

# Convergence analysis of Crank–Nicolson and Rannacher time-marching

**Michael B. Giles**

Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD, UK

**Rebecca Carter**

Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD, UK

This paper presents a convergence analysis of Crank–Nicolson and Rannacher time-marching methods which are often used in finite difference discretizations of the Black–Scholes equations. Particular attention is paid to the important role of Rannacher’s startup procedure, in which one or more initial timesteps use backward Euler timestepping, to achieve second-order convergence for approximations of the first and second derivatives. Numerical results confirm the sharpness of the error analysis which is based on asymptotic analysis of the behavior of the Fourier transform. The relevance to Black–Scholes applications is discussed in detail, with numerical results supporting recommendations on how to maximize the accuracy for a given computational cost.

## 1 Introduction

Many applications in mathematical finance require the numerical solution of variants of the Black–Scholes equation (Wilmott *et al* 1995)

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

which is solved backwards in time, from a given payoff at the terminal time  $t = T$ , to an initial time  $t = 0$ . Switching to the new coordinate  $x \equiv \log S$  gives the transformed equation

$$\frac{\partial V}{\partial t} = rV - \left( r - \frac{1}{2} \sigma^2 \right) \frac{\partial V}{\partial x} - \frac{1}{2} \sigma^2 \frac{\partial^2 V}{\partial x^2}$$

Using a uniform grid with spacing  $h$  and timestep  $k$ , second-order central space differencing and Crank–Nicolson time integration results in the discrete equations

$$(I + \frac{1}{2}D)V_j^{n+1} = (I - \frac{1}{2}D)V_j^n$$

where

$$D = \frac{k}{2h^2} \sigma^2 \delta_x^2 - \frac{k}{2h} \left( r - \frac{1}{2} \sigma^2 \right) \delta_{2x} - rk$$

with  $\delta_x^2$  and  $\delta_{2x}$  being the standard second difference and central first difference operators, respectively.

For European call options, the payoff function at the terminal time is

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

The top-left plot in Figure 1 shows the numerical solution  $V(S, 0)$  at time  $t = 0$  for parameter values  $r = 0.05$ ,  $\sigma = 0.2$ ,  $K = 1$ ,  $T = 2$ . The agreement between the numerical solution and the analytic solution (Wilmott *et al* 1995) appears quite good, but the other two left-hand plots show much poorer agreement for the finite difference approximations to  $\Delta \equiv \partial V / \partial S$  and  $\Gamma \equiv \partial^2 V / \partial S^2$ . In particular, note that the maximum error in the computed value for  $\Gamma$  occurs at  $S = 1$ , which is the location of the discontinuity in the first derivative of the initial data.

Figure 2 shows corresponding results for a digital call option for which the payoff is

$$V(S, T) = H(S - K)$$

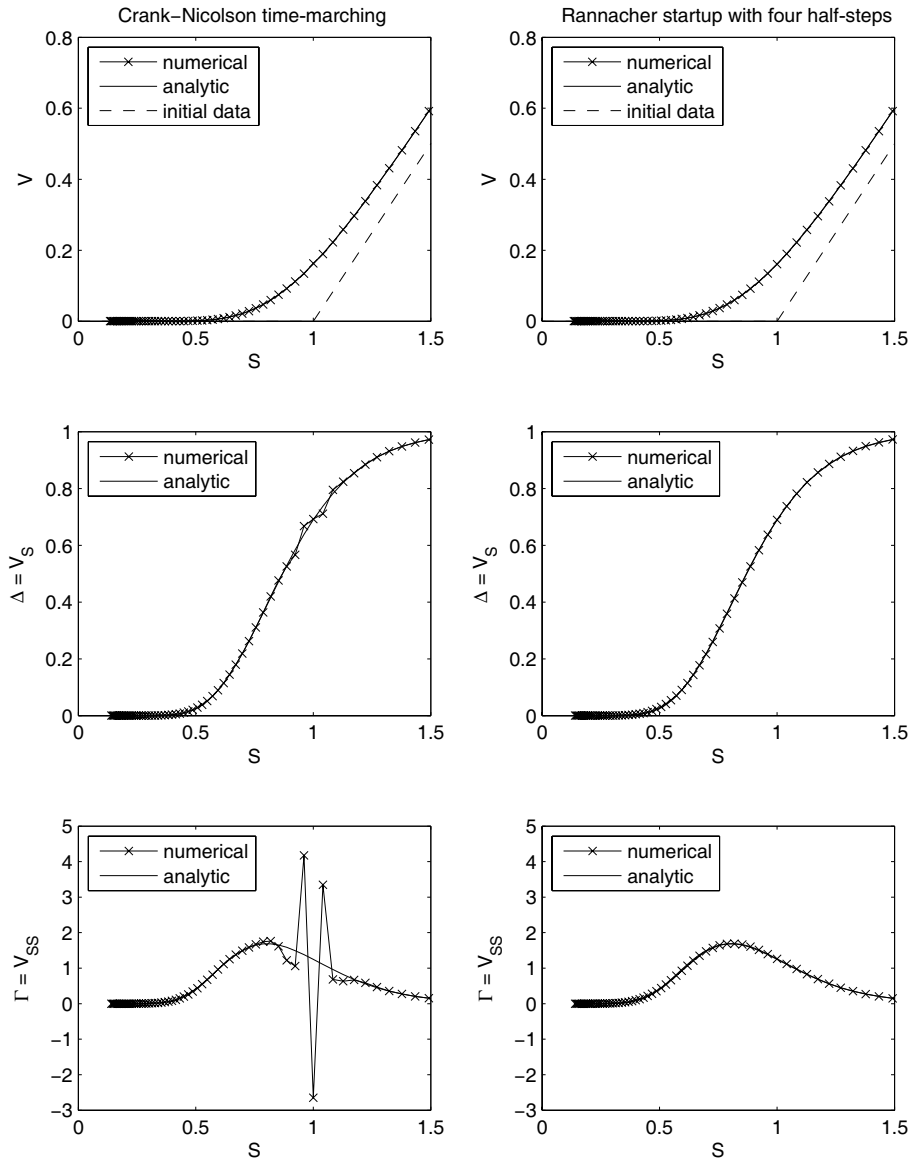
where  $H(x)$  is the Heaviside function. For this case, the accuracy of the Crank–Nicolson approximation is noticeably poorer, due to the reduced regularity in the initial data.

The left-hand plots in Figure 3 show the behavior of the maximum error for the European call option as the computational grid is refined, keeping fixed the ratio  $\lambda \equiv k/h$ . It can be seen that the numerical solution  $V_j$  exhibits first-order convergence, while the discrete approximation to  $\Delta$  does not converge, and the approximation to  $\Gamma$  diverges. The left-hand plots in Figure 4 show that for the digital call option there is no convergence even for the option value  $V$ .

At first sight, this may appear surprising. Textbooks introducing the Crank–Nicolson method for parabolic equations almost always describe it as an unconditionally stable, convergent approximation. However, this statement is a little misleading in its simplicity. It is unconditionally stable in the  $L_2$  norm. This, together with consistency, ensures convergence in the  $L_2$  norm for initial data which lies in  $L_2$  (Richtmyer and Morton 1967), although 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 Black–Scholes application, the initial data for  $V$  lies in  $L_2$ , as does its first derivative, but the second derivative is the Dirac delta function which does not lie in  $L_2$ . This then is the root cause of the observed failure to converge as the grid is refined. Furthermore, it is the maximum error, the  $L_\infty$  error, which is most relevant in financial applications.

Rannacher (1984) analyzed 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 time integration. This solution, often referred to as Rannacher timestepping, has been used with success in approximations of

FIGURE 1  $V$ ,  $\Delta$  and  $\Gamma$  for a European call option.



**FIGURE 2**  $V$ ,  $\Delta$  and  $\Gamma$  for a digital call option.

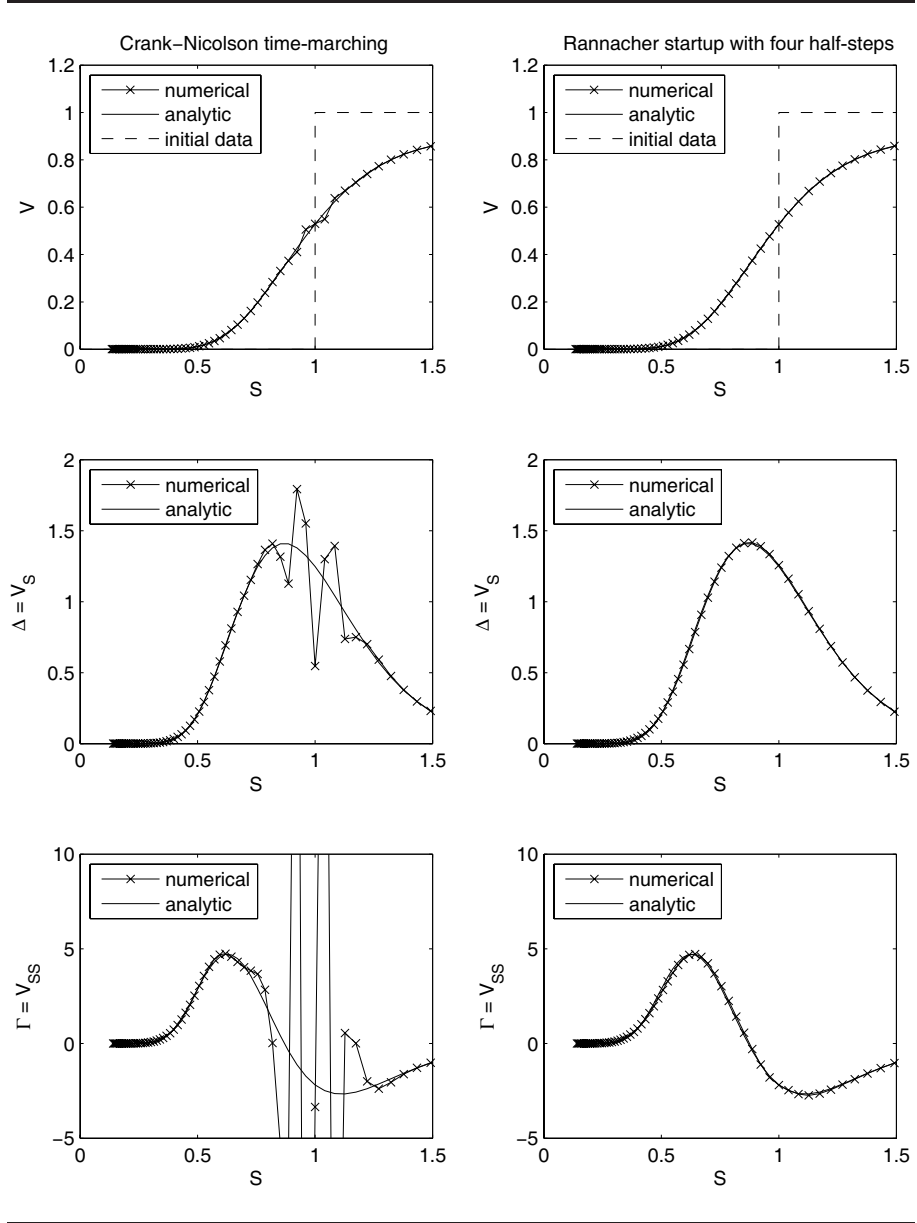


FIGURE 3 Grid convergence for a European call option.

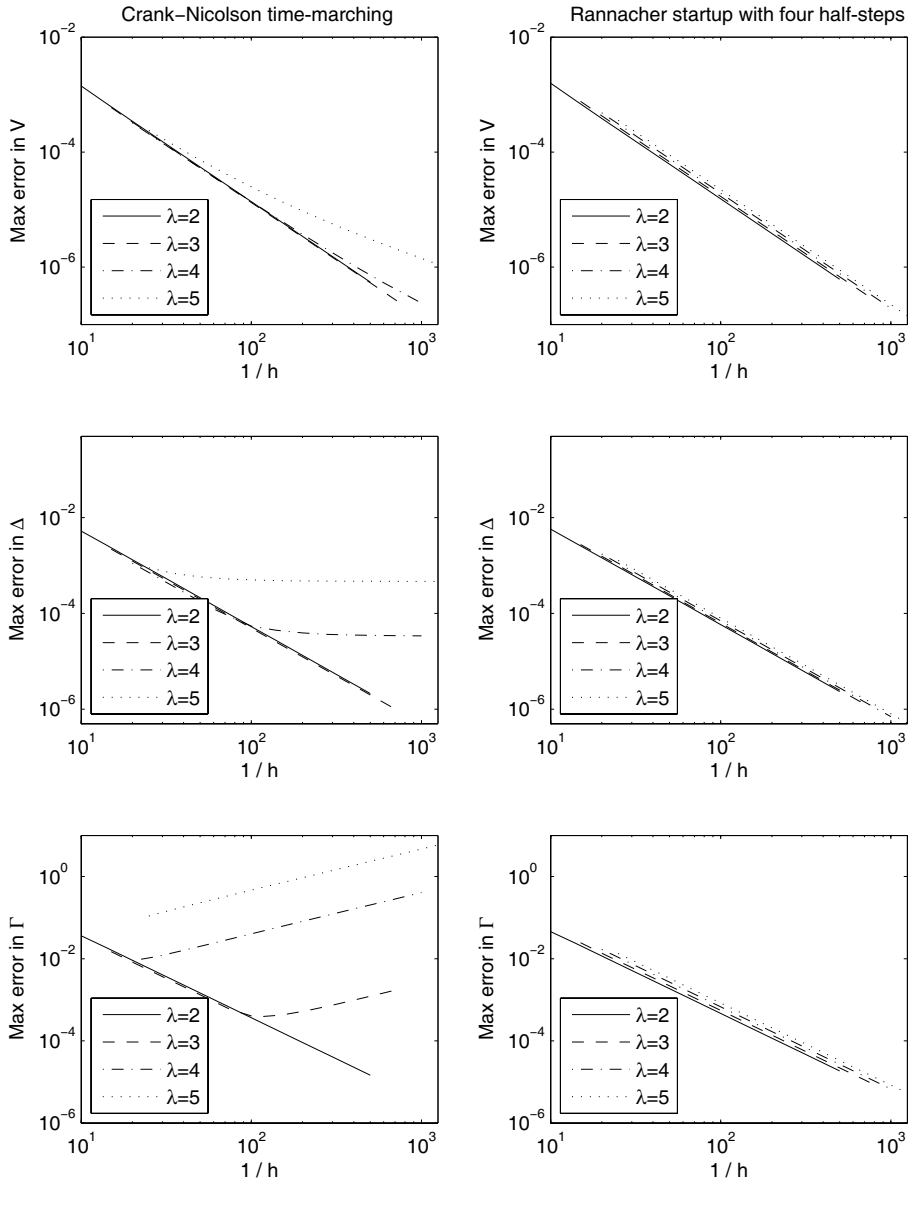
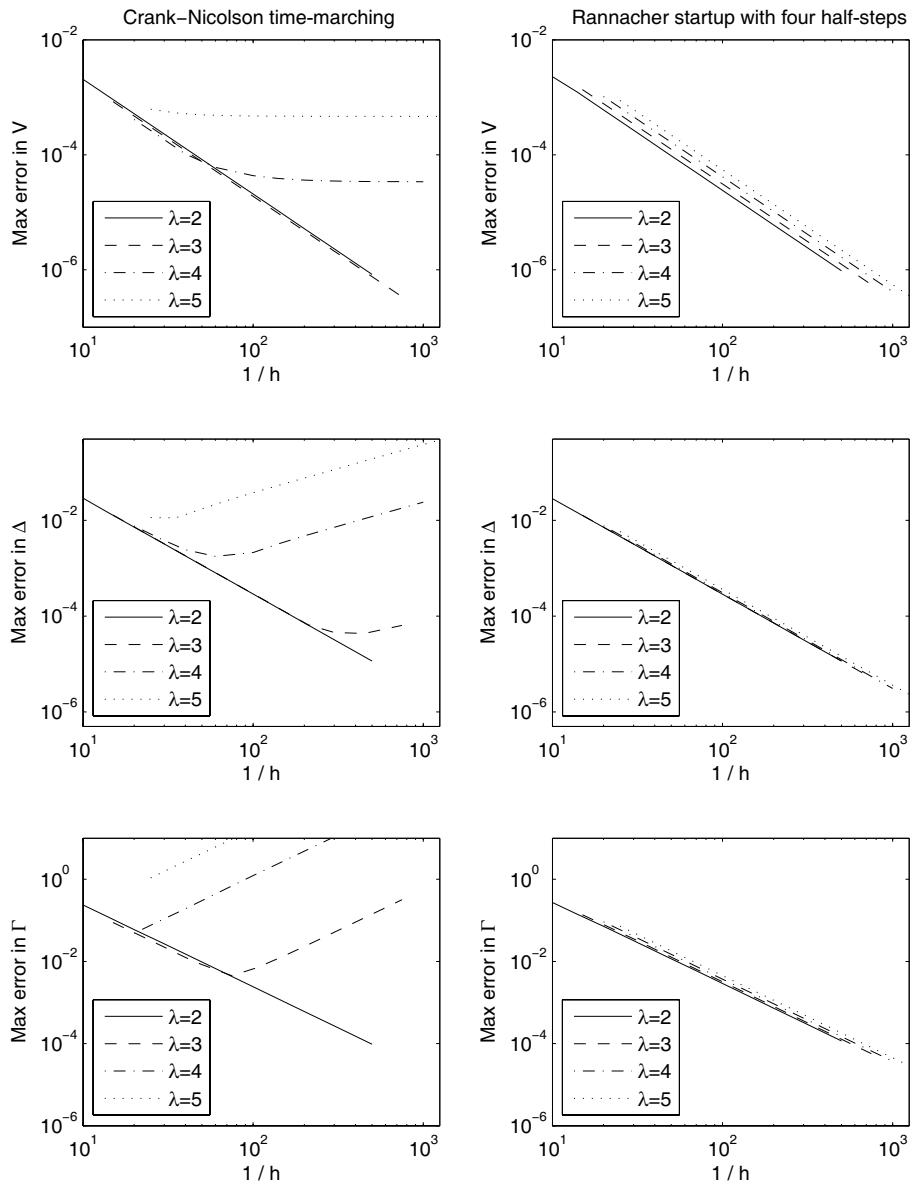


FIGURE 4 Grid convergence for a digital call option.



the Black–Scholes equations (Ikonen and Toivanen 2004; Pooley *et al* 2003a, 2003b; Raahauge 2004; Zvan *et al* 2003). The right-hand plots in Figures 1–4 show that replacing the first two Crank–Nicolson timesteps by four half-timesteps of backward Euler, for which

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

results in second-order convergence for  $V$ ,  $\Delta$  and  $\Gamma$ .

As an example to show that the behavior of these simple Black–Scholes approximations is representative of more realistic multi-factor applications, Figure 5 shows similar results for Heston’s stochastic volatility model (Heston 1993) for which the partial differential equation is

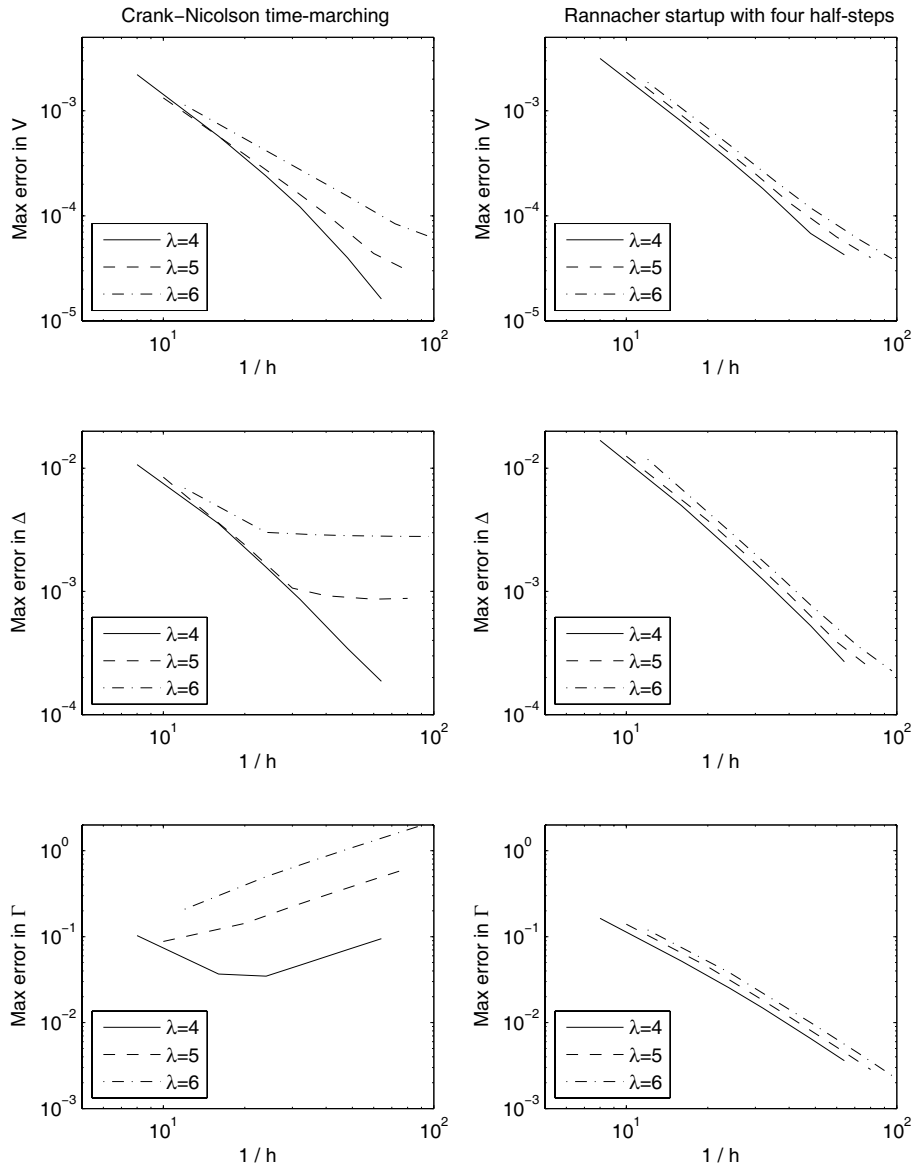
$$\frac{\partial V}{\partial t} = rV - rS \frac{\partial V}{\partial S} - \kappa(\theta - v) \frac{\partial V}{\partial v} - \frac{1}{2}vS^2 \frac{\partial^2 V}{\partial S^2} - \rho\omega vS \frac{\partial^2 V}{\partial S \partial v} - \frac{1}{2}\omega^2 v \frac{\partial^2 V}{\partial v^2}$$

In two dimensions, the standard Crank–Nicolson method would require the solution of a very large linear system of equations, making the computational cost approximately quadratic in the number of grid points. This is avoided here by using the Alternative Direction Implicit (ADI) method which is a factored approximation to the Crank–Nicolson method (Duffy 2006) but has a computational cost that is proportional to the number of grid points, as in one dimension. To maintain second-order accuracy it is necessary to use the Craig–Sneyd treatment of the cross-derivative term (Craig and Sneyd 1988). The numerical results are for a European call option using the following parameter values:  $r = 0.05$ ,  $\kappa = 2$ ,  $\theta = 0.04$ ,  $\omega = 0.25$ ,  $\rho = -0.5$ ,  $K = 1$ ,  $T = 2$ . The analytic reference solution is obtained using numerical integration (Kahl and Jackel 2005). The error plots, which show the maximum error along the line  $v = \theta$ , are very similar to those obtained for the Black–Scholes equation, with the Rannacher timestepping restoring second-order accuracy for  $V$ ,  $\Delta$  and  $\Gamma$ .

The purpose of this paper is to explain the behavior of Rannacher timestepping by performing a detailed error analysis of a simpler problem, the discretization of the constant coefficient one-dimensional convection–diffusion equation. This reveals that four half-timesteps of backward Euler time-marching is the minimum required to recover second-order convergence for these two problems. The use of more than four half-timesteps will lead to an increase in the overall error, due to the lower-order discretization error inherent in the backward Euler discretization, and therefore four half-timesteps can be regarded as optimal.

The approach which is followed is similar to that of Giles (2004), performing an asymptotic analysis of the behavior of the Fourier transform and then using this to bound the error resulting from the inverse transform (Strang 1986). Numerical results confirm the sharpness of the error bounds that are derived. A number of extensions to the analysis are discussed and investigated numerically, and guidance is offered to practitioners wanting to achieve the greatest accuracy for a given computational cost.

**FIGURE 5** Grid convergence for a European call option with the Heston model.





There are other ways of obtaining high accuracy implicit solutions. One option is to follow the approach of Broadie and Detemple (1996) and replace the first timestep calculation by the specification of an approximate analytic solution at the end of the first timestep. This yields the same improvement in accuracy as Rannacher timestepping, provided the approximate solution is sufficiently accurate. However, when generalizing this approach to multi-factor models with general payoffs, simple analytic approximations exist for some problems but not for others. When developing a pricing engine for a large class of models, Rannacher timestepping offers both simplicity and generality.

## 2 Model problem and discretization

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

$$\frac{\partial v}{\partial t} + a \frac{\partial v}{\partial x} = \frac{\partial^2 v}{\partial x^2} \tag{1}$$

over  $-\infty < x < \infty$  and  $0 < t < 1$ . The generalization to non-unit diffusivity and terminal times other than  $t = 1$  will be discussed later.

If  $v(x, t)$  satisfies this equation subject to the initial data

$$v(x, 0) = \max(x, 0)$$

then its second derivative

$$u \equiv \frac{\partial^2 v}{\partial x^2}$$

satisfies the same differential equation subject to the initial data

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

where  $\delta(x)$  is the Dirac delta function. Defining the Fourier transform pair

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

then the Fourier transform of Equation (1) yields

$$\frac{d\widehat{u}}{dt} = -(ia\kappa + \kappa^2)\widehat{u}$$

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

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

and hence

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

The Crank–Nicolson discretization of Equation (1) on a uniform grid with spacing  $h$  and timestep  $k$  is

$$(I + \frac{1}{2}D)V_j^{n+1} = (I - \frac{1}{2}D)V_j^n \tag{2}$$

where

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

with  $\delta_x^2$  and  $\delta_{2x}$  being the usual second difference and central first difference operators, respectively. The corresponding half-timestep backward Euler discretization used in the Rannacher startup is

$$(I + \frac{1}{2}D)V_j^{n+1/2} = V_j^n \tag{3}$$

If  $V_j^n$  satisfies these equations with initial data

$$V_j^0 = \max(x_j, 0)$$

then its divided second difference

$$U_j^n = h^{-2}\delta_x^2 V_j^n$$

satisfies the same difference equations subject to the Dirac initial data

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

The objective of the error analysis will be to quantify the error  $U_j^N - u(x_j, 1)$  for  $N = 1/k$ .

### 3 Fourier analysis

Using the mixed discrete/continuous Fourier transform pair (Strang 1986)

$$U_j^n = \frac{1}{2\pi h} \int_{-\pi}^{\pi} \widehat{U}^n(\theta) \exp(ij\theta) \, d\theta$$

$$\widehat{U}^n(\theta) = h \sum_{j=-\infty}^{\infty} U_j^n \exp(-ij\theta)$$

the Fourier transform of Equation (2) gives

$$\widehat{U}_m^{n+1} = \frac{1 - \frac{1}{2}ir \sin \theta - 2d \sin^2 \frac{1}{2}\theta}{1 + \frac{1}{2}ir \sin \theta + 2d \sin^2 \frac{1}{2}\theta} \widehat{U}_m^n$$

for  $n \geq R$ , 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 Equation (3) gives

$$\widehat{U}_m^{n+1} = \frac{1}{(1 + \frac{1}{2}ir \sin \theta + 2d \sin^2 \frac{1}{2}\theta)^2} \widehat{U}_m^n$$

These can be combined to give

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

where

$$z_1(\theta) = (1 - \frac{1}{2}ir \sin \theta - 2d \sin^2 \frac{1}{2}\theta)(1 + \frac{1}{2}ir \sin \theta + 2d \sin^2 \frac{1}{2}\theta)^{-1}$$

$$z_2(\theta) = (1 - \frac{1}{2}ir \sin \theta - 2d \sin^2 \frac{1}{2}\theta)^{-1}(1 + \frac{1}{2}ir \sin \theta + 2d \sin^2 \frac{1}{2}\theta)^{-1}$$

For the Dirac initial data,  $\widehat{U}^0 = 1$  and hence at the final iteration level  $N$  (assumed to be greater than  $R$ )

$$U_j^N = \frac{1}{2\pi h} \int_{-\pi}^{\pi} z_1^N(\theta) z_2^R(\theta) e^{ij\theta} d\theta \tag{4}$$

By making the substitutions  $\theta = \kappa h, x_j = jh$  the integral can also be expressed as

$$U_j^N = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} z_1^N(\kappa h) z_2^R(\kappa h) e^{i\kappa x_j} d\kappa$$

This is to be compared to the analytic solution  $u(x, 1)$  for which

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

with

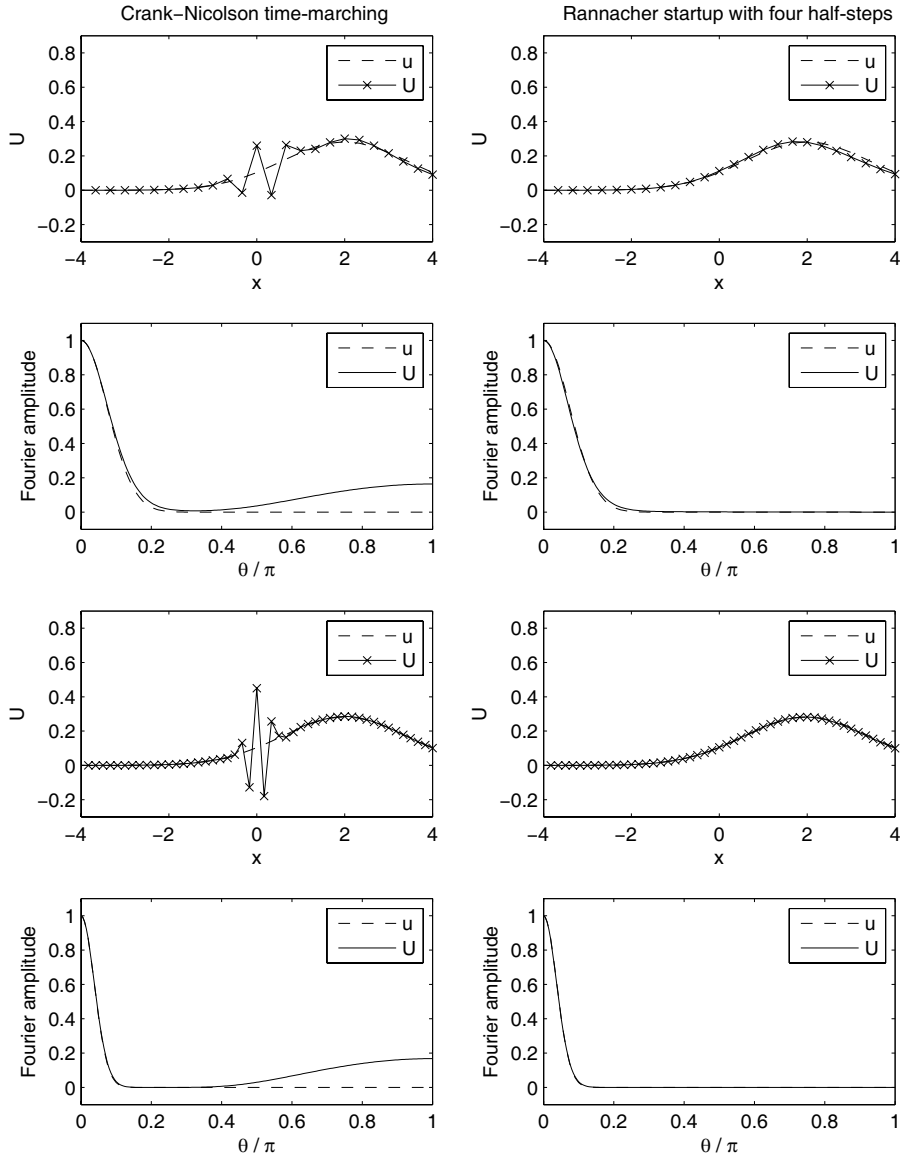
$$\widehat{u}(\kappa, 1) = \exp(-ia\kappa - \kappa^2)$$

Figure 6 plots comparisons between the numerical and analytic solutions to the convection–diffusion problem with  $a = 2$  at  $t = 1$  for two grid resolutions  $h = 1/3$  for the upper half of each 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.

Looking at the results in physical space (ie, the plots of  $U$  and  $u$  versus  $x$ ), the main feature to note is the high-wavenumber error near  $x = 0$  for the Crank–Nicolson time-marching. Its width appears proportional to  $h$ , while its magnitude appears proportional to  $h^{-1}$ ; this will be confirmed by the asymptotic analysis.

Looking at the comparison in Fourier space (ie, the plots of  $|\widehat{U}|$  and  $|\widehat{u}|$  versus  $\theta$ ) the main feature to note for the Crank–Nicolson results is that there appear to be three regions: one on the left of width proportional to  $h$  in which

**FIGURE 6** Numerical solution for the convection–diffusion equation with  $a = 2$ .



there is very good agreement between the numerical and analytic solution, one on the right with a width independent of  $h$  in which  $\widehat{u}$  is extremely small but  $\widehat{U}$  is not, and a central region in which both  $\widehat{u}$  and  $\widehat{U}$  are small. This separation into three regions is the basis for the asymptotic analysis, which considers a low-wavenumber range defined by  $|\kappa| < h^{-m}$ , a high-wavenumber range defined by  $h^{-r} < |\kappa|$  and the intermediate range  $h^{-m} < |\kappa| < h^{-r}$ . The constants  $m$  and  $r$  satisfy the constraints  $0 < m < \frac{1}{3}$  and  $\frac{1}{2} < r < 1$ . The reasons for these constraints will become clear in the asymptotic analysis.

The convergence analysis considers the limit  $h, k \rightarrow 0$  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  $O(k^2 + h^2)$ , and so keeping  $k = O(h)$  keeps the spatial and temporal approximation errors of the same order. We now analyze the Fourier error  $\widehat{U} - \widehat{u}$  in each of the three regions.

**PROPOSITION 3.1 (Low-wavenumber region)** For  $|\kappa| < h^{-m}$ , as  $h \rightarrow 0$  with  $\lambda = k/h$  held fixed,

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

where

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

**PROOF** Expressing each variable as a function of  $h$ ,

$$\theta = \kappa h, \quad N = \frac{1}{k} = \frac{1}{\lambda h}, \quad r = a \lambda, \quad d = \frac{\lambda}{h}$$

a Taylor series expansion gives

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

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  $h \rightarrow 0$ . Hence,

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

and so we obtain the result in the Proposition 3.1. □

**PROPOSITION 3.2 (High-wavenumber region)** For  $h^{-r} < |\kappa|$ , as  $h \rightarrow 0$  with  $\lambda = k/h$  held fixed,

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

PROOF  $z_1(\theta)$  can be re-expressed as

$$\begin{aligned} z_1(\theta) &= (1 - \frac{1}{2}ir \sin \theta - 2d \sin^2 \frac{1}{2}\theta)(1 + \frac{1}{2}ir \sin \theta + 2d \sin^2 \frac{1}{2}\theta)^{-1} \\ &= \left( \frac{1}{2d \sin^2 \frac{1}{2}\theta} - \frac{ir}{2d} \cot \frac{1}{2}\theta - 1 \right) \left( \frac{1}{2d \sin^2 \frac{1}{2}\theta} + \frac{ir}{2d} \cot \frac{1}{2}\theta + 1 \right)^{-1} \\ &\rightarrow -1 \quad \text{as } d \rightarrow \infty \end{aligned}$$

and similarly

$$\begin{aligned} z_2(\theta) &= (2d \sin^2 \frac{1}{2}\theta)^{-2} \left( \frac{1}{2d \sin^2 \frac{1}{2}\theta} - \frac{ir}{2d} \cot \frac{1}{2}\theta - 1 \right)^{-1} \\ &\quad \times \left( \frac{1}{2d \sin^2 \frac{1}{2}\theta} + \frac{ir}{2d} \cot \frac{1}{2}\theta + 1 \right)^{-1} \\ &\rightarrow -(2d \sin^2 \frac{1}{2}\theta)^{-2} \quad \text{as } d \rightarrow \infty \end{aligned}$$

Hence, expressing  $d$  and  $N$  as functions of  $h$  as in the proof of Proposition 3.1, Taylor series analysis gives

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

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

PROPOSITION 3.3 (Intermediate region) For  $h^{-m} < |\kappa| < h^{-r}$ , as  $h \rightarrow 0$  with  $\lambda = k/h$  held fixed,  $\widehat{U}^N(\kappa) = o(h^q)$ , for any  $q > 0$ .

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

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

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

Defining

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

and

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

then since  $\widehat{E}^{\text{low}} \ll \widehat{E}^{\text{high}}$  in the high-wavenumber region, and  $\widehat{E}^{\text{high}} \ll \widehat{E}^{\text{low}}$  in the low-wavenumber 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-wavenumber error is

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

with  $N^{(m)}(x)$  being the  $m$ th derivative of the normal distribution with zero mean and unit variance, and the high-wavenumber error is

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

where

$$\begin{aligned} f_j &= \frac{h}{2\pi} \int_{-\pi/h}^{\pi/h} \frac{e^{i\kappa x_j}}{\sin^{4R} \frac{1}{2}\theta} \exp\left(-\frac{1}{\lambda^2 \sin^2 \frac{1}{2}\theta}\right) d\kappa \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{ij\theta}}{\sin^{4R} \frac{1}{2}\theta} \exp\left(-\frac{1}{\lambda^2 \sin^2 \frac{1}{2}\theta}\right) d\theta \\ &= \frac{1}{\pi} \int_0^\pi \frac{\cos j\theta}{\sin^{4R} \frac{1}{2}\theta} \exp\left(-\frac{1}{\lambda^2 \sin^2 \frac{1}{2}\theta}\right) d\theta \end{aligned}$$

$E_j^{\text{high}}$  clearly has a width of  $O(h)$ , and has a maximum magnitude at  $j = 0$  where  $x_j = 0$ . This explains the observed behavior in Figure 6. The integral for  $j = 0$  can be evaluated analytically (see Appendix A) giving

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

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

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-s^2} ds$$

The fact that the low-wavenumber error is  $O(h^2)$  and the high-wavenumber error is  $O(h^{2R-1})$  is confirmed by the results in the upper plots of Figure 7 which have convergence results for the convection–diffusion equation with  $a = 2$ . It can be seen that for the standard Crank–Nicolson time-marching, the results exhibit  $O(h^2)$  convergence until  $h$  reaches a sufficiently small value that the  $O(h^{-1})$  high-wavenumber error becomes dominant. The plots show the sensitive dependence of the high-wavenumber error on the value of  $\lambda$ . For large values of  $\lambda$ ,  $\text{erfc}(\lambda^{-1}) \approx 1$  and so  $E_j^{\text{high}}$  becomes significant for quite large values of  $h$ . On the other hand, for small values of  $\lambda$ ,  $\text{erfc}(\lambda^{-1})$  is extremely small, and so  $E_j^{\text{high}}$  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-wavenumber error is  $O(h^3)$  and so the low-wavenumber error remains dominant for all  $h$ . The sharpness of the error analysis is demonstrated by the lower plots that compare the numerical error with the maximum magnitude of  $E_j^{\text{low}}$  and  $E_j^{\text{high}}$ .

## 4 Extensions

### 4.1 Alternative initial data

The analysis so far has assumed that there is a grid point at  $x = 0$ , so the grid is perfectly aligned with the discrete Dirac initial data, but if the grid is not perfectly aligned care must be taken in representing the initial data, as discussed by Pooley *et al* (2003).

If the grid points  $x_j$  are still taken to be located at  $jh$ , but instead of  $j$  taking integer values it is  $j + \alpha$  which takes integer values (with  $0 < \alpha < 1$ ) then the appropriate discretization of the Dirac initial data is

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

This gives

$$\widehat{U}_m^0 = (1 - \alpha) e^{-i\alpha\theta_m} + \alpha e^{i(1-\alpha)\theta_m}$$

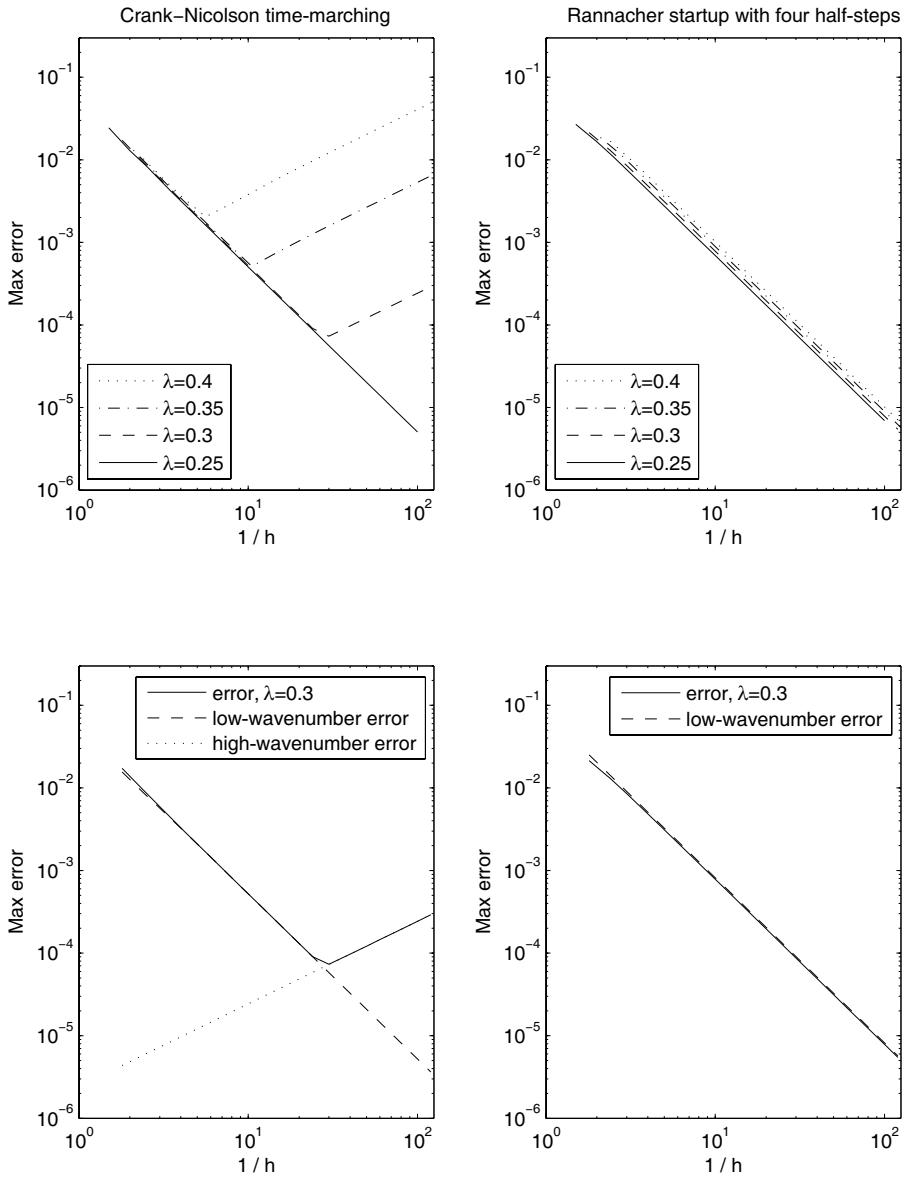
Putting  $\theta_m = \kappa h$ , an asymptotic expansion with respect to  $h$  gives

$$\widehat{U}_m^0 = 1 + O(\kappa^2 h^2)$$

which leads to the result that the low-wavenumber error remains second order. It can also be shown that the convergence order of the high-wavenumber error is also unaffected.



**FIGURE 7** Grid convergence for the convection–diffusion equation with  $a = 2$ .



**TABLE 1** Order of convergence of the high-wavenumber error using Rannacher timestepping with  $2R$  half-timesteps.

	$V$	$\Delta$	$\Gamma$
European call	$2R + 1$	$2R$	$2R - 1$
Digital call	$2R$	$2R - 1$	$2R - 2$

Although the focus of our analysis so far has been on Dirac initial data, there are other sets of initial data that are also of interest. One is the first difference of the discrete Dirac initial data; this is relevant to the approximation of  $\Gamma$  for the digital option. Another is a discrete equivalent of  $H(x) - \frac{1}{2}$ , where  $H(x)$  is the Heaviside step function; this is relevant to the approximation of  $V$  for the digital option, and  $\Delta$  for the European option.

For both of these sets of alternative initial data, the low-wavenumber error will still be  $O(h^2)$ . However, the high-wavenumber error will be one order worse in the first case,  $O(h^{2R-2})$ , 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,  $O(h^{2R})$ .

Table 1 summarizes the consequences of the analysis and its generalizations for the convergence of the high-wavenumber error in computing  $V$ ,  $\Delta$  and  $\Gamma$  for European and digital call options. The low-wavenumber error is  $O(h^2)$  in all cases, and using  $R = 2$  ensures that the high-wavenumber error is also at worst  $O(h^2)$ .

#### 4.2 Diffusion coefficient and terminal time

The model problem that has been analyzed has unit diffusivity, and the error is analyzed at the terminal time  $t = 1$ . Suppose now that the convection–diffusion equation is

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

and the error is to be analyzed at the terminal time  $t = T$ . The non-dimensionalization

$$t^* = \frac{t}{T}, \quad x^* = \frac{x}{\sqrt{\epsilon T}}, \quad k^* = \frac{k}{T}, \quad h^* = \frac{h}{\sqrt{\epsilon T}}$$

$$\lambda^* = \frac{k^*}{h^*} = \frac{k}{h} \sqrt{\frac{\epsilon}{T}} = \lambda \sqrt{\frac{\epsilon}{T}}, \quad a^* = a \sqrt{\frac{T}{\epsilon}}$$

reduces the more general problem to the one that has already been analyzed. Applying this non-dimensionalization to the Black–Scholes calculations in the introduction, with  $\epsilon = \frac{1}{2}\sigma^2$ , then the calculations for  $\lambda = 2, 3, 4, 5$  correspond to  $\lambda^* \approx 0.2, 0.3, 0.4, 0.5$ , respectively.

Having determined that  $\lambda^*$  is the key non-dimensional parameter, the next question is whether there is an optimal value for this parameter providing the maximum accuracy for a given computational cost. The error analysis gives a leading-order error of the form

$$E = ah^{*2} + bk^{*2}$$

while the computational cost is proportional to the product of the number of grid points and the number of timesteps, so

$$C = ch^{*-1}k^{*-1}$$

Hence, for a given computational cost  $C$ ,

$$E = h^*k^* \left( \frac{ah^*}{k^*} + \frac{bk^*}{h^*} \right) = \frac{c}{C} (a\lambda^{*-1} + b\lambda^*)$$

Too small a value for  $\lambda^*$  gives a large error due to large  $h^*$ , while too large a value of  $\lambda^*$  gives a large error due to large  $k^*$ . This tradeoff is clearly seen in the results shown in Figure 8. The curves labeled RT2/2 present the maximum errors for the European and digital call options as a function of  $\lambda^*$ , keeping fixed the computational work by increasing  $h$  as  $k$  is decreased. The optimum is seen to be around  $\lambda^* = 0.5$ . The curves labeled CN correspond to the results given by the standard Crank–Nicolson method. For small values of  $\lambda^*$  these give more accurate results because the Rannacher timestepping introduces additional low-wavenumber errors. However, as  $\lambda^*$  increases and  $\text{erfc}(1/\sqrt{\lambda^*})$  is no longer extremely small, the high-wavenumber error becomes dominant and the Crank–Nicolson error becomes very large.

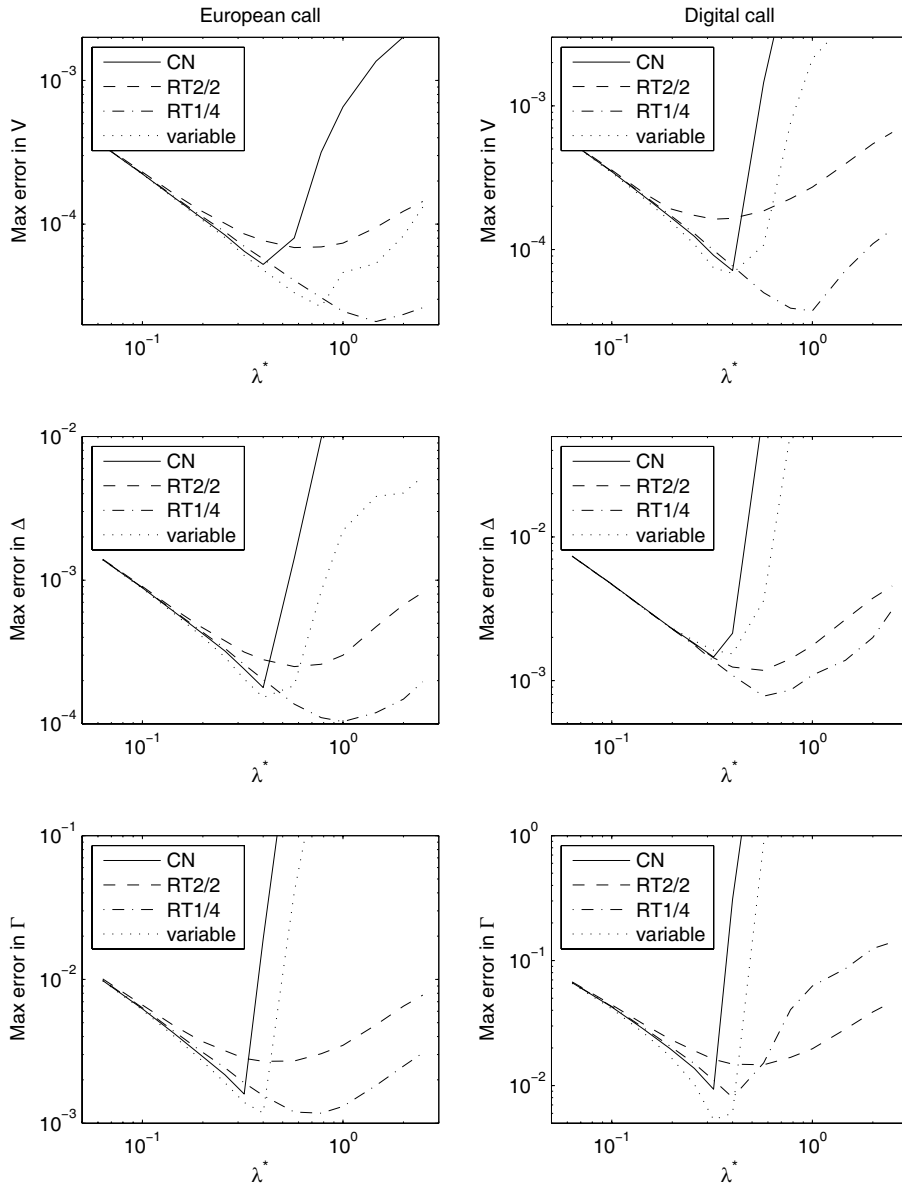
### 4.3 Alternative Rannacher treatments

Although Rannacher (1984) originally suggested the approach analyzed here, namely replacing Crank–Nicolson timesteps by two backward Euler half-timesteps, there are other options.

One could replace the Crank–Nicolson timesteps by full timestep backward Euler approximations, but in this case one would need to replace the first four Crank–Nicolson timesteps to get the same improvement in the order of convergence, and the larger timestep in the backward Euler time integration would increase the low-wavenumber error.

A better option is to replace the first Crank–Nicolson timestep by four quarter-timesteps of backward Euler time integration. This gives the same desired improvement in the order of convergence, but with a reduced low-wavenumber error. The curves labeled RT1/4 in Figure 8 present the results obtained in this way. The results are clearly superior in all cases except perhaps for the digital option  $\Gamma$  for which there is a second-order high-wavenumber error in addition to the second-order low-wavenumber error, and RT1/4 does not eliminate the high-wavenumber error as effectively as RT2/2.

**FIGURE 8** Comparison of four numerical methods: Crank–Nicolson (CN), Ran- nacher timestepping with four half-timesteps (RT2/2) with four quarter-timesteps (RT1/4), and with four quarter-timesteps and non-uniform timesteps (variable).



#### 4.4 Non-uniform timesteps

Another possibility is not to use uniform timesteps, but instead use smaller timesteps initially. Given the existence of an optimal value for  $\lambda^*$ , one way of choosing the timestep might be to keep fixed the value of  $\lambda^*(t)$  based on the current time  $t$  rather than the final time  $T$ . This requires  $k \propto \sqrt{t}$  which is accomplished by defining

$$k_n = t_{n+1} - t_n, \quad t_n = \frac{n^2}{N^2}T$$

so that  $t_0 = 0$ ,  $t_N = T$  and  $k_n \propto n \propto \sqrt{t_n}$ .

The curves labeled “variable” in Figure 8 show that this does not produce very good results. The problem is that, as with the basic Crank–Nicolson method, the error rises very sharply when  $\lambda^*$  increases. This is because the very small initial timesteps greatly reduce the effectiveness of the backward Euler smoothing of the high-wavenumber error. In fact, additional results, backed by numerical analysis, show that one does not obtain second-order convergence in any of the cases. This contrasts slightly with the results of Forsyth and Vetzal (2002) who found that for American options this variable timestep gives second-order convergence for the option value whereas the uniform timestep does not, probably due to its inadequate resolution of the initial behavior of the exercise boundary.

#### 4.5 Richardson extrapolation

The final extension to be considered is Richardson extrapolation (Dahlquist and Björck 1974; Geske and Johnson 1984). Given that the leading-order term in the low-wavenumber error is of the form

$$E_h^{\text{low}} = ah^2 + bk^2$$

then if one performs calculations with spacing  $2h$  and  $h$ , keeping the ratio  $\lambda$  fixed, then

$$4E_h^{\text{low}} - E_{2h}^{\text{low}} = 0$$

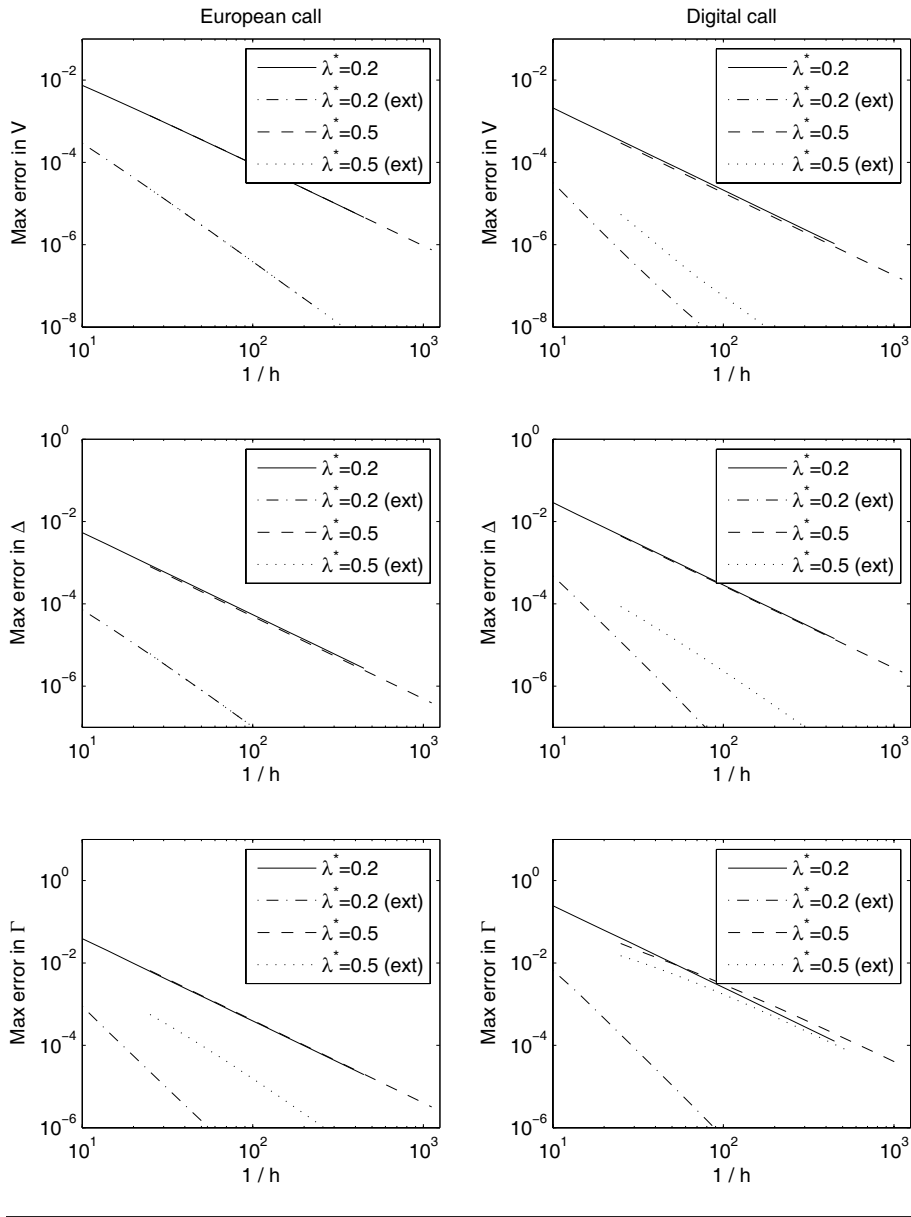
Hence, the extrapolated solution

$$U_{\text{ext}} = \frac{4}{3}U_h - \frac{1}{3}U_{2h}$$

will have a low-wavenumber error that does not have a leading second-order term.

Figure 9 shows the improved accuracy of this extrapolated solution. Further numerical analysis reveals that the next order low-wavenumber error term is due to the Rannacher time-marching and is proportional to  $Rk^3$ . This explains the third-order convergence that is apparent in some of the plots for  $\lambda^* = 0.5$ . For  $\lambda^* = 0.2$ ,  $k^3$  is sufficiently small so that the third-order error is relatively insignificant. The second-order convergence of the digital option  $\Gamma$  is due to the high-wavenumber error which is second order in this case and is not cancelled by the extrapolation procedure.

**FIGURE 9** Convergence comparison for Rannacher time-marching using four quarter-timesteps, with and without Richardson extrapolation.



### 5 Conclusions

In this paper we have analyzed the convergence of Crank–Nicolson approximations of the one-dimensional convection–diffusion equation, with and without the Rannacher startup procedure in which an initial  $R$  Crank–Nicolson timesteps are each replaced by two half-timesteps of backward Euler integration. The analysis proves, and numerical results confirm, that there is a low-wavenumber error of  $O(h^2)$ , and a high-wavenumber error of  $O(h^{2R-1})$ . Hence  $R = 2$  is the minimum to give  $O(h^2)$  convergence, and in general it is the optimum since larger values will increase the low-wavenumber error.

In considering extensions to this analysis and its relevance to Black–Scholes applications, it was shown that it is better to replace just the first Crank–Nicolson timestep by four quarter-timesteps of backward Euler time integration. This reduces the low-wavenumber error introduced by the Rannacher startup. In addition, the accuracy is maximized for a given computational cost by choosing the uniform timestep  $k$  and grid spacing  $h$  so that

$$\lambda^* = \frac{k}{h} \sqrt{\frac{\sigma^2}{2T}}$$

lies between 0.5 and 1.0. Using a variable timestep does not improve the accuracy, and in fact spoils the second-order convergence. The final comment is that numerical results show that very significant improvement in accuracy (or decrease in computational cost) can be achieved using Richardson extrapolation.

### Appendix A Evaluation of the integral

Consider the integral

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

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

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

and hence

$$\frac{dI_0}{d\alpha} = -\frac{4\alpha}{\pi} \int_0^\infty \exp(-\alpha^2(t^2 + 1)) dt = -\frac{2}{\sqrt{\pi}} \exp(-\alpha^2)$$

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

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

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

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

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

## REFERENCES

- Broadie M., and Detemple, J. (1996). American option valuation: new bounds, approximations, and a comparison of existing methods. *Review of Financial Studies* **9**(4), 1211–1250.
- Craig, I. J. D., and Sneyd, A. D. (1988). An alternating-direction implicit scheme for parabolic equations with mixed derivatives. *Computers & Mathematics with Applications* **16**(4), 341–350.
- Dahlquist, G., and Björck A. (1974). *Numerical Methods*. Prentice-Hall, Englewood Cliffs, NJ.
- Duffy, D. (2006). *Finite Difference Methods in Financial Engineering: A Partial Differential Equation Approach*. Wiley.
- Forsyth, P. A., and Vetzal, K. R. (2002). Quadratic convergence for valuing American options using a penalty method. *SIAM Journal on Scientific Computing* **23**(6).
- Geske, R., and Johnson, H. E. (1984). The American put option valued analytically. *Journal of Finance* **39**(5), 1511–1524.
- 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 University Computing Laboratory.
- Heston, S. I. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* **6**, 327–343.
- Ikonen, S., and Toivanen, J. (2004). Pricing American options using LU decomposition. Technical Report B 4/2004, Department of Mathematical Information Technology, University of Jyväskylä.
- Kahl, C., and Jäckel, P. (2005). Not-so-complex logarithms in the Heston model. *Wilmott* September.
- Pooley, D. M., Forsyth, P. A., and Vetzal, K. R. (2003a). Numerical convergence properties of option pricing PDEs with uncertain volatility. *IMA Journal of Numerical Analysis* **23**, 241–267.
- Pooley, D. M., Vetzal, K. R., and Forsyth, P. A. (2003b). Convergence remedies for non-smooth payoffs in option pricing. *Journal of Computational Finance* **6**(4).
- Raahauge, P. (2004). Higher-order finite element solutions of option prices. Working Paper Series No. 184, University of Aarhus Center for Analytical Finance.
- Rannacher, R. (1984). Finite element solution of diffusion problems with irregular data. *Numerische Mathematik* **43**, 309–327.
- Richtmyer, R. D., and Morton, K. W. (1967). *Difference Methods for Initial-Value Problems*, 2nd edn. Wiley-Interscience. Reprint (1994), Krieger Publishing Company, Malabar.
- Strang, G. (1986). *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, Wellesley, MA.
- Wilmott P., Howison, S., and Dewynne, J. (1995). *The Mathematics of Financial Derivatives*. Cambridge University Press, 1995.
- Zvan, R., Forsyth, P. A., and Vetzal, K. R. (2003). Negative coefficients in two factor option pricing models. *Journal of Computational Finance* **7**(1).