IMA Journal of Numerical Analysis (2007) **27**, 406−425
doi:10.1093/imanum/drl043
Advance Access publication on January 27, 2007

Sharp error estimates for discretizations of the 1D convection–diffusion equation with Dirac initial data

REBECCA CARTER AND MICHAEL B. GILES
Oxford University Computing Laboratory, Oxford OX1 3QD, UK

[Received on 9 July 2004; revised on 19 September 2006]

This paper derives sharp estimates of the error arising from explicit and implicit approximations of the constant-coefficient 1D convection–diffusion equation with Dirac initial data. The error analysis is based on Fourier analysis and asymptotic approximation of the integrals resulting from the inverse Fourier transform. This research is motivated by applications in computational finance and the desire to prove convergence of approximations to adjoint partial differential equations.

Keywords: convection–diffusion equation; Crank–Nicolson time marching; Rannacher startup; Dirac initial data.

1. Introduction

This paper is concerned with a detailed error analysis of two different discretizations of the 1D constant-coefficient convection–diffusion equation on an infinite domain. Both are based on a second-order central space discretization; one uses the forward Euler time discretization and the other uses Crank–Nicolson, with or without a Rannacher startup in which one or more Crank–Nicolson timesteps are replaced by two half-timesteps of backward Euler discretization to improve the convergence.

The novelty in this paper is in the focus on Dirac initial data. One reason for this focus is the concern with the convergence of adjoint discretizations. Adjoint methods are being used heavily for optimal design (Jameson, 1988; Jameson *et al.*, 1998), error analysis and correction for integral outputs (Barth & Deconinck, 2002; Giles & Süli, 2002; Pierce & Giles, 2004) and optimal grid adaptation (Becker & Rannacher, 2001; Darmofal & Venditti, 2003). In applications in which the original partial differential equation (PDE) is nonlinear, the adjoint discretization is usually obtained in one of two ways, either as a discretization of the adjoint PDE corresponding to the linearization of the original PDE or as the transposed equation corresponding to the linearized discretization of the original PDE. In either case, if the original nonlinear solution is smooth, then the coefficients of the adjoint discretization will be smooth, and it is possible to prove convergence in both steady and unsteady applications as the mesh spacing and timestep approach zero (Ulbrich, 2002, 2003). However, when the underlying nonlinear solution is discontinuous, as in the case of shocks in compressible flow, there is numerical evidence (Giles, 2003) showing that one must be careful in the treatment of the discontinuity to obtain convergence for the adjoint discretization.

To understand the connection between Dirac initial data and adjoint equations, consider the following system of linear equations:

$$U^{n+1} = A^n U^n$$

arising from the discretization of an unsteady linear 1D PDE. Here, U^n represents the approximation to a scalar variable $u(x, t)$ on a 1D grid with uniform spacing h at time $t^n = nk$. If one is interested in the

value of an integral output

$$J = \int_{-\infty}^{\infty} g(x)u(x,T)\mathrm{d}x,$$

this may be approximated as

$$J_h = h\sum_j g(x_j)U_j^N,$$

where $T = Nk$. Alternatively, but equivalently, it can be evaluated as

$$J_h = h\sum_j V_j^0 U_j^0,$$

where the adjoint solution V_j^n satisfies the adjoint discrete equations

$$V^n = (A^n)^{\mathrm{T}} V^{n+1},$$

subject to the final data

$$V_j^N = g(x_j).$$

The equivalence follows immediately from the identity

$$(V^0)^{\mathrm{T}}U^0 = (V^N)^{\mathrm{T}}A^{N-1}A^{N-2}\cdots A^1A^0U^0 = (V^N)^{\mathrm{T}}U^N.$$

This adjoint approach to evaluating the output functional is advantageous when there is a single output functional of interest, but many different sets of initial data. Under these circumstances, the standard approach would require a separate forward analysis for each set of initial data, whereas the adjoint approach requires just one adjoint calculation plus an inexpensive inner product evaluation for each set of initial data.

In the particular case of Dirac initial data with

$$U_j^0 = h^{-1}\delta_{j,0} \equiv \begin{cases} h^{-1}, & j = 0, \\ 0, & \text{otherwise}, \end{cases}$$

one obtains

$$V_0^0 = h\sum_j g(x_j)U_j^N.$$

Thus, convergence of the integral output for Dirac initial data is equivalent to pointwise convergence of the adjoint discretization. The results for the explicit forward Euler discretization in this paper will be used in future research to prove the pointwise convergence of adjoint discretizations when there are discontinuities in the solution of the underlying nonlinear PDE.

A second motivation for the analysis in this paper is the applications in mathematical finance which require the numerical solution of variants of the Black–Scholes equation (Wilmott *et al.*, 1995)

$$\frac{\partial V}{\partial t} = rV - rS\frac{\partial V}{\partial S} - \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}.$$

This is an equation which is solved backwards in time, from a terminal time $t = T$ to an initial time $t = 0$. The value for σ is sufficiently large so that the diffusion plays a significant role in the evolution of the solution; it is not a convection-dominated problem. Hence, second-order central space differencing and Crank–Nicolson time integration are widely used to approximate this equation. On a uniform grid with spacing h and timestep k, this results in the discrete equations

$$\left(I + \frac{1}{2}D_j\right) V_j^{n+1} = \left(I - \frac{1}{2}D_j\right) V_j^n,$$

where

$$D_j = -\frac{k}{2h^2}\sigma^2 (S_j)^2 \delta_s^2 - \frac{k}{2h} r S_j \delta_{2s} + rk$$

with δ_s^2 and δ_{2s} being the standard second-difference and central first-difference operators, respectively, and $t^{n+1} = t^n - k$, where n is the time level index which increases from $n = 0$ at time $t = T$ to $n = N$ at time $t = 0$.

For European call options, the 'initial' data at the terminal time are

$$V(S, T) = \max(S - K, 0).$$

The topleft plot in Fig. 1 shows the numerical solution $V(S, 0)$ at time $t = 0$ obtained on a uniform grid $0 \leqslant S \leqslant S_{\max} = 5$, using parameter values $r = 0.05$, $\sigma = 0.2$, $K = 1$ and $T = 2$. The boundary conditions which were used are $V_j = 0$ at $S = 0$ and $\delta^2 V_j = 0$ at $S = S_{\max}$. The agreement between the numerical solution and the analytic solution (Wilmott *et al.*, 1995) appears quite good, but in the financial application the first derivative, $\Delta \equiv \partial V/\partial S$, and the second derivative, $\Gamma \equiv \partial^2 V/\partial S^2$, are both important quantities. Their numerical values obtained by central differencing are in much poorer agreement with the analytic solution, as shown in the other two left-hand plots in Fig. 1. In particular, note that the maximum error in the computed value for Γ occurs at $S = 1$, which is the location of the discontinuity in the first derivative of the initial data.

The left-hand plots in Fig. 2 show the behaviour of the maximum error as the computational grid is refined, keeping the ratio $\lambda \equiv k/h$ fixed. It can be seen that the numerical solution V_j exhibits first-order convergence, while the discrete approximation to Δ does not converge and the approximation to Γ diverges.

At first sight, this may appear surprising as the Crank–Nicolson method is well-known to be consistent and unconditionally stable, and hence one expects convergence. However, it is unconditionally stable only in the L_2-norm, and this, together with consistency, ensures convergence in L_2 only for initial data which lie in L_2 (Richtmyer & Morton, 1967), and the order of convergence may be less than the second order achieved for smooth initial data. For example, the L_2 order of convergence for discontinuous initial data is $\frac{1}{2}$. With the European call, the initial data for V lie in L_2, as does its first derivative, but the second derivative does not. This then is the root cause of the observed failure to converge as the grid is refined. Furthermore, it is the maximum error, the L_∞-error, which is most relevant in financial applications.

Rannacher (1984) analysed this problem from the perspective of L_2-convergence of convection–diffusion approximations with discontinuous initial data. His objective was to recover second-order convergence in the context of Crank–Nicolson time marching (he also considered higher-order time integration schemes), and using energy methods, he proved that this could be achieved by replacing the Crank–Nicolson approximation for the very first timestep by two half-timesteps using backward Euler

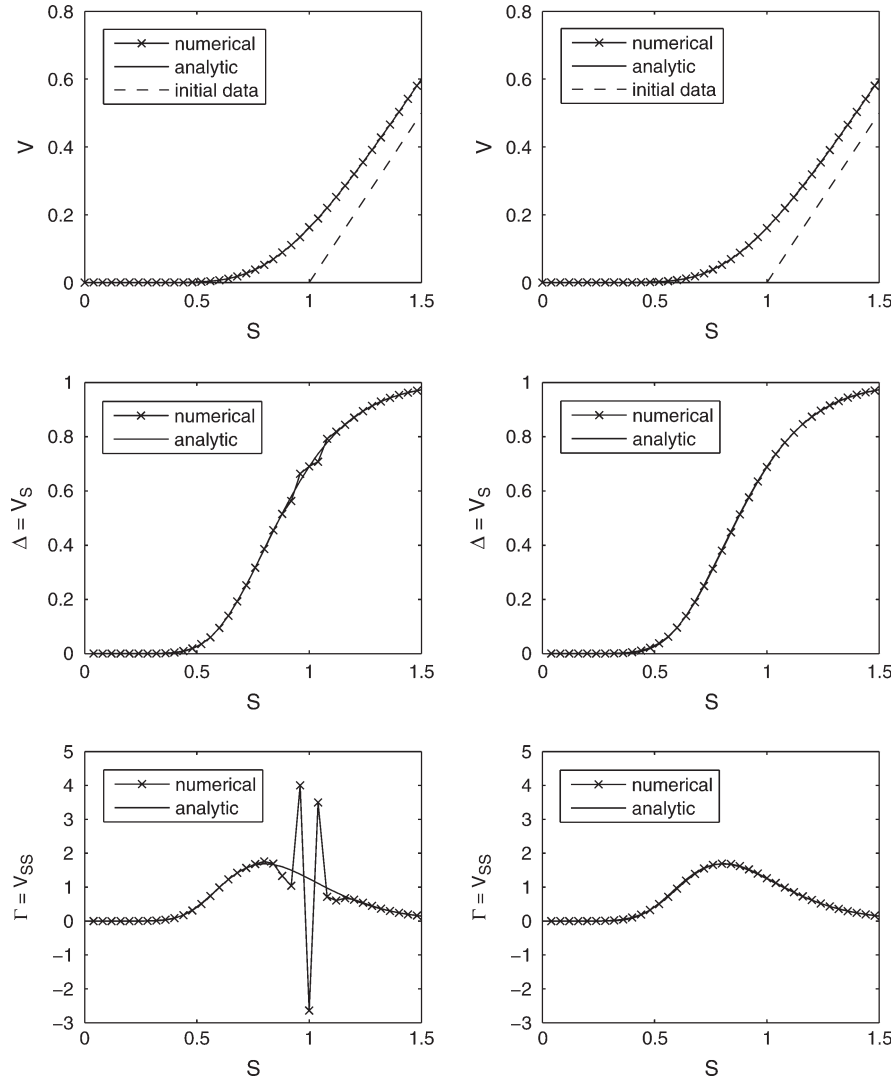

FIG. 1. V, Δ and Γ for European call option.

time integration. This solution, often referred to as Rannacher time-stepping, has been used with success in approximations of the Black–Scholes equations (Pooley *et al.*, 2003a,b). The right-hand plots in Figs 1 and 2 show that replacing the first two Crank–Nicolson timesteps by four half-timesteps of backward Euler, for which

$$\left(I + \frac{1}{2}D\right) V_j^{n+1/2} = V_j^n,$$

results in second-order convergence for V, Δ and Γ. The purpose of the analysis in this paper is to explain this behaviour by analysing the implicit discretization of the convection–diffusion equation subject

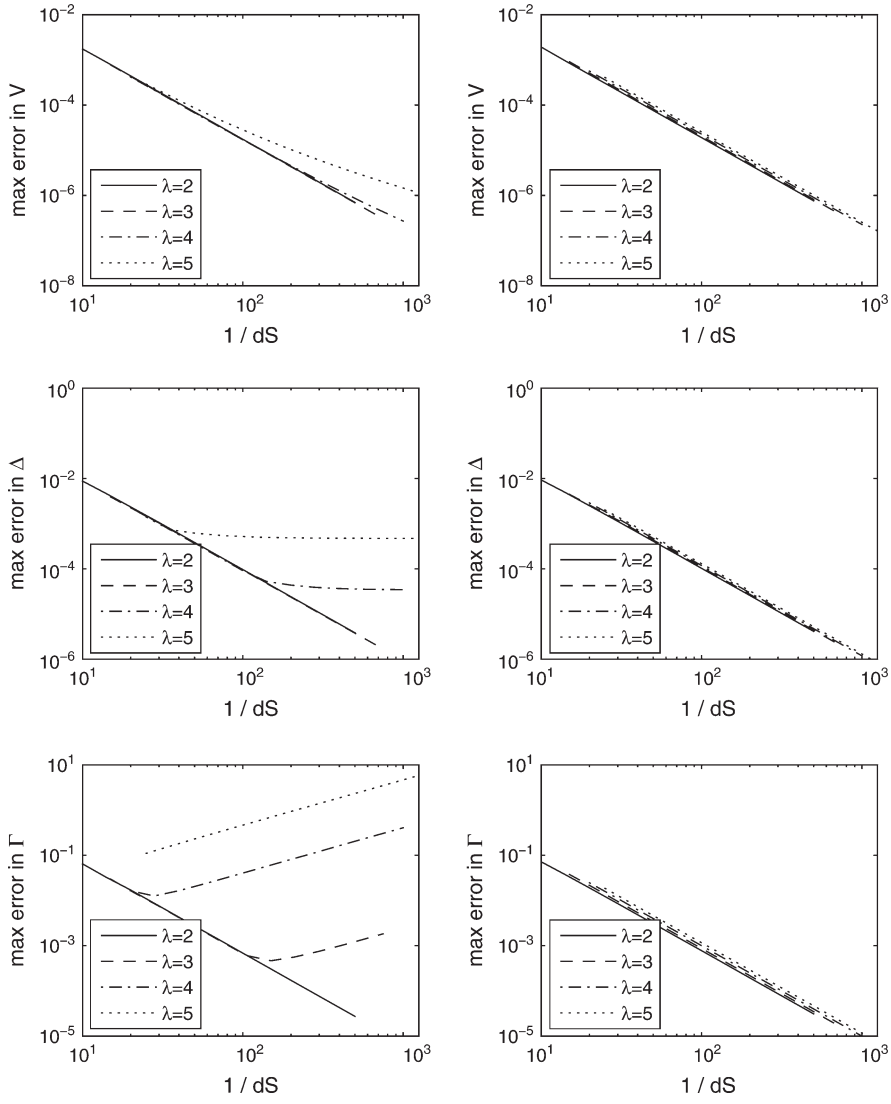

FIG. 2. Grid convergence for European call option.

to Dirac initial data, corresponding to the initial data for Γ. This will prove that four half-timesteps of backward Euler time marching are the minimum required to recover second-order convergence; the use of more than four half-timesteps will probably lead to an increase in the overall error, and therefore four half-timesteps can be regarded as optimal.

The numerical analysis is based on the Fourier transform of the discrete equations (Strang, 1986; Strikwerda, 1989) and asymptotic approximation of the inverse Fourier transform to bound the resulting discretization error. Numerical results confirm the sharpness of the error bounds which are derived.

2. Model problem and discretizations

The model problem to be analysed is the convection–diffusion equation

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = \frac{\partial^2 u}{\partial x^2}, \tag{2.1}$$

on $-\infty < x < \infty$ and $0 < t < 1$, subject to the Dirac initial data

$$u(x, 0) = \delta(x).$$

The generalization to nonunit diffusivity and terminal times other than $t = 1$ will be discussed later.

Defining the Fourier transform pair

$$\widehat{u}(\kappa, t) = \int_{-\infty}^{\infty} u(x, t)\mathrm{e}^{-\mathrm{i}\kappa x}\,\mathrm{d}x,$$

$$u(x, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{u}(\kappa, t)\mathrm{e}^{\mathrm{i}\kappa x}\,\mathrm{d}\kappa,$$

the Fourier transform of (2.1) yields

$$\frac{\mathrm{d}\widehat{u}}{\mathrm{d}t} = -(\mathrm{i}a\kappa + \kappa^2)\widehat{u},$$

subject to the initial data $\widehat{u}(\kappa, 0) = 1$. The solution to this is

$$\widehat{u}(\kappa, t) = \exp(-(\mathrm{i}a\kappa + \kappa^2)t),$$

and hence

$$u(x, t) = \frac{1}{\sqrt{4\pi t}}\exp\left(-\frac{(x - at)^2}{4t}\right) = \frac{1}{\sqrt{2t}}N\left(\frac{x - at}{\sqrt{2t}}\right),$$

where

$$N(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$$

is the standard normal distribution with zero mean and unit variance.

The forward Euler central space discretization on a uniform grid with spacing h and timestep k is

$$U_j^{n+1} = (I - D)U_j^n, \tag{2.2}$$

where

$$D = -d\delta_x^2 + \frac{r}{2}\delta_{2x}, \quad d = \frac{k}{h^2}, \quad r = \frac{ak}{h},$$

with δ_x^2 and δ_{2x} being the usual second-difference and central first-difference operators, respectively.

The Crank–Nicolson discretization is

$$\left(I + \frac{1}{2}D\right)U_j^{n+1} = \left(I - \frac{1}{2}D\right)U_j^n, \tag{2.3}$$

and the half-timestep backward Euler discretization used in the Rannacher startup is

$$\left(I + \frac{1}{2}D\right) U_j^{n+1/2} = U_j^n. \tag{2.4}$$

Assuming the grid points are at $x_j = jh$, the discrete approximation to the Dirac initial data in both the cases is

$$U_j^0 = h^{-1}\delta_{j,0} \equiv \begin{cases} h^{-1}, & j = 0, \\ 0, & \text{otherwise.} \end{cases}$$

The objective of the error analysis will be to quantify the error $U_j^N - u(x_j, 1)$ for $N = 1/k$. First, however, there is an important point to clarify, which is (2.3) and (2.4) do not have unique solutions because the homogeneous equation

$$\left(I + \frac{1}{2}D\right) V_j = 0$$

has nontrivial solutions of the form $V_j = w^j$, where w satisfies the quadratic equation

$$w - \frac{1}{2}d(w^2 - 2w + 1) + \frac{1}{2}r(w^2 - 1) = 0.$$

It can be shown that one root has magnitude greater than unity, leading to exponential growth as $j \to \infty$, while the other has magnitude less than unity, leading to exponential growth as $j \to -\infty$. Hence, there is at most one solution of (2.3) or (2.4) which remains bounded. We will now show that such a bounded solution does exist, and thus by requiring boundedness we obtain a unique solution.

To construct this bounded solution, we consider the use of periodic boundary conditions $U_{j+J}^n = U_j^n$ for $j = -J/2, -J/2 + 1$ (with J even) and all n. Using the discrete Fourier transform pair

$$U_j^n = \frac{1}{hJ} \sum_{m=-J/2+1}^{J/2} \widehat{U}_m^n \, \mathrm{e}^{\mathrm{i}j\theta_m},$$

$$\widehat{U}_m^n = h \sum_{j=-J/2+1}^{J/2} U_j^n \, \mathrm{e}^{-\mathrm{i}j\theta_m},$$

where $\theta_m = m\,\Delta\theta = 2\pi m/J$, the Fourier transform of (2.3) gives

$$\widehat{U}_m^{n+1} = \frac{1 - \frac{1}{2}\mathrm{i}r\sin\theta_m - 2d\sin^2\frac{\theta_m}{2}}{1 + \frac{1}{2}\mathrm{i}r\sin\theta_m + 2d\sin^2\frac{\theta_m}{2}} \widehat{U}_m^n$$

for $n \geqslant R$, where R is the number of initial Crank–Nicolson timesteps replaced by $2R$ half-timesteps of backward Euler time integration, while for $n < R$ the Fourier transform of (2.4) gives

$$\widehat{U}_m^{n+1} = \frac{1}{\left(1 + \frac{1}{2}\mathrm{i}r\sin\theta_m + 2d\sin^2\frac{\theta_m}{2}\right)^2} \widehat{U}_m^n.$$

These can be combined to give

$$\widehat{U}_m^n = z_1^n(\theta_m) z_2^{\min(n,R)}(\theta_m) \widehat{U}_m^0,$$

where

$$z_1(\theta) = \left(1 - \frac{1}{2}\mathrm{i}r\sin\theta_m - 2d\sin^2\frac{\theta_m}{2}\right)\left(1 + \frac{1}{2}\mathrm{i}r\sin\theta_m + 2d\sin^2\frac{\theta_m}{2}\right)^{-1},$$

$$z_2(\theta) = \left(1 - \frac{1}{2}\mathrm{i}r\sin\theta_m - 2d\sin^2\frac{\theta_m}{2}\right)^{-1}\left(1 + \frac{1}{2}\mathrm{i}r\sin\theta_m + 2d\sin^2\frac{\theta_m}{2}\right)^{-1}.$$

For the Dirac initial data, $\widehat{U}_m^0 = 1$ and hence

$$\begin{aligned} U_j^n &= \frac{1}{2\pi h}\sum_{m=-J/2+1}^{J/2} z_1^n(\theta_m) z_2^{\min(n,R)}(\theta_m)\mathrm{e}^{\mathrm{i}j\theta_m}\,\Delta\theta \\ &\longrightarrow \frac{1}{2\pi h}\int_{-\pi}^{\pi} z_1^n(\theta) z_2^{\min(n,R)}(\theta)\mathrm{e}^{\mathrm{i}j\theta}\,\mathrm{d}\theta \end{aligned}$$

as $J \to \infty$ with h held fixed; the limit clearly exists because of the continuity of $z_1(\theta)$ and $z_2(\theta)$. By making the substitutions $\theta = \kappa h$ and $x_j = jh$, the last integral can also be expressed as

$$U_j^n = \frac{1}{2\pi}\int_{-\pi/h}^{\pi/h} \widehat{U}^n(\kappa)\mathrm{e}^{\mathrm{i}\kappa x_j}\,\mathrm{d}\kappa, \tag{2.5}$$

where

$$\widehat{U}^n(\kappa) = z_1^n(\kappa h) z_2^{\min(n,R)}(\kappa h).$$

This has the correct initial data and satisfies the discrete equations on an infinite domain since for each j the periodic solution satisfies the discrete equations for all $J > j$. It is also bounded because of the discrete Parseval identity

$$h\sum_{j=-J/2+1}^{J/2} |U_j^n|^2 = \frac{1}{2\pi h}\sum_{m=-J/2+1}^{J/2} |\widehat{U}_m^n|^2\,\Delta\theta,$$

which in the limit $J \to \infty$ becomes

$$h\sum_j |U_j^n|^2 = \frac{1}{2\pi}\int_{-\pi/h}^{\pi/h} |\widehat{U}^n(\kappa)|^2\,\mathrm{d}\kappa.$$

A final comment is that the same transform pair of (2.5) together with

$$\widehat{U}^n(\kappa) = h\sum_{j=-\infty}^{\infty} U_j^n\,\mathrm{e}^{-\mathrm{i}\kappa x_j} \tag{2.6}$$

applies equally to the forward Euler error analysis in Section 3. Note that $\widehat{U}^n(\kappa)$ is periodic with period $2\pi/h$, which corresponds to the integral in (2.5) being over the single period $[-\pi/h, \pi/h]$.

3. Forward Euler error analysis

3.1 *Analysis of Fourier transform error*

The Fourier transform of (2.2) yields

$$\widehat{U}^N(\kappa) = z^N(\kappa),$$

where

$$z(\kappa) = 1 - \mathrm{i}ad h \sin\kappa h - 4d \sin^2 \frac{\kappa h}{2}.$$

We now compare this to the analytic solution $u(x, 1)$, whose Fourier transform is

$$\widehat{u}(\kappa, 1) = \exp(-\mathrm{i}a\kappa - \kappa^2).$$

The comparison is split over two wave number regions: a low wave number region in which $|\kappa| < h^{-m}$ for some constant m satisfying the constraint $0 < m < \frac{1}{2}$ and a high wave number region $h^{-m} < |\kappa| < \pi/h$.

PROPOSITION 3.1 (Low wave number region) For $|\kappa| < h^{-m}$, as $h \to 0$ with $Nk = 1$ and $d = k/h^2$ held fixed,

$$\widehat{U}^N(\kappa) - \widehat{u}(\kappa, 1) = h^2 \exp(-\mathrm{i}a\kappa - \kappa^2)\{p(a, d; \kappa) + \mathrm{O}(h^2(\kappa^2 + \kappa^8))\},$$

where

$$p(a, d; \kappa) = \frac{1}{2}da^2\kappa^2 + \left(\frac{1}{6} - d\right)\mathrm{i}a\kappa^3 + \left(\frac{1}{12} - \frac{1}{2}d\right)\kappa^4. \tag{3.1}$$

Proof. Performing a Taylor series expansion,

$$\log \widehat{U}^N = N \log z = -\mathrm{i}a\kappa - \kappa^2 + h^2 p(a, d; \kappa) + \mathrm{O}(h^4(\kappa^2 + \kappa^8)).$$

The restriction $m < \frac{1}{2}$ ensures that both the leading-order error term and the remainder term tend to zero as $h \to 0$, and hence

$$\begin{aligned}\widehat{U}^N &= \exp(-\mathrm{i}a\kappa - \kappa^2)\left\{1 + h^2 p(a, d; \kappa) + \mathrm{O}(h^4(\kappa^2 + \kappa^8))\right\} \\ &= \widehat{u}(\kappa, 1) + h^2 \exp(-\mathrm{i}a\kappa - \kappa^2)\left\{p(a, d; \kappa) + \mathrm{O}(h^2(\kappa^2 + \kappa^8))\right\}.\end{aligned}$$

A more detailed proof (Giles, 2004) gives precise bounds on the remainder term. □

PROPOSITION 3.2 (High wave number region) For $h^{-m} < |\kappa| < \pi/h$, as $h \to 0$ with $Nk = 1$ and $d = k/h^2$ held fixed with $d < \frac{1}{2}$, $\widehat{U}^N = \mathrm{o}(h^q)$ for any $q > 0$.

Proof. For h sufficiently small so that $|a|h < 1$,

$$\begin{aligned}|z|^2 &= 1 - 8d \sin^2 \frac{\kappa h}{2}\left(1 - 2d \sin^2 \frac{\kappa h}{2} - \frac{1}{2}(ah)^2 d \cos^2 \frac{\kappa h}{2}\right) \\ &\leqslant 1 - 8d(1 - 2d) \sin^2 \frac{\kappa h}{2}.\end{aligned}$$

Setting $\theta = \kappa h$, then since $\sin^2(\theta/2) \geqslant (\theta/\pi)^2$ for $\theta \in [0, \pi]$, it follows that for $|\kappa| \leqslant \pi/h$,

$$|z|^2 \leqslant 1 - \frac{8d(1-2d)\kappa^2 h^2}{\pi^2} \leqslant \exp\left(-\frac{8d(1-2d)\kappa^2 h^2}{\pi^2}\right),$$

and hence for fixed $Nk = Ndh^2 = 1$, we obtain

$$|\widehat{U}^N(\kappa)| \leqslant \exp\left(-\frac{4(1-2d)\kappa^2}{\pi^2}\right) \leqslant \exp\left(-\frac{4(1-2d)h^{-2m}}{\pi^2}\right).$$

□

3.2 *l_∞ and l_1 error estimates*

The inverse Fourier transform of $\exp(-\mathrm{i}a\kappa - \kappa^2)\,p(a, d; \kappa)$ is

$$e(x) = -\frac{d}{4\sqrt{2}}a^2 N^{(2)}\left(\frac{x-a}{\sqrt{2}}\right) - \left(\frac{1}{24} - \frac{d}{4}\right)aN^{(3)}\left(\frac{x-a}{\sqrt{2}}\right) + \left(\frac{1}{48\sqrt{2}} - \frac{d}{8\sqrt{2}}\right)N^{(4)}\left(\frac{x-a}{\sqrt{2}}\right), \tag{3.2}$$

where $N^{(m)}$ denotes the mth derivative of the normal distribution $N(x)$. Using the notation $q(h) \simeq r(h)$ to denote that

$$\frac{q(h)}{r(h)} - 1 = \mathrm{O}(h) \quad \text{as } h \to 0,$$

we obtain the following bounds for the l_∞- and l_1-norms of the error $U_j^N - u(x_j, 1)$ at the discrete grid points.

PROPOSITION 3.3 For the discretization (2.2), with fixed $d = k/h^2 < \frac{1}{2}$ as $h \to 0$,

$$\|U_j^N - u(x_j, 1)\|_{l_\infty} \simeq h^2 \|e\|_{L_\infty}$$

and

$$\|U_j^N - u(x_j, 1)\|_{l_1} \simeq h^2 \|e\|_{L_1},$$

except for the specific case $a = 0$ and $d = \frac{1}{6}$ for which $e(x)$ is identically zero.

Proof. We outline the proof; for additional details see Giles (2004). For any grid point x_j, the inverse Fourier transform gives

$$U_j^N - u(x_j, 1) = \frac{1}{2\pi}\int_{|\kappa|<h^{-m}} (\widehat{U}^N(\kappa) - \widehat{u}(\kappa, 1))\mathrm{e}^{\mathrm{i}\kappa x_j}\,\mathrm{d}\kappa + \frac{1}{2\pi}\int_{h^{-m}<|\kappa|<\pi/h} (\widehat{U}^N(\kappa) - \widehat{u}(\kappa, 1))\mathrm{e}^{\mathrm{i}\kappa x_j}\,\mathrm{d}\kappa - \frac{1}{2\pi}\int_{\pi/h<|\kappa|} \widehat{u}(\kappa, 1)\mathrm{e}^{\mathrm{i}\kappa x_j}\,\mathrm{d}\kappa.$$

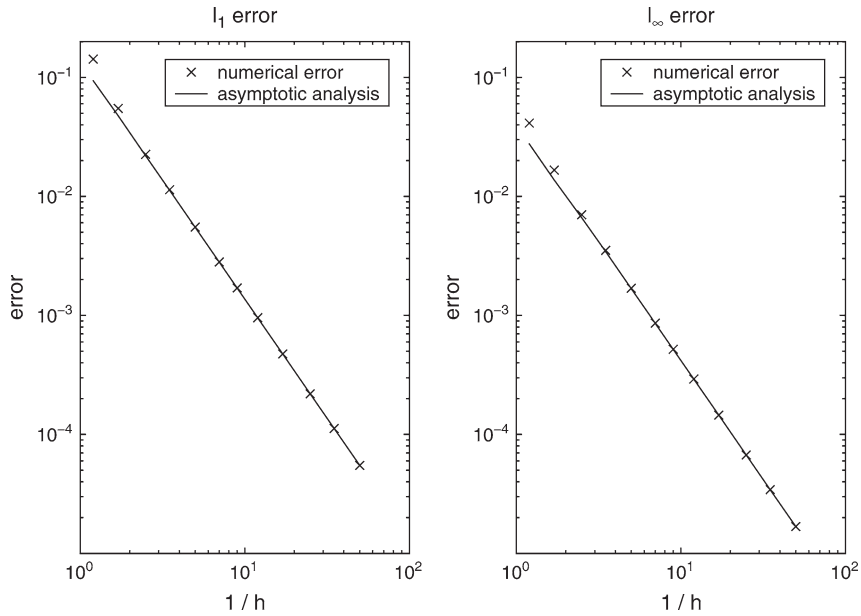

FIG. 3. Convergence results for the explicit discretization.

Both the last two integrals give contributions which are $o(h^q)$ for any $q > 0$, and hence

$$\begin{aligned} U_j^N - u(x_j, 1) &= \frac{1}{2\pi} \int_{|\kappa|<h^{-m}} h^2 \exp(-\mathrm{i}a\kappa - \kappa^2)\, p(a, d; \kappa) \mathrm{e}^{\mathrm{i}\kappa x_j}\, \mathrm{d}\kappa + \mathrm{O}(h^4) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} h^2 \exp(-\mathrm{i}a\kappa - \kappa^2)\, p(a, d; \kappa) \mathrm{e}^{\mathrm{i}\kappa x_j}\, \mathrm{d}\kappa + \mathrm{O}(h^4) \\ &= h^2 e(x_j) + \mathrm{O}(h^4). \end{aligned}$$

The l_∞ error bound follows immediately; the l_1 error bound requires the additional observation that due to the explicit nature of the discretization, U_j^N has compact support in a region of width $2Nh = \mathrm{O}(h^{-1})$, and $u(x, 1)$ is negligibly small outside this region. □

Figure 3 presents convergence results obtained on the truncated domain $-10 < x < 10$ and the time interval $0 < t < 1$ using $a = 2$. As the grid spacing h is reduced, the timestep is related to the grid spacing h through $k = dh^2$ with $d = 1/8$. For all but the very largest values of h, there is very good agreement between the numerical errors and the asymptotic analysis of Proposition 3.3.

4. Crank–Nicolson and Rannacher error analysis

The Fourier analysis in Section 2 gives the solution at the final iteration level $N = 1/k$ (assumed to be greater than R, the number of Crank–Nicolson timesteps replaced by two half-timesteps of backward Euler time marching) as

$$U_j^N = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} \widehat{U}^N(\kappa) \mathrm{e}^{\mathrm{i}\kappa x_j}\, \mathrm{d}\kappa,$$

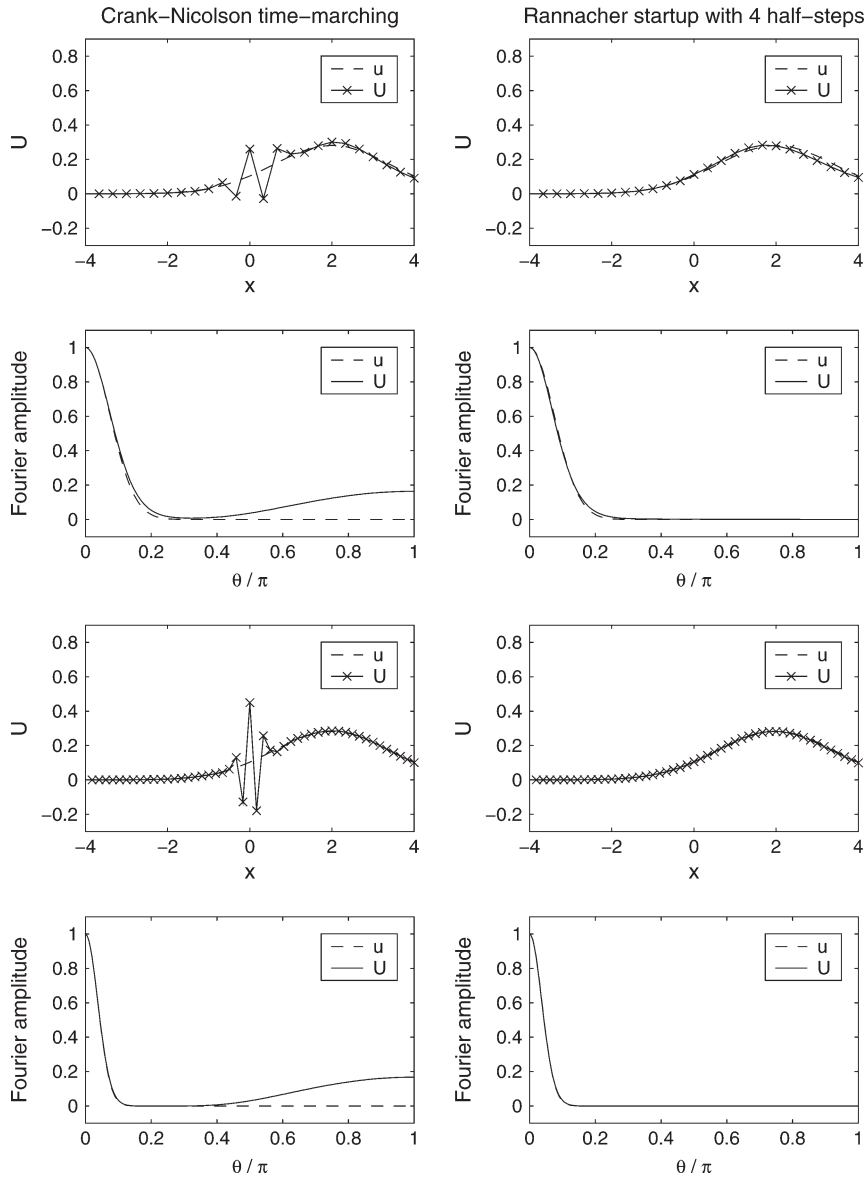

FIG. 4. Numerical results for the implicit discretizations.

where

$$\widehat{U}^N(\kappa) = z_1^N(\theta) z_2^R(\theta),$$

with $\theta = \kappa h$ as before.

Figure 4 plots comparisons between the numerical and analytic solutions to the convection–diffusion problem with $a = 2$ at the final time $t = 1$ for two grid resolutions, $h = 1/3$ for the upper half of the

figure and $h = 1/6$ for the lower half. The timestep is chosen so that $\lambda = k/h = 3/4$ in each case. The plots on the left are for Crank–Nicolson without any Rannacher startup, whereas the plots on the right are for $R = 2$, replacing the first two Crank–Nicolson timesteps by four half-timesteps of backward Euler integration.

The main feature of the results in physical space (i.e. the plots of U and u versus x) is the high wave number error near $x = 0$ for the Crank–Nicolson time marching. Asymptotic analysis will show that its width is proportional to h and its magnitude is proportional to h^{-1}. Looking at the comparison in Fourier space (i.e. the plots of $|\widehat{U}|$ and $|\widehat{u}|$ versus $\theta = \kappa h$), in the Crank–Nicolson results there appear to be three regions: an $O(h)$ region on the left in which $\widehat{u} \approx \widehat{U}$, an $O(1)$ region on the right in which $\widehat{u} \ll 1$ but $\widehat{U} = O(1)$ and a central region in which both $\widehat{u} \ll 1$ and $\widehat{U} \ll 1$. This is the basis for the asymptotic analysis, which considers a low wave number range defined by $|\kappa| < h^{-m}$, a high wave number range defined by $h^{-q} < |\kappa| < \pi/h$ and the intermediate range $h^{-m} < |\kappa| < h^{-q}$, with m and q satisfying the constraints $0 < m < \frac{1}{3}$ and $\frac{1}{2} < q < 1$ for reasons to be explained later.

The convergence analysis considers the limit $h, k \to 0$ with $\lambda = k/h$ held fixed. The reason for this choice of limit is that the truncation error due to the spatial central differencing and the Crank–Nicolson time integration is $O(k^2 + h^2)$, and so $k = \lambda h$ keeps the spatial and temporal approximation errors of the same order. We now analyse the Fourier error $\widehat{U} - \widehat{u}$ in each of the three regions.

PROPOSITION 4.1 (Low wave number region) For $|\kappa| < h^{-m}$, as $h \to 0$ with $\lambda = k/h$ held fixed,

$$\widehat{U}^N(\kappa) - \widehat{u}(\kappa, 1) = h^2 \exp(-\mathrm{i}a\kappa - \kappa^2) \left\{ p(a, \lambda, R; \kappa) + O(h(\kappa^3 + \kappa^9)) \right\},$$

where

$$p(a, \lambda, R; \kappa) = \frac{1}{6}\mathrm{i}a\kappa^3 + \frac{1}{12}\kappa^4 - \frac{1}{12}\lambda^2\kappa^3(\mathrm{i}a + \kappa)^3 + \frac{1}{4}R\lambda^2\kappa^2(\mathrm{i}a + \kappa)^2.$$

Proof. Setting $\theta = \kappa h$, $N = \frac{1}{k} = \frac{1}{\lambda h}$, $r = a\lambda$ and $d = \frac{\lambda}{h}$, a Taylor series expansion in h gives

$$\log \widehat{U}^N = N \log z_1 + R \log z_2 = -\mathrm{i}a\kappa - \kappa^2 + h^2 p(a, \lambda, R; \kappa) + O(h^3(\kappa^3 + \kappa^9)).$$

The restriction that $m < \frac{1}{3}$ ensures that the $h^2\kappa^6$ term and the $h^3\kappa^9$ remainder both tend to zero as $h \to 0$, and hence

$$\widehat{U}^N = \exp(-\mathrm{i}a\kappa - \kappa^2) \left\{ 1 + h^2 p(a, \lambda, R; \kappa) + O(h^3(\kappa^3 + \kappa^9)) \right\}.$$

□

PROPOSITION 4.2 (High wave number region) For $h^{-q} < |\kappa| < \pi/h$, as $h \to 0$ with $\lambda = k/h$ held fixed and with $\theta = \kappa h$,

$$\widehat{U}^N = (-1)^{N-R} \frac{h^{2R}}{\left(2\lambda \sin^2 \frac{\theta}{2}\right)^{2R}} \exp\left(-\frac{1}{\lambda^2 \sin^2 \frac{\theta}{2}} \right) (1 + O(h\theta^{-2})).$$

Proof.

$$z_1(\theta) = \left(1 - \frac{1}{2}\mathrm{i}r\sin\theta - 2d\sin^2\frac{\theta}{2}\right)\left(1 + \frac{1}{2}\mathrm{i}r\sin\theta + 2d\sin^2\frac{\theta}{2}\right)^{-1}$$

$$= \left(\frac{1}{2d\sin^2\frac{\theta}{2}} - \frac{\mathrm{i}r}{2d}\cot\frac{\theta}{2} - 1\right)\left(\frac{1}{2d\sin^2\frac{\theta}{2}} + \frac{\mathrm{i}r}{2d}\cot\frac{\theta}{2} + 1\right)^{-1}$$

$$\longrightarrow -1 \quad \text{as } d \to \infty,$$

and similarly

$$z_2(\theta) = \left(2d\sin^2\frac{\theta}{2}\right)^{-2}\left(\frac{1}{2d\sin^2\frac{\theta}{2}} - \frac{\mathrm{i}r}{2d}\cot\frac{\theta}{2} - 1\right)^{-1}\left(\frac{1}{2d\sin^2\frac{\theta}{2}} + \frac{\mathrm{i}r}{2d}\cot\frac{\theta}{2} + 1\right)^{-1}$$

$$\longrightarrow -\left(2d\sin^2\frac{\theta}{2}\right)^{-2} \quad \text{as } d \to \infty.$$

Hence, expressing d and N as functions of h as in the proof of Proposition 4.1, Taylor series analysis gives

$$\log\{(-1)^{N-R}\widehat{U}^N\} = 2R\log\frac{h}{2\lambda\sin^2\frac{\theta}{2}} - \frac{1}{\lambda^2\sin^2\frac{\theta}{2}} + \mathrm{O}\left(\frac{h}{\sin^2\frac{\theta}{2}}\right).$$

The restriction that $q > \frac{1}{2}$ ensures that the remainder term tends to zero as $h \to 0$, and therefore we obtain the result in the proposition. □

PROPOSITION 4.3 (Intermediate region) For $h^{-m} < |\kappa| < h^{-q}$, as $h \to 0$ with $\lambda = k/h$ held fixed, $\widehat{U}^N(\kappa) = \mathrm{o}(h^r)$ for any $r > 0$.

Proof. Defining $s = \sin^2\frac{\theta}{2}$,

$$|z_1|^2 = \frac{(1-ds)^2 + r^2 s(1-s)}{(1+ds)^2 + r^2 s(1-s)}.$$

Differentiating, one finds that $\mathrm{d}|z_1|^2/\mathrm{d}s = 0$ when $s^2 = (d^2 - r^2)^{-1}$. Substituting $r = a\lambda$ and $d = \frac{\lambda}{h}$, this shows that as $h \to 0$, $|z_1|$ has a maximum at $s = 0, 1$ and a minimum at $s \approx d^{-1}$, corresponding to $\kappa = \mathrm{O}(h^{-1/2})$ which lies within the intermediate region. Noting that for any $r > 0$, Propositions 4.1 and 4.2 prove that $|z_1|^N = \mathrm{o}(h^r)$ at both $\kappa = h^{-m}$ and $\kappa = h^{-q}$, it follows that $|z_1|^N = \mathrm{o}(h^r)$ within the entire intermediate region. Since $|z_1^N z_2^R| < |z_1|^{N-R}$, it follows that $\widehat{U}^N = \mathrm{o}(h^r)$ for any $r > 0$. □

Defining

$$\widehat{E}^{\mathrm{low}} = h^2\exp(-\mathrm{i}a\kappa - \kappa^2)p(a, \lambda, R; \kappa)$$

and

$$\widehat{E}^{\mathrm{high}} = (-1)^{N-R}\frac{h^{2R}}{\left(2\lambda\sin^2\frac{\theta}{2}\right)^{2R}}\exp\left(\frac{1}{\lambda^2\sin^2\frac{\theta}{2}}\right),$$

then since $\widehat{E}^{\text{low}} \ll \widehat{E}^{\text{high}}$ in the high wave number region and $\widehat{E}^{\text{high}} \ll \widehat{E}^{\text{low}}$ in the low wave number region, the results above can be combined to give

$$\widehat{U}^N(\kappa) - \widehat{u}(\kappa, 1) \approx \widehat{E}^{\text{low}} + \widehat{E}^{\text{high}}, \quad |\kappa| < \pi/h.$$

The inverse Fourier transform then gives

$$U_j^N - u(x_j, 1) \approx E_j^{\text{low}} + E_j^{\text{high}},$$

where the low wave number error is

$$E_j^{\text{low}} = h^2 \left\{ \frac{Ra^2\lambda^2}{8\sqrt{2}} N^{(2)}\left(\frac{x_j - a}{\sqrt{2}}\right) - \frac{2a + a^3\lambda^2 + 6Ra\lambda^2}{48} N^{(3)}\left(\frac{x_j - a}{\sqrt{2}}\right) \right.$$
$$\left. + \frac{1 + 3a^2\lambda^2 + 3R\lambda^2}{48\sqrt{2}} N^{(4)}\left(\frac{x_j - a}{\sqrt{2}}\right) - \frac{a\lambda^2}{32} N^{(5)}\left(\frac{x_j - a}{\sqrt{2}}\right) + \frac{\lambda^2}{96\sqrt{2}} N^{(6)}\left(\frac{x_j - a}{\sqrt{2}}\right) \right\}$$

and the high wave number error is

$$E_j^{\text{high}} = (-1)^{N-R} h^{2R-1} (2\lambda)^{-2R} f_j,$$

where

$$f_j = \frac{h}{2\pi} \int_{-\pi/h}^{\pi/h} \frac{\mathrm{e}^{-\mathrm{i}\kappa x_j}}{\sin^{4R}\frac{\theta}{2}} \exp\left(-\frac{1}{\lambda^2 \sin^2\frac{\theta}{2}}\right) \mathrm{d}\kappa$$
$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\mathrm{e}^{-\mathrm{i}j\theta}}{\sin^{4R}\frac{\theta}{2}} \exp\left(-\frac{1}{\lambda^2 \sin^2\frac{\theta}{2}}\right) \mathrm{d}\theta$$
$$= \frac{1}{\pi} \int_{0}^{\pi} \frac{\cos j\theta}{\sin^{4R}\frac{\theta}{2}} \exp\left(-\frac{1}{\lambda^2 \sin^2\frac{\theta}{2}}\right) \mathrm{d}\theta.$$

E_j^{high} clearly has a width which is $\mathrm{O}(h)$, and has a maximum magnitude at $j = 0$ where $x_j = 0$, which explains the observed behaviour in Fig. 4. The integral for $j = 0$ can be evaluated analytically (see Appendix) giving

$$\max_j |E_j^{\text{high}}| = |E_0^{\text{high}}| = h^{2R-1}(2\lambda)^{-2R} \frac{\mathrm{d}^{2R}}{\mathrm{d}\beta^{2R}} \operatorname{erfc}(\sqrt{\beta}),$$

where $\beta = \lambda^{-2}$ and $\operatorname{erfc}(x)$ is the complementary error function.

The fact that the low wave number is $\mathrm{O}(h^2)$ and the high wave number error is $\mathrm{O}(h^{2R-1})$ is confirmed by the results in the upper plots of Fig. 5, which show convergence results for the convection–diffusion case with $a = 2$. It can be seen that for the standard Crank–Nicolson time marching, the results exhibit $\mathrm{O}(h^2)$ convergence until h reaches a sufficiently small value so that the $\mathrm{O}(h^{-1})$ high wave number error becomes dominant. The plots show the sensitive dependence of the high wave number error

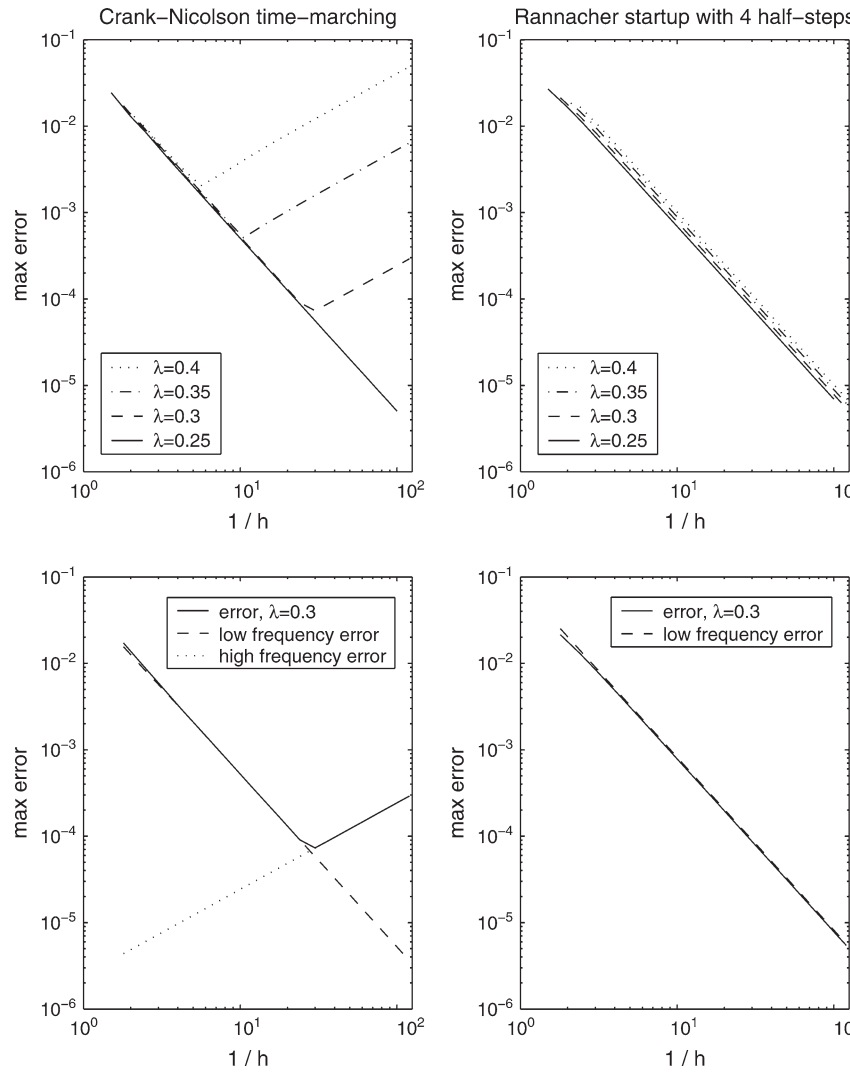

FIG. 5. Convergence results for the implicit discretizations.

on the value of λ. For large values of λ, $\operatorname{erfc}(\lambda^{-1}) \approx 1$ and so E_j^{high} becomes significant for quite large values of h. On the other hand, for small values of λ, $\operatorname{erfc}(\lambda^{-1})$ is extremely small, and so E_j^{high} does not become dominant until h is extremely small. With the Rannacher startup with four half-timesteps of backward Euler integration ($R = 2$), the high wave number error is $O(h^3)$ and so the low wave number error remains dominant for all h. The sharpness of the error analysis is demonstrated by the lower plots in the figure, which compare the numerical error with the maximum magnitude of E_j^{low} and E_j^{high}.

5. Extensions

5.1 *Diffusion coefficient and terminal time*

The error analysis of the convection–diffusion equation

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = \epsilon \frac{\partial^2 u}{\partial x^2}$$

and the terminal time $t = T$ is handled through the nondimensionalization

$$\bar{t} = \frac{t}{T}, \quad \bar{x} = \frac{x}{\sqrt{\epsilon T}}, \quad \bar{k} = \frac{k}{T}, \quad \bar{h} = \frac{h}{\sqrt{\epsilon T}}, \quad \bar{a} = \sqrt{\frac{T}{\epsilon}}\, a, \quad \overline{U}_j^n = \sqrt{\epsilon T}\, U_j^n,$$

which reduces the more general problem to the one which has already been analysed.

5.2 *Alternative initial data*

The analysis so far has assumed that the grid is perfectly aligned with the Dirac initial data at $x = 0$. Suppose instead that the grid points are at $x_j = (j-\alpha)h$ with $0 < \alpha < 1$. The appropriate discretization of the initial data in this case is

$$U_j^0 = \begin{cases} (1-\alpha)h^{-1}, & j = 0, \\ \alpha h^{-1}, & j = 1, \\ 0, & \text{otherwise.} \end{cases}$$

The Fourier transform pair (2.5) and (2.6) remains valid for these displaced grid points, and therefore

$$\widehat{U}^0(\kappa) = (1-\alpha)\mathrm{e}^{-\mathrm{i}\alpha\kappa h} + \alpha\, \mathrm{e}^{\mathrm{i}(1-\alpha)\kappa h} = 1 + \mathrm{O}(\kappa^2 h^2).$$

This leads to the result that the low wave number error remains second order. Further analysis shows that the convergence order of the high wave number error in the implicit discretizations is also unaffected.

Although the focus of our analysis so far has been on Dirac initial data, there are other sets of initial data which are also of interest. One is the first difference of the discrete Dirac initial data. If we take the grid points to be at $x_j = (j - \frac{1}{2})h$, then we have

$$U_j^0 = \begin{cases} -h^{-2}, & j = 0, \\ h^{-2}, & j = 1, \\ 0, & \text{otherwise,} \end{cases}$$

for which

$$\widehat{U}^0(\kappa) = 2\mathrm{i}\, h^{-1} \sin\frac{\kappa h}{2},$$

leading to

$$U_j^n = \frac{\mathrm{i}}{\pi h^2} \int_{-\pi/h}^{\pi/h} \sin\frac{\kappa h}{2}\, z^n(\kappa h)\, \mathrm{e}^{\mathrm{i}\kappa x_j}\, \mathrm{d}\kappa,$$

for the forward Euler discretization.

Another is a discrete equivalent of $H(x) - \frac{1}{2}$, where $H(x)$ is the Heaviside step function. If we again use a grid with $x_j = (j - \frac{1}{2})h$, then using the initial data

$$U_j^0 = \begin{cases} -\frac{1}{2}, & j \leqslant 0, \\ \frac{1}{2}, & j > 0, \end{cases}$$

leads to

$$U_j^n = \frac{1}{4\pi \mathrm{i}} \mathrm{PV} \int_{-\pi/h}^{\pi/h} \left(\sin \frac{\kappa h}{2} \right)^{-1} z^n(\kappa h) \mathrm{e}^{\mathrm{i}\kappa x_j} \, \mathrm{d}\kappa,$$

for the forward Euler discretization, where PV denotes the Cauchy principal value about $\kappa = 0$ where the integrand is singular.

For both these sets of alternative initial data, the error in the forward Euler discretization remains $\mathrm{O}(h^2)$. With the implicit discretization, the low wave number error will still be $\mathrm{O}(h^2)$ but the high wave number error will be one order worse in the first case, $O(h^{-2+2R})$ where R is again the number of Crank–Nicolson timesteps replaced by two half-timesteps of backward Euler integration, and one order better in the second case, $\mathrm{O}(h^{2R})$.

6. Conclusions

In this paper, we have derived sharp estimates of the error arising from explicit and implicit discretizations of the constant-coefficient 1D convection–diffusion equation subject to Dirac initial data.

The extension of the Fourier analysis to multiple dimensions would pose no particular difficulties. To extend the analysis to varying coefficients would not be so easy, but could be performed using a matched inner and outer asymptotic analysis, with the inner analysis in the neighbourhood of the Dirac initial data being performed using the analysis in this paper, treating the coefficients as being locally approximately constant. The inner solution would then have to be matched to an outer solution describing the subsequent evolution of the solution and the discretization error in the outer region in which the solution is well resolved, at least asymptotically.

Regarding the use of Rannacher time-stepping, replacing each of the first R Crank–Nicolson timesteps by two half-timesteps of backward Euler integration, the analysis proves, and the numerical results confirm, that there is a low wave number error which is $\mathrm{O}(h^2)$ and a high wave number error which is $\mathrm{O}(h^{2R-1})$. Hence, $R = 2$ is the minimum to give $\mathrm{O}(h^2)$ convergence, and it is likely to be the optimum in general since larger values will increase the low wave number error.

References

BARTH, T. & DECONINCK, H. (eds) (2002) *Error Estimation and Adaptive Discretization Methods in Computational Fluid Dynamics*. Lecture Notes in Computational Science and Engineering, vol. 25. New York: Springer.

BECKER, R. & RANNACHER, R. (2001) An optimal control approach to error control and mesh adaptation. *Acta Numerica 2001* (A. Iserles ed.). Cambridge: Cambridge University Press, pp. 1–102.

DARMOFAL, D. & VENDITTI, D. (2003) Anisotropic grid adaptation for functional outputs: application to two-dimensional viscous flows. *J. Comput. Phys.*, **187**, 22–46.

GILES, M. B. (2003) Discrete adjoint approximations with shocks. *Hyperbolic Problems: Theory, Numerics, Applications* (T. Hou & E. Tadmor eds). New York: Springer, pp. 185–194.

GILES, M. B. (2004) Sharp error estimates for a discretisation of the 1D convection/diffusion equation with Dirac initial data. *Technical Report NA04/17*. Oxford: Oxford University Computing Laboratory.

GILES, M. B. & SÜLI, E. (2002) Adjoint methods for PDEs: *a posteriori* error analysis and postprocessing by duality. *Acta Numerica 2002* (A. Iserles ed.). Cambridge: Cambridge University Press, pp. 145–236.

JAMESON, A. (1988) Aerodynamic design via control theory. *J. Sci. Comput.*, **3**, 233–260.

JAMESON, A., PIERCE, N. & MARTINELLI, L. (1998) Optimum aerodynamic design using the Navier-Stokes equations. *J. Theor. Comput. Fluid Mech.*, **10**, 213–237.

PIERCE, N. A. & GILES, M. B. (2004) Adjoint and defect error bounding and correction for functional estimates. *J. Comput. Phys.*, **200**, 769–794.

POOLEY, D. M., FORSYTH, P. A. & VETZAL, K. R. (2003a) Numerical convergence properties of option pricing PDEs with uncertain volatility. *IMA J. Numer. Anal.*, **23**, 241–267.

POOLEY, D. M., VETZAL, K. R. & FORSYTH, P. A. (2003b) Convergence remedies for non-smooth payoffs in option pricing. *J. Comput. Financ.*, **6**, 25–40.

RANNACHER, R. (1984) Finite element solution of diffusion problems with irregular data. *Numer. Math.*, **43**, 309–327.

RICHTMYER, R. D. & MORTON, K. W. (1967) *Difference Methods for Initial-Value Problems*, 2nd edn. Wiley-Interscience. (Reprint edn, 1994, New York: Krieger Publishing Company, Malabar.)

STRANG, G. (1986) *Introduction to Applied Mathematics*. Wellesley, MA: Wellesley College.

STRIKWERDA, J. C. (1989) *Finite Difference Schemes and Partial Differential Equations*. California: Wadsworth & Brooks.

ULBRICH, S. (2002) A sensitivity and adjoint calculus for discontinuous solutions of hyperbolic conservation laws with source terms. *SIAM J. Control Optim.*, **41**, 740–797.

ULBRICH, S. (2003) Adjoint-based derivative computations for the optimal control of discontinuous solutions of hyperbolic conservation laws. *Syst. Control Lett.*, **48**, 313–328.

WILMOTT, P., HOWISON, S. & DEWYNNE, J. (1995) *The Mathematics of Financial Derivatives*. Cambridge: Cambridge University Press.

Appendix. Evaluation of the integral

Consider the integral

$$I_0 = \frac{1}{\pi}\int_0^\pi \exp\left(-\frac{1}{\lambda^2 \sin^2\frac{\theta}{2}}\right)\mathrm{d}\theta.$$

Making the substitutions $t = \cot\frac{\theta}{2}$ and $\alpha = \lambda^{-1}$, one obtains

$$I_0 = \frac{2}{\pi}\int_0^\infty \frac{1}{t^2+1}\exp(-\alpha^2(t^2+1))\mathrm{d}t,$$

and hence

$$\frac{\mathrm{d}I_0}{\mathrm{d}\alpha} = -\frac{4\alpha}{\pi}\int_0^\infty \exp(-\alpha^2(t^2+1))\mathrm{d}t = -\frac{2}{\sqrt{\pi}}\exp(-\alpha^2).$$

Since $I_0 \to 0$ as $\alpha \to \infty$, integration gives

$$I_0 = \frac{2}{\sqrt{\pi}} \int_{\lambda^{-1}}^{\infty} \exp(-s^2) \mathrm{d}s \equiv \operatorname{erfc}(\lambda^{-1}),$$

where $\operatorname{erfc}(x)$ is the complementary error function.

Switching to a new variable $\beta = \lambda^{-2} = \alpha^2$, $I_0(\beta) = \operatorname{erfc}(\sqrt{\beta})$ and

$$I_R(\beta) \equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\left(\sin^2 \frac{\theta}{2}\right)^{2R}} \exp\left(-\frac{\beta}{\sin^2 \frac{\theta}{2}}\right) \mathrm{d}\theta = \frac{\mathrm{d}^{2R} I_0}{\mathrm{d}\beta^{2R}}.$$