Jeremiah Mcdowell
MAT 275 Honors

Numerical and Analytic Analysis of the Heat Equation

I. Introduction

The heat equationis perhaps the most fundamental tool for modeling distribution of heat in an object. As such, it has become familiar and widely applicable in many scientific fields. In addition to its modeling uses, for example, it is of interest to mathematicians as a prototypical example of a parabolic partial differential equation. In other fields, it is used to study or solve other partial differential equations, such as the Black-Scholes model in financial mathematics.

Its relative simplicity and specificity also qualify it as an excellent subject for a novice's foray into partial differential equations. Because its basic concepts are intuitive and relatively clear to understand. The heat equation represents a simple but useful case study that can provide insight into basic aspects of partial differential equations that can then be generalized to more complicated equations.

The heat equation expresses the change in the temperature as a function of time and one spatial dimension given by $u(t, x)$ in terms of the function's partial derivatives. The heat equation has the form

$$u_t = u_{xx}$$

or, equivalently,

$$\mathcal{L}(u) = u_{xx} - u_t = 0.$$

II. Derivation

The heat equation comes about as a consequence of Fourier's law. Fourier proclaimed in *TheorieAnalytique de la Chaleur* that the unit rate of flow of heat energy q is proportional to the gradient of the temperature u,

$$q = -k * \nabla u,$$

Where k is called the conductivity of the system. In one dimension,

$$\nabla u = \frac{\partial u}{\partial x} = u_x,$$

Hence the equation becomes

$$q = -ku_x.$$

Additionally, the change ΔQ of an object's internal energy can be given by means of the formula

$$\Delta Q = c\rho\Delta u,$$

Where the constant of proportionality c is called the object's capacity, ρ is the object's density, and Δu is the change in temperature. Given an initial condition of zero energy at zero temperature, this then becomes

$$Q = c\rho u.$$

If we consider (x, t) to represent a point in the Cartesian product space of time and space, we consider a rectangle

$$R = \{(\varepsilon, \tau) : x - \Delta x \leq \varepsilon \leq x + \Delta x \text{ and } t - \Delta t \ \leq \tau \leq t + \Delta t$$

Then the increase in internal energy in a small spatial region

$$x - \Delta x \ \leq \varepsilon \leq x + \Delta x$$

over the small time interval $$t - \Delta t \leq \tau \leq t + \Delta t$$

is given by

$$c\rho \int_{x-\Delta x}^{x+\Delta x} \{u(\varepsilon, t + \Delta t) - u(\varepsilon, t - \Delta t)\}\, d\varepsilon = \iint_R \frac{\partial u}{\partial \tau}\, d\varepsilon\, d\tau = \iint_R u_\tau\, d\varepsilon\, d\tau.$$

where the fundamental theorem of calculus is employed. In the absence of work, heat sources, or heat sinks, this change is wholly accounted for by the total flow of heat energy into the region $[x - \Delta x, x + \Delta x]$, which is given by

$$k \int_{t-\Delta t}^{t+\Delta t} \left\{\frac{\partial u}{\partial x}(x + \Delta x, \tau) - \frac{\partial u}{\partial x}(x - \Delta x, \tau)\right\} d\tau = \iint_R \frac{\partial^2 u}{\partial \varepsilon^2}\, d\varepsilon\, d\tau = \iint_R u_{\varepsilon\varepsilon}\, d\varepsilon\, d\tau$$

according to Fourier's law. By conservation of energy,

$$\iint_R c\rho u_\tau - k u_{\varepsilon\varepsilon}\, d\varepsilon\, d\tau = 0$$

for any Δx and Δt. By the fundamental lemma of calculus of variations, we have that

$$c\rho u_t - k u_{xx} = 0,$$

thus,

$$u_t = \frac{k}{c\rho} u_{xx},$$

where the constant $$\alpha = \frac{k}{c\rho}$$

is the thermal diffusivity of the material. If we set $\tau = \alpha t$ and relabel τ as t, we have the classical equation

$$u_{xx} - u_t = 0.$$

III. Difference Quotient

We proceed in our solution to examine the heat equation using difference quotients for U_t and U_{xx}. Using the second-order central difference quotient with respect to space to represent U_{xx} and the first order forward difference quotient with respect to time to represent U_t, we write

$$\frac{U_{k+1,i} - u_{k,i}}{\Delta t} = \frac{U_{k+1,i-1} - 2U_{k+1,i} + U_{k+1,i+1}}{(\Delta x)^2}$$

(Note that we choose to center our central difference quotient about the time U_{k+1}, as we have shown that this will yield a stable solution, the analysis of which is in the following section)

We now say $D = \frac{\Delta t}{(\Delta x)^2}$ for the purposes of visually simplifying the equation, which then becomes

$$U_{k+1,i} - u_{k,i} = DU_{k+1,i-1} - 2DU_{k+1,i} + DU_{k+1,i+1}$$

$$U_{k,i} = -DU_{k+1,i-1} + (1 + 2D)U_{k+1,i} - DU_{k+1,i+1}.$$

Now in this form, we can examine this equation for all values of i from 1 to n to construct a profile of the temperature of the object along its entire length. This can in fact be represented by a matrix equation, thusly:

$$\begin{bmatrix} 1+2D & -D & 0 & 0 & \cdots & 0 \\ -D & 1+2D & -D & 0 & \cdots & 0 \\ 0 & -D & 1+2D & -D & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \cdots & \vdots \\ 0 & \cdots & 0 & -D & 1+2D & -D \\ 0 & \cdots & 0 & 0 & -D & 1+2D \end{bmatrix} \begin{bmatrix} U_{k+1,1} \\ U_{k+1,2} \\ U_{k+1,3} \\ \vdots \\ U_{k+1,n-1} \\ U_{k+1,n} \end{bmatrix} = \begin{bmatrix} U_{k,1} \\ U_{k,2} \\ U_{k,3} \\ \vdots \\ U_{k,n-1} \\ U_{k,n} \end{bmatrix}$$

Note that the difference quotient form breaks for the 1st and nth rows, as $U_{k+1,i-1}$ and $U_{k+1,i+1}$are not properly evaluable in these respective cases. One option for alleviating this problem is to factor in the boundary conditions for the equation:

$$-DU_{k+1,0} + (1+2D)U_{k+1,1} - DU_{k+1,2} = U_{k,1}$$

$$\rightarrow (1+2D)U_{k+1,1} - DU_{k+1,2} = U_{k,1} + \left(DU_{k+1,0}\right) = U_{k,1} + DU_L,$$

Where U_L represents the boundary condition at the left. Similarly,

$$-DU_{k+1,n-1} + (1+2D)U_{k+1,n} - DU_{k+1,n+1} = U_{k,n}$$

$$\rightarrow -DU_{k+1,n-1} + (1+2D)U_{k+1,n} = U_{k,n} + \left(DU_{k+1,n+1}\right) = U_{k,n} + DU_R,$$

With U_R the boundary condition at the right. For the purposes of our model, the boundary conditions of the plate do not change over time. Taking this into consideration, our matrix equation becomes:

$$\begin{bmatrix} 1+2D & -D & 0 & 0 & \cdots & 0 \\ -D & 1+2D & -D & 0 & \cdots & 0 \\ 0 & -D & 1+2D & -D & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \cdots & \vdots \\ 0 & \cdots & 0 & -D & 1+2D & -D \\ 0 & \cdots & 0 & 0 & -D & 1+2D \end{bmatrix} \begin{bmatrix} U_{k+1,1} \\ U_{k+1,2} \\ U_{k+1,3} \\ \vdots \\ U_{k+1,n-1} \\ U_{k+1,n} \end{bmatrix} = \begin{bmatrix} U_{k,1} + DU_L \\ U_{k,2} \\ U_{k,3} \\ \vdots \\ U_{k,n-1} \\ U_{k,n} + DU_R \end{bmatrix}$$

We then multiply both sides by the inverse of our coefficient matrix to get a form that can generate the values of the equation at each step from the previous.

$$\begin{bmatrix} U_{k+1,1} \\ U_{k+1,2} \\ U_{k+1,3} \\ \vdots \\ U_{k+1,n-1} \\ U_{k+1,n} \end{bmatrix} = \begin{bmatrix} 1+2D & -D & 0 & 0 & \cdots & 0 \\ -D & 1+2D & -D & 0 & \cdots & 0 \\ 0 & -D & 1+2D & -D & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \cdots & \vdots \\ 0 & \cdots & 0 & -D & 1+2D & -D \\ 0 & \cdots & 0 & 0 & -D & 1+2D \end{bmatrix}^{-1} \begin{bmatrix} U_{k,1}+DU_L \\ U_{k,2} \\ U_{k,3} \\ \vdots \\ U_{k,n-1} \\ U_{k,n}+DU_R \end{bmatrix}$$

IV. Stability

For any differential equation$U_{n,p,r\ldots}$ examined with respect to *n*, we say that the function value at some future point *n+1* is equal to some factor times the value at the current point*n*- that is,

$$U_{n+1} = G * U_n$$

with G being the growth factor of the equation. In particular, this applies to the error function *e*:

$$e_{n+1} = G * e_n$$

In this case, if the absolute value of G is less than 1, we say that the equation is stable, as the error decreases. In order to assess when this is true for our equation of two variables *t* and *x*, we say that

$$U_{n,j} = e^{ikx_{j+1}}$$

Then, advancing with respect to *n*, we say that the advanced term can be found by adding the finite space representation of the equation to the current term:

$$U_{n+1,j} = U_{n,j} + \Delta t(\frac{U_{n,j+1} - U_{n,j-1}}{2\Delta x})$$

Then, making a substitution with respect to our earlier equation, this becomes

$$GU_{n,j} = U_{n,j} + (\frac{\Delta t}{2\Delta x})(U_{n,j+1} - U_{n,j-1})$$

$$Ge^{ikx_j} = e^{ikx_j} + (\frac{\Delta t}{2\Delta x})(e^{ik(x+\Delta x)} - e^{ik(x-\Delta x)})$$

$$G = 1 + \left(\frac{\Delta t}{\Delta x}\right)\left(\frac{e^{ik\Delta x} - e^{-ik\Delta x}}{2}\right)$$

$$G = 1 + \left(\frac{\Delta t}{\Delta x}\right) i(\sin(k\Delta x))$$

Since stability necessitates that the absolute value of *G* be less than 1, we also know that the absolute value of G times its complex conjugate must also be less than 1:

$$|G * G^*| < 1$$

$$\left|\left(1 + \frac{\Delta t}{\Delta x} i * \sin(k\Delta x)\right)\left(1 - \frac{\Delta t}{\Delta x} i * si\,n(k\Delta x)\right)\right| < 1$$

$$\left|1 + \left(\frac{\Delta t}{\Delta x} i * \sin(k\Delta x)\right)^2\right| < 1:$$

This is never true- thus, this explicit solution is never stable.

However, this is not the only solution we can use to evaluate our function, if we change our stencil by choosing a different reference point, then the results may change in response.

Suppose we select the future point $U_{n+1,j}$ as our stencil instead of $U_{n,j}$. Then, we have

$$U_{n+1,j} = U_{n,j} + \Delta t * \left(\frac{U_{n+1,j+1} - U_{n+!,j-1}}{2\Delta x}\right)$$

$$Ge^{ikx_j} = e^{ikx_j} + \left(\frac{\Delta t}{\Delta x}\right)(\frac{Ge^{ik(x+\Delta x)} - Ge^{ik(x-\Delta x)}}{2})$$

$$G = \left(\frac{\Delta t}{\Delta x}\right) G \left(\frac{e^{ik\Delta x} - e^{-ik\Delta x}}{2}\right) + 1$$

$$G\left(1 - \left(\frac{\Delta t}{\Delta x}\right) i * \sin(k\Delta x)\right) = 1$$

$$G = \frac{1}{\left(1 - \left(\frac{\Delta t}{\Delta x}\right) i * si\,n(k\Delta x)\right)}$$

$$|G * G^{*}| < 1$$

$$\left|\frac{1}{\left(1 + \left(\frac{\Delta t}{\Delta x} * si\,n(k\Delta x)\right)^{2}\right)}\right| < 1$$

This equation is true when the denominator is greater than 1, which is true for all values of t and x. Hence, this solution is always stable.

V. Analytic Solution

While numerical solutions given by these sorts of matrices are guaranteed to be accurate, the computational time required for the generation of matrices when *n* becomes very large grows increasingly prohibitive. As such, another option for analyzing the behavior of this equation is to use analytical approaches to examine an approximation accurate to within some degree of error. In particular, we can utilize a Fourier series approximation to examine the behavior of our PDE if we decompose it into three ODEs.

For this derivation, let us assume a plate, linearized to have sides of length 1 with boundaries of

$$u(0, y, t) = u(1, y, t) = u(x, 0, t) = u(x, 1, t) = 0$$

And an initial function $f(x, y)$ that gives the temperature distribution at $t = 0$.

Let us assume that our temperature function decomposes into three functions, one with respect to x,one with respect to y, andone with respect to t, as follows:

$$U(x, y, t) = X(x)Y(y)T(t)$$

Converting to the form of the heat equation, we have

$$T'(t)X(x)Y(y) = T(t)X''(x)Y(y) + T(t)X(x)Y''(y)$$

And consequently,

$$\frac{X''(x)}{X(x)} + \frac{Y''(y)}{Y(y)} = \frac{T'(t)}{T(t)}$$

For a function purely of time and one purely of space to be equal to one another across all values of x, y, and t is, reasonably, only possible when both sides of the equation are equal to some constant *C*. We call this constant $C = -\lambda^2 - m^2$ by convention. Then, applying our original PDE setup yields

$$T'(t) + (\lambda^2 + m^2)\, T(t) \; = \; 0 \text{ and } X''Y + Y''X + \lambda^2 + m^2 = 0$$

for our two ODEs. Considering the second equation, we write it as

$$-\left(\frac{Y''}{Y} + \lambda^2\right) = \frac{X''}{X} + m^2$$

Again, we say that both sides are equal to some constant μ^2, since the left side is solely a function of y and the right solely a function of x. This then gives us

$$\frac{Y''}{Y} + \lambda^2 = -\mu^2 \text{and} \frac{X''}{X} + m^2 = \mu^2$$

For which we can set up the standard-form equations

$$Y'' + (-\lambda^2 + \mu^2)Y = 0 \ \text{and} X'' + (-m^2 - \mu^2)X = 0$$

Now, let $(-\lambda^2 + \mu^2) = l^2$ and $(-m^2 - \mu^2) = k^2$. We can now solve the equations for X and Y:

$$X(x) = Asin(lx) + Bcos(lx)$$

$$Y(y) = Csin(ky) + Dcos(ky)$$

And, looking back to the equation for time, we find that

$$T(t) = T_0 e^{-\lambda^2 - m^2}$$

Thus, our equation for heat can be written as

$$U(x, y, t) = T_0 e^{(-\lambda^2 - m^2)t}\big(Gsin(lx) + Hcos(lx)\big)\big(Csin(ky) + Dcos(ky)\big).$$

From here, we can utilize our boundary conditions:

$$0 \ = \ U(0, y, 0) \ = \ H\big(Csin(ky) + Dcos(ky)\big) \ \Rightarrow \ H \ = \ 0$$

$$0 \ = \ U(x, 0,0) = \ D\big(Gsin(lx) + Hcos(lx)\big) \ \Rightarrow \ D \ = \ 0$$

At this point, we find it easier to combine our constants into a single term:

$$U(x, y, t) = e^{(-\lambda^2 - m^2)t} Asin(lx)\sin(ky)$$

Now, resuming applying the boundary conditions, we have that

$$U(1, y, 0) = Asin(l)\sin(ky) = 0 \text{ and } U(x, 1,0) = Asin(lx)\sin(k) = 0$$

Then l and $k = L\pi$ and $K\pi$ respectively, where $L, K = 1,2,3 \ldots$;

Thus, we find that our coefficient A, and indeed our entire equation, depends on the values we choose for L and K;

$$U_{L,K}(x,y,t) = e^{((L\pi)^2+(K\pi)^2)t} A_{L,K} \sin(L\pi x) \sin(K\pi y)$$

Finally, then,

$$U_{L,K}(x,y,0) = A_{L,K} \sin(L\pi x) \sin(K\pi y)$$

And if we take the sums of $U_{L,K}(x,y,0)$ across all L, K,

$$\sum_{K=0}^{\infty} \sum_{L=0}^{\infty} U_{L,K}(x,y,0) = \sum_{K=0}^{\infty} \sum_{L=0}^{\infty} A_{L,K} \sin(L\pi x) \sin(K\pi y) = f(x,y) = U_0$$

Which is our initial heat distribution.

From here, we can begin to construct our Fourier series approximation. Take

$$\sum_{L,K=0}^{\infty} \sin(n\pi x) A_{L,K} \sin(L\pi x) \sin(K\pi y) = \sum_{L,K=0}^{\infty} \sin(n\pi x) U_0$$

Then, converting both sides into integrals gives us

$$\int_0^1 \sin(n\pi x) A_{L,K} \sin(L\pi x) \sin(K\pi y) dx = \int_0^1 \sin(n\pi x) U_0 dx$$

$$\Rightarrow A \sin(K\pi y) \left[\left. \frac{-\cos(n\pi x)\sin(L\pi x)}{n\pi} \right|_0^1 + \frac{L}{n\pi} \int_0^1 \cos(n\pi x)\cos(L\pi x) dx \right]$$

$$\Rightarrow A \sin(K\pi y) \left[\left. \frac{-\cos(n\pi x)\sin(L\pi x)}{n\pi} \right|_0^1 + \frac{L}{n\pi} \left[\left. \frac{\sin(n\pi x)\cos(L\pi x)}{n\pi} \right|_0^1 + \frac{L}{n\pi} \int_0^1 \sin(n\pi x)\sin(L\pi x) dx \right] \right]$$

$$\Rightarrow A \sin(K\pi y) \left[(\frac{L}{n\pi})^2 \int_0^1 \sin(n\pi x)\sin(L\pi x) dx \right]$$

If $n \neq L$, this is zero always; hence, we presume $n = L$. Then we have

$$\Rightarrow A\sin(K\pi y)\left[(\frac{1}{\pi})^2 \int_0^1 \sin^2(L\pi x)dx \right]$$
$$\Rightarrow A\sin(K\pi y) * \frac{1}{2\pi^2}\int_0^1 1-\cos(2L\pi x)dx$$
$$\Rightarrow A\sin(K\pi y) = \frac{1}{2}\int_0^1 \sin(L\pi x)U_0 dx$$

We then perform a similar technique with respect to y:

$$\sum_{L,K=0}^{\infty} \sin(m\pi y)A_{L,K}\sin(L\pi x)\sin(K\pi y) = \sum_{L,K=0}^{\infty} \sin(m\pi y)U_0$$
$$\Rightarrow \int_0^1 \sin(m\pi y)A_{L,K}\sin(L\pi x)\sin(K\pi y)dy = \int_0^1 \sin(m\pi y)U_0 dy$$
$$\Rightarrow A\sin(L\pi x)\left[\frac{-\cos(m\pi y)\sin(K\pi y)}{m\pi}\bigg|_0^1 + \frac{K}{m\pi}\int_0^1 \cos(m\pi y)\cos(K\pi y)dy \right]$$
$$\Rightarrow A\sin(L\pi x)\left[\frac{-\cos(m\pi y)\sin(K\pi y)}{m\pi}\bigg|_0^1 + \frac{K}{m\pi}\left[\frac{\sin(m\pi y)\cos(K\pi y)}{m\pi}\bigg|_0^1 + \frac{K}{m\pi}\int_0^1 \sin(m\pi y)\sin(K\pi y)dy \right]\right]$$
$$\Rightarrow A\sin(L\pi x)\left[(\frac{K}{m\pi})^2 \int_0^1 \sin(m\pi y)\sin(K\pi y)dy \right]$$

Then once again, we let $m = K$;

$$\Rightarrow A\sin(L\pi x)\left[(\frac{1}{\pi})^2 \int_0^1 \sin^2(K\pi y)dy \right]$$
$$\Rightarrow A\sin(L\pi x) * \frac{1}{2\pi^2}\int_0^1 1-\cos(2K\pi y)dy$$
$$\Rightarrow A\sin(L\pi x) = \frac{1}{2}\int_0^1 \sin(K\pi y)U_0 dy$$

Combining these two gives us that

$$\frac{A_{L,K}}{4} = \int_0^1\int_0^1 \sin(n\pi x)\sin(m\pi y)U_0 dydx = \int_0^1\int_0^1 \sin(L\pi x)\sin(K\pi y)U_0 dydx$$

$$\Rightarrow A_{L,K} = 4\int_0^1\int_0^1 \sin(L\pi x)\sin(K\pi y)U_0 dydx$$

Then, finally, we have our complete Fourier series in terms of L and K:

$$U(x,y,t) = \sum_{L,K=0}^{\infty} e^{((L\pi)^2+(K\pi)^2)t} A_{L,K}\sin(L\pi x)\sin(K\pi y)$$

$$= \sum_{L,K=0}^{\infty} e^{((L\pi)^2+(K\pi)^2)t}\sin(L\pi x)\sin(K\pi y)(\int_0^1\int_0^1 \sin(L\pi x)\sin(K\pi y)U_0 dydx)$$