We would like to solve the heat (diffusion) equation
\[ u_t - ku_{xx} = 0, \quad (1) \]
and obtain a solution formula depending on the given initial data, similar to the case of the wave equation. However, the methods that we used to arrive at d’Alambert’s solution for the wave IVP do not yield much for the heat equation. To see this, recall that the heat equation is of parabolic type, and hence, it has only one family of characteristic lines. If we rewrite the equation in the form
\[ ku_{xx} + \cdots = 0, \]
where the dots stand for the lower order terms, then you can see that the coefficients of the leading order terms are
\[ A = k, \quad B = C = 0. \]
The slope of the characteristic lines are then
\[ \frac{dt}{dx} = \frac{B \pm \sqrt{\Delta}}{2A} = \frac{B}{2A} = 0. \]
Consequently, the single family of characteristic lines will be given by
\[ t = c. \]
These characteristic lines are not very helpful, since they are parallel to the \( x \) axis. Thus, one cannot trace points in the \( xt \) plane along the characteristics to the \( x \) axis, along which the initial data is defined. Notice that there is also no way to factor the heat equation into first order equations, either, so the methods used for the wave equation do not shed any light on the solutions of the heat equation.

Instead, we will study the properties of the heat equation, and use the gained knowledge to devise a way of reducing the heat equation to an ODE, as we have done for every PDE we have solved so far.

### 8.1 The maximum principle

The first properties that we need to make sure of, are the uniqueness and stability for the solution of the problem with certain auxiliary conditions. This would guarantee that the problem is wellposed, and the chosen auxiliary conditions do not break the physicality of the problem. We begin by establishing the following property, that will be later used to prove uniqueness and stability.

**Maximum Principle.** If \( u(x, t) \) satisfies the heat equation (1) in the rectangle \( R = \{0 \leq x \leq l, 0 \leq t \leq T \} \) in space-time, then the maximum value of \( u(x, t) \) over the rectangle is assumed either initially \( (t = 0) \), or on the lateral sides \( (x = 0, \text{ or } x = l) \).

Mathematically, the maximum principle asserts that the maximum of \( u(x, t) \) over the three sides must be equal to the maximum of \( u(x, t) \) over the entire rectangle. If we denote the set of points comprising the three sides by \( \Gamma = \{(x, t) \in R \mid t = 0 \text{ or } x = 0 \text{ or } x = l\} \), then the maximum principle can be written as
\[
\max_{(x, t) \in \Gamma} \{u(x, t)\} = \max_{(x, t) \in R} \{u(x, t)\}. \quad (2)
\]

If you think of the heat conduction phenomena in a thin rod, then the maximum principle makes physical sense, since the initial temperature, as well as the temperature at the endpoints will dissipate through conduction of heat, and at no point the temperature can rise above the highest initial or endpoint temperature. In fact, a stronger version of the maximum principle holds, which asserts that the maximum over the rectangle \( R \) can not be attained at a point not belonging to \( \Gamma \), unless \( u \equiv \text{constant} \), i.e. for nonconstant solutions the following strict inequality holds
\[
\max_{(x, t) \in R \backslash \Gamma} \{u(x, t)\} < \max_{(x, t) \in R} \{u(x, t)\},
\]
where \( R \setminus \Gamma \) is the set of all points of \( R \) that are not in \( \Gamma \) (difference of sets). This makes physical sense as well, since the heat from the point of highest initial or boundary temperature will necessarily transfer to points of lower temperature, thus decreasing the highest temperature of the rod.

We finally note, that the maximum principle also implies a minimum principle, since one can apply it to the function \(-u(x, t)\), which also solves the heat equation, and make use of the following identity,

\[
\min \{u(x, t)\} = -\max \{-u(x, t)\}.
\]

Thus, the minima points of the function \( u(x, t) \) will exactly coincide with the maxima points of \(-u(x, t)\), of which, by the maximum principle, there must necessarily be in \( \Gamma \).

**Proof of the maximum principle.** If the maximum of the function \( u(x, t) \) over the rectangle \( R \) is assumed at an internal point \((x_0, t_0)\), then the gradient of \( u \) must vanish at that point, i.e. \( u_t(x_0, t_0) = u_x(x_0, t_0) = 0 \). If in addition we had the strict inequality \( u_{xx}(x_0, t_0) < 0 \), then one would get a contradiction by plugging the point \((x_0, t_0)\) into the heat equation. Indeed, we would have

\[
u_t(x_0, t_0) - ku_{xx}(x_0, t_0) = -ku_{xx}(x_0, t_0) > 0.
\]

This contradicts the heat equation (1), which must hold for all values of \( x \) and \( t \). Thus, the contradiction would imply that the maximum point \((x_0, t_0)\) cannot be an internal point. However, from the second derivative test we have the weaker inequality \( u_{xx}(x_0, t_0) \leq 0 \) (the point would not be a maximum if \( u_{xx}(x_0, t_0) > 0 \)), which is not enough for this argument to go through.

The way out, is to recycle the above argument with a slight modification to the function \( u \). Define a new function

\[
v(x, t) = u(x, t) + \epsilon x^2, \tag{3}
\]

where \( \epsilon > 0 \) is a constant that can be taken as small as one wants. Now let \( M \) be the maximum value of \( u \) over the three sides, which we denoted by \( \Gamma \) above. That is

\[
M = \max_{(x,t) \in \Gamma} \{u(x,t)\}.
\]

To prove the maximum principle, we need to establish (2). The maximum over \( \Gamma \) is always less than or equal to the maximum over \( R \), since \( \Gamma \subset R \). So we only need to show the opposite inequality, which would follow from showing that

\[
u(x, t) \leq M, \quad \text{for all the points } (x, t) \in R.
\]

(4)

Notice that from the definition of \( v \), we have that at the points of \( \Gamma \), \( v(x, t) \leq M + \epsilon l^2 \), since the maximum value of \( \epsilon x^2 \) on \( \Gamma \) is \( \epsilon l^2 \). Then, instead of proving inequality (4), we will prove that

\[
v(x, t) \leq M + \epsilon l^2, \quad \text{for all the points } (x, t) \in R, \tag{5}
\]

which implies (4). Indeed, from the definition of \( v \) in (3), we have that in the rectangle \( R \)

\[
u(x, t) \leq u(x, t) + \epsilon x^2 \leq M + \epsilon l^2 - x^2,
\]

where we used (5) to bound \( v(x, t) \). Now, since the point \((x, t)\) is taken from the rectangle \( R \), we have that \( 0 \leq x \leq l \), and the difference \( l^2 - x^2 \) is bounded. But then the right hand side of the above inequality can be made as close to \( M \) as possible by taking \( \epsilon \) small enough, which implies the bound (4).

If we formally apply the heat operator to the function \( v \), and use the definition (3), we will get

\[
v_t - kv_{xx} = u_t - k(u_{xx} + 2\epsilon) = (u_t - ku_{xx}) - 2k\epsilon < 0,
\]

since both \( k, \epsilon > 0 \), and \( u \) satisfies the heat equation (1) on \( R \). Thus, \( v \) satisfies the heat inequality in \( R \)

\[
v_t - kv_{xx} < 0. \tag{6}
\]
We now recycle the above argument, which barely failed for \( u \), applying it to \( v \) instead. Suppose \( v(x, t) \) attains its maximum value at an internal point \((x_0, t_0)\). Then necessarily \( v_t(x_0, t_0) = 0 \), and \( v_{xx}(x_0, t_0) \leq 0 \). Hence, at this point we have
\[
v_t(x_0, t_0) - kv_{xx}(x_0, t_0) = -kv_{xx}(x_0, t_0) \geq 0,
\]
which contradicts the heat inequality (6). Thus, \( v \) cannot have an internal maximum point in \( R \).

Similarly, suppose that \( v(x, t) \) attains its maximum value at a point \((x_0, t_0)\) on the fourth side of the rectangle \( R \), i.e. when \( t_0 = T \). Then we still have that \( v_x(x_0, t_0) = 0 \), and \( v_{xx}(x_0, t_0) \leq 0 \), but \( v_t(x_0, t_0) \) does not have to be zero, since \( t_0 = T \) is an endpoint in the \( t \) direction. However, from the definition of the derivative, and our assumption that \((x_0, t_0)\) is a point of maximum, we have
\[
v_t(x_0, t_0) = \lim_{\delta \to 0^+} \frac{v(x_0, t_0) - v(x_0, t_0 - \delta)}{\delta} \geq 0.
\]
So at this point we still have
\[
v_t(x_0, t_0) - kv_{xx}(x_0, t_0) \geq 0,
\]
which again contradicts the heat inequality (6).

Now, since the continuous function \( v(x, t) \) must attain its maximum value somewhere in the closed rectangle \( R \), this must happen on one of the remaining three sides, which comprise the set \( \Gamma \). Hence,
\[
v(x, t) \leq \max_{(x,t) \in R} \{v(x, t)\} = \max_{(x,t) \in \Gamma} \{v(x, t)\} \leq M + \epsilon l^2,
\]
which finishes the proof of (5).

### 8.2 Uniqueness

Consider the Dirichlet problem for the heat equation,
\[
\begin{cases}
  u_t - ku_{xx} = f(x, t) & \text{for } 0 \leq x \leq l, \quad t > 0 \\
  u(x, 0) = \phi(x), & u(0, t) = g(t), \quad u(l, t) = h(t),
\end{cases}
\]
for given functions \( f, \phi, g, h \). We will use the maximum principle to show uniqueness and stability for the solutions of this problem (recall that last time we used the energy method to prove uniqueness for the same problem).

**Uniqueness of solutions.** There is at most one solution to the Dirichlet problem (7).

Indeed, arguing from the inverse, suppose that there are two functions, \( u \), and \( v \), that both solve the inhomogeneous heat equation and satisfy the initial and Dirichlet boundary conditions of (7). Then their difference, \( w = u - v \), satisfies the homogeneous heat equation with zero initial-boundary conditions, i.e.
\[
\begin{cases}
  w_t - kw_{xx} = 0 & \text{for } 0 \leq x \leq l, \quad t > 0 \\
  w(x, 0) = 0, & w(0, t) = 0, \quad w(l, t) = 0,
\end{cases}
\]

But from the maximum principle, we know that \( w \) assumes its maximum and minimum values on one of the three sides \( t = 0, \ x = 0, \) and \( x = l \). And since \( w = 0 \) on all of these three sides from the initial and boundary conditions in (8), we have that for \( x \in [0, l], t > 0 \)
\[
0 \leq w \leq 0 \quad \Rightarrow \quad w(x, t) \equiv 0.
\]
Hence,
\[
u - v = w \equiv 0, \quad \text{or} \quad u \equiv v,
\]
and the solution must indeed be unique.

Notice again that all of the above arguments hold for the case of the infinite interval \(-\infty < x < \infty\) as well. In this case one ignores the effect of the infinitely far endpoints and considers an IVP. And the maximum principle simply asserts that the maximum of the solutions must be attained initially. We will use this in the next lecture when deriving the solution for the IVP for the heat equation on the entire real line \( x \in \mathbb{R} \).
8.3 Stability

Stability of solutions with respect to the auxiliary conditions is the third ingredient of well-posedness, after existence and uniqueness. It asserts that close auxiliary conditions lead to close solutions. There are, however, different ways of measuring closeness of functions, which initial and boundary data, as well as the solutions are.

Consider two solutions, \( u_1, u_2 \), of the heat equation (1) for \( x \in [0, l], t > 0 \), which satisfy the following initial-boundary conditions

\[
\begin{align*}
  u_1(x, 0) &= \phi_1(x), & u_2(x, 0) &= \phi_2(x), \\
  u_1(0, t) &= g_1(t), & u_2(0, t) &= g_2(t), \\
  u_1(l, t) &= h_1(t), & u_2(l, t) &= h_2(t).
\end{align*}
\]  
(9)

Stability of solutions means that closeness of \( \phi_1 \) to \( \phi_2 \), \( g_1 \) to \( g_2 \) and \( h_1 \) to \( h_2 \) implies the closeness of \( u_1 \) to \( u_2 \). Notice that the difference \( w = u_1 - u_2 \) solves the heat equation as well, and satisfies the following initial-boundary conditions

\[
\begin{align*}
  w_1(x, 0) &= \phi_1(x) - \phi_2(x), \\
  w_1(0, t) &= g_1(t) - g_2(t), & w_1(l, t) &= h_1(t) - h_2(t).
\end{align*}
\]

But then the maximum and minimum principles imply

\[
- \max_{(x,t) \in \Gamma} \{ |w(x,t)| \} \leq \max_{0 \leq x \leq l, \ t \geq 0} \{ w(x,t) \} \leq \max_{(x,t) \in \Gamma} \{ |w(x,t)| \},
\]

and hence, the absolute value of the difference \( u_1 - u_2 \) will be bounded by

\[
\max_{0 \leq x \leq l, \ t \geq 0} \{ |u_1(x,t) - u_2(x,t)| \} = \max_{0 \leq x \leq l, \ t \geq 0} \{ |w(x,t)| \} \leq \max_{(x,t) \in \Gamma} \{ |w(x,t)| \} = \max_{0 \leq x \leq l, \ t \geq 0} \{ |\phi_1(x) - \phi_2(x)|, |g_1(t) - g_2(t)|, |h_1(t) - h_2(t)| \}.
\]

Thus, the smallness of the maximum of the differences \( |\phi_1 - \phi_2|, |g_1 - g_2| \) and \( |h_1 - h_2| \) implies the smallness of the maximum of the difference of solutions \( |u_1 - u_2| \). In this case the stability is said to be in the uniform sense, i.e. smallness is understood to hold uniformly in the \((x,t)\) variables.

An alternate way of showing the stability is provided by the energy method. Suppose \( u_1 \) and \( u_2 \) solve the heat equation with initial data \( \phi_1 \) and \( \phi_2 \) respectively, and zero boundary conditions. This would be the case for the problem over the entire real line \( x \in \mathbb{R} \), or if \( g_1 = g_2 = h_1 = h_2 = 0 \) in (9). In this case the energy method for the difference \( w = u_1 - u_2 \) implies that \( E[w](t) \leq E[w](0) \) for all \( t \geq 0 \), or

\[
\int_0^t |u_1(x,t) - u_2(x,t)|^2 \, dx \leq \int_0^t |\phi_1(x) - \phi_2(x)|^2 \, dx, \quad \text{for all } \ t \geq 0.
\]

Thus the closeness of \( \phi_1 \) to \( \phi_2 \) in the sense of the square integral of the difference implies the closeness of the respective solutions in the same sense. This is called stability in the square integral \( (L^2) \) sense.

8.4 Conclusion

As expected, the method of characteristics is inefficient for solving the heat equation. We then need to find an alternative method of reducing the equation to an ODE. But before embarking on this path, we first study the properties of the heat equation, which will serve as beacons in the later reduction to an ODE. Today we established the maximum principle for the heat equation, which immediately implied the uniqueness and stability for the solution. Next time we will look at the invariance properties of the equation and derive the solution using these properties.
9 Heat equation: solution

Equipped with the uniqueness property for the solutions of the heat equation with appropriate auxiliary conditions, we will next present a way of deriving the solution to the heat equation

\[ u_t - ku_{xx} = 0. \tag{10} \]

Considering the equation on the entire real line \( x \in \mathbb{R} \) simplifies the problem by eliminating the effects of the boundaries. We will first concentrate on this case, which corresponds to the dynamics of the temperature in a rod of infinite length. We want to solve the IVP

\[
\begin{cases}
  u_t - ku_{xx} = 0 & (-\infty < x < \infty, 0 < t < \infty), \\
  u(x,0) = \phi(x). 
\end{cases}
\tag{11}
\]

Since the solution to the above IVP is not easy to derive directly, unlike the case of the wave IVP, we will first derive a particular solution for a special simple initial data, and try to produce solutions satisfying all other initial conditions by exploiting the invariance properties of the heat equation.

9.1 Invariance properties of the heat equation

The heat equation (10) is invariant under the following transformations

(a) Spatial translations: If \( u(x,t) \) is a solution of (10), then so is the function \( u(x - y, t) \) for any fixed \( y \).

(b) Differentiation: If \( u \) is a solution of (10), then so are \( u_x, u_t, u_{xx} \) and so on.

(c) Linear combinations: If \( u_1, u_2, \ldots, u_n \) are solutions of (10), then so is \( u = c_1u_1 + c_2u_2 + \cdots + c_nu_n \) for any constants \( c_1, c_2, \ldots, c_n \).

(d) Integration: If \( S(x,t) \) is a solution of (10), then so is the integral

\[ v(x,t) = \int_{-\infty}^{\infty} S(x - y, t)g(y) \, dy \]

for any function \( g(y) \), as long as the improper integral converges (we will ignore the issue of the convergence for the time being).

(e) Dilation (scaling): If \( u(x,t) \) is a solution of (10), then so is the dilated function \( v(x,t) = u(ax, at) \) for any constant \( a > 0 \) (compare this to the scaling property of the wave equation, which is invariant under the dilation \( u(x,t) \mapsto u(ax, at) \) for all \( a \in \mathbb{R} \)).

Properties (a), (b) and (c) are trivial (check by substitution), while property (d) is the limiting case of property (c). Indeed, if we use the notation \( u^y(x,t) = S(x - y, t) \), and \( c^y = g(y)\Delta y \), then \( u^y \) is also a solution by property (a), and we have the formal limit

\[ \int_{-\infty}^{\infty} S(x - y, t)g(y) \, dy = \lim_{\Delta y \to 0} \sum_y c^y u^y. \]

To make this precise, we need to consider a finite interval of integration, which is partitioned by points \( \{y_i\}_{i=1}^n \) into subintervals of length \( \Delta y \), and use the definition of the integral as the limit of the corresponding Riemann sum to write

\[ \int_{-\infty}^{\infty} S(x - y, t)g(y) \, dy = \lim_{b \to \infty} \lim_{n \to \infty} \sum_{i=1}^{n} S(x - y_i)g(y_i)\Delta y, \]

where \( -b = y_1 < y_2 < \ldots y_n = b \) is a partition of the interval \([-b, b]\).

Finally, property (e) can be checked by direct substitution as well. Notice that we cannot formally reverse the time by dilating with the factor \( a = -1 \), as was the case for the wave equation, since the \( \sqrt{\alpha} \) factor in front of the \( x \) argument would make the dilated function complex, which is not allowed in the theory of real PDEs (what is the meaning of complex valued temperature?!). We will later see that the heat equation is indeed time irreversible.
9.2 Solving a particular IVP

As a special initial data we take the following function

\[ H(x) = \begin{cases} 
1, & x > 0, \\
0, & x < 0,
\end{cases} \]

which is called the Heaviside step function. We consider the IVP

\[
\begin{aligned}
Q_t - kQ_{xx} &= 0 & (\infty < x < \infty, 0 < t < \infty), \\
Q(x,0) &= H(x),
\end{aligned}
\]  

which we solve in successive steps.

**Step 1: Reduction to an ODE.** Notice that the Heaviside function (12) is invariant under the dilation \( x \mapsto \sqrt{a}x \), i.e. \( H(\sqrt{a}x) = H(x) \). From the dilation property of the heat equation, we know that \( Q(\sqrt{a}x, at) \) also solves the heat equation. But \( Q(\sqrt{a}x, 0) = H(\sqrt{a}x) = H(x) \), thus \( Q(\sqrt{a}x, at) \) and \( Q(x, t) \) both solve the IVP (13). The uniqueness of solutions then implies that \( Q(\sqrt{a}x, at) = Q(x, t) \) for all \( x \in \mathbb{R}, t > 0 \), so \( Q \) is invariant under the dilation \((x, t) \mapsto (\sqrt{a}x, at)\) as well.

Due to this invariance, \( Q \) can depend only on the ratio \( \frac{x}{\sqrt{t}} \), that is \( Q(x,t) = q(\frac{x}{\sqrt{t}}) \). To see this, define the function \( q \) in the following way

\[ q(z) = Q(1, 1) = Q(z, 1) = \frac{x}{\sqrt{t}}. \]

Thus \( Q \) is completely determined by the function of one variable \( q \).

For convenience of future calculations we pass to the function \( g(z) = q(\sqrt{4kt}) \), so that

\[ Q(x,t) = \frac{x}{\sqrt{t}} = g\left(\frac{x}{\sqrt{4kt}}\right) = g(p), \]

where we used the notation \( p = x/\sqrt{4kt} \). We next compute the derivatives of \( Q \) in terms of \( g \), and substitute them into the heat equation in order to obtain an ODE for \( g \). Using the chain rule, one gets

\[
\begin{aligned}
Q_t &= \frac{dg}{dp} \frac{dp}{dt} = \frac{4k}{2} \frac{x}{(\sqrt{4kt})^3} g'(p) = \frac{1}{2t} \frac{x}{\sqrt{4kt}} g'(p), \\
Q_x &= \frac{dg}{dp} \frac{dp}{dx} = \frac{1}{\sqrt{4kt}} g'(p), \\
Q_{xx} &= \frac{d^2g}{dp^2} \frac{dp}{dt} = \frac{1}{4kt} g''(p).
\end{aligned}
\]

The heat equation then implies

\[ 0 = Q_t - kQ_{xx} = \frac{1}{4t} [-2pg'(p) - g''(p)], \]

which gives the following equation for \( g \)

\[ g'' + 2pg' = 0. \]  

**Step 2: Solving the ODE.** Using the integrating factor \( \exp(\int 2p \, dp) = e^{p^2} \), the ODE (14) reduces to

\[ [e^{p^2} g'(p)]' = 0. \]
Thus, we have
\[ e^{p^2} g'(p) = c_1. \]

Solving for \( g'(p) \), and integrating, we obtain
\[ g(p) = c_1 \int e^{-p^2} \, dp + c_2. \]

**Step 3: Checking the initial condition.** Recalling that \( Q(x, t) = g(p) \), where \( p = x/\sqrt{4kt} \), we obtain the following explicit formula for \( Q \)
\[ Q(x, t) = c_1 \int_{\infty}^{x/\sqrt{4kt}} e^{-p^2} \, dp + c_2. \]  
(15)

Notice that we chose a particular antiderivative, which we are free to do due to the presence of the arbitrary constants. Also note that the above formula is only valid for \( t > 0 \), so to check the initial condition, we need to take the limit \( t \to 0^+ \). Recalling the initial condition from (13), we have that,

if \( x > 0 \), \[ 1 = \lim_{t \to 0^+} Q(x, t) = c_1 \int_{-\infty}^{\infty} e^{-p^2} \, dp + c_2 = c_1 \frac{\sqrt{\pi}}{2} + c_2, \]

if \( x < 0 \), \[ 0 = \lim_{t \to 0^+} Q(x, t) = c_1 \int_{-\infty}^{\infty} e^{-p^2} \, dp + c_2 = -c_1 \frac{\sqrt{\pi}}{2} + c_2, \]

where we used the fact that \( \int_{0}^{\infty} e^{-p^2} \, dp = \sqrt{\pi}/2 \) to compute the improper integrals. The above identities give
\[ c_1 \frac{\sqrt{\pi}}{2} + c_2 = 1, \quad -c_1 \frac{\sqrt{\pi}}{2} + c_2 = 0. \]

Solving for \( c_1 \) and \( c_2 \), we get \( c_1 = 1/\sqrt{\pi} \) and \( c_2 = 1/2 \). Substituting these into (15) gives the unique solution of the IVP (13),
\[ Q(x, t) = \frac{1}{2} + \frac{1}{\sqrt{\pi}} \int_{0}^{x/\sqrt{4kt}} e^{-p^2} \, dp, \quad \text{for} \quad t > 0. \]  
(16)

## 9.3 Solving the general IVP

Returning to the general IVP (11), we would like to derive a solution formula, which will express the solution to the IVP in terms of the initial data (similar to d’Alambert’s solution for the wave equation).

We first define the function
\[ S(x, t) = \frac{\partial Q}{\partial x}(x, t), \]  
(17)

where \( Q(x, t) \) is the solution to the particular IVP (13), and is given by (16). Then, by the invariance properties of the heat equation, \( S(x, t) \) also solves the heat equation, and so does

\[ u(x, t) = \int_{-\infty}^{\infty} S(x - y, t) \phi(y) \, dy, \quad \text{for} \quad t > 0. \]  
(18)

We claim that this \( u \) is the unique solution of the IVP (11). To verify this claim one only needs to check the initial condition of (11). Notice that using \( S = Q_x \), we can rewrite \( u \) as follows
\[ u(x, t) = \int_{-\infty}^{\infty} \frac{\partial Q}{\partial x}(x - y, t) \phi(y) \, dy = -\int_{-\infty}^{\infty} \frac{\partial}{\partial y} [Q(x - y, t)] \phi(y) \, dy \]

Integrating by parts in the last integral, we get
\[ u(x, t) = -Q(x - y, t) \phi(y) \bigg|_{y = -\infty}^{y = \infty} + \int_{-\infty}^{\infty} Q(x - y, t) \phi'(y) \, dy. \]
We assume that the boundary terms vanish, which can be guaranteed for example by assuming that $\phi(y)$ vanishes for large $|y|$ (this is not strictly necessary, since $S(x-y,t)$ decays rapidly as $|y-x|$ becomes large, as we will shortly see). Now plugging in $t = 0$, and using that $Q$ has the Heaviside function (12) as its initial data, we have

$$u(x,0) = \int_{-\infty}^{\infty} Q(x-y,0)\phi'(y) \, dy = \int_{-\infty}^{\infty} \phi'(y) \, dy = \phi(y)|_{y=x} = \phi(x).$$

So $u(x,t)$ indeed satisfies the initial condition of (11).

We can compute $S(x,t)$ from (17), which will give

$$S(x,t) = \frac{1}{\sqrt{4\pi kt}} e^{-x^2/4kt},$$

Using this expression of $S(x,t)$, we can now rewrite the solution given by (18) in the following explicit form

$$u(x,t) = \frac{1}{\sqrt{4\pi kt}} \int_{-\infty}^{\infty} e^{-(x-y)^2/4kt} \phi(y) \, dy, \quad \text{for} \quad t > 0. \quad \text{(19)}$$

The function $S(x,t)$ is known as the heat kernel, fundamental solution, source function, Green’s function, or propagator of the heat equation. Notice that it gives a way of propagating the initial data $\phi$ to later times, giving the solution at any time $t > 0$.

It is clear that formula (19) does not make sense for $t = 0$, although one can compute the limit of $u(t,x)$ as $t \to 0+$ in that formula, which will give an alternate way of checking the initial condition of (11).

### 9.4 Conclusion

We derived the solution to the heat equation by first looking at a particular initial data, which was invariant under dilation. This guaranteed that the solution corresponding to this initial data is also dilation invariant, which reduced the heat equation to an ODE. After solving this ODE, and obtaining the solution, we saw that the solution to the general heat IVP can be written in an integral form using this particular solution. Next time we will explore the solution given by formula (19), and will study its qualitative behavior.
10 Heat equation: interpretation of the solution

Last time we considered the IVP for the heat equation on the whole line

\[
\begin{align*}
    \left\{ \begin{array}{l}
    u_t - ku_{xx} = 0 \quad (-\infty < x < \infty, 0 < t < \infty), \\
    u(x, 0) = \phi(x),
    \end{array} \right.
\end{align*}
\]

and derived the solution formula

\[
u(x, t) = \int_{-\infty}^{\infty} S(x - y, t)\phi(y) \, dy, \quad \text{for} \quad t > 0,
\]

where \( S(x, t) \) is the heat kernel,

\[
S(x, t) = \frac{1}{\sqrt{4\pi kt}} e^{-x^2/4kt}.
\]

Substituting this expression into (21), we can rewrite the solution as

\[
u(x, t) = \frac{1}{\sqrt{4\pi kt}} \int_{-\infty}^{\infty} e^{-(x-y)^2/4kt} \phi(y) \, dy, \quad \text{for} \quad t > 0.
\]

Recall that to derive the solution formula we first considered the heat IVP with the following particular initial data

\[
Q(x, 0) = H(x) = \begin{cases} 
1, & x > 0, \\
0, & x < 0.
\end{cases}
\]

Then using dilation invariance of the Heaviside step function \( H(x) \), and the uniqueness of solutions to the heat IVP on the whole line, we deduced that \( Q \) depends only on the ratio \( x/\sqrt{t} \), which lead to a reduction of the heat equation to an ODE. Solving the ODE and checking the initial condition (24), we arrived at the following explicit solution

\[
Q(x, t) = \frac{1}{2} + \frac{1}{\sqrt{\pi}} \int_{0}^{x/\sqrt{4kt}} e^{-p^2} \, dp, \quad \text{for} \quad t > 0.
\]

The heat kernel \( S(x, t) \) was then defined as the spatial derivative of this particular solution \( Q(x, t) \), i.e.

\[
S(x, t) = \frac{\partial Q}{\partial x}(x, t),
\]

and hence it also solves the heat equation by the differentiation property.

The key to understanding the solution formula (21) is to understand the behavior of the heat kernel \( S(x, t) \). To this end some technical machinery is needed, which we develop next.

10.1 Dirac delta function

Notice that, due to the discontinuity in the initial data of \( Q \), the derivative \( Q_x(x, t) \), which we used in the definition of the function \( S \) in (26), is not defined in the traditional sense when \( t = 0 \). So how can one make sense of this derivative, and what is the initial data for \( S(x, t) \)?

It is not difficult to see that the problem is at the point \( x = 0 \). Indeed, using that \( Q(x, 0) = H(x) \) is constant for any \( x \neq 0 \), we will have \( S(x, 0) = 0 \) for all \( x \) different from zero. However, \( H(x) \) has a jump discontinuity at \( x = 0 \), as is seen in Figure 1, and one can imagine that at this point the rate of growth of \( H \) is infinite. Then the “derivative”

\[
\delta(x) = H'(x)
\]

is zero everywhere, except at \( x = 0 \), where it has a spike of zero width and infinite height. Refer to Figure 2 below for an intuitive sketch of the graph of \( \delta \). Of course, \( \delta \) is not a function in the traditional sense, but is
rather a *generalized function*, or *distribution*. Unlike regular functions, which are characterized by their finite values at every point in their domains, distributions are characterized by how they act on regular functions.

To make this rigorous, we define the set of *test functions* \( \mathcal{D} = C^\infty_c \), the elements of which are smooth functions with compact support. So \( \phi \in \mathcal{D} \), if and only if \( \phi \) has continuous derivatives of any order \( k \in \mathbb{N} \), and the closure of the support of \( \phi \),

\[
\text{supp}(\phi) = \{ x \in \mathbb{R} | \phi(x) \neq 0 \},
\]

is compact. Recall that compact sets in \( \mathbb{R} \) are those that are closed and bounded. In particular for any test function \( \phi \) there is a rectangle \( [-R, R] \), outside of which \( \phi \) vanishes. Notice that derivatives of test functions are also test functions, as are sums, scalar multiples and products of test functions.

Distributions are continuous linear *functionals* on \( \mathcal{D} \), that is, they are continuous linear maps from \( \mathcal{D} \) to the real numbers \( \mathbb{R} \). Notice that for any regular function \( f \), we can define the functional

\[
f[\phi] = \int_{-\infty}^{\infty} f(x) \phi(x) \, dx,
\]

which makes \( f \) into a distribution, since to every \( \phi \in \mathcal{D} \) it assigns the number \( \int_{-\infty}^{\infty} f(x) \phi(x) \, dx \). This integral will converge under very weak conditions on \( f \) (\( f \in L^1_{\text{loc}} \)), due to the compact support of \( \phi \). In particular, \( f \) can certainly have jump discontinuities. Note that we committed an abuse of notation to identify the distribution associated with \( f \) by the same letter \( f \). The particular notion in which we use the function will be clear from the context.

One can also define the *distributional derivative* of \( f \) to be the distribution, which acts on the test functions as follows

\[
f'[\phi] = -\int_{-\infty}^{\infty} f(x) \phi'(x) \, dx.
\]

Notice that integration by parts and the compact support of test functions makes this definition consistent with the regular derivative for differentiable functions (check that the distribution formed as in (28) by the derivative of \( f \) coincides with the distributional derivative of \( f \)).

We can also apply the notion of the distributional derivative to the Heaviside step function \( H(x) \), and think of the definition (27) in the sense of distributional derivatives. Let us now compute how \( \delta \), called the *Dirac delta function*, acts on test functions. By the definition of the distributional derivative,

\[
\delta[\phi] = -\int_{-\infty}^{\infty} H(x) \phi'(x) \, dx.
\]

Recalling the definition of \( H(x) \) in (24), we have that

\[
\delta[\phi] = -\int_{0}^{\infty} \phi'(x) \, dx = -\phi(x)\big|_{0}^{\infty} = \phi(0).\tag{29}
\]

Thus, the Dirac delta function maps test functions to their values at \( x = 0 \). We can make a translation in the \( x \) variable, and define \( \delta(x - y) = H'(x - y) \), i.e. \( \delta(x - y) \) is the distributional derivative of the distribution
formed by the function $H(x - y)$. Then it is not difficult to see that $\delta(x - y)[\phi] = \phi(y)$. That is, $\delta(x - y)$ maps test functions to their values at $y$. We will make the abuse of notation mentioned above, and write this as

$$\int_{-\infty}^{\infty} \delta(x - y)\phi(x) \, dx = \phi(y).$$

We also note that $\delta(x - y) = \delta(y - x)$, since $\delta$ is even, if we think of it as a regular function with a centered spike (one can prove this from the definition of $\delta$ as a distribution).

Using these new notions, we can make sense of the initial data for $S(x, y)$. Indeed,

$$S(x, 0) = \delta(x). \tag{30}$$

Since the initial data is a distribution, one then thinks of the equation to be in the sense of distributions as well, that is, treat the spatial derivatives appearing in the equation as distributional derivatives. This requires the generalization of the idea of a distribution to smooth (in the time variable) 1-parameter family of distributions. We call this type of solutions weak solutions (recall the solutions of the wave equation with discontinuous data).

Thus $S(x, t)$ is a week solution of the heat equation, if

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(x, t)[\phi_t(x, t) - k\phi_{xx}(x, t)] \, dx \, dt = 0,$$

for any test function $\phi$ of two variables. This means that the distribution $(\partial_t - k\partial^2_x)S$, with the derivatives taken in the distributional sense, is the zero distribution. Notice that the weak solution $S(x, t)$ arising from the initial data (30) has the form (22), which is an infinitely differentiable function of $x$ and $t$. This is in stark contrast to the case of the wave equation, where, as we have seen in the examples, the discontinuity of the initial data is preserved in time.

Having the $\delta$ function in our arsenal of tools, we can now give an alternate proof that (21) satisfies the initial conditions of (20). Directly plugging in $t = 0$ into (21), which we are now allowed to do by treating it as a distribution, and using (30), we get

$$u(x, 0) = \int_{-\infty}^{\infty} \delta(x - y)\phi(y) \, dy = \phi(x).$$

10.2 Interpretation of the solution

Let us look at the solution (23) in detail, and try to understand how the heat kernel $S(x, t)$ propagates the initial data $\phi(x)$. Notice that $S(x, t)$, given by (22), is a well-defined function of $(x, t)$ for any $t > 0$. Moreover, $S(x, t)$ is positive, is even in the $x$ variable, and for a fixed $t$ has a bell-shaped graph. In general, the function

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}$$

is called Gaussian function or Gaussian. In the probability theory, it gives the density of the normal distribution with mean $\mu$ and standard deviation $\sigma$. The graph of the Gaussian is a bell-curve with its peak of height $1/\sqrt{2\pi\sigma^2}$ at $x = \mu$ and the width of the bell at mid-height roughly equal to $2\sigma$. Thus, for some fixed time $t$ the height of $S(x, t)$ at its peak $x = 0$ is $1/\sqrt{4\pi kt}$, which decays as $t$ grows.

Notice that as $t \to 0+$, the height of the peak becomes arbitrarily large, and the width of the bell-curve, $\sqrt{2kt}$ goes to zero. This, of course, is expected, since $S(x, t)$ has the initial data (30). One can think of $S(x, t)$ as the temperature distribution at time $t$ that arises from the initial distribution given by the Dirac delta function. With passing time the highest temperature at $x = 0$ gets gradually transferred to the other points of the rod. It also makes sense, that points closer to $x = 0$ will have higher temperature than those farther away. Graphs of $S(x, t)$ for three different times are sketched in Figure 3 below.

From the initial condition (30), we see that initially the temperature at every point $x \neq 0$ is zero, but $S(x, t) > 0$ for any $x$ and $t > 0$. This means that heat is instantaneously transferred to all points of the rod (closer
points get more heat), so the speed of heat conduction is infinite. Compare this to the finite speed of propagation for the wave equation. One can also compute the area below the graph of $S(x, t)$ at any time $t > 0$ to get

$$\int_{-\infty}^{\infty} S(x, t) \, dx = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-p^2} \, dp = 1,$$

where we used the change of variables $p = x/\sqrt{4kt}$. At $t = 0$, we have

$$\int_{-\infty}^{\infty} S(x, 0) \, dx = \int_{-\infty}^{\infty} \delta(x) \, dx = 1,$$

where we think of the last integral as the $\delta$ distribution applied to the constant function 1 (more precisely, a test function that is equal to 1 in some open interval around $x = 0$). This shows that the area below the graph of $S(x, t)$ is preserved in time and is equal to 1, so for any fixed time $t \geq 0$, $S(x, t)$ can be thought of as a probability density function. At time $t = 0$ its the probability density that assigns probability 1 to the point $x = 0$, as was seen in (29), and for times $t > 0$ it is a normal distribution with mean $x = 0$ and standard deviation $\sigma = \sqrt{2kt}$ that grows with time. As we mentioned earlier, $S(x, t)$ is smooth, in spite of having a discontinuous initial data. We will see in the next lecture that this is true for any solution of the heat IVP (20) with general initial data.

We now look at the solution (23) with general data $\phi(x)$. First, notice that the integrand in (21),

$$S(x - y, t)\phi(y),$$

measures the effect of $\phi(y)$ (the initial temperature at the point $y$) felt at the point $x$ at some later time $t$. The source function $S(x - y, t)$, which has its peak precisely at $y$, weights the contribution of $\phi(y)$ according to the distance of $y$ from $x$ and the elapsed time $t$.

Since the value of $u(x, t)$ (temperature at the point $x$ at time $t$) is the total sum of contributions from the initial temperature at all points $y$, we have the formal sum

$$u(x, t) \approx \sum_y S(x - y, t)\phi(y),$$

which in the limit gives formula (21). So, the heat kernel $S(x, t)$ gives a way of propagating the initial data $\phi$ to later times. Of course the contribution from a point $y_1$ closer to $x$ has a bigger weight $S(x - y_1, t)$, than the contribution from a point $y_2$ farther away, which gets weighted by $S(x - y_2, t)$.

The function $S(x, t)$ appears in various other physical situations. For example in the random (Brownian) motion of a particle in one dimension. If the probability of finding the particle at position $x$ initially is given by the density function $\phi(x)$, then the density defining the probability of finding the particle at position $x$ at time $t$ is given by the same formula (21).

**Example 10.1.** Solve the heat equation with the initial condition $u(x, 0) = e^x$.

Using the solution formula (23), we have

$$u(x, t) = \frac{1}{\sqrt{4\pi kt}} \int_{-\infty}^{\infty} e^{-(x-y)^2/4kt} e^y \, dy = \frac{1}{\sqrt{4\pi kt}} \int_{-\infty}^{\infty} e^{-x^2 + 2xy - y^2 + 4kt y}/4kt \, dy$$
We can complete the squares in the numerator of the exponent, writing it as

\[
\frac{-x^2 + 2xy - y^2 + 4kt}{4kt} = \frac{-x^2 + 2(x + 2kt)y - y^2}{4kt} = \frac{-(y - 2kt - x)^2 + 4ktx + 4k^2t^2}{4kt} = -\left(\frac{y - 2kt - x}{\sqrt{4kt}}\right)^2 + x + kt.
\]

We then have

\[
u(x,t) = \frac{1}{\sqrt{4\pi kt}} \int_{-\infty}^{\infty} e^{x+kt} e^{-\left[(y-2kt-x)/\sqrt{4kt}\right]^2} dy = \frac{1}{\sqrt{\pi}} e^{x+kt} \int_{-\infty}^{\infty} e^{-p^2} dp = e^{x+kt}.
\]

Notice that \(u(x, t)\) grows with time, which may seem to be in contradiction with the maximum principle. However, thinking in terms of heat conduction, we see that the initial temperature \(u(x, 0) = e^x\) is itself infinitely large at the far right end of the rod \(x = +\infty\). So the temperature does not grow out of nowhere, but rather gets transferred from right to left with the “speed” \(k\). Thus the initial exponential distribution of the temperature “travels” from right to left with the speed \(k\) as \(t\) grows. Compare this to the example in Strauss, where the initial temperature \(u(x, 0) = e^{-x}\) “travels” from left to right, since the initial temperature peaks at the far left end \(x = -\infty\).

In the above example we were able to compute the solution explicitly, however, the integral in (23) may be impossible to evaluate completely in terms of elementary functions for general initial data \(\phi(x)\). Due to this, the answers for particular problems are usually written in terms of the error function in statistics,

\[
\mathcal{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-p^2} dp.
\]

Notice that \(\mathcal{Erf}(0) = 0\), and \(\lim_{x \to \infty} \mathcal{Erf}(x) = 1\). Using this function, we can rewrite the function \(Q(x, t)\) given by (25), which solves the heat IVP with Heaviside initial data, as follows

\[
Q(x, t) = \frac{1}{2} + \frac{1}{2} \mathcal{Erf}\left(\frac{x}{\sqrt{4kt}}\right).
\]

### 10.3 Conclusion

Using the notions of distribution and distributional derivative, we can make sense of the heat kernel \(S(x, t)\) that has the Dirac \(\delta\) function as its initial data. Comparing the expression of the heat kernel (22) with the density function of the normal (Gaussian) distribution, we saw that the solution formula (21) essentially weights the initial data by the bell-shaped curve \(S(x, t)\), thus giving the contribution from the initial heat at different points towards the temperature at point \(x\) at time \(t\).