Control of the Heat Equation

Thomas I. Seidman
Department of Mathematics and Statistics
University of Maryland Baltimore County
Baltimore, MD 21228, USA
e-mail: ⟨seidman@math.umbc.edu⟩

ABSTRACT:

Key Words: .

1. Introduction

We must begin by making an important distinction between the considerations appropriate to a great variety of practical problems for controlled heat transfer and those more theoretical considerations which arise when we wish to apply the essential ideas developed for the control theory of ordinary differential equations in the context of systems governed by partial differential equations — here, the linear *heat equation*

$$\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} \quad \text{for } t > 0,\ \mathbf{x} = (x, y, z) \in \Omega \subset \mathbb{R}^3. \tag{1.1}$$

Many of the former set of problems are ones of optimal design, rather than of dynamic control and many of the essential concerns are related to fluid flow in a heat exchanger or to phase changes (e.g., condensation) or to other issues which go well beyond the physical situations described by (1.1). Some properties of (1.1) are relevant for these problems and we shall touch on these, but the essential concerns which dominate them are outside the scope of this article.

The primary focus of this article will be, from the point of view of control theory, on the inherent distinctions one must make between 'lumped parameter systems' (with finite-dimensional state space, governed by ordinary differential equations) and 'distributed parameter systems' governed by partial differential equations such as (1.1) so that the state, for each t, is a function of position in the spatial region Ω. While (1.1) may be viewed abstractly as an ordinary differential equation[1]

$$\frac{dv}{dt} = \mathbf{\Delta} v + \psi \quad \text{for } t > 0, \tag{1.3}$$

it is important to realize that abstract ordinary differential equations such as (1.3) are quite different in nature from the more familiar ordinary differential equations with finite-dimensional state so one's intuition must be attuned

[1]Now $v(t)$ denotes the state, viewed as an element of an infinite-dimensional space of functions on Ω, and $\mathbf{\Delta} = \vec{\nabla}^2$ is the *Laplace operator*, given in the 3-dimensional case by

$$\mathbf{\Delta} : v \mapsto \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} \tag{1.2}$$

with specification of the relevant boundary conditions.

to this situation. Further, the intuition appropriate to consideration of the parabolic partial differential equation (1.3) is quite different form what would be appropriate, say, for the *wave equation*

$$\frac{d^2w}{dt^2} = \mathbf{\Delta}w + \psi_2 \quad \text{for } t > 0, \tag{1.4}$$

which describes a very different set of physical phenomena with very different properties (although in subsection 5:C we do describe an interesting relation for the corresponding theories of observation and control).

The first two sections of this article provide, as background, some relevant properties of (1.1), including the presentation of some examples and implications of these general properties for practical heat conduction problems. We then turn to the discussion of system-theoretic properties of (1.1) or (1.3). We will emphasize, in particular, the considerations which arise when the input/output occurs in a way which has no direct analog in the theory of lumped parameter systems — not through the equation itself, but through the boundary conditions which are appropriate to the partial differential equation (1.1). This mode of interaction is, of course, quite plausible for physical implementation since it is typically difficult to influence or to observe directly the behavior of the system in the interior of the spatial region.

2. Background: physical derivation

Unlike situations involving ordinary differential equations with finite-dimensional state space, it is almost impossible to work with partial differential equations without developing a deep appreciation for the characteristic properties of the particular kind of equation. For the classical equations, such as (1.1), this is closely related to physical interpretations. Thus, we begin with a discussion of some interpretations of (1.1) and only then note the salient properties which will be needed to understand its control.

While we speak of (1.1) as the *heat equation*, governing conductive heat transfer, our intuition will be aided by noting also that this same equation also governs molecular diffusion for dilute solutions and certain dispersion phenomena as well as the evolution of the probability distribution in the stochastic theory of Brownian motion.

For heat conduction, we begin with the fundamental notions of *heat con-*

tent $\mathcal{Q}$ and *temperature*, related[2] by

(2.2) [heat content] = [heat capacity] · [temperature].

or, in symbols,

(2.3) $$Q = \rho c\, T$$

where Q is here the heat density (per unit volume), ρ is the mass density, T is the temperature, and c is the 'incremental heat capacity' [amount of heat needed to raise the temperature of a unit mass by, say, 1°]. The well-known physics of the situation is that *heat will flow by conduction from one body to another at a rate proportional to the difference of their temperatures.* Within a continuum one has a *heat flux* vector $\vec{q}$ describing the heat flow: $\vec{q}\cdot\vec{n}\,dA$ is the rate (per unit time) at which heat flows through any (imaginary) surface element dA, oriented by its unit normal $\vec{n}$. This is now given by *Fourier's Law*:

(2.4) $$\vec{q} = -k\, grad T = -k\vec{\nabla}T$$

with a (constant[3]) *coefficient of heat conduction* $k > 0$.

For any (imaginary) region $\mathcal{B}$ in the material, the total rate of heat flow out of $\mathcal{B}$ is then $\int_{\partial\mathcal{B}} \vec{q}\cdot\vec{n}\,dA$ (where $\vec{n}$ is the outward normal to the bounding surface $\partial\mathcal{B}$) and, by the Divergence Theorem, this equals the volume integral of $\operatorname{div}\vec{q}$. Combining this with (2.3) and (2.4) — and using the arbitrariness

[2]More precisely, since the *mass density* ρ and the *incremental heat capacity* c (i.e., the amount of heat needed to raise the temperature of a unit mass of material by, e.g., 1°C when it is already at temperature ϑ) are each temperature-dependent, the heat content in a region $\mathcal{R}$ with temperature distribution $T(\cdot)$ is given by

(2.1) $$Q = Q(\mathcal{B}) = \int_{\mathcal{B}}\int_0^T [\rho c](\vartheta)\, d\vartheta\, dV.$$

For our present purposes we are assuming that (except, perhaps, for the juxtaposition of regions with dissimilar materials) we may take ρc to be constant. Essentially, this means that we assume the temperature variation is not so large as to force us to work with the more complicated nonlinear model implied by (2.1). In particular, it means that we will not treat situations involving phase changes such as condensation or melting.

[3]This coefficient k is, in general, also temperature dependent as well as a material property. Our earlier assumption in connection with ρc is relevant here also to permit us to take k to be a constant.

of $\mathcal{B}$ — this gives the governing[4] heat equation

$$\rho c \frac{\partial T}{\partial t} = \vec{\nabla} \cdot k \vec{\nabla} T + \psi \tag{2.5}$$

where ψ is a possible source term for heat.

Let us now derive the equation governing molecular diffusion, so we are considering the spread of some substance in another (e.g., a 'solute' in a 'solvent') caused, as discussed in one of Einstein's famous papers of 1905, by the random collisions of molecules. Assuming a dilute enough solution that one can neglect the volume fraction occupied by the solute in comparison with the solvent, we may present our analysis simply in terms of the concentration (relative density) C of the relevant chemical component. One has, entirely analogous to the previous derivation, a *material flux* vector $\vec{J}$ which is now given by *Fick's Law*:

$$\vec{J} = -D \vec{\nabla} C \tag{2.6}$$

where $D > 0$ is the *diffusion coefficient*[5]. As in deriving (2.5), this law for the flux immediately leads to the conservation equation

$$\frac{\partial C}{\partial t} = \vec{\nabla} \cdot D \vec{\nabla} C + \psi \tag{2.7}$$

where ψ is now a source term for this component — say, by some chemical reaction.

A rather different mechanism for the spread of some substance in another depends on the effect of comparatively small relative velocity fluctuations of the medium — e.g., gusting in the atmospheric spread of the plume from a smokestack or the effect of path variation through the interstices of packed soil in considering the spread of a pollutant in groundwater flow. Here one again has a material flux for the concentration — given now by *Darcy's Law*, which appears identical to (2.6). The situation may well be more complicated

[4]It is essential to realize that $\vec{q}$, as given in (2.4), refers to heat flow *relative to the material*. If there is spatial motion of the material itself, then this argument remains valid provided the regions $\mathcal{B}$ are taken as moving correspondingly — i.e., (1.3) holds in material coordinates. When this is referred to stationary coordinates, we view the heat as transported in space by the material motion — i.e., we have convection.

[5]More detailed treatments might consider the possibility that D depends on the temperature, etc., of the solvent and is quite possibly also dependent on the existing concentration, even for dilute concentrations. As earlier, we neglect these effects as insignificant for the situations under consideration and take D to be constant.

here, however, since one may well have anisotropy (D is then a matrix) and/or various forms of degeneracy (e.g., D becoming 0 when $C = 0$); nevertheless, we still get (2.7) with this *dispersion coefficient* D. Here, as earlier, we will assume we may take a constant scalar $D > 0$.

As we are assuming constant coefficients in each case, we may simplify (2.5) or (2.7) by writing these as

$$v_t = D\boldsymbol{\Delta} v + \psi \tag{2.8}$$

where v stands either for the temperature T or the concentration C, the subscript t denotes a partial derivative, and, in considering (2.5), D stands for the *thermal diffusivity* $\alpha = k/\rho c$. We may, of course, always choose units to make $D = 1$ in (2.8) so it becomes precisely (1.3). It is interesting and important for applications to have some idea of the range of magnitudes of the coefficient D in (2.8) in fixed units — say, $cm^2/sec.$ — for various situations. For heat conduction, typical values of the coefficient $D = \alpha$ are, quite approximately:

> 8.4 for heat conduction in diamond; 1.1 for copper; .2–.6 for steam (rising with temperature); .17 for cast iron and .086 for bronze; .011 for ice; 7.8×10^{-3} for glass; 4×10^{-3} for soil; 1.4–1.7×10^{-3} for water; 6.2×10^{-4} for hard rubber; etc.

For molecular diffusion, typical figures for D might be

> around .1 for many cases of gaseous diffusion; .28 for the diffusion of water vapor in air and 2×10^{-5} for air dissolved in water; 2×10^{-6} for a dilute solution of water in ethanol and 8.4×10^{-6} for ethanol in water; 1.5×10^{-8} for solid diffusion of carbon in iron and 1.6×10^{-10} for hydrogen in glass, etc.

Finally, for, e.g., the dispersion of a smoke plume in mildly stable atmosphere (say, a 15 *mph* breeze) one might, on the other hand, have D approximately $10^6\ cm^2/sec.$ — as might be expected, dispersion is a far more effective spreading mechanism than molecular diffusion.

Assuming one knows the initial state of the physical system

$$v(x, t = 0) = v_0(x) \qquad \text{on } \Omega \tag{2.9}$$

where Ω is the region of $\mathbb{R}^3$ we wish to consider, we still cannot expect to determine the system evolution unless we also know (or can determine)

the source term ψ and, unless Ω would be all of $\mathbb{R}^3$, can furnish adequate information about the interaction at the boundary $\partial\Omega$. The simplest setting is that there is to be no such interaction at all: the physical system is to be *insulated* from the rest of the universe so there is no flux across the boundary. Formally, this requires $\vec{q} \cdot \vec{n} = 0$ or, from (2.7) or (2.4) with the scaling of (2.8),

$$-D\frac{\partial v}{\partial n} = -D\vec{\nabla} v \cdot \vec{n} = 0. \tag{2.10}$$

More generally, the flux might be more arbitrary but known so we have the *inhomogeneous Neumann condition*:

$$-D\frac{\partial v}{\partial n} = g_1 \qquad \text{on } \Sigma = (0,T) \times \partial\Omega. \tag{2.11}$$

An alternative[6] set of data would involve knowing the temperature (concentration) at the boundary, i.e., having the *Dirichlet condition*:

$$v = g_0 \qquad \text{on } \Sigma = (0,T) \times \partial\Omega. \tag{2.13}$$

The mathematical theory supports our physical interpretation:

> *If we have* (1.1) *on* $\mathcal{Q} = (0,T) \times \Omega$ *with* ψ *specified on* $\mathcal{Q}$ *and the initial condition* (2.9) *specified on* Ω*, then either*[7] *of the boundary conditions (2.11) or (2.13) suffices to determine the evolution of the system on* $\mathcal{Q}$*, i.e., for* $0 < t \leq T$.

We refer to either of these as the *direct problem*. An important property of this problem is that it is *well-posed*, i.e., a unique solution exists for each

[6]Slightly more plausible physically would be to assume that the ambient temperature or concentration would be known or determinable to be g 'just outside' $\partial\Omega$ and then to use the flux law (proportionality to the difference) directly:

$$-D\frac{\partial v}{\partial n} = \vec{q} \cdot \vec{n} = \lambda(v - g) \qquad \text{on } \Sigma \tag{2.12}$$

with a *flux transfer coefficient* $\lambda > 0$. Note that, if $\lambda \approx 0$ (negligible heat or material transport), then we effectively get (2.10). On the other hand, if λ is very large ($v - g = -(D/\lambda)\partial v/\partial n$ with $D/\lambda \approx 0$), then v will immediately tend to match g at $\partial\Omega$, giving (2.13); see subsection 4A.

[7]We may also have, more generally, a partition of $\partial\Omega$ into $\Gamma_0 \cup \Gamma_1$ with data given in the form (2.13) on $\Sigma_0 = (0,T) \times \Gamma_0$ and in the form (2.11) on $\Sigma_1 = (0,T) \times \Gamma_1$.

choice of the data and small changes in the data produce[8] correspondingly small changes in the solution.

3. Background: significant properties

In this section we note some of the characteristic properties of the 'direct problem' for the partial differential equation (1.1) and, related to these, introduce the representation formulas underlying the mathematical treatment.

3:A. The Maximum Principle and conservation

One characteristic property, going back to the physical derivation, is that (1.1) is a *conservation equation*. In the simplest form, when heat or material is neither created nor destroyed in the interior ($\psi \equiv 0$) and if the region is insulated (2.10), then [total heat or material] $= \int_\Omega v\, dV$ is constant in time. More generally, we have

$$\frac{d}{dt}\left[\int_\Omega v\, dV\right] = \int_{\partial\Omega} g_1\, dA + \int_\Omega \psi\, dV \tag{3:A.1}$$

for v satisfying (1.3)-(2.11).

Another important property is the Maximum Principle:

> *Let v satisfy* (1.3) *with $\psi \geq 0$ on $\mathcal{Q}_\tau := (0,\tau)\times\Omega$. Then the minimum value of $v(t,\mathbf{x})$ on $\overline{\mathcal{Q}_\tau}$ is attained either initially ($t = 0$) or at the boundary ($\mathbf{x} \in \partial\Omega$). Unless v is a constant, this value cannot also occur in the interior of $\mathcal{Q}_\tau$; if it is a boundary minimum with $t > 0$, then one must have $\partial v/\partial\vec{n} > 0$ at that point. Similarly, if v satisfies* (1.3) *with $\psi \leq 0$, then its maximum is attained for $t = 0$ or at $\mathbf{x} \in \partial\Omega$, etc.*

One simple argument for this rests on the observation that at an interior minimum one would necessarily have $v_t = \partial v/\partial t \leq 0$ and also $\Delta v \geq 0$.

The Maximum Principle shows, for example, that the mathematics of (1.1) is consistent with the requirement for physical interpretation that a

[8] We note that making this precise — i.e., specifying the appropriate meanings of 'small' — becomes rather technical and, unlike the situation for ordinary differential equations, can be done in several ways which each may be useful for different situations. We will see, on the other hand, that some other problems which arise in system-theoretic analysis turn out to be 'ill-posed', i.e., not to have this well-posedness property; see, e.g., subsection 4:C.

concentration cannot become negative and the fact that, since heat flows 'from hotter to cooler', it is impossible to develop a 'hot spot' except by providing a heat source.

3:B. Smoothing and localization

Perhaps the dominant feature of (1.1) is that solutions rapidly smooth out, with peaks and valleys of the initial data averaging out. We will see this in more mathematical detail later, but comment now on three points:

- approach to steady state
- infinite propagation speed
- localization and geometric reduction

The first simply means that if neither ψ nor the data g_0 would vary in time, then the solution v of (1.3)-(2.13) on $(0,\infty)\times\Omega$ would tend, as $t\to\infty$, to the unique solution $\bar{v}$ of the (elliptic) *steady-state equation*

$$-\left[\frac{\partial^2\bar{v}}{\partial x^2}+\frac{\partial^2\bar{v}}{\partial y^2}+\frac{\partial^2\bar{v}}{\partial z^2}\right]=\psi,\quad \bar{v}\Big|_{\partial\Omega}=g_0. \tag{3:B.1}$$

Essentially the same would hold if we were to use (2.11) rather than (2.13) except that, as is obvious from (3:A.1), we must then impose a consistency condition that

$$\int_{\partial\Omega} g_1\,dA+\int_\Omega \psi\,dV=0$$

for there to be a steady state at all — and then must note that the solution of the steady-state equation

$$-\left[\frac{\partial^2\bar{v}}{\partial x^2}+\frac{\partial^2\bar{v}}{\partial y^2}+\frac{\partial^2\bar{v}}{\partial z^2}\right]=\psi,\quad \partial\bar{v}/\partial\vec{n}=g_1 \tag{3:B.2}$$

only becomes unique when one supplements (3:B.2) by specifying, from the initial conditions (2.9), the value of $\int_\Omega \bar{v}\,dV$.

Unlike the situation with the wave equation (1.4), the mathematical formulation (1.1), etc., implies an infinite propagation speed for disturbances — e.g., the effect of a change in the boundary data $g_0(t,\mathbf{x})$ at some point $\mathbf{x}_*\in\partial\Omega$ occurring at a time $t=t_*$ is immediately felt throughout the region, affecting the solution for every $\mathbf{x}\in\Omega$ at every $t>t_*$. One can see that this is necessary to have the Maximum Principle, for example, but it is certainly non-physical. This phenomenon is a consequence of idealizations in our derivation and becomes consistent with our physical intuition when

we note that this 'immediate influence' is extremely small: there is, indeed, a noticeable delay before a perturbation will have a noticeable effect at a distance.

Consistent with the last observation, we note that the behavior in any subregion will, to a great extent, be affected only very slightly (in any fixed time) by what happens at parts of the boundary which may be very far away; this is a sort of 'localization' principle. For example, if we are only interested in what is happening close to one part of the boundary, then we may effectively treat the far boundary as 'at infinity'. To the extent that there is little spatial variation in the data at the nearby part of the boundary, we may then approximate the solution quite well by looking at the solution of the problem considered on a half-space with spatially constant boundary data, dependent only on time. Taking coordinates so the boundary becomes the plane '$x = 0$', one easily sees that this solution will be independent of the variables y, z if the initial data and source term are. The equation (1.1) then reduces to a one-dimensional form

$$\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial x^2} + \psi(t, x) \tag{3:B.3}$$

for $t > 0$ and, now, $x > 0$ with, e.g., specification of $v(t, 0) = g_0(t)$ and of $v(0, x) = v_0(x)$. Similar dimensional reductions occur in other contexts — one might get (3:B.3) for $0 < x < L$ where L gives the thickness of a slab in appropriate units or one might get a two-dimensional form corresponding to a body which is long compared to its constant constant cross-section and with data which is relatively constant longitudinally. In any case, our equation will be (1.3), with the dimensionally suitable interpretation of the Laplace operator. Even if the initial data does depend on the variables to be omitted, our first property asserts that this variation will tend to disappear so we may still get a good approximation after waiting through an initial transient. On the other hand, one usually cannot accept this approximation near, e.g., the ends of the body where 'end effects' due to those boundary conditions may become significant.

3:C. Linearity

We follow Fourier in using the *linearity* of the heat equation, expressed as a 'superposition principle' for solutions, to obtain a general representation for solutions as an infinite series. Let $\{[e_k,\, \lambda_k] :\ k = 0, 1, \ldots\}$ be the pairs of

eigenfunctions and *eigenvalues* for $-\boldsymbol{\Delta}$ on Ω, i.e.,

$$-\boldsymbol{\Delta} e_k = \lambda_k e_k \quad \text{on } \Omega \text{ (with BC) for } k = 0, 1, \ldots \tag{3:C.1}$$

where "BC" denotes one of the homogeneous conditions

$$e_k = 0 \quad \text{or} \quad \frac{\partial e_k}{\partial \vec{n}} = 0 \qquad \text{on } \partial\Omega \tag{3:C.2}$$

according as we are considering (2.13) or (2.11). It is always possible to take these so that

$$\int_\Omega |e_k|^2 \, dV = 1, \quad \int_\Omega e_i e_k \, dV = 0 \text{ for } i \neq k, \tag{3:C.3}$$

with $0 \leq \lambda_0 < \lambda_1 \leq \ldots \to \infty$; we have $\lambda_0 > 0$ for (2.13) and $\lambda = 0$ for (2.11).

One sees immediately from (3:C.1) that each function $e^{-\lambda_k t} e_k(\mathbf{x})$ satisfies (1.1) so, superposing, we see that

$$v(t, \mathbf{x}) = \sum_k c_k e^{-\lambda_k t} e_k(\mathbf{x}) \tag{3:C.4}$$

gives the 'general solution' with the coefficients (c_k) obtained from (2.9) by

$$c_k = \langle e_k, v_0 \rangle \qquad \text{so } v_0(\cdot) = \sum_k c_k e_k(\cdot), \tag{3:C.5}$$

assuming (3:C.3). Note that $\langle \cdot, \cdot \rangle$ denotes the $L^2(\Omega)$ inner product: $\langle f, g \rangle = \int_\Omega f(\mathbf{x}) g(\mathbf{x}) \, d^m\mathbf{x}$ (for m-dimensional Ω — with, physically, $m = 1, 2, 3$). The expansion (3:C.5), and so (3:C.4), is valid if the function v_0 is in the Hilbert space $L^2(\Omega)$, i.e., if $\int_\Omega |v_0|^2 < \infty$. Note that the series (3:C.5) need not converge pointwise unless one assumes more smoothness for v_0 but, since it is known that, asymptotically as $k \to \infty$, one has

$$\lambda_k \sim C k^{2/m} \qquad \text{with } C = C(\Omega), \tag{3:C.6}$$

the factors $e^{-\lambda_k t}$ decrease quite rapidly for any fixed $t > 0$ and (3:C.4) then converges nicely to a smooth function. Indeed, this is just the 'smoothing' noted above: this argument can be used to show that solutions of (1.1) are analytic (representable locally by convergent power series) in the interior of Ω for any $t > 0$ and we note that this does not depend on having homogeneous boundary conditions.

Essentially the same approach can be used when there is a source term ψ as in (2.5) but we still have homogeneous boundary conditions as, e.g., $g_0 = 0$ in (2.13). We can then obtain the more general representation

$$\begin{aligned} &v(t,\mathbf{x}) = \textstyle\sum_k \gamma_k(t) e_k(\mathbf{x}) \qquad \text{where} \\ &\gamma_k(t) = c_k e^{-\lambda_k t} + \textstyle\int_0^t e^{-\lambda_k(t-s)} \psi_k(s)\, ds, \\ &c_k = \langle e_k, v_0 \rangle, \qquad \psi_k(t) = \langle e_k, \psi(t,\cdot) \rangle \end{aligned} \tag{3:C.7}$$

for the solution of (2.5). When ψ is constant in t this reduces to

$$\gamma_k(t) = \psi_k/\lambda_k + [c_k - \psi_k/\lambda_k]\, e^{-\lambda_k t} \longrightarrow \psi_k/\lambda_k$$

which not only shows that $v(t,\cdot) \to \bar{v}$, as in (3:B.1) with $g_0 = 0$, but also demonstrates the exponential rate of convergence with the transient dominated by the principal terms, corresponding to the smaller eigenvalues. This last must be modified slightly when using (2.11), since one then has $\lambda_0 = 0$.

Another consequence of linearity is that the effect of a perturbation is simply additive: if $\hat{v}$ is the solution of (2.5) with data $\hat{\psi}$ and $\hat{v}_0$ and one perturbs this to obtain a new perturbed solution $\tilde{v}$ for the data $\hat{\psi} + \psi$ and $\hat{v}_0 + v_0$ (and unperturbed boundary data), then the solution perturbation $v = \tilde{v} - \hat{v}$ itself satisfies (2.5) with data ψ and v_0 and homogeneous boundary conditions. If we now multiply the partial differential equation by v and integrate, we obtain

$$\frac{d}{dt}\left(\tfrac{1}{2}\int_\Omega |v|^2\right) + \int_\Omega |\vec{\nabla} v|^2 = \int_\Omega v\psi,$$

using the Divergence Theorem to see that $\int v\mathbf{\Delta} v = -\int |\vec{\nabla} v|^2$ with no boundary term since the boundary conditions are homogeneous. The Cauchy-Schwartz Inequality gives $|\int v\psi| \le \|v\|\,\|\psi\|$ where $\|\cdot\|$ is the $L^2(\Omega)$-norm: $\|v\| = [\int_\Omega |v|^2]^{1/2}$ and we can then apply the Gronwall Inequality[9] to get, for example, the *energy inequality*

$$\|v(t)\|^2, \quad 2\int_0^t \|\vec{\nabla} v\|^2\, ds \le \left(\|v_0\|^2 + \int_0^t \|\psi\|^2\, ds\right) e^t. \tag{3:C.8}$$

This is one form of the well-posedness property asserted at the end of the last section.

[9] *If a function $\varphi \ge 0$ satisfies: $\varphi(t) \le C + M\int_0^t \varphi(s)\, ds$ for $0 \le t \le T$, then it satisfies: $\varphi(t) \le Ce^{Mt}$ there.*

3:D. Autonomy, similarity and scalings

Two additional useful properties of the heat equation are *autonomy* and *causality*. The first just means that the equation itself is time-independent so a time-shifted setting just gives the time-shifted solution. For the pure initial-value problem — i.e., (1.1) with $g = 0$ in (2.13) or (2.11) — 'causality' means that $v(t, \cdot)$ is determined by its 'initial data' at *any* previous time t_0 so we may write

$$v(t, \cdot) = \mathbf{S}(t - t_0)\, v(t_0, \cdot) \tag{3:D.1}$$

where $\mathbf{S}(\tau)$ is the *solution operator* for (1.1) for elapsed time $\tau \geq 0$. This operator $\mathbf{S}(\tau)$ is a nice linear operator in a variety of settings, e.g., $L^2(\Omega)$ or the space $\mathcal{C}(\bar{\Omega})$ of continuous functions with the topology of uniform convergence. A comparison with (3:C.4) shows that

$$\mathbf{S}(t) : e_k \mapsto e^{-\lambda_k t} e_k \quad \text{so } \mathbf{S}(t) \left[\sum_k c_k e_k \right] = \left[\sum_k c_k e^{-\lambda_k t} e_k \right]. \tag{3:D.2}$$

From (3:D.2) one obtains the fundamental 'semigroup property'

$$\mathbf{S}(s + t) = \mathbf{S}(t) \circ \mathbf{S}(s) \qquad \text{for } t, s \geq 0. \tag{3:D.3}$$

This only means that, if one initiates (1.1) with any initial data v_0 at time 0 and so obtains $v(s, \cdot) = \mathbf{S}(s)v_0$ after a time s and $v(s + t, \cdot) = \mathbf{S}(s + t)v_0$ after a longer time interval of length $s + t$, as in (3:D.1), 'causality' gives $v(s + t, \cdot) = \mathbf{S}(t)v(s, \cdot)$. It is possible to verify that this operator function is strongly continuous at $t = 0$:

$$\mathbf{S}(t)v_0 \to v_0 \text{ as } t \to 0 \text{ for each } v_0$$

and differentiable for $t > 0$: the equation (1.1) just tells us that

$$\frac{d}{dt}\mathbf{S}(t) = \mathbf{\Delta}\mathbf{S}(t) \tag{3:D.4}$$

where the Laplace operator $\mathbf{\Delta}$ here includes specification of the appropriate boundary conditions; we refer to $\mathbf{\Delta}$ in (3:D.4) as 'the infinitesimal generator of the semigroup $\mathbf{S}(\cdot)$'.

In terms of $\mathbf{S}(\cdot)$ we obtain a new solution representation for (2.5):

$$v(t, \cdot) = \mathbf{S}(t)v_0 + \int_0^t \mathbf{S}(t - s)\psi(s, \cdot)\, ds. \tag{3:D.5}$$

Note that $\mathbf{S}$ precisely corresponds to the 'Fundamental Solution of the homogeneous equation' for ordinary differential equations and (3:D.5) is just the usual 'variation of parameters' solution for the inhomogeneous equation (2.5); compare also with (3:C.4). We may also treat the system with inhomogeneous boundary conditions by introducing the *Green's operator* $\mathbf{G} : g \mapsto w$, defined by solving

$$-\mathbf{\Delta} w = 0 \text{ on } \Omega, \qquad \mathbf{B}w = g \text{ at } \partial\Omega. \tag{3:D.6}$$

with $\mathbf{B}w$ either w or $\partial w/\partial\vec{n}$, according as we consider (2.13) or (2.11). Using the fact that $u = v - w$ then satisfies $u_t = \mathbf{\Delta} u + (\psi - w_t)$ with homogeneous boundary conditions, we may use (3:D.5) and (3:D.4) to get, after an integration by parts,

$$\begin{aligned} v(t,\cdot) = \;& \mathbf{S}(t)v_0 + \mathbf{G}[g_0(t) - g_0(0)] \\ & + \textstyle\int_0^t \mathbf{S}(t-s)\psi(s,\cdot)\,ds - \int_0^t \mathbf{\Delta}\mathbf{S}(t-s)\mathbf{G}g_0(s)\,ds. \end{aligned} \tag{3:D.7}$$

The autonomy/causality above corresponds to the invariance of (1.1) under time-shifting and we now note the invariance under some other transformations. For this, we temporarily ignore considerations related to the domain boundary and take Ω to be the whole 3-dimensional space $\mathbb{R}^3$.

It is immediate that in considering (2.8) (with $\psi = 0$) with constant coefficients we have ensured that we may shift solutions arbitrarily in space. Not quite as obvious mathematically is the physically obvious fact that we may rotate in space. In particular, we may consider solutions which, spatially, depend only on the distance from the origin so $v = v(t,r)$ with $r = |\mathbf{x}| = \sqrt{x^2+y^2+z^2}$. The equation (2.8) with $\psi = \psi(t,r)$ is then equivalent to the equation

$$\frac{\partial v}{\partial t} = D\left[\frac{\partial^2 v}{\partial r^{\,2}} + \frac{2}{r}\frac{\partial v}{\partial r}\right] + \psi \tag{3:D.8}$$

which involves only a single spatial variable. For the two-dimensional setting $\mathbf{x} = (x,y)$ as in subsection 3.B, this becomes

$$\frac{\partial v}{\partial t} = D\left[\frac{\partial^2 v}{\partial r^{\,2}} + \frac{1}{r}\frac{\partial v}{\partial r}\right] + \psi. \tag{3:D.9}$$

More generally, for the d-dimensional case, (3:D.8) and (3:D.9) can be written as

$$v_t = r^{-(d-1)}\left(r^{d-1}v_r\right)_r + \psi. \tag{3:D.10}$$

The apparent singularity of these equations as $r \to 0$ is, of course, only an effect of the use of polar coordinates. As in subsection 3.C, we may seek a series representation like (3:C.4) for solutions of (3:D.10) with the role of the eigenfunction equation (3:C.1) now played by Bessel's equation; we then obtain an expansion in Bessel functions with the exponentially decaying time dependence $e^{-\lambda_k t}$ as earlier.

Finally, we may also make a combined scaling of both time and space. If, for some constant c, we set

$$\hat{t} = c^2 D t, \qquad \hat{\mathbf{x}} = c\mathbf{x}, \tag{3:D.11}$$

then, for any solution v of (2.8) with $\psi = 0$, the function $\hat{v}(\hat{t}, \hat{\mathbf{x}}) = v(t, \mathbf{x})$ will satisfy (1.1) in the new variables. This corresponds to the earlier comment that we may make $D = 1$ by appropriate choice of units.

Closely related to the above is the observation that the function

$$k(t, \mathbf{x}) = (4\pi D t)^{-d/2} e^{-|\mathbf{x}|^2/4Dt} \tag{3:D.12}$$

satisfies (3:D.10) for $t > 0$ while a simple computation shows[10] that

$$\int_{\mathbb{R}^d} k(t, \mathbf{x})\, d_d\mathbf{x} = 1 \quad \text{for each } t > 0 \tag{3:D.13}$$

so $k(t, \cdot)$ becomes a δ-function as $t \to 0$. Thus, $k(t - s, \mathbf{x} - \mathbf{y})$ is the *impulse response function* for an impulse at $(s, \mathbf{y})$. Taking $d = 3$, we note that

$$v(t, \mathbf{x}) = \int_{\mathbb{R}^3} k(t, \mathbf{x} - \mathbf{y}) v_0(\mathbf{y})\, d_3\mathbf{y} \tag{3:D.14}$$

is a superposition of solutions (now by integration, rather than by summation) so linearity ensures that v is itself a solution; we also have

$$v(t, \cdot) \longrightarrow v_0 \text{ as } t \to 0, \tag{3:D.15}$$

where the specific interpretation of this convergence depends on how smooth v_0 is assumed to be. Thus, (3:D.14) provides another solution representation — although, as noted, it ignores the effect of the boundary for a physical region which is not all of $\mathbb{R}^3$. For practical purposes, following the ideas of subsection 3.B, the formula (3:D.14) will be a good approximation to the

[10] We may observe that $k(t, \cdot)$ is a multivariate normal distribution (Gaussian) with standard deviation $\sqrt{2Dt} \to 0$ as $t \to 0$.

solution so long as $\sqrt{2Dt}$ is quite small[11] as compared to the distance from the point $\mathbf{x}$ to the boundary of the region.

4. Some control-theoretic problems

In this section we provide three comparatively elementary examples to see how the considerations above apply to some control-theoretic questions. The first relates to a simplified version of a quite practical heat transfer problem and is treated with the use of rough approximations, essentially to see how such heuristic treatment can be used to obtain practical results. The second describes the problem of control to a specified terminal state — which would be a standard problem in the case of ordinary differential equations but which involves some new considerations in this distributed parameter setting. The final example is a 'coefficient identification' problem: using interaction (input/output) boundary to determine the function $q = q(x)$ in an equation of the form $u_t = u_{xx} - qu$, generalizing (3:B.3).

4:A. A simple heat transfer problem

We consider a slab of thickness a and diffusion coefficient D within which heat is generated at constant rate ψ. On one side this is insulated ($v_x = 0$) and on the other it is in contact with a stream of coolant (diffusion coefficient D') moving in an adjacent duct with constant flow rate F in the y-direction. Thus, we are considering the slab as occupying $\{(x, y) : 0 < x < a,\ 0 < y < L\}$ and the duct as occupying $\{(x, y) : a < x < \bar{a},\ 0 < y < L\}$ with $a, \bar{a} \ll L$ and no dependence on z.

If the coolant enters the duct at $y = 0$ with input temperature u_0, our problem is to determine how hot the slab will become. For this purpose,

[11] When $\mathbf{x}$ is too close to the boundary for this to work well, it is often plausible to think of $\partial\Omega$ as 'almost flat' on the relevant spatial scale and then to extend v_0 by reflection across it — as an odd function if one were using (2.13) with $g_0 = 0$ or as an even function if one were using (2.11) with $g_1 = 0$. For (2.13) with, say, $g_0 = g_0(t)$ locally, there would then be a further correction by adding

$$\int_0^t \hat{k}(t-s,x)g_0(s)\,ds, \qquad \hat{k}(\tau,x) = \frac{x}{\tau}k_1(\tau,x) = 2D\frac{\partial k_1}{\partial x} \tag{3:D.16}$$

where $k_1 = k_1(\tau, x)$ is as in (3:D.12) for $d = 1$ and x is here the distance from $\mathbf{x}$ to the boundary; compare (3:D.7). There are also comparable correction formulas for more complicated settings.

we assume that we are operating in steady-state, that the coolant flow is turbulent enough to ensure perfect mixing (and so constant temperature) across the duct, and that — to a first approximation — the longitudinal transfer of heat is entirely by the coolant flow so we may consider conduction in the slab only in the transverse direction ($0 < x < a$).

The source term ψ in the slab gives heat production $a\psi$ per unit distance in y and this must be carried off by the coolant stream to have a steady-state. We might, as noted earlier, shift to material coordinates in the stream to obtain an equation there but, more simply, we just observe that when the coolant has reached the point y it must have absorbed the amount $a\psi y$ of heat per second and, for a flow rate F (choosing units so ρc in (2.3) is 1) this will have raised the coolant temperature from u_0 to $[u_0 + a\psi y/F] =: u(y)$.

Now consider the transverse conduction in the slab. We have there $v_t = Dv_{xx} + \psi$ with $v_t = 0$ for steady-state. As $v_x = 0$ at the outer boundary $x = 0$, the solution has the form $v = v^* - (\psi/2D)x^2$ where v^* is exactly what we wish to determine. If we assume a simple temperature match of slab to coolant ($v = u(y)$ at $x = a$), then this gives $v^*(y) - (\psi/2D)a^2 = v(a, y) = u(y) = u_0 + a\psi y/F$ so

$$\text{(4:A.1)} \qquad \begin{aligned} v^* = v^*(y) &= u_0 + \left[\tfrac{y}{F} + \tfrac{a}{2D}\right] a\psi \\ v &= u_0 + \left[\tfrac{y}{F} + \tfrac{a}{2D}\left(1 - \left[\tfrac{x}{a}\right]^2\right)\right] a\psi. \end{aligned}$$

A slight correction of this derivation is worth noting: for the coolant flow we expect a boundary layer (say, of thickness δ) of 'stagnant' coolant at the duct wall and within this layer we have $u_x \approx \text{constant} = -\left[v(a) - u\Big|_{\text{flow}}\right]/\delta$ while also $-D'u_x = \text{flux} = a\psi$ by Fourier's Law so, instead of matching $v(a) = u(y)$, we get $v(a) = u(y) + (\delta/D')a\psi$ which also increases v^*, v by $(\delta/D')a\psi$ as a correction to (4:A.1); effectively, this correction notes the reduction of heat transfer through replacing the boundary conditions (2.13) by (2.12) with $\lambda = D'/\delta$. Much more complicated corrections would be needed if one would have to consider conduction within the duct, especially with a velocity profile other than the plug flow assumed here.

We also note that in this derivation we neglected longitudinal conduction in the slab, essentially omitting the v_{yy} term in (2.8). Since (4:A.1) gives $v_{yy} = 0$, this is consistent with the equation. It is, however, inconsistent with reasonable boundary conditions at the ends of the slab ($y = 0, L$) and one would expect 'end effects' as well as some evening out of v^*.

We note that, although this was derived in steady-state, we could think of using (4:A.1) for an optimal control problem (especially if ψ would be time dependent, but slowly varying) with the flow rate F as control.

4:B. Exact control

We consider the problem of using ψ as control to reach a specified 'target state' $\omega = \omega(x)$ at time T. We base the discussion on the representation[12] (3:C.7), which permits us to treat each component independently: the condition that $v(T,\cdot) = \omega$ becomes the sequence of 'moment equations'

$$\begin{aligned} \gamma_k(T) &= c_k e^{-\lambda_k T} + \textstyle\int_0^T e^{-\lambda_k(T-s)}\psi_k(s)\,ds \\ &= \omega_k := \langle e_k, \omega\rangle \end{aligned} \tag{4:B.2}$$

for each k. This does not determine the control uniquely, when one exists, so we select by optimality, minimizing the norm of ψ in $L^2(\mathcal{Q})$ with $\mathcal{Q} = (0,T)\times\Omega$. This turns out to be equivalent to requiring that $\psi_k(t)$ should be a constant times $e^{\lambda_k(T-t)}$ so, noting that $\int_0^T \left|e^{-\lambda_k(T-s)}\right|^2 ds = \left[1 - e^{-2\lambda_k T}\right]/2\lambda_k$, the conditions (4:B.2) give us the formula

$$\psi(t,\mathbf{x}) = 2\sum_k \lambda_k \left(\frac{\omega_k - c_k e^{-\lambda_k T}}{1 - e^{-2\lambda_k T}}\right) e^{-\lambda_k(T-t)} e_k(\mathbf{x}). \tag{4:B.3}$$

This formula converges if (and only if) the specified target state ω is, in fact, attainable by some control in $L^2(\mathcal{Q})$. So far, so good!

Let us now see what happens when we actually attempt to implement the use of (4:B.3). Adding a touch of realism, one must truncate the expansion (say, at $k = K$) and one must then find each coefficient $\alpha_k = \omega_k - c_k e^{-\lambda_k T}$ with an error bound ε_k by using some algorithm of numerical integration on Ω to compute the inner products $\langle e_k, \omega\rangle$. [For simplicity we assume that we would already know exactly the relevant eigenvalues and eigenfunctions, as is the case for (3:B.3) and for a variety of higher-dimensional geometries.]

[12] For definiteness, one may think of the 1-dimensional heat equation (3:B.3) with homogeneous Dirichlet boundary conditions at $x = 0, 1$. The eigenvalues and normalized eigenfunctions are then

$$\lambda_k = k^2\pi^2, \qquad e_k(x) = \sqrt{2}\sin k\pi x \tag{4:B.1}$$

so the expansions, starting at $k = 1$, for convenience, are standard Fourier sine series.

Denoting the optimal control by Ψ and the approximation obtained by Ψ_K, we can bound the total error by

$$\|\Psi - \Psi_K\|_{\mathcal{Q}}^2 \leq 2\sum_{k=1}^{K}\left[\frac{\lambda_k\varepsilon_k^2}{1-e^{-2\lambda_k T}}\right] + 2\sum_{k>K}\left[\frac{\lambda_k\alpha_k^2}{1-e^{-2\lambda_k T}}\right]. \tag{4:B.4}$$

For an attainable target ω the second sum is small for large K, corresponding to convergence of (4:B.3). The use of a fixed error bound $|\varepsilon_k| \leq \varepsilon$ for the coefficient computation would make the first sum of the order of $K^{1+(2/d)}\varepsilon$ by (3:C.6) so this sum would become large as K increased. To make the total error (4:B.4) small requires picking K and then choosing ε dependent on this choice — or using a relative error condition: $|\varepsilon_k| \leq \varepsilon|\alpha_k|$. This last seems quite plausible for numerical integration with floating point arithmetic — but one trap remains! Neglecting v_0, a plausible form of the error estimate for a method of numerical integration might be

$$|\varepsilon_k| \leq C_\nu h^\nu \|\omega e_k\|_{[\nu]} \sim C'_\nu h^\nu \lambda_k^{\nu/2}\|\omega\|_{[\nu]}$$

where h characterizes a mesh size and the subscript on $\|\cdot\|_{[\nu]}$ indicates consideration of derivatives of order up to ν, with ν depending on the choice of integration method; we have noted that $\|e_k\|_{[\nu]} \sim \lambda_k^{\nu/2}$ since the differential operator $\boldsymbol{\Delta}$ is already of order 2. This means that one might have to refine the mesh progressively to get such a uniform relative error for large k.

4:C. System identification

Finally, we consider a 1-dimensional example governed by an equation known to have the form[13]

$$\frac{\partial v}{\partial t} = D\frac{\partial^2 v}{\partial x^2} - q(x)v \tag{4:C.1}$$

but with D and the specific coefficient function $q(\cdot)$ unknown or known with inadequate accuracy. We will assume here that $u = 0$ at $x = 1$, but that interaction is possible at the end $x = 0$: there one can both manipulate the temperature and observe the resulting heat flux; for simplicity, we assume

[13]For example, such an equation might arise for a rod with heat loss to the environment — appearing through a boundary condition at the surface of the rod as in (2.12), with $g =$ constant and λ varying along the rod — if one then reduced to a simplified 1-dimensional form.

that $v_0 = 0$. Thus, we consider the input/output pairing: $g \mapsto f$, defined through (4:C.1) with

$$\begin{gathered} u(t,1) = 0 \qquad u(t,0) = g(t) \\ f(t) := -Du_x(t,0). \end{gathered} \tag{4:C.2}$$

By linearity, causality, and autonomy of the equation (4:C.1), we see that this pairing takes the convolution form

$$f(t) = \int_0^t \sigma(t-s)g(s)\,ds = \int_0^t \sigma(s)g(t-s)\,ds \tag{4:C.3}$$

where $\sigma(\cdot)$ is a kind of impulse response function. Much as we obtained (3:C.4) and (3:D.7), we get

$$\begin{aligned} \sigma(t) &= \textstyle\sum_k \sigma_k e^{-\lambda_k t} \\ \text{with} \qquad & \sigma_k := -D\lambda_k e_k'(0)\langle z, e_k\rangle \\ Dz'' - qz = 0 \quad & z(0) = 1,\ z(1) = 0, \\ -De_k'' + qe_k = \lambda_k e_k \quad & e_k(0) = 0 = e_k(1), \end{aligned} \tag{4:C.4}$$

noting that z and $\{(\lambda_k, e_k)\}$ are unknown since q is unknown.

Viewing (4:C.3) as an integral equation for σ, it can be shown that (4:C.3) determines σ for appropriate choices of the input $g(\cdot)$ — simplest would be if we could take g to be a δ-function (impulse) so the observed f would just be σ: otherwise we must first solve a Volterra equation of first kind, which is already an ill-posed problem. The function $\sigma(\cdot)$ contains all the information about the unknown q which we can get and it is possible to show that quite large differences for q may produce only very small perturbations of σ; thus, this identification problem cannot be 'well-posed', regardless of $g(\cdot)$.

None of the coefficients σ_k will be 0 so, given $\sigma(\cdot)$, (4:C.4) uniquely determines the eigenvalues $\{\lambda_k\}$ which appear there as exponents. We note that $\lambda_k \sim D\pi^2 k^2$ so $D = \lim_k \lambda_k/\pi^2 k^2$ is then determined. It is then possible to show (by an argument involving analytic continuation, Fourier transforms, and properties of the corresponding wave equation) that $\sigma(\cdot)$ uniquely determines $q(\cdot)$, as desired.

The discussion above gives no suggestion as to how to compute D, $q(\cdot)$ from the observations. Typically, one seeks nodal values for a discretization of q. This can be done, for example, by *history matching*, an approach

often used for such identification problems, in which one solves the direct problem with a guessed q to obtain a resulting '$f = f(q)$' and proceeds to find the q which makes this best match the observed f. With some further *a priori* information about the function q — say, a known bound on the derivative q' — the uniqueness result, although itself non-constructive, serves to ensure convergence for these computations to the correct result as the discretization is refined. Note that it is the auxiliary *a priori* information which converts this to a well-posed problem, although one which will be quite badly conditioned so the practical difficulties do not entirely disappear.

5. More advanced system theory

In this section we consider the system-theoretic results available for the heat equation, especially regarding observability and controllability. Our emphasis is on how, although the relevant questions are quite parallel to those standard in 'lumped parameter' control theory, one has new technical difficulties which can occur only because of the infinite-dimensional state space; this will also mean that this section will depend more heavily on results of Functional Analysis[14] and special mathematical results for the partial differential equations involved. One new consideration is that the geometry of the region Ω is relevant here. We will here concentrate primarily on problems in which input/output interaction (for control and for observation) is restricted to the boundary — partly because this is physically reasonable and partly because it is only for a system governed by a partial differential equation that one could even consider 'control via the boundary conditions'.

5:A. The duality of observability/controllability

For the finite-dimensional case, controllability for a problem and observability for the adjoint problem are dual — essentially, one can control $\dot{x} = Ax + Bg$ ($g(\cdot) =$control) from one arbitrary state to another if and only if only the trivial solution of the adjoint equation $-\dot{y} = A^*y$ can give [observation]$= B^*y(\cdot) \equiv 0$. Something similar holds for the heat equation with boundary I/O, but we must be rather careful in our statement.

We begin by computing the relevant adjoint problem, taking the boundary

[14]We note [1] as a possible general refence for Functional Analysis, specifically directed toward distributed parameter system theory.

control problem as

$$u_t = \mathbf{\Delta} u \text{ on } \mathcal{Q} \text{ with } \mathbf{B}u = g \text{ on } \Sigma \text{ and } u\Big|_{t=0} = u_0 \tag{5:A.1}$$

in which the control function g is the data for the boundary conditions, defined on $\Sigma = (0,T) \times \partial\Omega$. As for (3:D.6), the operator $\mathbf{B}$ will correspond to either (2.13) or (2.11); we may now further include in the specification of $\mathbf{B}$ a requirement that $g(\cdot)$ must be 0 outside some fixed 'patch' — i.e., a relatively open subset $\mathcal{U} \subset \Sigma$, viewed as an 'accessible' portion of Σ — and then refer to this as a problem of 'patch control'. Note that

$$u_T := u(T,\cdot) = \mathbf{S}(T)u_0 + \mathbf{L}g(\cdot) \tag{5:A.2}$$

where (3:D.7) gives

$$\mathbf{L} : g(\cdot) \mapsto \mathbf{G}[g(T) - g(0)] - \int_0^T \mathbf{\Delta}\mathbf{S}(T-s)\mathbf{G}g(s)\, ds.$$

For the adjoint problem, we consider

$$-v_t = \mathbf{\Delta} v \qquad \mathbf{B}v = 0;\ \varphi = [\hat{\mathbf{B}}v]\Big|_{\mathcal{U}} \tag{5:A.3}$$

where $\hat{\mathbf{B}}$ gives the 'complementary' boundary data: $\hat{\mathbf{B}}v := \partial v/\partial\vec{n}$ if $\mathbf{B}$ corresponds to (2.13) and $\hat{\mathbf{B}}v := v\Big|_{\partial\Omega}$ if $\mathbf{B}$ corresponds to (2.11). We then have, using the Divergence Theorem,

$$\begin{aligned} \left[\int_\Omega u_T v_T\right] - \left[\int_\Omega u_0 v_0\right] &= \int_{\mathcal{Q}} (uv)_t \quad = \int_Q [(\vec{\nabla}^2 u)v - u(\vec{\nabla}^2 v)] \\ &= \int_\Sigma [u_{\vec{n}} v - u v_{\vec{n}}] \quad = -\int_{\mathcal{U}} g\varphi \end{aligned}$$

where we write v_T, v_0 for $v(T,\cdot), v(0,\cdot)$, respectively, and set $\mathcal{Q} = (0,T) \times \Omega$. Thus, with subscripts indicating the domain for the inner product of $L^2(\cdot)$, we have the identity

$$\langle u_T, v_T\rangle_\Omega + \langle g, \varphi\rangle_{\mathcal{U}} = \langle u_0, v_0\rangle_\Omega \tag{5:A.4}$$

from which we wish to draw conclusions.

First, consider the *reachable set* $\mathcal{R} = \{u_T : g = \text{any} \in L^2(\mathcal{U}); u_0 = 0\}$, which is just the range of the operator $\mathbf{L} : L^2(\mathcal{U}) \to L^2(\Omega)$. If this were not dense, i.e., if we did not have $\overline{\mathcal{R}} = L^2(\Omega)$, then (by the Hahn–Banach Theorem) there would be some nonzero v_T^* orthogonal to all $u_T \in \overline{\mathcal{R}}$ so (5:A.4)

would give $\langle g, \varphi^* \rangle_{\mathcal{U}} = 0$ for all g, whence $\varphi^* = 0$, violating *detectability* (i.e., that $\varphi^* = 0$ only if $v^* = 0$). Conversely, a violation of detectability would give a nonzero v^* with $v_T^* \neq 0$ orthogonal to $\overline{\mathcal{R}}$. Thus, detectability is equivalent to *approximate controllability*. This last means that one could control arbitrarily closely to any target state, even if it cannot be reached exactly — a meaningless distinction for finite-dimensional linear systems although significant for the heat equation, as we have already noted that solutions of (1.1) are analytic in the interior of Ω so only very special targets can be exactly reachable.

Detectability means that the map $v_T \mapsto v \mapsto \varphi$ is 1–1 so, inverting, $\varphi \mapsto v_T \mapsto v_0$ is well-defined: one can predict (note the time-reversal in (5:A.3)) v_0 from observation of φ on $\mathcal{U}$. In the finite-dimensional case, any linear map such as $\mathbf{A} : \varphi \mapsto v_0$ would necessarily be continuous (bounded), but here this is not automatically the case; note that the natural domain of $\mathbf{A}$ is the range of $v_T \mapsto \varphi$ and, if one had continuity, this would extend to the closure $\mathcal{M} = \mathcal{M}_{\mathcal{U}} \subset L^2(\mathcal{U})$. For bounded $\mathbf{A} : \mathcal{M} \to L^2(\Omega)$ there is a bounded adjoint operator $\mathbf{A}^* : L^2(\Omega) \to \mathcal{M}$ and, if we were to set $g = \mathbf{A}^* u_0$ in (5:A.1), we would get

$$\langle u_T, v_T \rangle_\Omega = \langle u_0, \mathbf{A}\varphi \rangle_\Omega - \langle \mathbf{A}^* u_0, \varphi \rangle_{\mathcal{U}} = 0 \text{ for every } v_T \in L^2(\Omega).$$

This would imply $u_T = 0$ so $g = \mathbf{A}^* u_0$ is a nullcontrol from u_0 — indeed, it turns out that this g is the *optimal* nullcontrol in the sense of minimizing the $L^2(\mathcal{U})$-norm. Conversely, if there is some nullcontrol $\tilde{g}$ for each u_0, there will be a minimum-norm nullcontrol g and the linear map $\mathbf{C} : u_0 \mapsto g$ is then continuous by the Closed Graph Theorem; further, its adjoint $\mathbf{A} = \mathbf{C}^*$ is just the observation operator: $\varphi \mapsto v_T$. Thus, bounded observability for the adjoint problem is equivalent to nullcontrollability for (5:A.1) which, from (5:A.2), is equivalent to the consideration that the range of $\mathbf{L}$ contains the range of $\mathbf{S}(T)$.

Suppose we have nullcontrollability for arbitrarily small $T > 0$, always taking $\mathcal{U} = [0, T] \times U$ for some fixed patch $U \subset \partial\Omega$. A simple argument shows that the reachable set $\mathcal{R}$ must then be entirely independent of T and of the initial state u_0. No satisfactory characterization of $\mathcal{R}$ is available, although there are various known sufficient considerations to have some $\omega \in \mathcal{R}$.

5:B. The 1-dimensional case

From the discussion above, it will clearly be sufficient to prove bounded observability for the one-dimensional heat equation to have nullcontrollability

also. It is interesting to note that this was not realized at the time these results were first proved so each was originally proved independently. We will consider specifically the observability problem with $\Omega = (0,1)$, $\mathcal{U} = (0,T) \times \{0\}$ and

$$v_t = v_{xx} \qquad v(t,0) = v(t,1) = 0; \qquad \varphi(t) := v_x(t,0), \tag{5:B.1}$$

for which we explicitly know (4:B.1). From (3:C.4), we get

$$\varphi(t) = \sum_k \tilde{c}_k e^{-\lambda_k t}, \tag{5:B.2}$$

$$v(T,\cdot) = \sum_k \frac{\tilde{c}_k}{k\pi} e^{-\lambda_k T} e_k(\cdot). \tag{5:B.3}$$

[Note that, for convenience, we have re-reversed time in comparison with (5:A.3) and that, from (3:C.4), we have $\tilde{c}_k = \sqrt{2}k\pi \int_0^1 v_0(x) \sin k\pi x \, dx$ — although we will have no need of any explicit information about v_0.]

The form of (5:B.2) is a *Dirichlet series*; note that this becomes a power series in $\xi = e^{-\pi^2 t}$ with only the k^2 powers appearing: $e^{-\lambda_k t} = e^{-k^2\pi t} = \xi^{k^2}$. The theory of such series centers on the Müntz–Szász Theorem (extending the Weierstrass Approximation Theorem) which, for our purposes, shows that only quite special functions can have L^2-convergent expansions (5:B.2) when $\Sigma 1/\lambda_k < \infty$. One has estimates for (5:B.2) of the form

$$|\tilde{c}_k| \le \beta_k \|\varphi\|_{L^2(0,\infty)} \tag{5:B.4}$$

with the values of β_k explicitly computable as an infinite product

$$\beta_k = \sqrt{1+2\lambda_k} \prod_{i \ne k} \left| 1 + \frac{1+2\lambda_k}{\lambda_i - \lambda_k} \right| \tag{5:B.5}$$

(convergent when $\sum_k 1/\lambda_k$ is convergent); note that $1/\beta_k$ is the distance in $L^2(0,\infty)$ from $\exp[-\lambda_k t]$ to span $\{\exp[-\lambda_i t] : i \ne k\}$. L. Schwartz has further shown that for functions given as in (5:B.2) one has

$$\|\varphi\|_{L^2(0,\infty)} \le \Gamma_T \|\varphi\|_{L^2(0,T)}. \tag{5:B.6}$$

Combining these estimates shows that

$$\|v(T,\cdot)\|_{L^2(0,1)} \le C_T \Gamma_T \|\varphi\|_{L^2(0,T)} \quad \left(C_T^2 := \sum_k \left[\frac{\beta_k}{k\pi} \right]^2 e^{-2k^2\pi^2 T} \right). \tag{5:B.7}$$

The sequence β_k increases moderately rapidly as $k \to \infty$ but the exponentials $\exp[-k^2\pi^2 T]$ decay even more rapidly so the sum giving C_T^2 is always convergent and (5:B.7) provides a bound ($\|\mathbf{A}\| \leq \Gamma_T C_T < \infty$) for the observation operator[15] $\mathbf{A} : \mathcal{M} = \mathcal{M}_{[0,T]} \to L^2(\Omega) : \varphi \mapsto v(T, \cdot)$, when $T > 0$ is arbitrarily small.

A somewhat different way of looking at this is to note that the linear functional: $\mathcal{M} \to \mathbb{R} : \varphi \mapsto \tilde{c}_k$ must be given by a function $g_k \in L^2(0,T)$ such that

$$\int_0^T g_k(t) e^{-\lambda_i t}\, dt = \delta_{i,k} := \begin{cases} 0 & \text{if } i \neq k, \\ 1 & \text{if } i = k. \end{cases} \tag{5:B.8}$$

If we think of $g_k(\cdot)$ as defined on $\mathbb{R}$ (0 off $[0,T]$), we may take the Fourier transform and note that (5:B.8) just asserts the 'interpolation conditions'

$$\hat{g}_k(-j\lambda_i) = \sqrt{2\pi}\delta_{i,k} \qquad (j = \sqrt{-1}) \tag{5:B.9}$$

so it is sufficient to construct functions $\hat{g}_k$ satisfying (5:B.9), together with the properties required by the Paley–Wiener Theorem to get the inverse Fourier transform in $L^2(0,T)$ with $\|g_k\| = \beta_k$. This approach leads to the sharp asymptotic estimate

$$\ln \|\mathbf{A}\| = \mathcal{O}(1/T) \quad \text{as } T \to 0, \tag{5:B.10}$$

showing how much more difficult[16] observability or controllability becomes for small T, even though one does have these for every $T > 0$.

A variant on this would consider the interior point observation $\varphi(t) := v(t,a)$. The observability properties now depend on number-theoretic properties of $0 < a < 1$ — e.g., for rational $a = m/n$ one gets no information at all about c_k when k is a multiple of n, since then $\sin k\pi a = 0$. It can be shown that one has bounded observability (with arbitrarily small $T > 0$) for a in a set of full measure whereas the complementary set for which this fails is uncountable in each subinterval.[17] Finally, we note that an essentially

[15] We note at this point that a recent estimate by Borwein and Erdélyi makes it possible to obtain comparable results when $\mathcal{U}$ has the form $U = \mathcal{E} \times \{0\}$ with $\mathcal{E}$ any subset of $[0,T]$ having positive measure; one consequence of this is a bang-bang principle for time-optimal constrained boundary control.

[16] This may be compared to the corresponding estimate: $\|\mathbf{A}\| = \mathcal{O}\left(T^{-(K+1/2)}\right)$ for the finite-dimensional setting, with K the minimal index giving the rank condition there.

[17] Since the 'bad set' has measure 0, one might guess that observation using local integral averages (as a 'generalized thermometer') should always work but, somewhat surprisingly, this turns out to be false.

identical treatment for all of the material of this subsection would work more generally for (3:B.3) and with other boundary conditions.

5:C. Higher-dimensional geometries

For higher-dimensional cases, we note first that we can obtain observability for any 'cylindrical' region $\Omega := (0,1) \times \hat{\Omega} \subset \mathbb{R}^d$ with $\mathcal{U} = (0,T) \times [0 \times \hat{\Omega}]$ by using the method of 'separation of variables' to reduce this to a sequence of independent one-dimensional problems: noting that we have here

$$\lambda_{k,\ell} = k^2\pi^2 + \hat{\lambda}_\ell \qquad e_{k,\ell}(x,\hat{\mathbf{x}}) = [\sqrt{2}\sin k\pi x]\hat{e}_\ell(\hat{\mathbf{x}}),$$

we get

$$\begin{aligned} \mathbf{A}v_x(\cdot,0,\cdot) &= \textstyle\sum_\ell [\mathbf{A}_1\varphi_\ell](x) e^{-\hat{\lambda}_\ell T}\hat{e}_\ell(\hat{\mathbf{x}}), \\ \text{with} \qquad \varphi_\ell(t) &:= e^{\hat{\lambda}_\ell t}\langle v_x(t,0,\cdot), \hat{e}_\ell\rangle_{\hat{\Omega}} \end{aligned}$$

where $\mathbf{A}_1$ is the observability operator for (5:B.1). It is easy to check that this gives $\|\mathbf{A}\| \le \|\mathbf{A}_1\| < \infty$ and we have nullcontrollability by duality.

For more general regions, when $\mathcal{U}$ is all of $\Sigma := (0,T) \times \partial\Omega$ we may shift to the context of nullcontrollability for (5:A.1) and rely on a simple geometric observation. Suppose we have $\Omega \subset \tilde{\Omega} \subset \mathbb{R}^d$ where $\tilde{\Omega}$ is some conveniently chosen region (e.g., a cylinder, as above) for which we already know that we have nullcontrollability for (5:A.1)$\tilde{}$, i.e., with $\mathcal{Q}, \mathcal{U}$ replaced by $\tilde{\mathcal{Q}} = (0,T) \times \tilde{Om}$ and $\tilde{\mathcal{U}} = \tilde{\Sigma} = (0,T) \times \partial\tilde{\Omega}$, respectively. Given any initial data $u_0 \in L^2(\Omega)$ for (5:A.1), we extend it as 0 to all of $\tilde{\Omega}$ and, as has been assumed possible, let $\tilde{u}$ be a (controlled) solution of (5:A.1)$\tilde{}$, vanishing on all of $\tilde{\Omega}$ at time T. The operator $\mathbf{B}$ acting (at Σ) on $\tilde{u}$ will have *some* value, which we now call 'g' and using this in (5:A.1) necessarily (by uniqueness) gives the restriction to $\mathcal{Q}$ of $\tilde{u}$ which vanishes at T. Thus, this g is a nullcontrol for (5:A.1). As already noted, once we have a nullcontrol g for each u_0, it follows, as noted earlier, that the nullcontrol operator $\mathbf{C}$ for this setting is well-defined and continuous; by duality, one also has bounded observability. We do note that it is unnecessary that $\mathcal{U}$ be all of Σ here: this construction works if the 'unused' part of Σ would be contained in $\tilde{\Sigma} \setminus \tilde{\mathcal{U}}$.

At this point we note a deep connection between the theories for the wave and heat equations:

> *observability for the wave equation $w_{tt} = \mathbf{\Delta}w$ for some $[\Omega, U]$ and some $T^* > 0$ implies observability for the heat equation $u_t = \mathbf{\Delta}u$ for arbitrary $T > 0$ in the same geometric setting $[\Omega, U]$.*

We now have the condition $\int_{\mathcal{U}} g_k z_i \exp[-\lambda_i t] = \delta_{i,k}$ (with $z_k := \hat{\mathbf{B}} e_k$), corresponding to (5:B.8) for this higher-dimensional problem. When $\mathcal{U}$ would have the form $(0,T) \times U$ for a patch $U \subset \partial\Omega$, one can take the Fourier transform (in t only) and get, as with (5:B.9),

$$\langle \hat{g}_k(-j\lambda_i, \cdot), z_i \rangle_U = \sqrt{2\pi}\delta_{i,k}. \tag{5:C.1}$$

D. Russell observed [5] that, if h_k would be the function for the boundary observability problem for the wave equation $w_{tt} = \mathbf{\Delta} w$ on Ω corresponding to g_k, one would have instead that

$$\langle \hat{h}_{\pm k}(\pm j\sqrt{\lambda_i}, \cdot), z_i \rangle_U = \sqrt{2\pi}\delta_{i,k}\delta_{\pm} \tag{5:C.2}$$

(where $\delta_{\pm} = 1, 0$ according as the occurrences of $\pm$ on the left do or do not match), so the spatial dependence is identical and the time-dependence closely related. This suggests constructing $\hat{g}_k$ as

$$\hat{g}_k(\tau, \cdot) := \left[h_{+k}(\sigma) + h_{-k}(\sigma)\right]/2 \qquad \sigma^2 := j\tau,$$

(noting that the rhs is an even analytic function of σ and so is an analytic function of σ^2), so that (5:C.2) would imply the desired (5:C.1). This does not quite work, since it would not give $g_k \in L^2(\mathcal{U})$, but can be modified, multiplying by a suitable function $R = R(\tau)$ on the right, to get g_k as an inverse Fourier transform. We note that it is the construction of the mollifier $R(\cdot)$ which gives the asymptotics (5:B.10); it is also this multiplication which makes the implication above irreversible. We note that the relation used here is also usable to obtain uniqueness results for a higher-dimensional version of the identification problem of subsection 4:C from known corresponding results for the wave equation.

One gets the observability for the wave equation for the heat equation) for suitable[18]

$[\Omega, U]$ by an argument [5], [7] based on Scattering Theory (existence of a uniform decay rate for the portion of the total energy remaining in Ω when this is embedded in a larger region $\tilde{\Omega}$). In particular, one can take $\tilde{\Omega}$ to be the complement of a 'star-shaped' obstacle at which $\mathbf{B}w = 0$ provided $\partial\Omega \setminus U \subset \partial\tilde{\Omega} = \partial\,(\text{obstacle})$. On the other hand, there is a significant gemetric restriction on on $[\Omega, U]$ for which one can observe or control the

[18]Note that the finite propagation speed associated with the wave equation implies existence of a minimum time T_*, depending on Ω, for observability/nullcontrollability.

wave equation: there can be no 'trapped waves' (continually reflecting off the unused boundary $\Omega \setminus U$ without ever intersecting U) which, roughly, requires that $\partial\Omega \setminus U$ should be 'visible' from some single point outside $\overline{\Omega}$.

This argument then gives observability/controllability for the heat equation for a variety of geometric settings, but there is a price: the geometric restriction noted above. This is quite reasonable for the wave equation but seems irrelevant to the behavior of the heat equation and it was long conjectured that accessibility of an arbitrary patch $U \subset \partial\Omega$ would suffice for observation or control of (1.1). Quite recently, G. Lebeau and L. Robbiano have settled this conjecture by demonstrating — using an argument based on new Carleman-type estimates for a related elliptic equation — patch null-controllability using an arbitrary patch[19] $\mathcal{U}$ for an arbitrary Ω.

References

[1] R.F. Curtain and A.J. Pritchard, *Functional Analysis in Modern Applied Mathematics*, Academic Press, New York, 1977.

[2] W. Krabs, *On Moment Theory and Controllability of One-Dimensional Vibrating Systems and Heating Processes*, (Lecture Notes in Control and Inf. Sci. # 173), Springer-Verlag, Berlin, 1992.

[3] I. Lasiecka and R. Triggiani, *Differential and Algebraic Riccati Equations with Application to Boundary/Point Control Problems: Continuous Theory and Approximation Theory*, (Lecture Notes in Control and Inf. Sci. # 164), Springer-Verlag, Berlin, 1991.

[4] J.-L. Lions, *Exact controllability, stabilization, and perturbations*, SIAM Review **30** (1988), pp. 1–68.

[5] D.L. Russell, "A unified boundary controllability theory for hyperbolic and parabolic partial differential equations," Stud. Appl. Math. **LII** (1973), pp. 189–211.

[19] They have given this result both for the case of boundary controllability described here and also for the case of distributed control — (1.3) with homogeneous boundary conditions and control function ψ, vanishing outside a patch $\mathcal{U}$ open in $\mathcal{Q} = (0,T) \times \Omega$.

[6] T.I. Seidman, "Boundary observation and control for the heat equation," in *Calculus of Variations and Control Theory* (D.L. Russell, edit.), Academic Press, N.Y. (1976), pp. 321–351.

[7] T.I. Seidman, "Exact boundary controllability for some evolution equations," SIAM J. Control Opt. **16** (1978), pp. 979–999.