## 10 Heat equation: interpretation of the solution

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

$$\begin{cases} u_t - ku_{xx} = 0 & (-\infty < x < \infty, 0 < t < \infty), \\ u(x,0) = \phi(x), \end{cases} \tag{1}$$

and derived the solution formula

$$u(x,t) = \int_{-\infty}^{\infty} S(x-y,t)\phi(y)\,dy, \qquad \text{for} \quad t > 0, \tag{2}$$

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

$$S(x,t) = \frac{1}{\sqrt{4\pi kt}} e^{-x^2/4kt}. \tag{3}$$

Substituting this expression into (2), we can rewrite the solution as

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

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

$$Q(x,0) = H(x) = \begin{cases} 1, & x > 0, \\ 0, & x < 0. \end{cases} \tag{5}$$

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 (5), we arrived at the following explicit solution

$$Q(x,t) = \frac{1}{2} + \frac{1}{\sqrt{\pi}} \int_0^{x/\sqrt{4kt}} e^{-p^2}\,dp, \qquad \text{for} \quad t > 0. \tag{6}$$

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

$$S(x,t) = \frac{\partial Q}{\partial x}(x,t), \tag{7}$$

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

The key to understanding the solution formula (2) 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 (7), 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"

$$\delta(x) = H'(x) \tag{8}$$

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,

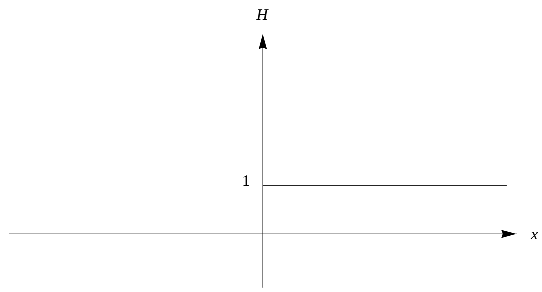


Figure 1: The graph of the Heaviside step function.

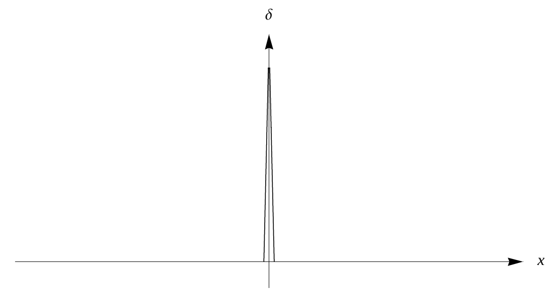


Figure 2: The sketch of the Dirac $\delta$ function.

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* $\mathcal{D} = C_c^\infty$, 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 $\phi$,

$$\operatorname{supp}(\phi) = \{x \in \mathbb{R} \mid \phi(x) \neq 0\},$$

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

$$f[\phi] = \int_{-\infty}^{\infty} f(x)\phi(x)\, dx, \tag{9}$$

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 $f$ ($f \in L^1_{loc}$), 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 (9) 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 (8) 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 (5), we have that

$$\delta[\phi] = -\int_0^{\infty} \phi'(x)\, dx = -\phi(x)\Big|_0^{\infty} = \phi(0). \tag{10}$$

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,

$$S(x, 0) = \delta(x). \tag{11}$$

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 distirbutions. 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

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

for any test function $\phi$ of two variables. This means that the distribution $(\partial_t - k\partial_x^2)S$, 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 (11) has the form (3), 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 (2) satisfies the initial conditions of (1). Directly plugging in $t = 0$ into (2), which we are now allowed to do by treating it as a distribution, and using (11), 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 (4) 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 (3), 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 $\frac{1}{\sqrt{4\pi kt}}$, 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 (11). 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.

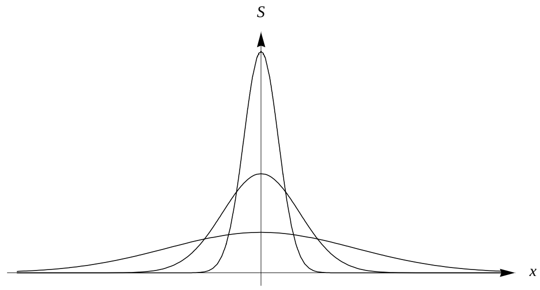


Figure 3: The graphs of the heat kernel at different times.

From the initial condition (11), 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 (10), 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 (1) with general initial data.

We now look at the solution (4) with general data $\phi(x)$. First, notice that the integrand in (2),

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

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 (2). 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 (2).

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

Using the solution formula (4), we have

$$u(x,t) = \frac{1}{\sqrt{4\pi kt}} \int_{-\infty}^{\infty} e^{-(x-y)^2/4kt} e^y \, dy = \frac{1}{\sqrt{4\pi kt}} \int_{-\infty}^{\infty} e^{[-x^2+2xy-y^2+4kty]/4kt} \, dy$$

We can complete the squares in the numerator of the exponent, writing it as

$$\begin{aligned} \frac{-x^2+2xy-y^2+4kty}{4kt} &= \frac{-x^2+2(x+2kt)y-y^2}{4kt} \\ &= \frac{-(y-2kt-x)^2+4ktx+4k^2t^2}{4kt} = -\left(\frac{y-2kt-x}{\sqrt{4kt}}\right)^2 + x + kt. \end{aligned}$$

We then have

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

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 (4) 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,

$$\mathcal{E}\text{rf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-p^2} \, dp.$$

Notice that $\mathcal{E}\text{rf}(0) = 0$, and $\lim_{x\to\infty} \mathcal{E}\text{rf}(x) = 1$. Using this function, we can rewrite the function $Q(x,t)$ given by (6), which solves the heat IVP with Heaviside initial data, as follows

$$Q(x,t) = \frac{1}{2} + \frac{1}{2}\mathcal{E}\text{rf}\left(\frac{x}{\sqrt{4kt}}\right).$$

### 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 (3) with the density function of the normal (Gaussian) distribution, we saw that the solution formula (2) 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 $t$.