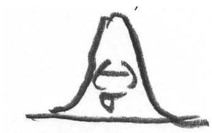
**Proof of the Fourier reconstruction formula (48)** The central idea is to approximate $f \in L^1(\mathbb{R}^n)$ by a convolution with a nice function which sharply peaks at 0 and has integral 1, and choose the nice function to be one whose Fourier transform is explicitly known. (The first part of the idea, approximation by convolution with a sharply peaked function, is familiar to us from the Proof of Lemma 1.2 3).) We choose the nice function to be a normalized Gaussian with standard deviation $\sigma > 0$,

$$G_\sigma(x) := \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-|x|^2/2\sigma^2} \tag{54}$$




(this is convenient because of the simple behaviour of Gaussians under Fourier transformation, cf. Example 2).

**Step 1:** Proof for Gaussians, $f = G_\sigma$. We calculate

$$\begin{aligned} G_\sigma(x) &= (2\pi)^{-n}\widehat{\widehat{G_\sigma}}(x) \text{ (by Example 2)} \\ &= (2\pi)^{-n}\int_{\mathbb{R}^n} e^{-ik\cdot x}\widehat{G_\sigma}(k)\,dk \text{ (by the def. of the F.T.)} \\ &= (2\pi)^{-n}\int_{\mathbb{R}^n} e^{ik'\cdot x}\widehat{G_\sigma}(k')\,dk' \text{ (by the substitution } k = -k'). \end{aligned}$$

In the last line we have used that $\widehat{G_\sigma}(k) = \widehat{G_\sigma}(-k)$, as is evident from the explicit formula in (53).

**Step 2:** Proof for convolutions of general functions with Gaussians, $f = G_\epsilon * u$ with $u \in L^1$, $\hat{u} \in L^1$, $\epsilon > 0$. By Step 1 we have for any $x$, $y \in \mathbb{R}^n$

$$G_\epsilon(x-y) = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{ik\cdot(x-y)}\widehat{G_\epsilon}(k)\,dk.$$

Multiplying by $u(y)$ and integrating over $y$ gives

$$\begin{aligned}(G_\epsilon * u)(x) &= (2\pi)^{-n} \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^n} e^{ik\cdot x} e^{-ik\cdot y} \widehat{G_\epsilon}(k)\, u(y)\, dk \right) dy \\ &= (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ik\cdot x} \underbrace{\int_{\mathbb{R}^n} e^{-ik\cdot y} u(y)\, dy}_{=\hat{u}(k)}\, \widehat{G_\epsilon}(k)\, dk \text{ (by Fubini)} \\ &= (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ik\cdot x} \widehat{G_\epsilon * u}(k)\, dk \text{ (by Theorem 2.1 3)).}\end{aligned} \tag{55}$$

**Step 3:** Proof for general functions $u$ with $u \in L^1$, $\hat{u} \in L^1$. We now let $\epsilon \to 0$ in (55). First we claim that the right hand side converges to

$$(2\pi)^{-n} \int_{\mathbb{R}^d} e^{ik\cdot x} \hat{u}(k)\, dk. \tag{56}$$

This follows because the integrand converges pointwise to that of (56) (note that $\widehat{G_\epsilon}(k) = e^{-\epsilon^2|k|^2/2} \to 1$ as $\epsilon \to 0$) and the convergence is dominated (since the absolute value of the integrand is bounded from above by $|\hat{u}| \in L^1(\mathbb{R}^n)$).

It remains to analyze the left hand side of (55). Intuitively, $G_\epsilon(x-y)$ peaks at $y = x$ and has integral 1, so – just as in our heuristic analysis of partial Fourier sums (4) in Section 1 – we expect

$$\int_{\mathbb{R}^n} G_\epsilon(x-y) u(y)\, dy \approx \int_{\mathbb{R}^n} G_\epsilon(x-y)\, f(x)\, dy = u(x).$$

Rigorously: we claim that
(a) $G_\epsilon * u \to u$ in $L^1(\mathbb{R}^n)$ as $\epsilon \to 0$
(b) $G_{\epsilon_j} * u \to u$ a.e. for some subsequence $\epsilon_j \to 0$.
Assertion (a) follows from a general measure-theoretic result concerning mollification of $L^1$ functions, and would continue

to hold if $G_\epsilon$ were replaced by any function of form $\phi_\epsilon(x) = \epsilon^{-n}\phi(\epsilon^{-1}x)$ with $\phi : \mathbb{R}^n \to \mathbb{R}$ nonnegative, integrable, and $\int_{\mathbb{R}^n} \phi = 1$. See Lemma 4.1 a) in Appendix B. Assertion (b) follows from another general measure-theoretic result, namely that any sequence of functions which converges in $L^1(\mathbb{R}^d)$ contains a subsequence which converges almost everywhere (see Lemma 4.3 in Appendix B). This completes the proof of the Fourier representation formula (48).

## 2 Solving the heat equation

Fourier's original motivation for initiating the branch of mathematics that is now called 'Fourier analysis' was to solve his newly introduced partial differential equation model for propagation of heat.

Armed with the calculus developed in Theorem 2.1, we are now in a position to understand how Fourier achieved this task.[24]

A basic version of Fourier's model is the following:

$$\begin{array}{ll} \text{(H)} & \frac{\partial u}{\partial t}(x,t) = \kappa \Delta u(x,t) \quad \text{for all } x \in \mathbb{R}^n,\ t > 0 \\ \text{(IC)} & u(x,0) = u_0(x), \quad \text{for all } x \in \mathbb{R}^n \\ \text{(BC)} & u(x,t) \to g \quad \text{as } |x| \to \infty,\ \text{for all } t > 0. \end{array}$$

Here $u = u(x,t)$ is a function from $\mathbb{R}^n \times [0,\infty)$ to $\mathbb{R}$, and $\Delta = \sum_{j=1}^n \frac{\partial^2}{\partial x_j^2}$. Physically, $u(x,t)$ is the temperature at the point $x \in \mathbb{R}^n$ at time $t$, the spatial region $\mathbb{R}^n$ idealizes a large bounded region occupied by a conducting material, and the coefficient $\kappa$ reflects the conductivity of this material (it is large for conductors and small for insulators).

[24]The overall strategy is not limited to any particular partial differential equation; applications to other problems are described in Sections 6.4 and 6.5.

Equation (H) is called the *heat equation*, and is one of the fundamental linear partial differential equations of mathematics. It also arises as a basic model in various other contexts, such as diffusion of reactants in chemical physics, Brownian motion in stochastic analysis, or option pricing in financial mathematics.

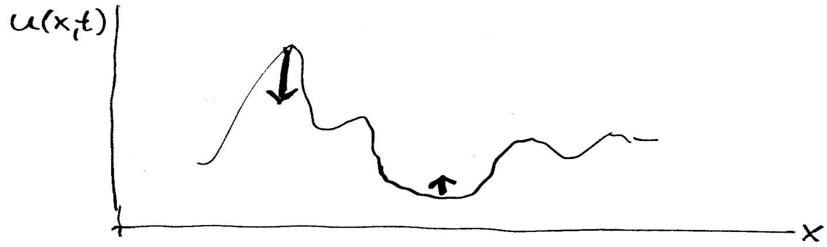
Fourier's goal was to determine the (unknown) function $u$ from (assumed to be known) *initial data* $u_0$ and *boundary data* $g$.

**Example of a counterintuitive prediction Fourier made about heat** At approximately 2 to 3 metres depth below the earth's surface, it is colder in summer than in winter. As F. remarks, this suggests a good depth for the construction of cellars.[25] See Problem 4.

To understand the mathematical meaning of the heat equation, it is useful to look at the case of one space dimension, $n = 1$, and to consider a point $x$ which is a local spatial maximum of $u$. By the necessary condition for maxima, the second derivative $\frac{\partial^2}{\partial x^2}u(x,t)$ is negative (assuming the maximum is nondegenerate, i.e. the second derivative does not vanish). Eq. (H) says that at such a point, $u$ *decreases* with time. (This is in line with our intuition about temperature.) Analogously, if $x$ is a nondegenerate local spatial minimum of $u$, then $\frac{\partial^2}{\partial x^2}u(x,t) > 0$, and so eq. (H) makes $u$ *increase* with time. Moreover, the size of the second derivative (which reflects how sharply $u$ peaks at $x$) is by (H) proportional to the speed of increase/decrease. Thus (H) says that sharp peaks are levelled out quicker than mild peaks.[26]

[25] He was French – was he thinking of wine cellars (also called *une cave*)?

[26] You can test this experimentally at home: take two ice cubes of equal size, chop one of them up into small pieces, and watch them melt in a glass of water!




To solve (H), (IC), (BC), we assume without loss of generality $g = 0$ (otherwise consider the function $u - g$), and we assume – as needs to be jusified later – that there exists a smooth solution $u$ which decays sufficiently rapidly to 0 as $|x| \to \infty$ so that $u(\cdot, t) \in L^1(\mathbb{R}^n)$ for all $t$. We introduce the partial Fourier transform with respect to $x$,

$$(\mathcal{F}_x u)(k,t) := \int_{\mathbb{R}^n} e^{-ix\cdot k} u(x,t)\, dx. \tag{57}$$

Taking the partial Fourier transform with respect to $x$ of (H) and using the rule for the Fourier transform of a derivative $(\widehat{\partial f/\partial x_j})(k) = ik_j \widehat{f}(k)$, Theorem 2.1 7)) gives

$$\frac{\partial}{\partial t}(\mathcal{F}_x u)(k,t) = \kappa \underbrace{\sum_{j=1}^{n} (ik_j)^2}_{=-|k|^2} (\mathcal{F}_x u)(k,t). \tag{58}$$

(Here we have assumed that $u$ is sufficiently well behaved so that $\mathcal{F}_x u$ is differentiable with respect to $t$ and this differentiation can be carried out underneath the integral sign in (57), and that the rule for the Fourier transform of second spatial derivatives is applicable.)

Key observation: *The PDE for $u$ has turned into an ODE for $\mathcal{F}_x u$.*

More precisely, (58) is a system of decoupled ODE's, one for each fixed $k \in \mathbb{R}^n$. Each ODE is of the form $\dot{y}(t) = \alpha y(t)$, which is easy to solve: $y(t) = y(0)e^{\alpha t}$. This together with the initial condition (IC) gives

$$(\mathcal{F}_x u)(k,t) = e^{-\kappa|k|^2 t}(\mathcal{F}_x u)(k,0) = e^{-\kappa|k|^2 t}\widehat{u_0}(k). \qquad (59)$$

To find $u$, it remains to determine the inverse Fourier transform of the right hand side. In light of the rule for the Fourier transform of a convolution, it suffices to determine the inverse Fourier transform of the multiplier $e^{-\kappa|k|^2 t}$. This was already carried out in Example 2; consequently

$$e^{-\kappa|k|^2 t} = \widehat{G_t}(k) \text{ for } G_t(x) := \frac{1}{\sqrt{4\pi\kappa t}^{\,n}} e^{-\frac{|x|^2}{4\kappa t}} \;\; (t > 0).$$

Hence by the rule for the Fourier transform of a convolution (Theorem 2.1 3))

$$(\mathcal{F}_x u)(\cdot,t) = \mathcal{F}_x(G_t * u_0)$$

and thus, by the fact that two functions whose Fourier transforms conincide must be identical (Theorem 2.1 (48))

$$u(x,t) = \frac{1}{\sqrt{4\pi\kappa t}^{\,n}} \int_{\mathbb{R}^n} e^{-\frac{|x-y|^2}{4\kappa t}} u_0(y)\, dy \;\; (x \in \mathbb{R}^n,\, t > 0) \qquad (60)$$

(Fourier, 1811).

Our derivation relied on assuming that there exists a smooth, nicely behaved solution to the heat equation. So it only shows that *if such a solution exists, then it must be given by (60).* But we can easily rid ourselves of this premise, by checking a posteriori that (60) indeed solves (H).[27] This leads to

**Theorem 2.2** *(Solution to the heat equation) Let* $u_0 \in L^1(\mathbb{R}^n)$*. The function* $u : \mathbb{R}^n \times (0,\infty) \to \mathbb{C}$ *defined by (60)*

[27] If you believe that the premise is sufficiently plausible not to require checking, you should skip the following theorem and its proof.

*is once continuously differentiable in $t$ and twice continuously differentiable in $x$,*[28] *and has the following properties:*
*a) $u$ solves (H) in $\mathbb{R}^n \times (0, \infty)$*
*b) $u$ solves (BC) with $g = 0$*
*c) $u(\cdot, t) \to u_0$ in $L^1(\mathbb{R}^n)$ $(t \to 0)$.*

**Proof** The main task is to prove the asserted differentiability properties of $u$ (which are needed for the derivatives appearing in (H) to be well defined). These follow by justifying differentiation underneath the integral sign in (60), via considering the relevant difference quotients and passing to the limit with the help of basic results about parameter-dependent integrals. For completeness we include the details. We use the elementary fact that if a sequence $g_j$ of bounded functions converges uniformly to $g$ and $u \in L^1(\mathbb{R}^n)$, then

$$|\int_{\mathbb{R}^n} g_j u - \int_{\mathbb{R}^n} g\, u| \le \sup_{y \in \mathbb{R}^n} |g_j(y) - g(y)|\, ||u||_1 \to 0. \quad (61)$$

Fix $x \in \mathbb{R}^n$, $t > 0$, and write $u(x,t)$ as $\int_{\mathbb{R}^n} G_t(x-y) u_0(y)\, dy$. As $h \to 0$, the difference quotient $\frac{1}{h}(G_{t+h}(x-y) - G_t(x-y))$ converges uniformly with respect to $y$ to $\frac{\partial}{\partial t} G_t(x-y)$, so by (61)

$$\begin{aligned} \frac{u(x,t+h) - u(x,t)}{h} &= \int_{\mathbb{R}^n} \frac{G_{t+h}(x-y) - G_t(x-y)}{h} u_0(y)\, dy \\ &\to \int_{\mathbb{R}^n} \frac{\partial}{\partial t} G_t(x-y) u_0(y)\, dy, \end{aligned}$$

i.e. $u$ is differentiable with respect to $t$, with derivative given by

$$\frac{\partial u}{\partial t}(x,t) = \int_{\mathbb{R}^n} \frac{\partial}{\partial t} G_t(x-y) u_0(y)\, dy. \quad (62)$$

[28] For more about differentiability of $u$ see Corollary 5.2.

Analogously, by uniform convergence of the spatial difference quotients

$$\frac{G_t(x+he_j-y) - G_t(x-y)}{h} \text{ and } \frac{\frac{\partial}{\partial x_j}G_t(x+he_k-y) - \frac{\partial}{\partial x_j}G_t(x-y)}{h}$$

(where $e_j$ denote the unit vector in $\mathbb{R}^n$ with $j^{th}$ component equal to 1 and the remaining components equal to zero), $u$ is once respectively twice differentiable with respect to $x$, with partial derivatives given by

$$\begin{aligned} \frac{\partial u}{\partial x_j}(x,t) &= \int_{\mathbb{R}^n} \frac{\partial}{\partial x_j} G_t(x-y)u_0(y)\,dy, \\ \frac{\partial^2 u}{\partial x_j \partial x_k}(x,t) &= \int_{\mathbb{R}^n} \frac{\partial^2}{\partial x_j \partial x_k} G_t(x-y)u_0(y)\,dy. \quad (63) \end{aligned}$$

Finally, continuity of the above partial derivatives of $u$ follows by noting that if $(x^{(\nu)}, t^{(\nu)})$ converges to $(x,t)$, then the corresponding prefactors of $u_0(y)$ in the above intergrals converge uniformly in $y$, and applying (61).

(a) now follows from (62), (63), and the fact (easily checked from the definition of $G_t$ by explicit computation) that

$$\frac{\partial}{\partial t}G_t(x-y) = \left(-\frac{n}{2t} + \frac{|x-y|^2}{4\kappa t^2}\right) G_t(x-y) = \kappa \Delta_x G_t(x-y).$$

(b) follows since as $|x| \to \infty$, the integrand in (60) tends pointwise to zero, and the convergence is dominated since the integrand is bounded in absolute value by $|u_0(y)| \in L^1(\mathbb{R}^n)$.

Finally, (c) is a consequence of Lemma 4.1 (a) in Appendix C. The proof of Theorem 2.2 is complete.

Let us now interpret the solution formula (60). The function $G_t$ appearing in the formula is a Gaussian of standard deviation $\sim \sqrt{t}$.


Thus the solution at point $x$ emerges from taking a "weighted average" of the initial data over a region around $x$ of diameter $\sim \sqrt{t}$. In particular, for small $t$, only a small neighbourhood of $x$ is relevant, but for large $t$, the average has to be taken over a larger and larger region, "wiping out" the detailed structure of the initial data below the averaging lengthscale.

On physical grounds one would expect the above solution to be unique. The natural setting for formulating and proving this uniqueness mathematically (that of Sobolev spaces) lies beyond the scope of this short section. See e.g. [Ev].

We single out one basic consequence of the solution formula:

**Corollary 2.1** *(Long time behaviour) Let* $u(\cdot, 0) \in L^1(\mathbb{R}^n)$*. Let* $u(x,t)$ *be the solution to the heat equation given by (60). Then*

$$\sup_{x \in \mathbb{R}^n} |u(x,t)| \le C t^{-n/2} ||u(\cdot, 0)||_1, \quad C = (4\pi\kappa)^{-n/2}$$

*for all* $t > 0$*. In particular,* $\sup_{x \in \mathbb{R}^n} |u(x,t)| \to 0$ *as* $t \to \infty$*.*

Physical interpretation: at long time, the temperature of a conducting body approaches the temperature prescribed at its boundary.[29]

[29] In the light of this finding, let us revisit the modelling assumption of a constant boundary condition made in (BC). It may be paraphrased, physically, as assuming a surrounding "infinite reservoir" whose behaviour is only negligibly affected by the conductor. Think of warm summer air, surrounding your favourite cold drink.