1 Heat Equation With Dirichlet Boundary

Let $q(x) = \frac{\sin(\pi x)}{2-\cos(2\pi x)}$. Consider the heat equation on $[0, 1]$.

$$u_t = \alpha u_{xx} + f, \quad u(x, 0) = g(x), \quad u(0, t) = u(1, t) = 0$$

with diffusion constant $\alpha = \frac{1}{100}$ and the following two sets of data:

(i) $f(x, t) = 0, \quad g(x) = q(x)$

(ii) $f(x, t) = q(x)\sin(\pi t), \quad g(x) = 0$

For each case, calculate the exact solution of $u(1/4, 1)$ and apply the Forward Time Centered Space (FTCS) scheme to obtain an approximate solution. Note that we use the notation $\nu := \frac{\alpha k}{h^2}$ to represent the relationship between the step sizes in time and space.

Solution:

(a) (i) We are given the exact solution to (i):

$$u(x, t) = \sum_{k}^{\infty} c_k e^{-\alpha k^2 \pi^2 t} \sin(k\pi x), \quad c_k = 2\int_0^1 q(x)\sin(k\pi x)dx$$

We use the trapezoidal rule to solve for c_k, experimenting with the number of quadrature points until the absolute error between *scipy.integrate* and our solution is less than 10^{-16}. We then use sufficient additional terms in the summation of basis functions till the result for $u(x, t)$ does not change within a 10^{-12} tolerance. The exact solution up to 12 digits using this method is: $\boxed{0.301848388303}$.

(ii) We are given the exact solution to (ii):

$$u(x, t) = \sum_{k}^{\infty} \left(\int_0^t e^{-\alpha k^2 \pi^2 (t-s)} \sin(\pi s) ds \right) c_k \sin(k\pi x), \quad c_k = 2\int_0^1 q(x)\sin(k\pi x)dx$$

First, we solve the time integral term analytically. Let $\beta_k := \alpha k^2 \pi^2$ for conciseness.

$$\int_0^t e^{-\beta_k (t-s)} \sin(\pi s) ds = e^{-\beta_k t} \underbrace{\int_0^t e^{\beta_k s} \sin(\pi s) ds}_{A}$$

We use integration by parts. Let $u = \sin(\pi s)$ and $dv = e^{\beta_k s}$. Then $du = \pi\cos(\pi s)ds$ and $v = \frac{1}{\beta_k} e^{\beta_k s}$.

$$\begin{aligned} A &= uv - \int v du \\ &= \frac{1}{\beta_k} \sin(\pi s) e^{\beta_k s} - \frac{\pi}{\beta_k} \underbrace{\int e^{\beta_k s} \cos(\pi s) ds}_{B} \end{aligned}$$

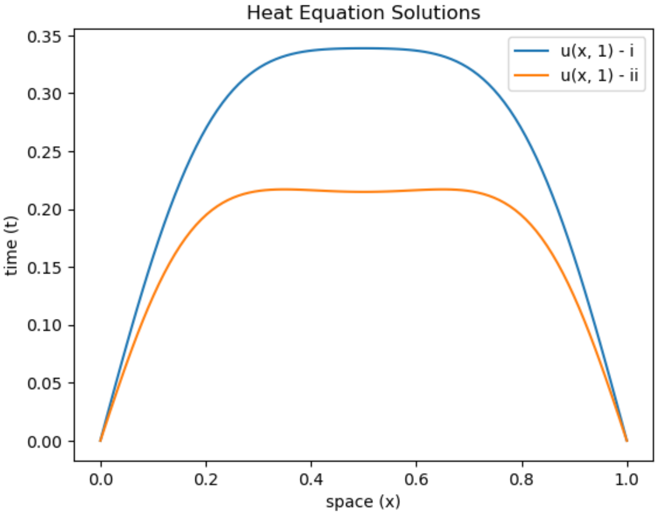
Let $u = \cos(\pi s)$ and $dv = e^{\beta_k s}$. Then $du = -\pi \sin(\pi s) ds$ and $v = \frac{1}{\beta_k} e^{\beta_k s}$.

$$\begin{aligned}
B &= uv - \int v du \\
&= \frac{1}{\beta_k}\cos(\pi s)e^{\beta_k s} + \frac{\pi}{\beta_k}\int e^{\beta_k s}\sin(\pi s) ds \\
&= \frac{1}{\beta_k}\cos(\pi s)e^{\beta_k s} + \frac{\pi}{\beta_k} A \\
A &= \frac{1}{\beta_k}\sin(\pi s)e^{\beta_k s} - \frac{\pi}{\beta_k}\left(\frac{1}{\beta_k}\cos(\pi s)e^{\beta_k s} + \frac{\pi}{\beta_k}A\right) \\
A &= \frac{\beta_k \sin(\pi s)e^{\beta_k s} - \pi\cos(\pi s)e^{\beta_k s}}{\beta_k^2 + \pi^2}
\end{aligned}$$

Applying the bounds of integration yields:

$$\int_0^t e^{-\beta_k(t-s)}\sin(\pi s) ds = \frac{\beta_k \sin(\pi t) - \pi\cos(\pi t) + \pi e^{-\beta_k t}}{\beta_k^2 + \pi^2}$$

Now to solve for the exact solution for the PDE, as in (i), we apply the trapezoidal rule to solve for c_k to machine precision and then add sequential terms until the approximation for $u(x,t)$ stays within a tolerance of 10^{-12}. The exact solution up to 12 digits using this method is: $\boxed{0.209395363856}$. Shown below is the exact solution to $u(x, 1)$ for both settings in (i) and (ii).



(b) We first compute the stability of the scheme. When $f = 0$, the scheme is $D_t^+ u = \alpha D_x^+ D_x^- u$. This expression corresponds to an update step map $\mathcal{B} : \mathbb{R} \to \mathbb{R}$ defined as follows:

$$u_j^{n+1} = \mathcal{B}(u_j^n) = \nu u_{j+1} + (1 - 2\nu)u_j + \nu u_{j-1}$$

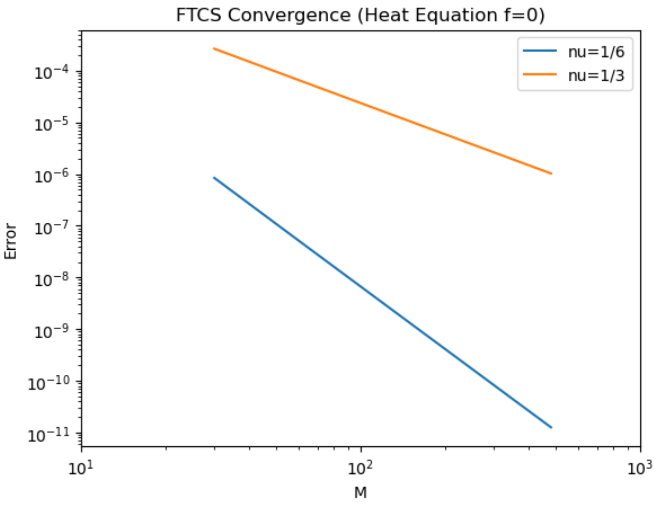
Though ν depends on the value of α, the form of this update map is the same as that of the case where $\alpha = 1$ which was studied in lecture. Therefore, we can take the conclusion from lecture that the scheme is stable if and only if $\boxed{\nu \leq \frac{1}{2}}$.

Now we compute the truncation error for the general case, including when $f \neq 0$ for use in part (c). The general scheme is $D_t^+ u = \alpha D_x^+ D_x^- u + f$. We represent the truncation error by τ.

$$\begin{aligned}
\underbrace{D_t^+ u}_{A} &= \alpha \underbrace{D_x^+ D_x^- u}_{B} + f + \tau \\
A &= u_t + \frac{k}{2} u_{tt} + O(k^2) \\
B &= u_{xx} + \frac{h^2}{12} u_{xxxx} + O(h^4) \\
\tau &= A - \alpha B - f \\
&= \cancel{(u_t - \alpha u_x x - f)} + \frac{1}{2}\left(k u_{tt} - \frac{\alpha h^2}{6} u_{xxxx}\right) + O(k^2 + h^4) \\
&= \frac{1}{2}\left(k u_{tt} - \frac{\alpha h^2}{6} u_{xxxx}\right) + O(k^2 + h^4) \\
&= \frac{1}{2}\left(\left(k\alpha^2 - \frac{\alpha h^2}{6}\right) u_{xxxx} + k(\alpha f_{xx} + f_t)\right) + O(k^2 + h^4)
\end{aligned}$$

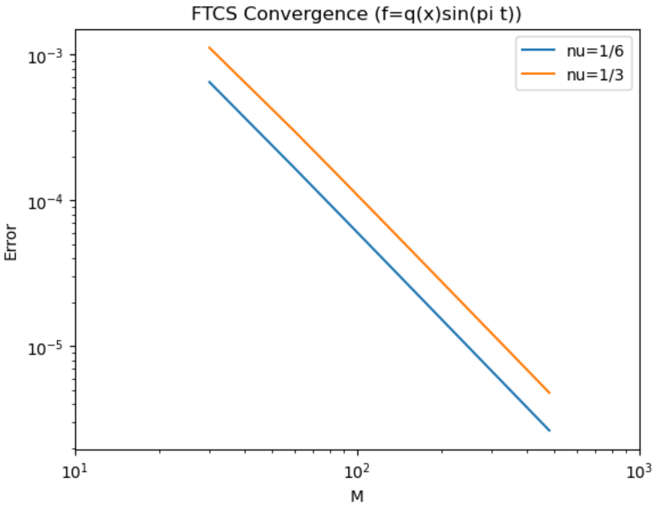
If $f = 0$, we can fully cancel the $O(k+h^2)$ term by letting $\boxed{\nu = \frac{1}{6}}$. The differing orders of convergence dependent on ν are demonstrated empirically below. Note that for the implementation of the Dirichlet boundaries, every iteration the grid points for $u(0,t)$ and $u(1,t)$ are separately set to 0. We use the following method for computing the error of each method:

$$\text{err} = \sqrt{h \sum_{j=0}^{M-1} \left(u_{\text{numerical}}(x_j, T - u_{\text{exact}}(x_j, T)\right)^2}, \quad h = 1/M,\ x_j = jh,\ T = 1.$$



The slope for $\nu = 1/6$ is ≈ -4 while the slope for $\nu = 1/3$ is ≈ -2. Because of how k and h relate via the constant ν, the results support the expected $O(k + h^2)$ accuracy for $\nu \neq 1/6$ and $O(k^2 + h^4)$ for $\nu = 1/6$.

(c) Recall that the $O(k+h^2)$ term in the truncation error of the given scheme only cancels fully if $f = 0$. Otherwise, we are left with an additional $O(k)$ term. Observe below how setting $\nu = 1/6$ no longer affects the slope of the error convergence.



The above convergence has a slope of ≈ -2 for both values of ν, however the error or $\nu = 1/6$ is still consistently lower because the additional $O(k)$ error only comes from f_t whereas the u_{xxxx} term does cancel. To achieve a higher order scheme when $f \neq 0$, instead consider the following alternative scheme:

$$\underbrace{D_t^+ u}_{A} = \alpha \underbrace{D_x^+ D_x^- u}_{B} + \underbrace{(1/3)f_j^n + (1/2)f_j^{n+1} + (1/12)f_{j+1}^n + (1/12)f_{j-1}^n}_{C} + \tau$$

The A and B terms are the same as in our earlier analysis. We expand C below:

$$\begin{aligned} C &= \frac{1}{3}f + \frac{1}{2}(f + kf_t + O(k^2)) + \frac{1}{12}(2f + h^2 f_{xx} + O(h^4)) \\ &= f + \frac{k}{2}f_t + \frac{h^2}{12}f_{xx} + O(k^2 + h^4) \end{aligned}$$

Finally, we compute the truncation error of the proposed scheme:

$$\begin{aligned} \tau &= A - \alpha B - C \\ &= \frac{1}{2}\left(ku_{tt} - \frac{\alpha h^2}{6}u_{xxxx} - kf_t - \frac{h^2}{6}f_{xx}\right) + O(k^2 + h^4) \end{aligned}$$

Observe that f relates to u by:

$$f_t = u_{tt} - \alpha u_{xxt}, \quad f_{xx} = u_{xxt} - \alpha u_{xxxx}$$

We can now substitute the expressions for f_t and f_{xx} into the truncation error to get an expression that depends only on u.

$$\begin{aligned} \tau = \frac{1}{2}&\left(\cancel{ku_{tt}} - \cancel{\frac{\alpha h^2}{6}u_{xxxx}} - \cancel{ku_{tt}} + k\alpha u_{xxt} - \frac{h^2}{6}u_{xxt} + \cancel{\frac{\alpha h^2}{6}u_{xxxx}}\right) + O(k^2 + h^4) \\ &= \left(k\alpha - \frac{h^2}{6}\right)u_{xxt} + O(k^2 + h^4) \end{aligned}$$

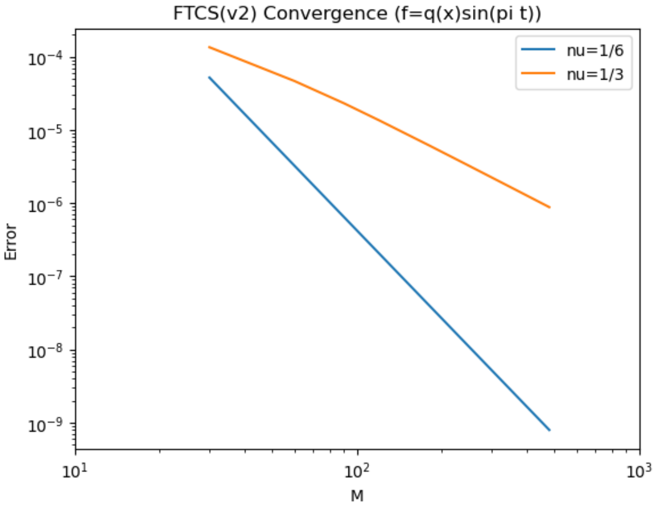
Under this new construction, once again if $\nu = \frac{1}{6}$ then the $O(k + h^2)$ term does in fact cancel and we are left with a truncation error of $O(k^2 + h^4)$ as desired. In fact, we can explicitly compute the leading order term as follows:

$$\tau = \underbrace{\frac{k^2}{6}u_{ttt}}_{\tau_A} - \underbrace{\frac{\alpha h^4}{360}u_{xxxxxx}}_{\tau_B} - \underbrace{\left(\frac{k^2}{4}f_{tt} + \frac{h^4}{144}f_{xxxx}\right)}_{\tau_C} + O(k^3 + h^6)$$

Let $M = \max\left\{\max_{x,t}\left|\partial_t^3 u\right|, \max_{x,t}\left|\partial_x^6 u\right|, \max_{x,t}\left|\partial_t^2 f\right|, \max_{x,t}\left|\partial_x^4 f\right|\right\}$. Then we have the following strict bound on the τ:

$$\begin{aligned}\tau &\leq M\left(\frac{k^2}{6} + \frac{\alpha h^4}{360} + \frac{k^2}{4} + \frac{h^4}{144}\right) \\ &= M\left(\frac{5k^2}{12} + \left(\frac{\alpha}{360} + \frac{1}{144}\right)h^4\right) \\ &= O(k^2 + h^4)\end{aligned}$$

The results below verify this expected order of convergence.



Indeed, the slope is ≈ -4 so we have successfully verified the $O(k^2 + h^4)$ convergence when $\nu = 1/6$.

2 Heat Equation With Neumann Boundary

Consider the PDE $u_t = \alpha u_{xx} + u$ with $\alpha = 1/20$ on the interval $0 \leq x \leq 1$ with insulating boundary conditions, $u_x(0,t) = 0, u_x(1,t) = 0$ and initial conditions $u(x,0) = g(x) = \sin^2(\pi x)$. Calculate the exact solution, evaluate the stability of the FTCS scheme for this PDE, and devise a new scheme with truncation error $O(k^2 + h^4)$.

Solution:

(a) We begin by solving for the exact solution of the PDE via Separation of Variables. The solution will be of the form $u(x,t) = X(x)T(t)$. Then, $XT' = \alpha X''T + XT$. We divide through by XT to get $\frac{T'}{T} = \alpha\frac{X''}{X} + 1 = \lambda$ where λ is a constant since the ratios are independently constant in space and time respectively. Suppose solutions of X are of the form $X(x) = e^{rx}$. Then $\alpha r^2 + 1 = \lambda \Rightarrow r = \pm\beta$ where $\beta = \sqrt{\frac{\lambda-1}{\alpha}}$. Thus we the write basis solutions of X:

$$X(x) = C_1 e^{\beta x} + C_2 e^{-\beta x}, \quad C_1, C_2 \in \mathbb{C}$$

First apply the insulating boundary at $x = 0$ to get $X'(0) = \beta(C_1 - C_2) = 0 \Rightarrow C_1 = C_2 = C$, assuming $\beta \neq 0$. Additionally, assume $C \neq 0$. Both of these assumptions are valid because a constant solution cannot satisfy the given initial condition. Next, apply the insulating boundary at $x = 1$ to get $X'(1) = C\beta(e^{\beta} - e^{-\beta}) = 0 \Rightarrow e^{\beta} = e^{-\beta}$. Then $e^{2\beta} = 1 = e^{k\cdot 2\pi i}$ for $k \in \mathbb{Z}$. Then we can define the kth basis element of the solution space for X, as $X_k(x) = C(e^{\beta_k x} + e^{-\beta_k x}) = c_k \cos(k\pi x)$ where $\beta_k = k\pi i$.

Now we can solve for a corresponding solution basis for T. Using the separation of variables from earlier, $T(t) = e^{\lambda t}$. By using the respective values of λ for each basis element, we get $T_k(t) = e^{(\alpha\beta_k^2+1)t} = e^{(-\alpha\pi^2k^2+1)t}$. Finally we put the solutions for T and X together to form a basis for $u(x,t)$: $u_k(x,t) = c_k\cos(k\pi x)e^{(-\alpha\pi^2k^2+1)t}$.

We proceed to find the coefficients c_k by applying the initial condition. Observe that $g(x) = \sin^2(\pi x) = \frac{1}{2} - \frac{1}{2}\cos(2\pi x)$. We can match this Fourier series to the general summation of basis elements u_k with corresponding weights c_k to yield $c_0 = \frac{1}{2}$, $c_2 = -\frac{1}{2}$, and all other coefficients set to 0. Thus, our final solution after plugging in the values for c_k and α is:

$$\boxed{u(x,t) = \frac{1}{2}e^t - \frac{1}{2}\cos(2\pi x)e^{\left(1-\frac{\pi^2}{5}\right)t}}$$

(b) We seek values of $\nu = \alpha k/h^2$ for which the FTCS scheme stable for the given PDE. The scheme is defined by $D_t^+ u = \alpha D_x^+ D_x^- u + u$. Applying these finite difference operators yields

$$\frac{1}{k}(u_j^{n+1} - u_j^n) = \frac{\alpha}{h^2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n) + u_j^n$$

By isolating u_j^{n+1}, we construct the update step $u_j^{n+1} \leftarrow \mathcal{B}(u_j^n)$:

$$\mathcal{B}(u_j) = \nu u_{j+1} + (1 + k - 2\nu)u_j + \nu u_{j-1}$$

For the scheme to be stable, we want to bound $\|\mathcal{B}^n\|$ by a constant that depends only on T. We can compute this norm by first diagonalizing $\mathcal{B}$ with the Z-transform so that $Z\mathcal{B} = \mathcal{G}Z$ where $\mathcal{G} : L^2(-\pi,\pi) \to L^2(-\pi,\pi)$ is diagonal. We know from lecture that for a finite difference operator $\mathcal{B}$,

$$ZBu(\xi) = \underbrace{\left(\sum_m c_m e^{im\xi}\right)}_{G(\xi)}\underbrace{\left(\sum_\ell u_\ell e^{-i\ell\xi}\right)}_{Zu(\xi)}$$

By substituting $\mathcal{G}$ and Z and using the fact that $\sqrt{\frac{h}{2\pi}}Z$ is a unitary operator, we have $\|\mathcal{B}\| = \left\|Z^{-1}\mathcal{G}Z\right\| = \|\mathcal{G}\|$. For the FTCS scheme, the nonzero coefficients are:

$$c_{-1} = \nu,\ c_0 = (1 + k - 2\nu),\ c_1 = \nu$$

Using these coefficients, we compute the corresponding amplification factor and norm.

$$\begin{aligned}
G(\xi) &= \nu e^{i\xi} + (1+k-2\nu)e^{0} + \nu e^{-i\xi} \\
&= 1 + k + \nu(e^{i\xi/2} - e^{-i\xi/2})^2 \\
&= \boxed{1 + k - 4\nu \sin^2(\xi/2)} \\
\|\mathcal{G}\| &= \|G(\xi)\|_\infty \\
&= |1 + k - 4\nu \sin^2(\xi/2)|_\infty
\end{aligned}$$

If $\nu \le \frac{1}{2}$ then $-2 \le -4\nu \sin^2(\xi/2) \le 0 \Rightarrow \|\mathcal{G}\| = 1 + k$. This is the norm of the finite difference operator on the infinite real line, however when we impose the Dirichlet boundary conditions, the result simply forms an upper-bound. Thus $\|\mathcal{B}\| \le 1 + k$. This bound is of the form $1 + Ck$ which is a sufficient condition for the Lax-Richtmyer theorem, so we conclude that FTCS is stable when $\nu \le 1/2$.

(c) We begin by continuing the computation of the FTCS scheme truncation error by substituting in $f = u$.

$$\begin{aligned}
\tau &= \frac{1}{2}\left(\left(k\alpha^2 - \frac{\alpha h^2}{6}\right)u_{xxxx} + k(\alpha u_{xx} + u_t)\right) + O(k^2 + h^4) \\
&= \frac{k}{2}(2\alpha u_{xx} + u) + O(k^2 + h^4) \\
&= k\alpha u_{xx} + \frac{k}{2}u + O(k^2 + h^4)
\end{aligned}$$

Motivated by the remaining $O(k)$ terms, we modify the FTCS scheme like so:

$$\underbrace{D_t^+ u}_{A} = \alpha(1+k)\underbrace{D_x^+ D_x^- u}_{B} + \left(1 + \frac{k}{2}\right)u$$

The leading order term of this method, assuming $\nu = 1/6$, is calculated below:

$$\begin{aligned}
A &= u_t + \frac{k}{2}u_{tt} + \frac{k^2}{6}u_{ttt} + O(k^3) \\
B &= u_{xx} + \frac{h^2}{12}u_{xxxx} + \frac{h^4}{360}u_{xxxxxx} + O(h^6) \\
\tau &= A - \alpha(1+k)B - \left(1 + \frac{k}{2}\right)u \\
&= \cancel{(u_t - \alpha u_{xx} - u)} \\
&\quad + \frac{k}{2}u_{tt} - \frac{\alpha h^2}{12}u_{xxxx} - \alpha k u_{xx} - \frac{\alpha k h^2}{12}u_{xxxx} - \frac{k}{2}u \\
&\quad + \frac{k^2}{6}u_{ttt} + \frac{h^4}{360}u_{xxxxxx} + O(kh^4 + k^3 + h^6) \\
&= \cancel{\frac{k\alpha^2}{2}u_{xxxx} - \frac{\alpha h^2}{12}u_{xxxx}} \\
&\quad + \frac{\alpha k}{2}u_{xx} + \frac{k}{2}u_t - \alpha k u_{xx} - \frac{k}{2}u \\
&\quad - \frac{\alpha k h^2}{12}u_{xxxx} + \frac{k^2}{6}u_{ttt} + \frac{h^4}{360}u_{xxxxxx} + O(kh^4 + k^3 + h^6) \\
&= \cancel{-\frac{\alpha k}{2}u_{xx}} + \cancel{\frac{\alpha k}{2}u_{xx}} + \cancel{\frac{k}{2}u} - \cancel{\frac{k}{2}u} \\
&\quad - \frac{\alpha k h^2}{12}u_{xxxx} + \frac{k^2}{6}u_{ttt} + \frac{h^4}{360}u_{xxxxxx} + O(kh^4 + k^3 + h^6) \\
&= -\frac{\alpha k h^2}{12}u_{xxxx} + \frac{k^2}{6}u_{ttt} + \frac{h^4}{360}u_{xxxxxx} + O(kh^4 + k^3 + h^6)
\end{aligned}$$

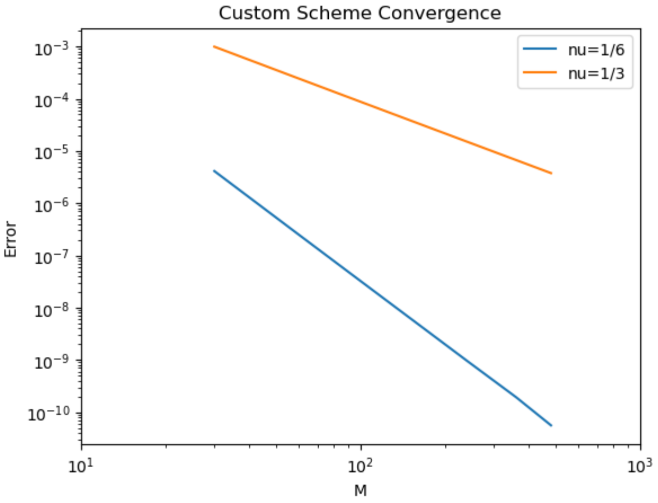
Let $M = \max\left\{\max_{x,t}\left|\partial_t^3 u\right|, \max_{x,t}\left|\partial_x^4 u\right|, \max_{x,t}\left|\partial_x^6 u\right|\right\}$. Then we have the following strict bound on the truncation error:

$$\begin{aligned}\tau &\leq M\left(\frac{\alpha k h^2}{12} + \frac{k^2}{6} + \frac{h^4}{360}\right)\\ &= O(kh^2 + k^2 + h^4).\end{aligned}$$

With the relationship that $k = O(h^2)$, this becomes 4th order accurate in space as we would like. Written below is this scheme explicitly in terms of spatial points:

$$u_j^{n+1} = \nu(1+k)(u_j^{n+1} - 2u_j^n + u_j^{n-1}) + \left(1 + k + \frac{k^2}{2}\right)u_j^n$$

The results below shows the convergence of this method. To handle the Neumann boundaries, we add two ghost nodes at u_{-1} and u_{M+1} with values u_1 and u_{M-1}. These ghost nodes are consistent with the PDE because of the symmetry of its solution, so they do not contribute any additional error.



When $\nu \neq 1/6$, the $O(k + h^2)$ terms do not cancel so we get 2nd order convergence. As a result, the slope of the $\nu = 1/3$ line is ≈ -2. Additionally, when $\nu = 1/6$, we expect a 4th order convergence. Indeed, the slope of the $\nu = 1/6$ line is ≈ -4.3. Interestingly, the slope magnitude is greater than 4, but this is likely because the 4th order term is a function of high order derivatives which may just be small for the given function at later time steps. For values of $T < 1$, we see a slope closer to -4; it is only as we increase T that the order of convergence appears faster.

3 Linear Functional Norms

Consider the linear functional $\rho(f) = \int_0^4 f(x)dx$ in two contexts, first acting on $L^1[0,4]$, and then on $L^2[0,4]$. Compute the norm of ρ in each of these cases, $\|\rho\|_1$ and $\|\rho\|_2$. Recall the norm of a linear functional:

$$\|\rho\| = \sup_{f\neq 0} |\rho(f)|/\|f\| = \sup_{\|f\|=1} |\rho(f)|.$$

Solution:

(a) First, $|\rho(f)| \leq \int_0^4 |f|dx = \|f\|_1$. Then $\|\rho\|_1 \leq \sup_{\|f\|=1} \|f\|_1 = 1$. This upper bound is achievable by $f(x) = \frac{1}{4}$ so we conclude that $\boxed{\|\rho\|_1 = 1}$.

(b) For the case of $\|\rho\|_2$, we can use the fact that the inner product on $L^2[0,4]$ is $\langle f, g\rangle = \int_0^4 fg$. Then, it is clear that $\rho(f) = \langle f, 1\rangle$. In fact, this is just an example of the Riesz Representation theorem applied to Hilbert spaces. By the Cauchy–Schwarz inequality, $|\rho(f)| \leq \|f\|_2 \cdot \|1\|_2$. It follows that $\|\rho\|_2 \leq \sup_{\|f\|=1} \|f\|_2 \cdot \|1\|_2 = \|1\|_2 = \sqrt{\int_0^4 1^2 dx} = 2$. This upper bound is achievable by $f(x) = \frac{1}{2}$ so we conclude that $\boxed{\|\rho\|_2 = 2}$.