1. Determine the values of θ for which the θ method is L-stable.

Solution:
For the theta method applied to $u' = \lambda u$ we have

$$u_{j+1} = u_j + \lambda h(\theta u_j + (1-\theta)u_{j+1})$$

Thus

$$u_{j+1} = \frac{1+z\theta}{1-(1-\theta)z}u_j = R(z)u_j$$

with $R(z) = (1+\theta z)/(1-(1-\theta)z$.

Recall that a one-step method is L-stable if it is A-stable and $R(z) \to 0$ as $z \to \infty$. The θ-methods are A-stable for $0 \le \theta$. Moreover

$$\lim_{z\to\infty} \frac{1+z\theta}{1-(1-\theta)z} = \frac{\theta}{\theta-1}$$

Thus only Backwards Euler ($\theta = 0$) is L-stable.

2. Consider the function $f(x) = x - x^2$ on the interval $[0,1]$.

a) Compute the Fourier sine series coefficients of $f(x)$.

b) The function $f(x)$ looks pretty darn smooth. In class we saw that smoothness should be reflected in the rate of decay of the Fourier coefficients. But the coefficients of $f(x)$ don't decay very fast, only $O(k^3)$. Why doesn't this contradict the theorems we saw in class?

c) It can be shown that $\sum_{k=N}^{\infty} \frac{1}{k^3} \le \frac{1}{2N^2}$. Use this result to show that if $s_N(x)$ is the partial sum of the Fourier sin series of $f(x) = x - x^2$ with N terms, then

$$\max_{0\le x\le 1} |f(x) - s_N(x)| \le \frac{4}{\pi^3 N^2}$$

d) Generate a convincing plot that shows that the error between $f(x)$ and the partial sum $s_N(x)$ of the Fourier sine series N terms converges $O(N^{-2})$. You must show the code used to generate the plot.

Solution, part a:
Recall that the Fourier sine coefficeint is defined by

$$c_k = 2\int_0^1 (x - x^2)\sin(k\pi x)\, dx$$

A routine computation using integration by parts shows

$$c_k = (-1)^{\frac{k-1}{2}} \frac{8}{(\pi k)^3}$$

if k is odd, and $c_k = 0$ otherwise.

Solution, part b:
The theorem presented in class had fine print: it applied to a C^ℓ function but required $u^{(2j)}(0) = 0$ and $u^{(2j)}(1) = 0$ for $2j < \ell$. Although $f(0) = f(1) = 0$, we have $f''(0) = f''(1) = -2$. Thus the theorem can be applied for $\ell = 1, 2, 3$ but not $\ell = 4$ Thus it only implies $O(k^{-3})$ decay and no faster. This is the decay we observe.

Solution, part c:
Observe that

$$\begin{aligned} |f(x) - s_N(x)| &= \left| \sum_{k=N+1}^{\infty} c_k \sin(k\pi x) \right| \\ &\leq \sum_{k=N+1}^{\infty} |c_k| \leq \sum_{k=N+1}^{\infty} \frac{8}{\pi^3} \frac{1}{k^3} \\ &= \frac{8}{\pi^3} \sum_{k=N+1}^{\infty} \frac{1}{k^3} \\ &\leq \frac{8}{\pi^3} \frac{1}{2} \frac{1}{(N+1)^2} \\ &\leq \frac{4}{\pi^3} \frac{1}{N^2}. \end{aligned} \tag{1}$$

This estimate does not depend on x, so

$$\max_{x \in [0,1]} |f(x) - s_N(x)| \leq \frac{4}{\pi^3 N^2}.$$

Solution, part d:
See worksheet.

3. Suppose u is a solution of $u_t = u_{xx}$ for $0 \leq x \leq 1$ and $t \geq 0$ with boundary condition $u|_{x=0,1} = 0$.

a) Suppose that u has j continuous time derivatives and $2j$ continuous space derivatives everywhere on its domain for some $j = 1, 2, 3, \ldots$. Show that $(\partial_x)^{2j} u = 0$ at $x = 0, 1$.

b) Suppose you solve this problem with initial data $u(x, 0) = x - x^2$. Does the solution have one continuous time derivative and two continuous space derivatives everywhere on its domain? Justify you answer briefly.

Solution, part a:
Since $u = 0$ at $x = 0$ we have $u_t = 0$. But $u_t = u_{xx}$, so $u_{xx} = 0$ at $x = 0$ as well. Similarly, $u_{tt} = 0$ at $x = 0$. But

$$u_{tt} = \partial_t u_t = \partial_t u_{xx} = \partial_x^2 u_t = \partial_x^2 u_{xx} = u_{xxxx}.$$

Thus, assuming u has two continuous time derivative and four continuous space derivatives, $u_{xxxx} = 0$ at $x = 0$. This same argument generalizes to show that if u has j continuous time derivatives and $2j$ continuous space derivatives, then $(\partial_x)^{2j} u = 0$ at $x = 0$.

A similar argument applies at $x = 1$.

Solution, part b:

Note that the initial data satisfies $f_{xx}(0) = -2 \neq 0$. Thus u cannot have one continuous time derivative and two continuous space derivatives in the domain.

The fundamental issue occurs as a mismatch between the boundary conditions and the PDE. The Dirichlet condition enforces $u_t = 0$ at $x = 0$. The initial data enforces $u_{xx} = -2$ everywhere. The PDE has $u_t = u_{xx}$. We cannot have all three of these statements hold at $t = 0$ and $x = 0$ since $-2 \neq 0$.

4. Recall the partial sums $s_N(x)$ from problem 2. Suppose $u_N(x,t)$ is the solution of the heat equation $u_t = u_{xx}$ for $0 \leq x \leq 1$ with $u_N(x,0) = s_N(x)$ with Dirichlet boundary conditions. Suppose $u(x,t)$ the the solution of the heat equation with the same boundary conditions but with $u(x,0) = x - x^2$. Show that $|u_N(x,t) - u(x,t)| < 10^{-7}$ for all $x \in [0,1]$ and all $t \geq 0$ if $N = 1200$.

Solution:

Note that

$$u(x,t) - u_N(x,t) = \sum_{k=N+1}^{\infty} c_k e^{-k^2\pi^2 t} \sin(k\pi x)$$

where the c_k's are the Fourier coefficients from problem 2. In particular, each $|c_k| \leq 8/(\pi^3 k^3)$. Thus, using the fact that $|e^{-k^2\pi^2 t}| \leq 1$ and $|\sin(k\pi x)| \leq 1$ we find

$$|u(x,t) - u_N(x,t)| \leq \sum_{k=N+1}^{\infty} |c_k|.$$

The argument of problem 2b shows

$$\sum_{k=N+1}^{\infty} |c_k| \leq \frac{8}{\pi^3 N^2}$$

If $N = 1200$ then $8/(\pi^3 N^2 \approx 0.9 \times 10^{-8}$ and hence $|u(x,t) - u_N(x,t) < 10^{-7}$.

5. Suppose we wish to solve $u_t = u_{xx}$ with homogeneous Dirichlet boundary conditions and $u(x,0) = x - x^2$. The aim of this exercise (and indeed this entire assignment) is to show that if the solution of the heat equation isn't smooth, then the order of accuracy of your numerical solution can be reduced from the accuracy expected using arguments that use smoothness.

We don't know the exact solution of the heat equation with this initial condition. But by the previous problem, we know that we can compute an approximate solution with a

known error by using the series solution with 1200 terms. So this will play the role of the "exact" solution, which is good enough until we see errors on the order of 10^{-7}.

We are going to compare solving the heat equation with homogeneous Dirichlet boundary conditions with initial condition $f_1(x) = x - x^2$ and with initial condition $f_2(x) = \sin(\pi x)/4$.

a) Generate a graph of $f_1(x)$ and $f_2(x)$ for $0 \le x \le 1$. This step is just to convince you that these initial conditions look "close" to each other.

b) For $N = 50, 100, 500, 1000, 5000$ and $M = 2N$, generate a solution of the heat equation using backwards Euler and initial condition $f_2(x) = \sin(\pi x)/4$. Then compute the error at the first time step (i.e. at the first time beyond $t = 0$). Generate a log-log plot of the error versus N and compute the order of convergence.

c) Repeat the above but measuring error at the final time step. Why do you see the two orders of convergence you observe in this part and in the previous part? Why are they different?

d) Repeat parts b) and c), but with Crank Nicolson.

e) Now generate the same log-log plots of error (with computed orders of convergence) when the initial condition is $f_1(x) = x - x^2$. There should be four log-log plots (first time step and last time step for each of backward Euler and Crank Nicolson).

f) Discuss the differences you see between the various convergence plots for f_1 and their corresponding plots for f_2.

Solution, part a:
See worksheet.

Solution, part b:
See worksheet.

Solution, part c:
See worksheet.

Solution, part d:
See worksheet.

Solution, part e:
For Backwards Euler and $u_0(x) = \sin(\pi x)/4$ we find that on the first time step the error is $O(k^2)$ and the final error is $O(k)$. In effect the final error results from compounding $O(1/k)$ errors of size $O(k^2)$ to get an $O(k)$ final error.

For Crank Nicolson and this same initial data there is a similar analysis except that the order of convergence is one better. The error at the initial timestep corresponds to k times

the local truncation error size, and is $O(k^3)$, and the error at the final timestep is $O(k^2)$ for the same reasons outlined above.

For $u_0(x) = x - x^2$ there is a difference at the first timestep. For both Backwards Euler and Crank Nicholson an error or size $O(k)$ has occured. Thus the global rate of convergence can't be any better than $O(k)$ for both of these methods; Crank Nicolson doesn't experience better overall convergence. Nevertheless, by the final timestep we still see $O(k)$ convergence for Backwards Euler and $O(k^2)$ convergence for Crank Nicolson. We got lucky! The error in the first timestep must arise from high-frequency contributions, which are scaled away by exponential factor that decay faster than $O(k^2)$ to preserve the final error.