k_-}\widehat{f_0})_+ + \widehat{f_+} = \widehat{g}_0. \tag{4.22}$$

Multiply by $e^{-ib\xi}$, take the $-$ part, thus eliminating $\widehat{f_+}$, which is a $+$ function with respect to b, and multiply by $e^{ib\xi}$, yielding

$$\lambda\widehat{f_0} - e^{ia\xi}(e^{-ia\xi}\widehat{k_-}\widehat{f_0})_+ - e^{ib\xi}(e^{-ib\xi}\widehat{k_+}\widehat{f_0})_- = \widehat{g}_0. \tag{4.23}$$

Define $\widehat{\varphi_1} = \big(\widehat{k_-} + \frac{1-\lambda}{2}\big)\widehat{f_0}$ and decompose it with respect to a, $\widehat{\varphi_1} = \widehat{\varphi_{1+}} + \widehat{\varphi_{1-}}$; define $\widehat{\varphi_2} = \big(\widehat{k_+} + \frac{1-\lambda}{2}\big)\widehat{f_0}$ and decompose it with respect to b, $\widehat{\varphi_2} = \widehat{\varphi_{2+}} + \widehat{\varphi_{2-}}$; this gives

$$\widehat{f_0} - \widehat{\varphi_{1+}} - \widehat{\varphi_{2-}} = \widehat{g_0}. \tag{4.24}$$

The two coupled Wiener-Hopf equations are now obtained multiplying once by $\widehat{k_-}e^{-ia\xi}$ and once by $\widehat{k_+}e^{-ib\xi}$,

$$\begin{pmatrix} 1-\widehat{k_-} & 0 \\ -e^{i(a-b)\xi}\widehat{k_+} & 1 \end{pmatrix}\begin{pmatrix} e^{-ia\xi}\widehat{\varphi_{1+}} \\ e^{-ib\xi}\widehat{\varphi_{2+}} \end{pmatrix} + \begin{pmatrix} 1 & -e^{i(b-a)\xi}\widehat{k_-} \\ 0 & 1-\widehat{k_+} \end{pmatrix}\begin{pmatrix} e^{-ia\xi}\widehat{\varphi_{1-}} \\ e^{-ib\xi}\widehat{\varphi_{2-}} \end{pmatrix} = \begin{pmatrix} e^{-ia\xi}\widehat{k_-}\widehat{g_0} \\ e^{-ib\xi}\widehat{k_+}\widehat{g_0} \end{pmatrix}, \tag{4.25}$$

for short

$$\widehat{\mathbf{M}_{\mathrm{r}-}}^{-1}\widehat{\boldsymbol{\varphi}_+} + \widehat{\mathbf{M}_{\mathrm{r}+}}\,\widehat{\boldsymbol{\varphi}_-} = \widehat{\mathbf{g}}. \tag{4.26}$$

Here one can see that the parameter λ has been inserted in the definition of $\widehat{\boldsymbol{\varphi}_1}$ and $\widehat{\boldsymbol{\varphi}_2}$ to avoid that it appears in place of the numbers 1 in the diagonal elements of $\widehat{\mathbf{M}_{\mathrm{r}-}}^{-1}$ and $\widehat{\mathbf{M}_{\mathrm{r}+}}$, which would make these matrices singular for $\lambda = 0$. Multiplying from the left by $\widehat{\mathbf{M}_{\mathrm{r}-}}$ yields

$$\widehat{\boldsymbol{\varphi}_+} + \widehat{\mathbf{M}}\,\widehat{\boldsymbol{\varphi}_-} = \widehat{\mathbf{M}_{\mathrm{r}-}}\,\widehat{\mathbf{g}}, \tag{4.27}$$

where

$$\begin{aligned} \widehat{\mathbf{M}} &= \widehat{\mathbf{M}_{\mathrm{r}-}}\widehat{\mathbf{M}_{\mathrm{r}+}} \\ &= \frac{1}{1-\widehat{k_-}}\begin{pmatrix} 1 & 0 \\ e^{i(a-b)\xi}\widehat{k_+} & 1-\widehat{k_-} \end{pmatrix}\begin{pmatrix} 1 & -e^{i(b-a)\xi}\widehat{k_-} \\ 0 & 1-\widehat{k_+} \end{pmatrix} \\ &= \frac{1}{1-\widehat{k_-}}\begin{pmatrix} 1 & -e^{i(b-a)\xi}\widehat{k_-} \\ e^{i(a-b)\xi}\widehat{k_+} & 1-\widehat{k} \end{pmatrix}. \end{aligned} \tag{4.28}$$

An equivalent result is obtained multiplying Eq. (4.26) from the left by $\widehat{\mathbf{M}_{\mathrm{r}+}}^{-1}$. If we knew how to factorise $\widehat{\mathbf{M}} = \widehat{\mathbf{M}_{\mathrm{l}+}}\widehat{\mathbf{M}_{\mathrm{l}-}}$, multiplying Eq. (4.27) from the left by $\widehat{\mathbf{M}_{\mathrm{l}+}}^{-1}$ would lead finally to

$$\widehat{\mathbf{M}_{\mathrm{l}+}}^{-1}\widehat{\boldsymbol{\varphi}_+} + \widehat{\mathbf{M}_{\mathrm{l}-}}\,\widehat{\boldsymbol{\varphi}_-} = \widehat{\mathbf{M}_{\mathrm{l}+}}^{-1}\widehat{\mathbf{M}_{\mathrm{r}-}}\,\widehat{\mathbf{g}}, \tag{4.29}$$

which, like Eq. (4.18), is a matrix version of Eq. (3.5) and is solved in a similar fashion decomposing its right-hand side.

Unfortunately we are stuck again: though formulas to convert left $(+-)$ factorisations of 2×2 matrices into right $(-+)$ ones and vice versa have been published by Jones (1991), Eqs. (8)–(11), their application to obtain $\widehat{\mathbf{M}_{\mathrm{l}+}}\widehat{\mathbf{M}_{\mathrm{l}-}}$ from $\widehat{\mathbf{M}_{\mathrm{r}-}}\widehat{\mathbf{M}_{\mathrm{r}+}}$ given by Eq. (4.28) is not straightforward.

4.6 *Iterative solution based on Voronin's approach*

An iterative solution is possible also with Voronin's approach. Write Eq. (4.24) multiplied once by $\widehat{k_-}/(1-\widehat{k_-})$ and once by $\widehat{k_+}/(1-\widehat{k_+})$,

$$\widehat{\varphi_{1+}} - \frac{1}{1-\widehat{k_-}}\widehat{\varphi_{1-}} - \frac{\widehat{k_-}}{1-\widehat{k_-}}\widehat{\varphi_{2-}} = \frac{\widehat{k_-}}{1-\widehat{k_-}}\widehat{g_0} \tag{4.30}$$

$$\widehat{\varphi_{2-}} - \frac{1}{1-\widehat{k_+}}\widehat{\varphi_{2+}} - \frac{\widehat{k_+}}{1-\widehat{k_+}}\widehat{\varphi_{1+}} = \frac{\widehat{k_+}}{1-\widehat{k_+}}\widehat{g_0}, \tag{4.31}$$

in Eq. (4.30) set

$$\widehat{c_1} = \frac{\widehat{k_-}}{1-\widehat{k_-}}(\widehat{g_0} + \widehat{\varphi_{2-}}), \tag{4.32}$$

start from the guess $\widehat{\varphi_{2-}} = 0$, decompose $\widehat{c_1} = \widehat{c_{1+}} + \widehat{c_{1-}}$ with respect to a, and compute the approximations

$$\widehat{\varphi_{1+}} = \widehat{c_{1+}} \tag{4.33}$$

$$\widehat{\varphi_{1-}} = (1-\widehat{k_-})\widehat{c_{1-}} \tag{4.34}$$

as $+$ and $-$ functions with respect to a, as well as

$$\widehat{f_0} = \frac{1}{\widehat{k_-} + \frac{1-\lambda}{2}}(\widehat{\varphi_{1+}} + \widehat{\varphi_{1-}}) = \frac{1}{\widehat{k_-} + \frac{1-\lambda}{2}}\left(\widehat{c_{1+}} + (1-\widehat{k_-})\widehat{c_{1-}}\right); \tag{4.35}$$

then turn to Eq. (4.31), set

$$\widehat{c_2} = \frac{\widehat{k_+}}{1-\widehat{k_+}}(\widehat{g_0} - \widehat{\varphi_{1+}}), \tag{4.36}$$

decompose $\widehat{c_2} = \widehat{c_{2+}} + \widehat{c_{2-}}$ with respect to b, and compute the new approximations

$$\widehat{\varphi_{2-}} = \widehat{c_{2-}} \tag{4.37}$$

$$\widehat{\varphi_{2+}} = (1-\widehat{k_+})\widehat{c_{2+}} \tag{4.38}$$

as $+$ and $-$ functions with respect to b, as well as

$$\widehat{f_0} = \frac{1}{\widehat{k_+} + \frac{1-\lambda}{2}}(\widehat{\varphi_{2-}} + \widehat{\varphi_{2+}}) = \frac{1}{\widehat{k_+} + \frac{1-\lambda}{2}}\left(\widehat{c_{2-}} + (1-\widehat{k_+})\widehat{c_{2+}}\right). \tag{4.39}$$

This is repeated until the difference between the values of $\widehat{f_0}$ at an iteration and the previous falls below a threshold. An equivalent result is obtained starting from the guess $\widehat{\varphi_{1+}} = 0$ in Eq. (4.31) and the computation of $\widehat{c_2}$. Similarly to Section 4 4.2, the iterations are performed looking for a fixed point on the variables $\widehat{\varphi_1}$ and $\widehat{\varphi_2}$, while $\widehat{f_0}$ is a side product output at each step, but not used to compute the next step. For details, see the function whf_gmf_v.m in the supplementary material.

Work on factorisation has found methods for particular matrix classes, often approximate and corresponding to specific applications. For example, after presenting an approximate solution of the scalar Wiener-Hopf equation based on the approximation of the kernel with a rational function (Kisil, 2013), Kisil (2015) has developed and analysed an approximate factorisation approach for Daniele-Khrapkov matrices; although several matrix classes can be reduced to them, this is not a general solution and does not include our case. Kisil also introduced an iterative Wiener-Hopf method for triangular matrices with exponential factors (Kisil, 2018), but our aim would be a general direct (i.e. non-iterative) numerical method. Rogosin & Mishuris (2016) and Kisil et al. (2021) have reviewed constructive analytical and numerical methods for the factorisation of several specific classes of matrices. These comprehensive reviews confirm that there is no general factorisation approach which can be used for all matrices, in particular no direct one.

5. Tests

As we present a general solution to the Fredholm equation, rather than one limited to a particular application, we provide several test cases to solve for f Eq. (1.1) with $\lambda = 1$. Although the methods developped herein can be applied to both Fredholm and Wiener-Hopf equations, for the numerical tests we have chosen to concentrate on solving examples of the Fredholm equation as it is the more challenging case and encompasses Wiener-Hopf as a special example when $a \to -\infty$ or $b \to +\infty$.

Solutions to Eq. (1.1) with simple closed-form expressions for $f(x)$, $g(x)$ and $k(x)$ are not readily available. However, if for simplicity we limit the requirement to $f(x)$ and $k(x)$, then closed-form expressions for $g(x)$ in Eq. (1.1) can be calculated. These $g(x)$ and $k(x)$ are used as inputs to our numerical method, whose accuracy is measured by comparing the result with $f(x)$. We selected $f(x)$ and $k(x)$ to give closed-form expressions for $g(x)$ and also to have Fourier transforms which are easily calculable. We derived three solutions, where both $f(x)$ and $k(x)$ are: 1. Gaussian (normal), 2. Cauchy (Lorentz), 3. Laplace (bilateral exponential). As discussed in Section 2.2, the decay of the functions as $x, \xi \to \infty$ can influence the error performance of Fourier-based methods. The functions were therefore selected to be exponentially decaying in both the state space and Fourier space (Gaussian), to be polynomially decaying in the state space and exponentially decaying in the Fourier space (Cauchy) and to be exponentially decaying in the state space and polynomially decaying in the Fourier space (Laplace). The derivation of $g(x)$ is described in the following sections.

5.1 *Gaussian*

We set $f(x) = k(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$. The expression for $g(x)$ is then

$$\begin{aligned} g(x) &= \frac{1}{\sqrt{\pi}} e^{-x^2} - \frac{1}{\pi} \int_a^b e^{-(x-y)^2} e^{-y^2} dy \\ &= \frac{1}{\sqrt{\pi}} e^{-x^2} - \frac{1}{\pi} e^{-\frac{x^2}{2}} \int_a^b e^{-2(y-\frac{x}{2})^2} dy \\ &= \frac{1}{\sqrt{\pi}} e^{-x^2} - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left[\Phi(2b-x) - \Phi(2a-x) \right], \quad x \in [a,b], \end{aligned} \tag{5.1}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

5.2 *Cauchy*

We set $f(x) = k(x) = \frac{1}{\pi(x^2+1)}$. The first step in the calculation of $g(x)$ is to solve the integral

$$g_{\text{int}}(x) = \frac{1}{\pi^2} \int_a^b \frac{1}{y^2+1} \frac{1}{(x-y)^2+1} dy. \tag{5.2}$$

Using partial fractions,

$$\begin{aligned}
g_{\text{int}}(x) =& \frac{1}{\pi^2}\int_a^b \frac{1}{y^2+1}\frac{1}{(x-y)^2+1}dy \\
=& \frac{1}{\pi^2 x(x^2+4)}\int_a^b \left(\frac{2y}{y^2+1}+\frac{x}{y^2+1}-\frac{2(y-x)}{(y-x)^2+1}+\frac{x}{(y-x)^2+1}\right)dy \\
=& \frac{1}{\pi^2 x(x^2+4)}\left[\log(y^2+1)+x\arctan(y)-\log[(x-y)^2+1]+x\arctan(y-x)\right]_a^b \\
=& \frac{1}{\pi^2 x(x^2+4)}\left\{\log\frac{(b^2+1)((a-x)^2+1)}{(a^2+1)((b-x)^2+1)}+\right. \\
& \left.+x\left[\arctan(b)-\arctan(a)+\arctan(b-x)-\arctan(a-x)\right]\right\}. \qquad (5.3)
\end{aligned}$$

This gives $g(x)$ in closed form,

$$\begin{aligned}
g(x) =& \frac{1}{\pi(x^2+1)}-\frac{1}{\pi^2 x(x^2+4)}\left\{\log\frac{(b^2+1)((a-x)^2+1)}{(a^2+1)((b-x)^2+1)}+\right. \\
& \left.+x\left[\arctan(b)-\arctan(a)+\arctan(b-x)-\arctan(a-x)\right]\right\}, \quad x\in[a,b]. \qquad (5.4)
\end{aligned}$$

5.3 *Laplace*

We set $f(x)=k(x)=\frac{1}{2}e^{-|x|}$. In order to make the calculation of $g(x)$ simpler, the values of a and b are restricted so that $0<a<b<\infty$. Then the formula for $g(x)$ in closed form is

$$\begin{aligned}
g(x) &= \frac{1}{2}e^{-x}-\frac{1}{4}\int_a^b e^{-|x-y|}e^{-y}dy \\
&= \frac{1}{2}e^{-x}-\frac{1}{4}\left[\int_x^b e^{(x-y)}e^{-y}dy+\int_a^x e^{-(x-y)}e^{-y}dy\right] \\
&= \frac{1}{2}e^{-x}-\frac{1}{4}e^{x}\left(\int_x^b e^{-2y}dy+e^{-x}\int_a^x dy\right) \\
&= \frac{1}{2}e^{-x}+\frac{1}{8}e^{x}\left[e^{-2y}\right]_x^b-\frac{1}{4}e^{-x}\left[y\right]_a^x \\
&= \frac{1}{2}e^{-x}+\frac{1}{8}e^{x}\left(e^{-2b}-e^{-2x}\right)-\frac{1}{4}e^{-x}(x-a) \\
&= e^{-x}\left[\frac{3}{8}+\frac{1}{8}e^{-2(b-x)}+\frac{1}{4}(a-x)\right], \quad x\in[a,b]. \qquad (5.5)
\end{aligned}$$

6. Results

We use the following methods to recover $f(x)$ and produce the detailed results shown in this section:

1. 4th order Newton-Cotes quadrature King (2009); Press et al. (2007) with preconditioner (Fusai et al., 2012); see the MATLAB functions quadrature.m and weights.m in the supplementary material.

2. Wiener-Hopf method using the sinc-based fast Hilbert transform with no zero padding. In order to counteract the oscillations on the recovered function, we used an exponential filter of order

8 on the final stage of the fixed-point algorithm. The maximum number of iterations of the fixed-point algorithm is set to 5. In fact, in most cases the final error level is achieved within 3 iterations. We discuss the use of the sinc-based fast Hilbert transform and a spectral filter in Subsection 6.1 (6.1.1) below.

3. Wiener-Hopf method using the symmetric sign function in the fast Hilbert transform, i.e. with zeros placed at both $\xi = 0$ and $\xi = \xi_{\min} = -\frac{N}{2}\Delta\xi$, similar to the method introduced by Rino (1970) and Henery (1974) and tested by Fusai et al. (2016).

4. Wiener-Hopf method with Voronin's variant using the symmetrical sign function for the Hilbert transform.

6.1 *Results for the Gaussian test case*

We first examine the performance of the different numerical methods with the Gaussian test case, with particular emphasis on the method used to implement the Hilbert transform.

6.1.1 *Sinc-based fast Hilbert transform and spectral filtering* In the financial pricing applications described by Feng & Linetsky (2008) and Fusai et al. (2016), the sinc-based fast Hilbert transform has shown excellent error convergence, especially when combined with a spectral filter as in Phelan et al. (2019). However, when we consider its use for this application we must take account of several ways in which the requirements differ from its general use for finding solutions to Wiener-Hopf or Fredholm equations.

Firstly the pricing methods that were implemented using the Wiener-Hopf method in Fusai et al. (2016), as devised by Green et al. (2010), use the analytic continuation of x, i.e. they give results for values of x both inside and outside the barriers (the integration limits of Fredholm equation). This means that for these applications there is no requirement to truncate the functions to the integration limits of a and b in the state space, unlike the problems presented as examples in this paper. The requirement to truncate the function means that there is a jump discontinuity introduced in the state space, meaning that, as described by Boyd (2001), the function in the Fourier space decays as a first order polynomial due to the Gibbs phenomenon. As explained in Stenger (1993) this polynomial decay means that the sinc-based fast Hilbert transform no longer obtains an error which is exponentially convergent with grid size but rather converges polynomially. This is in contrast with the aforementioned finance-based papers from Green et al. (2010) and Fusai et al. (2016) where the Fourier domain functions subject to the sinc-based fast Hilbert transform are exponentially decaying (or polynomially so in the case of the VG process) and so excellent error performance is achieved, especially in conjunction with a spectral filter to solve the issue with the fixed-point algorithm.

In contrast, here we solve the Fredholm and Wiener-Hopf equations as they were originally formulated, i.e. the function is only defined for the range of the integration $[a,b]$ and therefore the functions $g(x)$ and $k(x)$ must be truncated to the ranges $[a,b]$ and $[a-b,b-a]$ respectively. This truncation will introduce a jump in the functions which means that their Fourier transforms now have first order polynomial decay. Therefore the truncation error from the Hilbert transform will have a first order polynomial convergence unless we can exploit some symmetry between the Fourier domain functions for positive and negative ξ as in Phelan et al. (2018), in which case we may achieve second order polynomial convergence.

Moreover, there is a second important distinction to be made between the general solution presented here and the work in the above literature. In the finance literature the solutions to the Fredholm equation

are used to calculate the expectation of a further function, in this case the payoff function. Therefore the exact errors in the function for individual values of x are not particularly important. Rather, the finance literature is concerned with the average error, weighted according to the shape of the payoff function. This also has particular importance when we are considering the use of the sinc-based fast Hilbert transform described in Section 2 2.4 which was instrumental in achieving exponential error convergence with the number of FFT grid points N in Feng & Linetsky (2008) and Fusai et al. (2016).

In Figures 1 and 2, we show results using the sinc-based fast Hilbert transform with no filter for the Gaussian test case described in Section 5.1. It is immediately obvious that, even for high values of N, oscillations are visible in the numerical solution.

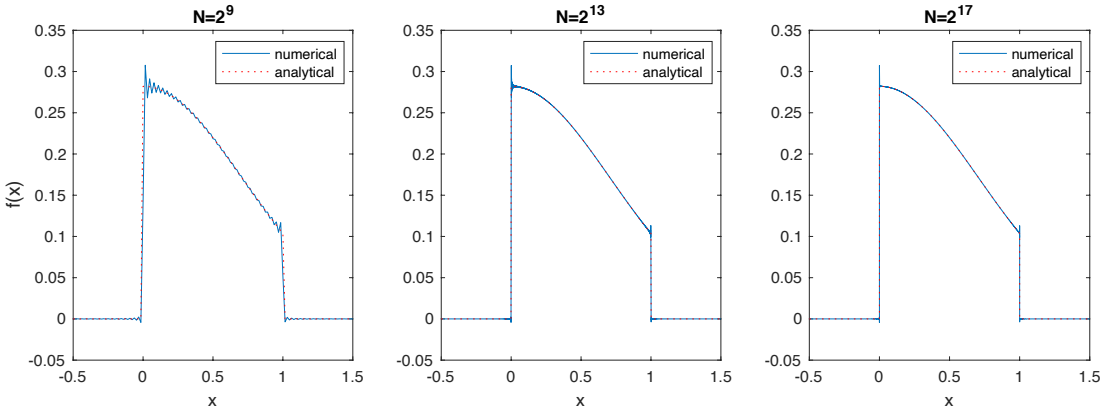

FIG. 1. Numerical and analytical $f(x)$ using the sinc-based fast Hilbert transform with no filter.

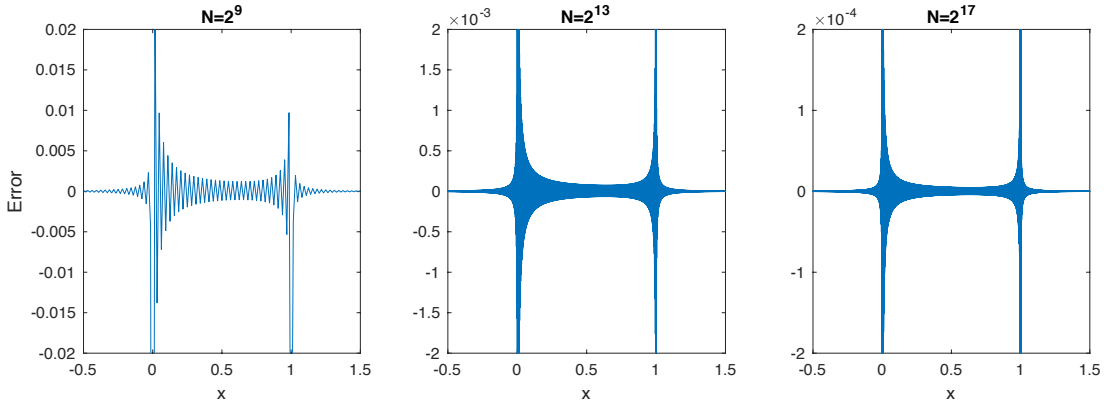

FIG. 2. Error in the numerical calculation of $f(x)$ using the sinc-based fast Hilbert transform with no filter.

The oscillations can be overcome with a spectral filter, but this can have a negative effect on the

accuracy of the numerical method, especially close to the discontinuities in the state space; this is illustrated in Figures 3–5. Figure 4 shows that the lower order filter gives a shallower slope at the discontinuity, but has a stronger effect on the oscillations. However, we can see from Figure 4 that, regardless of the order of the filter, the overshoot at the discontinuity remains approximately the same. Figure 5 shows that a spectral filter removes the oscillations away from the discontinuity and that the best results are achieved with a filter of order 8. Although the behaviour of the numerical method using the sinc-based fast Hilbert transform is not appropriate for a general solution to the Fredholm equation due to the high errors at function discontinuities, it remains the case that for applications where we are solely interested in a function value away from any jumps this may be an appropriate method to use.

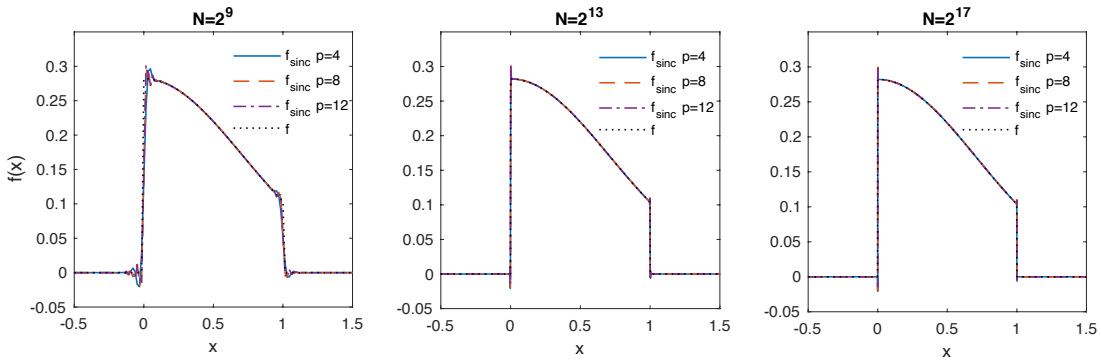

FIG. 3. Numerical and analytical $f(x)$ using the sinc-based fast Hilbert transform with an exponential filter. The parameter p describes the order of the filter.

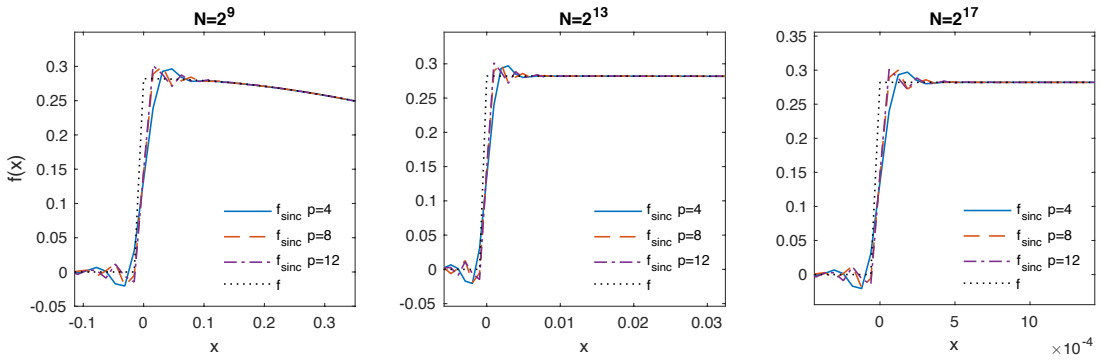

FIG. 4. Numerical and analytical $f(x)$ using the sinc-based fast Hilbert transform with an exponential filter, focussing on the discontinuity at $x = 0$. The parameter p describes the order of the filter.

6.1.2 *Sign-based fast Hilbert transform method* As an alternative to the sinc-based fast Hilbert transform, we examine the method of Rino (1970) and Henery (1974), which was used also by Fusai et al. (2016). It is based on the simple relationship between the Hilbert transform and the Fourier transform given in Eq. (2.21).

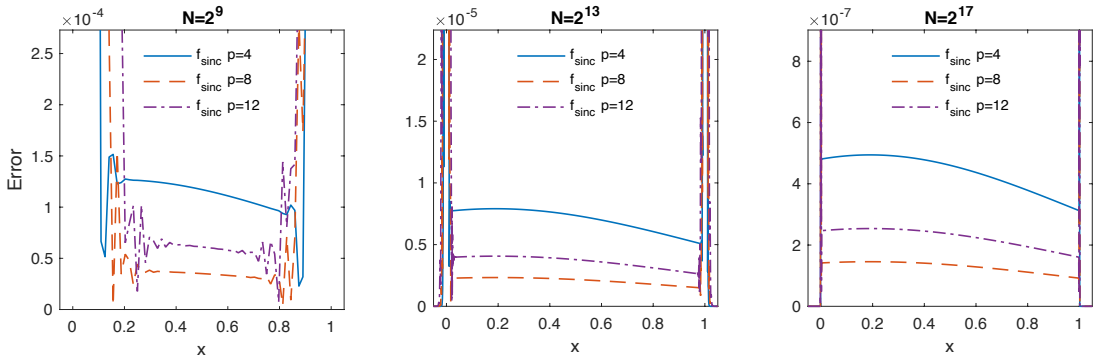

FIG. 5. Error between the numerical and analytical calculation of $f(x)$ using the sinc-based fast Hilbert transform with an exponential filter. The parameter p describes the order of the filter. The scale has been chosen to display the error away from the discontinuities of $f(x)$.

The results for the Gaussian test case are shown in Figures 6–8; f is the analytic solution, $f_{\mathrm{sgn}0}$ and f_{Vor} are the numerical solutions obtained with the Wiener-Hopf iterative method using the fast Hilbert transform implemented with the symmetric sign function, the latter in the Voronin variant. It is immediately apparent from Figure 6 that neither implementation suffers from the overshoot that was seen using the sinc-based fast Hilbert transform. However, looking at the discontinuity more closely in Figure 7, we can see that we will have a peak error at a single state-space grid point as the numerical solution increases to the final value of $f(x)$ more slowly than the analytic function. However, unlike the sinc-based function, where the extent of the oscillations depends not only on the filter, but also the shape the function used, we can state here that as long as the value of x is at least one grid step away from the discontinuity, the answer will be unaffected by the peak error. It is also interesting to note that the error is symmetrical around the discontinuity when the iterative Wiener-Hopf method is used, but not with the Voronin variant. This difference is likely to account for the better performance seen in Figure 8 compared to Figure 9. These display the error results away from the discontinuity and we can see that, although there is some variation in the error across x, the results for both methods are superior to those for the sinc-based fast Hilbert transform.

Although it is important to observe the functions which are calculated numerically, when assessing the performance of the numerical methods, the error convergence with CPU time and number of grid points N is also important. We measured this at 10%, 50% and 90% of the range between a and b; results for the Gaussian test case are shown in Figures 10–11.

The fastest converging method is the Wiener-Hopf iterative method using the sign-based fast Hilbert transform, achieving an error of $O(1/N^2)$. The other methods exhibited $O(1/N)$ error convergence, with the method using the sinc-based fast Hilbert transform with spectral filter achieving better absolute error performance versus N but converging with CPU time almost identically to the quadrature method. The $O(1/N)$ convergence achieved with the sinc-based fast Hilbert transform is consistent with the error bound described by Stenger (1993) for a function with a first order discontinuity, while the $O(1/N^2)$ convergence seen for the sign-based variant is consistent with that reported by Fusai et al. (2016).

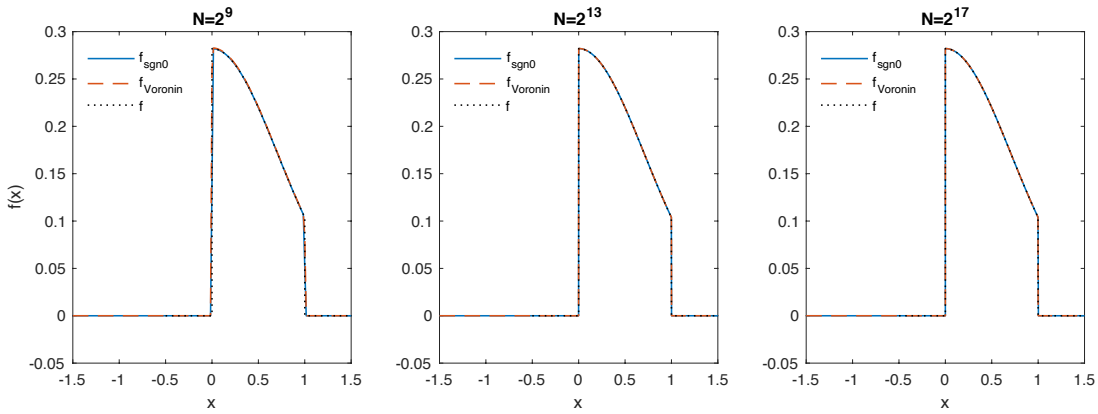

FIG. 6. Numerical and analytical $f(x)$ using the FFT based method with a symmetrical sign function.

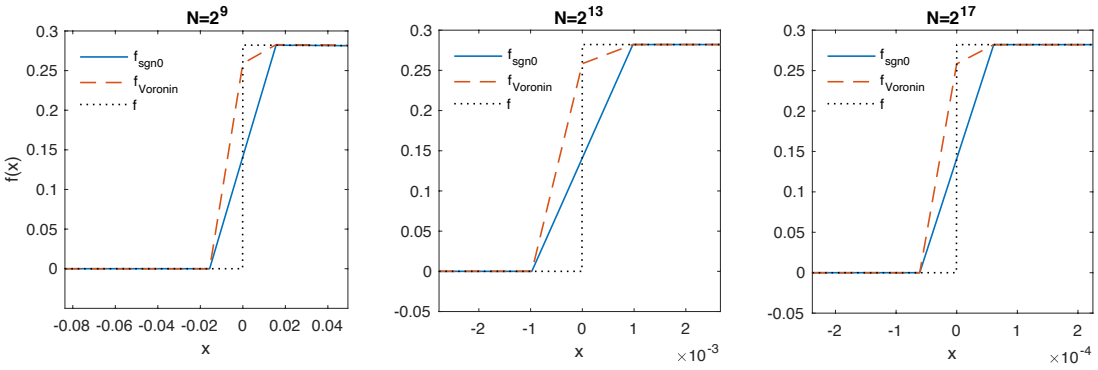

FIG. 7. Numerical and analytical $f(x)$ using the FFT based method with a symmetrical sign function, focussing on the discontinuity at $x = 0$.

6.2 *Results for Cauchy and Laplace test cases*

Figures 12–15 compare the results for the test cases in Sections 5.2 and 5.3 for the iterative Wiener-Hopf method with the sinc- and sign-based fast Hilbert transform methods; in the figures these are labelled f_{sinc} and f_{sgn}. We also compare the performance of the iterative Voronin method with the symmetrical sign function, labelled f_{Vor}. An 8th order exponential filter was used with the sinc-based fast Hilbert transform to counteract the oscillations, as described in Section 6.1.1. As a benchmark we include results from 4th order Newton-Cotes quadrature with a preconditioner (Press et al., 2007; Fusai et al., 2012), labelled f_{q}, which was the previous state of the art; using a Gaussian quadrature would imply to lose the Toeplitz structure of the matrix that is due to the convolution nature of the problem. The results for the Cauchy and Laplace test cases are consistent with those for the Gaussian test case; the use of the sinc-based fast Hilbert transform results in an overshoot at the function discontinuities and the sign-based method results in a spot error at function discontinuities. We also notice that the quadrature method has a spot error at the discontinuity, but this effects a smaller range of x than our new numerical methods. The reason for this smaller range is that the Fourier-based methods need a truncation in state

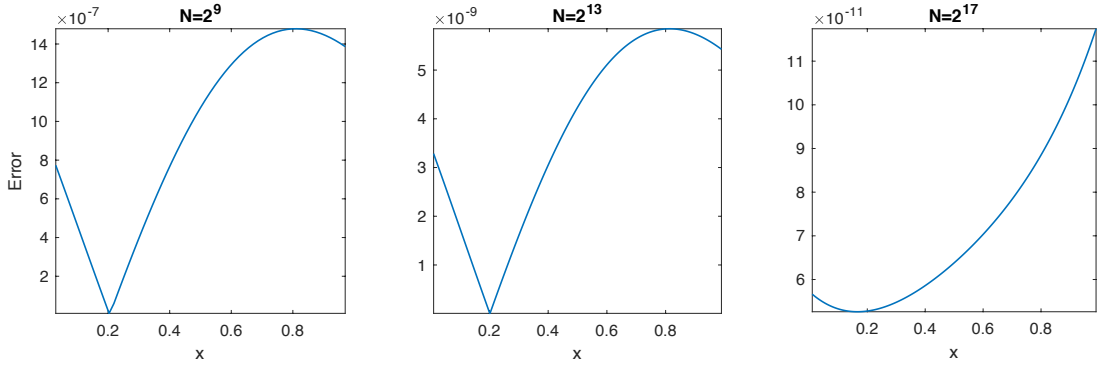

FIG. 8. Error of the numerical calculation of $f(x)$ for the new iterative Wiener-Hopf method using the sign-based fast Hilbert transform. The scale has been chosen to display the error away from the discontinuities of $f(x)$. The error is calculated by comparing the numerical calculation to the analytic solution.

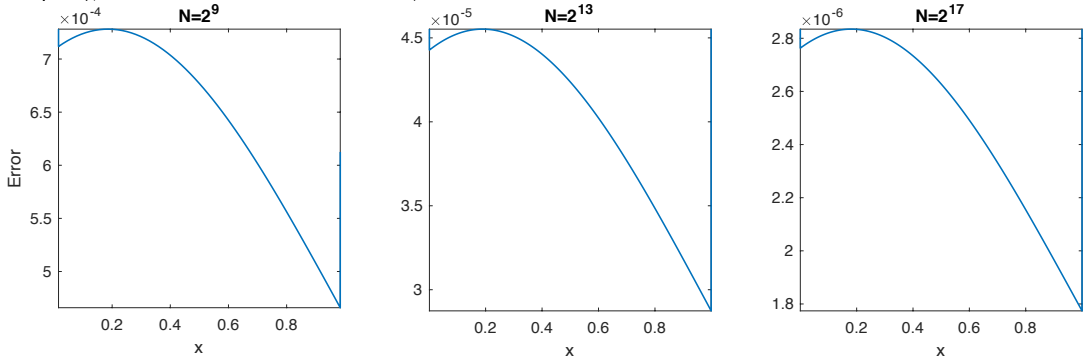

FIG. 9. Error of the numerical calculation of $f(x)$ for the Voronin method using the sign-based fast Hilbert transform. The scale has been chosen to display the error away from the discontinuities of $f(x)$. The error is calculated by comparing the numerical calculation to the analytic solution.

space at $\pm 4(b-a)$ in order to avoid wrap-round effects. In contrast, the range of x for quadrature only needs truncation at the integration limits a and b.

We also measured the error convergence with the Cauchy and Laplace test cases and the results are shown in Figures 16–19. These confirm the findings with the Gaussian test case in Section 5.1 which showed that the best performing method is the new iterative solution to the Wiener-Hopf equation with the Hilbert transform implemented using the FFT with the symmetrical sign function.

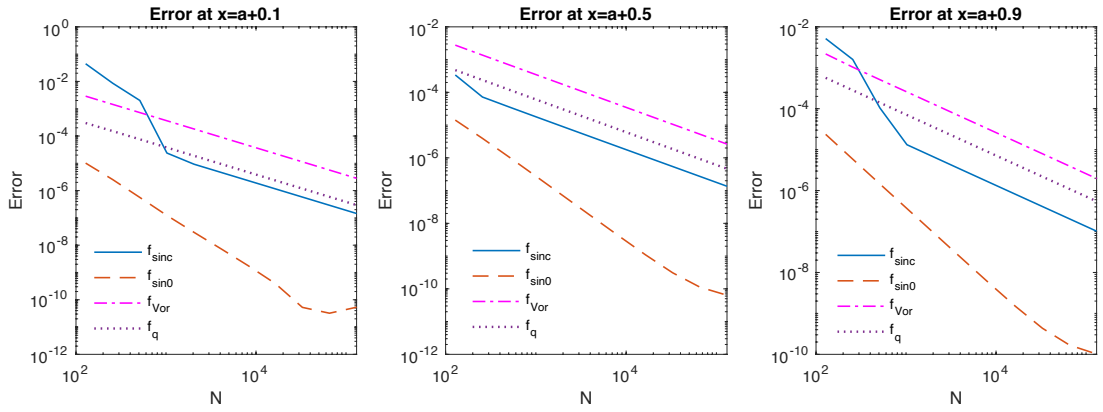

FIG. 10. Error convergence of the numerical methods against N with the Gaussian test case.

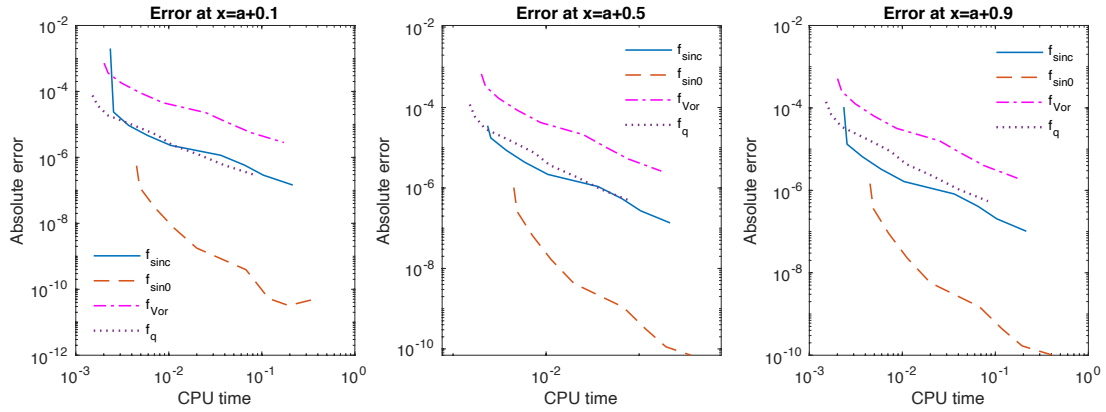

FIG. 11. Error convergence of the numerical methods against CPU time with the Gaussian test case.

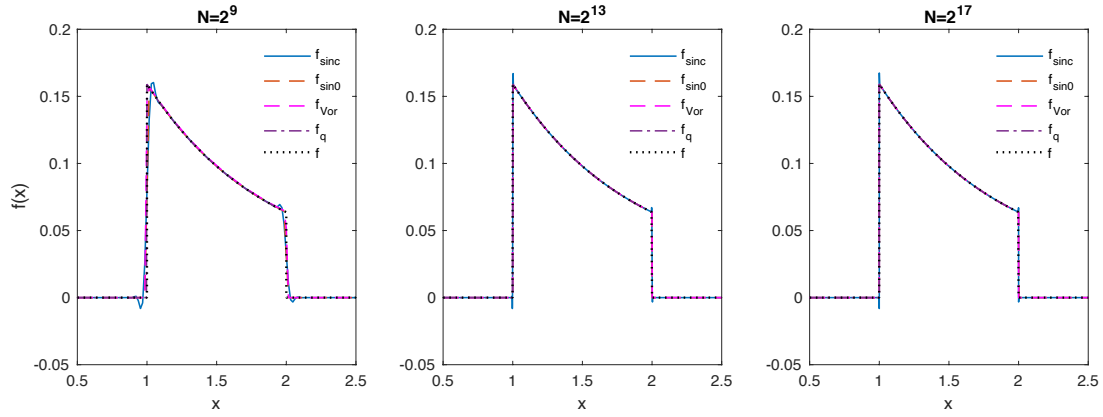

FIG. 12. Numerical and analytical $f(x)$ with the Cauchy test case.

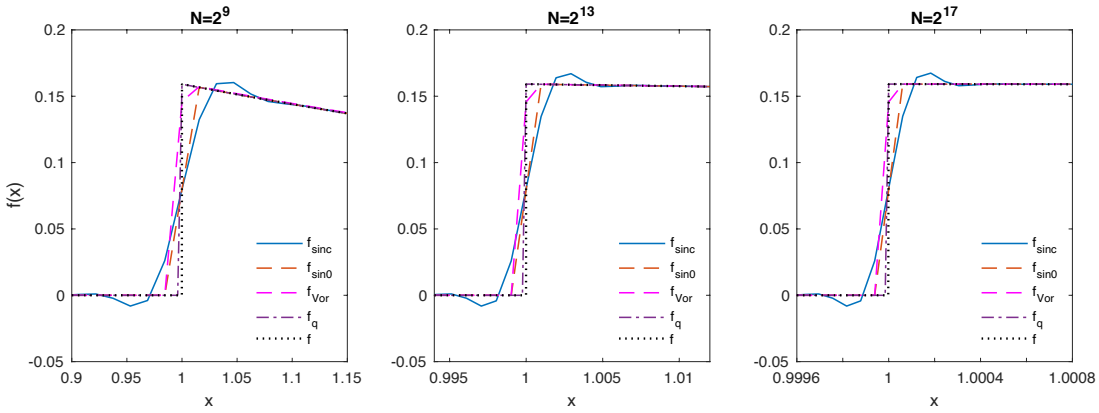

FIG. 13. Numerical and analytical $f(x)$ with the Cauchy test case focussing on the first discontinuity.

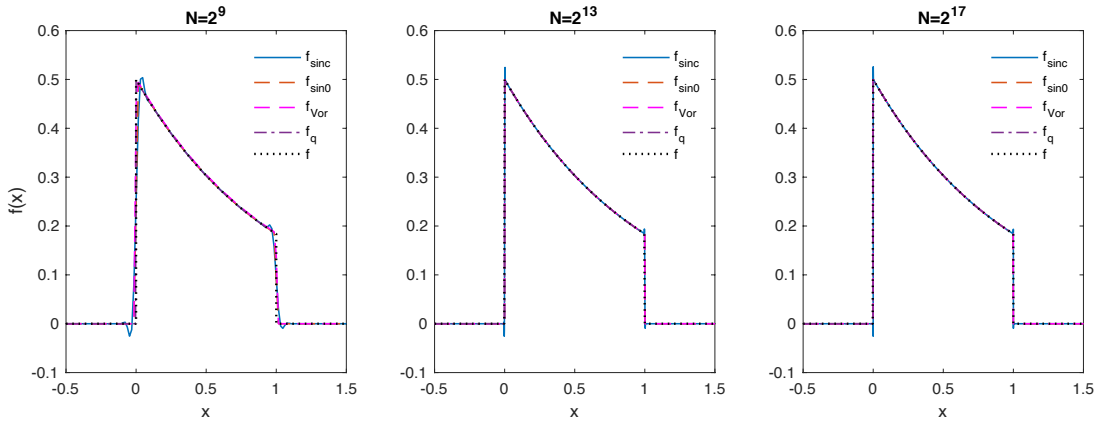

FIG. 14. Numerical and analytical $f(x)$ with the Laplace test case.

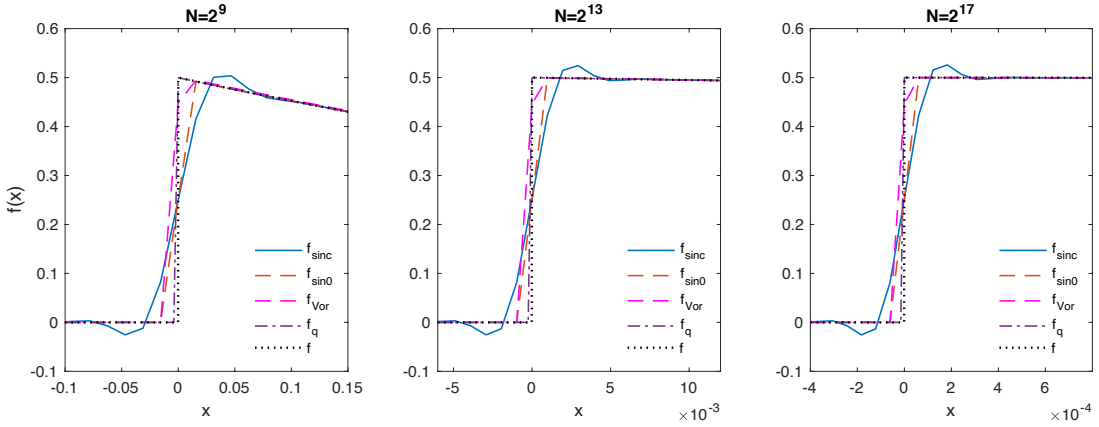

FIG. 15. Numerical and analytical $f(x)$ with the Laplace test case focussing on the first discontinuity.

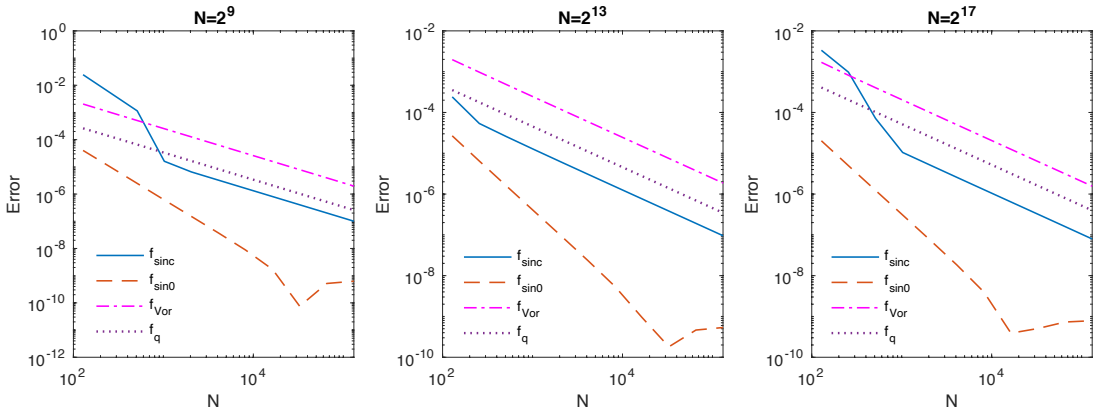

FIG. 16. Error convergence of the numerical methods against N with the Cauchy test case.

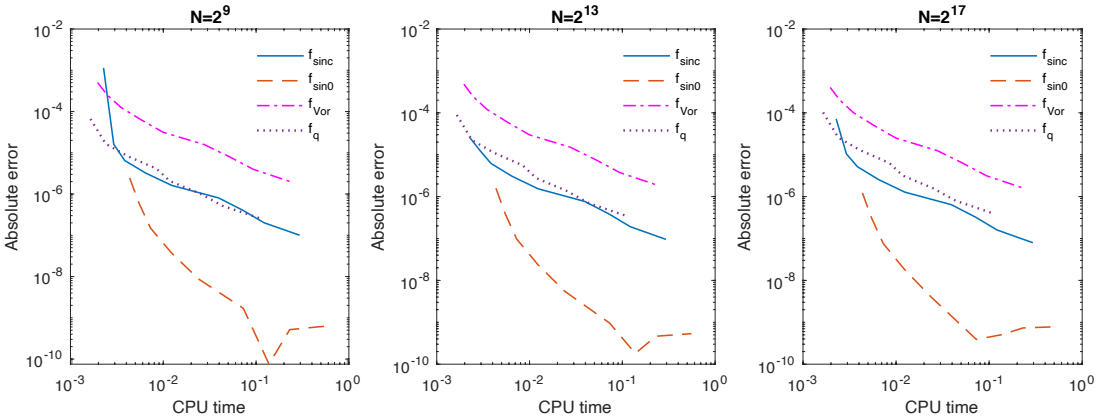

FIG. 17. Error convergence of the numerical methods against CPU time with the Cauchy test case.

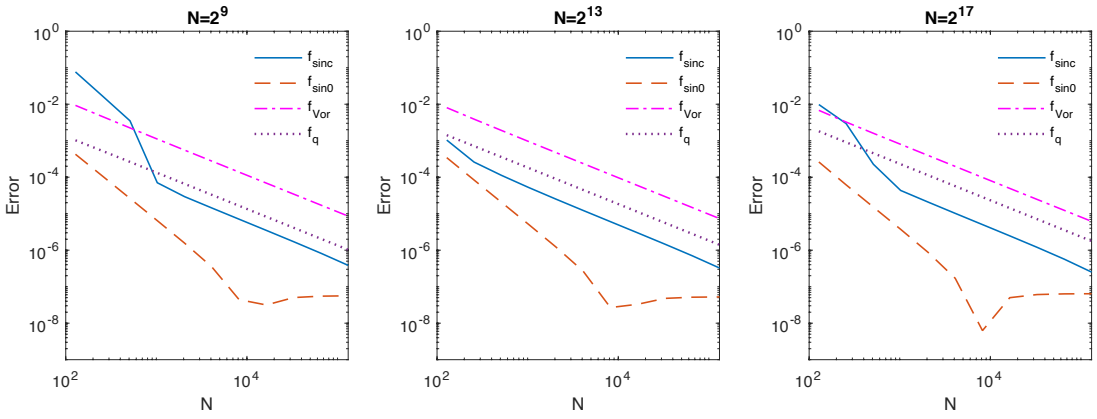

FIG. 18. Error convergence of the numerical methods against N with the Laplace test case.

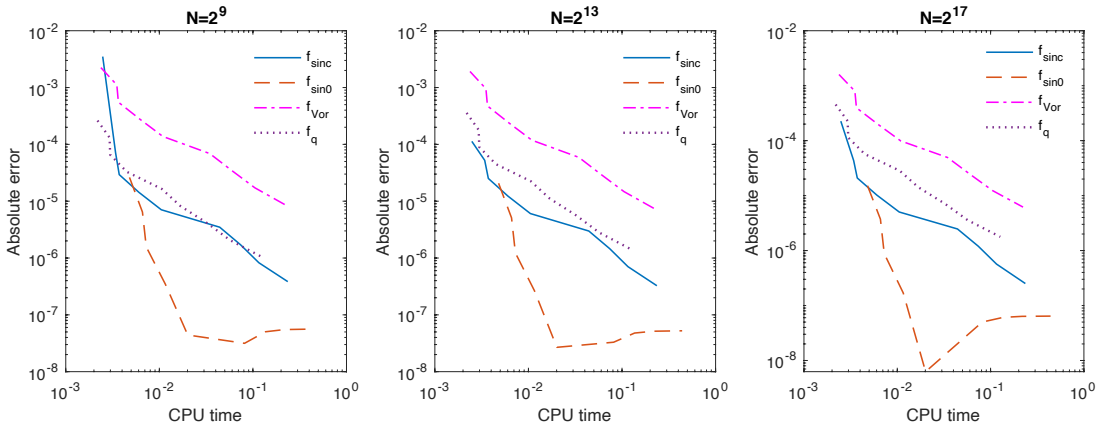

FIG. 19. Error convergence of the numerical methods against CPU time with the Laplace test case.

7. Conclusion

We developped numerical methods based on the FFT to solve convolution integral equations on a semi-infinite interval (Wiener-Hopf equation) or on a finite interval (Fredholm equation). We improved a previous method for the Wiener-Hopf equation based on the FFT and projection operators (Rino, 1970; Henery, 1974) by expressing the required Wiener-Hopf factorisation through a Hilbert transform via the Plemelj-Sokhotsky relations. This allowed us to compute the Hilbert transform with an accurate and efficient numerical method based on a sinc functions expansion and the FFT (Stenger, 1993, 2011), reducing the total number of required FFTs from 7 to 5. We further enhanced the error convergence using a spectral filter (Vandeven, 1991; Gottlieb & Shu, 1997). We resolved issues of a previous iterative extension to the Fredholm equation (Henery, 1977). Moreover, we devised a variant of our method inspired by the matrix factorisation approach of Voronin (2004).

We carried out extensive numerical tests on the Fredholm equation of the second kind using three kernels and provide operational open-source code. We implemented our new iterative Wiener-Hopf method with the sinc- or sign-based fast Hilbert transform, the latter also with the variant inspired by Voronin's matrix factorisation. For benchmark, we implemented a 4th order Newton-Cotes quadrature (Press et al., 2007; King, 2009) with preconditioner (Fusai et al., 2012).

Unlike an earlier application in option pricing with exponential convergence of a weighted average error, the iterative Wiener-Hopf method with the sinc-based fast Hilbert transform does not turn out optimal for a general solution of the Fredholm equation, having $O(1/N)$ convergence with the number of FFT grid points N and high errors close to the function discontinuities; this can be explained with the different requirements of the two problems, as here the solution over the whole interval is required. Instead, the iterative Wiener-Hopf method with the sign-based fast Hilbert transform has $O(1/N^2)$ convergence and therefore performs better than its sinc-based sibling, the quadrature method from the literature and the iterative method based on Voronin's partial solution, whose convergence is $O(1/N)$ even with the sign-based fast Hilbert transform. So in terms of error convergence, the iterative Wiener-Hopf method with the sign-based fast Hilbert transform reveals the new state of the art for the numerical solution of general Fredholm equations, achieving double the convergence speed of the known 4th order quadrature method.

The other aspect which we must compare for the different methods is the peak error at a discontinuity of $f(x)$ as shown in Figures 13 and 15. This error is wider for the Wiener-Hopf method than for the quadrature method because a wider x range is required to avoid the wrap-around or aliasing effect of Fourier transform methods. Therefore, if an accurate answer close to a discontinuity is required, the quadrature method may be best. However, the excellent CPU time versus error performance shown in Figures 11, 17 and 19 recommends using the Wiener-Hopf method with the sign-based fast Hilbert transform and a larger grid size to yield the required accuracy close to discontinuities.

Acknowledgments

Guido Germano is thankful to Vito Daniele for a useful discussion on the difficulty of factorising the matrices **L** of Section 4.4 and **M** of Section 4.5. Moreover, he gratefully acknowledges the Economic and Social Research Council (ESRC) for funding the Systemic Risk Centre (grant number ES/K002309/1).

Carolyn E. Phelan was supported by the Engineering and Physical Sciences Research Council (EPSRC) through the UK Centre for Doctoral Training in Financial Computing and Analytics (grant number EP/L015129/1).

Daniele Marazzina is member of the Gruppo Nazionale Calcolo Scientifico — Istituto Nazionale di

Alta Matematica (GNCS-INdAM), and his research is partially funded by the GNCS project "Numerical methods for addressing traditional and emerging challenges in finance".

Gianluca Fusai is member of the PNRR MUR (M4C2-1.1) project P20228CHNL, financed by the Italian Ministry of University and Research.

REFERENCES

Bart, H., Gohberg, I., Kaashoek, M. A. & Ran, A. C. M. (2004) *Factorization of matrix and operator functions: the state space method*. Birkhäuser, Basel. Operator Theory: Advances and Applications, Vol. 178.

Boyd, J. P. (2001) *Chebyshev and Fourier Spectral Methods*. Springer, Heidelberg.

Choi, J., Margetis, D., Squires, T. M. & Bazant, M. Z. (2005) Steady advection-diffusion around finite absorbers in two-dimensional potential flows. *Journal of Fluid Mechanics*, **536**, 155–184.

Daniele, V. & Lombardi, G. (2007) Fredholm factorization of Wiener-Hopf scalar and matrix kernels. *Radio Science*, **42**.

Daniele, V. G. (1984) On the solution of two coupled Wiener-Hopf equations. *SIAM Journal of Applied Mathematics*, **44**, 667–680.

Daniele, V. G. & Zich, R. S. (2014) *The Wiener-Hopf Method in Electromagnetics*. SciTech Publishing, Stevenage.

Duffy, D. G. (2008) *Mixed Boundary Value Problems*. CRC Press / Chapman & Hall (Taylor & Francis Group), Boca Raton, FL.

Feldman, I., Gohberg, I. & Krupnik, N. (2000) Convolution equations on finite intervals and factorization of matrix functions. *Integral Equations and Operator Theory*, **36**, 201–211.

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, New York, 3rd edition.

Feng, L. & Linetsky, V. (2008) Pricing discretely monitored barrier options and defaultable bonds in Lévy process models: a Hilbert transform approach. *Mathematical Finance*, **18**, 337–384.

Feng, L. & Linetsky, V. (2009) Computing exponential moments of the discrete maximum of a Lévy process and lookback options. *Finance and Stochastics*, **13**, 501–529.

Fredholm, I. (1903) Sur une classe d'équations fonctionelles. *Acta Mathematica*, **27**, 365–390.

Frigo, M. & Johnson, S. G. (2005) The design and implementation of FFTW3. *Proceedings of the IEEE*, **93**, 216–231.

Fusai, G., Abrahams, I. D. & Sgarra, C. (2006) An exact analytical solution for discrete barrier options. *Finance and Stochastics*, **10**, 1–26.

Fusai, G., Germano, G. & Marazzina, D. (2016) Spitzer identity, Wiener-Hopf factorisation and pricing of discretely monitored exotic options. *European Journal of Operational Research*, **251**, 124–134.

Fusai, G., Marazzina, D., Marena, M. & Ng, M. (2012) Z-transform and preconditioning techniques for option pricing. *Quantitative Finance*, **12**, 1381–1394.

Gohberg, I. C. & Fel'dman, I. A. (1974) *Convolution Equations and Projection Methods for their Solution*, volume 41 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence.

Gottlieb, D. & Shu, C. (1997) On the Gibbs phenomenon and its resolution. *SIAM Review*, **39**, 644–668.

Green, R., Fusai, G. & Abrahams, I. D. (2010) The Wiener-Hopf technique and discretely monitored path-dependent option pricing. *Mathematical Finance*, **20**, 259–288.

Henery, R. J. (1974) Solution of Wiener-Hopf integral equations using the fast Fourier transform. *Journal of the Institute of Mathematics and its Applications*, **13**, 89–96.

Henery, R. J. (1977) Solution of Fredholm integral equations with symmetric difference kernels. *Journal of the Institute of Mathematics and its Applications*, **19**, 29–37.

Jones, D. S. (1984) Factorization of a Wiener-Hopf matrix. *IMA Journal of Applied Mathematics*, **32**, 211–220.

Jones, D. S. (1991) Wiener-Hopf splitting of a 2×2 matrix. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **434**, 419–433.

King, F. W. (2009) *Hilbert Transforms*, volume 1. Cambridge University Press, Cambridge. Encyclopedia of Mathematics and its Applications, vol. 124.

Kisil, A. V. (2013) A constructive method for an approximate solution to scalar Wiener-Hopf equations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **469**, 20120721.

Kisil, A. V. (2015) Stability analysis of matrix Wiener-Hopf factorization of Daniele-Khrapkov class and reliable approximate factorization. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **471**, 20150146.

Kisil, A. V. (2018) An iterative Wiener-Hopf method for triangular matrix functions with exponential factors. *SIAM Journal on Applied Mathematics*, **78**, 45–62.

Kisil, A. V., Abrahams, I. D., Mishuris, G. & Rogosin, S. V. (2021) The Wiener-Hopf technique, its generalizations and applications: constructive and approximate methods. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **477**, 20210533.

Krein, M. G. (1962) Integral equations on a half-line with kernel depending on the difference of the arguments. In *Nine Papers on Analysis*, volume 22 of *2*, pages 163–288. American Mathematical Society. American Mathematical Society Translations.

Kress, R. & Martensen, E. (1970) Anwendung der Rechteckregel auf die reelle Hilberttransformation mit unendlichem Intervall. *Zeitschrift für Angewandte Mathematik und Mechanik*, **50**, T61–T64.

Lawrie, J. B. & Abrahams, I. D. (2007) A brief historical perspective of the Wiener-Hopf technique. *Journal of Engineering Mathematics*, **59**, 351–358.

Marazzina, D., Fusai, G. & Germano, G. (2012) Pricing credit derivatives in a Wiener-Hopf framework. In Cummins, M., Murphy, F. & Miller, J. J. H., editors, *Topics in Numerical Methods for Finance*, pages 139–154, New York. Springer Science+Business Media. Springer Proceedings in Mathematics & Statistics, vol. 19.

Margetis, D. & Choi, J. (2006) Generalized iteration method for first-kind integral equations. *Studies in Applied Mathematics*, **117**, 1–25.

McKechan, D. J. A., Robinson, C. & Sathyaprakash, B. S. (2010) A tapering window for time-domain templates and simulated signals in the detection of gravitational waves from coalescing compact binaries. *Classical and Quantum Gravity*, **27**, 084020.

Noble, B. (1958) *Methods Based on the Wiener-Hopf Technique for the Solution of Partial Differential Equations*. Pergamon Press, London. Reprinted by Chelsea, New York, 1988.

Pandey, J. N. (1996) *The Hilbert transform of Schwartz distributions and applications*. Wiley, New York.

Phelan, C. E., Fusai, G., Marazzina, D. & Germano, G. (2018) Fluctuation identities with continuous monitoring and their application to the pricing of barrier options. *European Journal of Operational Research*, **271**, 210–223.

Phelan, C. E., Marazzina, D., Fusai, G. & Germano, G. (2019) Hilbert transform, spectral filters and option pricing. *Annals of Operational Research*, **282**, 273–298.

Phelan, C. E., Marazzina, D. & Germano, G. (2020) Numerical pricing methods for α-quantile and perpetual early-exercise options. *Quantitative Finance*, **20**, 899–918.

Plancherel, M. (1910) Contribution à l'étude de la représentation d'une fonction arbitraire par des intégrales définies. *Rendiconti del Circolo Matematico di Palermo*, **30**(1), 289–335.

Polyanin, A. D. & Manzhirov, A. V. (1998) *Handbook of Integral Equations*. CRC Press, Boca Raton.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2007) *Numerical Recipes in C++*. Cambridge University Press, Cambridge, 3rd edition.

Rino, C. L. (1970) Factorization of spectra by discrete Fourier transforms. *IEEE Transactions on Information Theory*, **16**, 484–485.

Rogosin, S. & Mishuris, G. (2016) Constructive methods for factorization of matrix-functions. *IMA Journal of Applied Mathematics*, **81**, 365–391.

Ruijter, M. J., Versteegh, M. & Oosterlee, C. W. (2015) On the application of spectral filters in a Fourier option pricing technique. *Journal of Computational Finance*, **19**, 75–106.

Spitzer, F. (1957) The Wiener-Hopf equation whose kernel is a probability density. *Duke Mathematical Journal*, **24**, 327–343.

Stenger, F. (1973) The approximate solution of convolution-type integral equations. *SIAM Journal of Mathematical Analysis*, **4**, 536–555.

Stenger, F. (1993) *Numerical Methods Based on Sinc and Analytic Functions*. Springer, Berlin.

Stenger, F. (2011) *Handbook of Sinc Numerical Methods*. CRC Press / Chapman & Hall, Taylor & Francis Group, Boca Raton, FL.

Tadmor, E. (2007) Filters, mollifiers and the computation of the Gibbs phenomenon. *Acta Numerica*, **16**, 305–378.

Tadmor, E. & Tanner, J. (2005) Adaptive filters for piecewise smooth spectral data. *IMA Journal of Numerical Analysis*, **25**, 635–647.

Vandeven, H. (1991) Family of spectral filters for discontinuous problems. *Journal of Scientific Computing*, **6**, 159–192.

Vergara Caffarelli, G. & Loreti, P. (1999) Trasformata di Hilbert. Quaderni Didattici del Dipartimento di Metodi e Modelli Matematici per le Scienze Applicate, Università di Roma La Sapienza.

Voronin, A. F. (2004) A complete generalization of the Wiener-Hopf method to convolution integral equations with integrable kernel on a finite interval. *Differential Equations*, **40**, 1259–1267.

Weideman, J. A. C. (1995) Computing the Hilbert transform on the real line. *Mathematics of Computation.*, **64**, 745–762.

Weisstein, E. W. (2025) Fourier transform — Inverse function. MathWorld, http://mathworld.wolfram.com/FourierTransformInverseFunction.html.

Whittaker, E. T. & Watson, G. N. (1927) *A Course of Modern Analysis*. Cambridge University Press, Cambridge, 4th edition. Reprinted 2002.

Wiener, N. & Hopf, E. (1931) Über eine Klasse singulärer Integralgleichungen. *Sitzungsberichte der Preußischen Akademie der Wissenschaften, Mathematisch-Physikalische Klasse*, **31**, 696–706.