Wave Propagation: Odd is Better, but Three is Best

1. The Formal Solution of the Wave Equation

V Balakrishnan

V Balakrishnan is in the Department of Physics at the Indian Institute of Technology Madras. His research interests are dynamical systems and applications of stochastic processes.

The Best of all Possible Worlds

In Voltaire's classic satire *Candide*, the preceptor Pangloss is an unquestioning optimist who keeps insisting that we live in 'the best of all possible worlds', in spite of the most harrowing adversities he and his companions face. Pangloss' unbridled optimism is foolish, if not dangerously stupid, in the light of the events that take place – so much so, that 'panglossian' has come to describe a hopelessly idealistic view held in spite of direct evidence to the contrary.

At a more decidable level, however, we may ask whether the physical universe in which we live is, at least in some limited sense, 'the best of all possible worlds'. But we must be careful to define this sense, because it is not our intention to discuss any version of the so-called Anthropic Principle here. So we make our question more precise and our objective much more modest: we shall (merely) show that *it is impossible to send sharp signals in one- and two-dimensional spaces, in contrast to three-dimensional space.* This result, sometimes referred to as the 'strict form' of *Huygens' Principle*, means that the very possibility of communication, and thus the transmission of information from one location to another, is dependent on the dimensionality of space. In this sense, we are indeed fortunate to live in a space of three dimensions (although some may feel that mobile phones have made this a mixed blessing, at best).

Keywords
Wave propagation, Green's function, causality, contour integration.

Propagation of a Sharply Pulsed Signal

For the sake of definiteness, we choose a specific mathematical model of signal propagation. Even though it is a 'bare minimum' sort of model, it is general enough to establish the primary result. And this result remains unaffected in essence by the addition of various details and modifications. Substantiating the last statement would lead us into lengthy digressions, and so we shall not attempt to do so here.

We show that it is impossible to send sharp signals in one- and two-dimensional spaces, in contrast to three-dimensional space.

The signal we would like to transmit is a sharp pulse that has a definite beginning and a definite end in both space and time: in technical terms, it must be *localized in space and time.* This is essential for us to make out unambiguously that it emanated at *this* place at *this* time, and reached *that* place at *that* time. Let us denote the signal (or rather, some observable property of the signal, such as its amplitude) by $u(\mathbf{r}, t)$. The propagation of the signal in space-time is governed by the wave equation

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right) u(\mathbf{r}, t) = \rho(\mathbf{r}, t), \qquad (1)$$

where $\rho(\mathbf{r}, t)$ is a specified function of space and time that represents the *source* of the signal, and c is the speed of the signal. The differential operator on the LHS is called the wave operator or the d'Alembertian (see *Box* 1). It is the analogue, in space-time, of the Laplacian operator in space. In writing down (1), we have in mind an electromagnetic signal propagating in free space or vacuum. To keep matters as simple as possible, we assume that the signal is described by a single scalar function u rather than the electric and magnetic field vectors that an actual electromagnetic signal would comprise. We also assume that the space in which the signal propagates is ordinary Euclidean space of infinite extent. This helps us avoid complications arising from boundary conditions. As we would like to explore and compare signal propagation in spaces of different dimensions, we use the symbol D for the number of spatial dimensions: D may be 1, 2 or 3. Subsequently, we shall also look at what happens if $D > 3$.

The signal is a sharp pulse that has a definite beginning and a definite end in both space and time.

Box 1

The French mathematician, natural philosopher and encyclopaedist Jean le Rond d'Alembert (1717-1783) was among the first to understand the significance of, and study in some detail, several important differential equations of mathematical physics. Among other results, he showed that the general solution of the one-dimensional wave equation $\partial^2 u/\partial t^2 - c^2\, \partial^2 u/\partial x^2 = 0$ is of the form $u(x,t) = f_1(x+ct) + f_2(x-ct)$. This corresponds to the superposition of two different waveforms or pulses moving, respectively, to the left and right with speed c. It is the forerunner of *the method of characteristics* for a class of partial differential equations. D'Alembert's name is associated with many other discoveries as well, such as d'Alembert's Principle in Mechanics, d'Alembert's scheme in games of chance such as roulette, and d'Alembert's paradox: he showed that, in the streamlined, irrotational flow of a non-viscous fluid past a solid obstacle, the net drag force on the solid *vanishes*, contrary to what one would guess off-hand.

D'Alembert (along with the philosopher Denis Diderot) spent a good deal of time and effort on a massive project, the great French Encyclopaedia. (By the way, the popular hilarious story about the great mathematician Euler confounding Diderot with his spoof of a 'mathematical proof' of the existence of God appears to be – sadly enough – without foundation.) D'Alembert seems to have been a 'straight shooter'; according to W W Rouse Ball, "d'Alembert's style is brilliant but not polished, and faithfully reflects his character, which was bold, honest and frank. ⋯ with his dislike of sycophants and bores it is not surprising that during his life he had more enemies than friends."

To retain just the bare essentials, we go further. A sharply pulsed point source of unit strength is switched on at some point $\mathbf{r}_0$ at the instant of time t_0. We want to find the resulting signal $u(\mathbf{r},t)$ at an arbitrary point $\mathbf{r}$ at an arbitrary instant of time t. Here $\mathbf{r}_0$ and $\mathbf{r}$ stand for D-dimensional position vectors; when $D=1$, $\mathbf{r}$ reduces to a single coordinate (x, say). $u(\mathbf{r},t)$ is then a solution of the equation

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right) u(\mathbf{r},t) = \delta^{(D)}(\mathbf{r}-\mathbf{r}_0)\,\delta(t-t_0)\,. \qquad (2)$$

Here δ denotes the Dirac delta function, and $\delta^{(D)}$ its generalization to D dimensions.[1]

[1] For a simple discussion of the Dirac delta function and its properties, see *Resonance*, Vol.8, No.8, pp.48-58, 2003.

The natural boundary condition on $u(\mathbf{r},t)$ is simply $u(\mathbf{r},t) \to 0$ as $r \to \infty$, where r stands for $|\mathbf{r}|$, as usual. Moreover, as there is no disturbance anywhere before the source is switched on, we have $u(\mathbf{r}, t) = 0$ and $\partial u(\mathbf{r}, t)/\partial t = 0$ for all $t < t_0$ at all points. This requirement is called *causality*, which means that the effect cannot *precede* its cause. It implies that the solution $u(\mathbf{r}, t)$, which is naturally also dependent on t_0 and $\mathbf{r}_0$, must necessarily have a particular form. This is given by

Causality means that the effect cannot precede its cause.

$$u(\mathbf{r}, t) = \theta(t - t_0)\, K(\mathbf{r}, t\,;\, \mathbf{r}_0, t_0) \qquad (3)$$

where $\theta(t - t_0)$ denotes the unit step function (= 1 and 0, respectively, for $t > t_0$ and $t < t_0$.) Owing to the presence of the step function, it is clear that the quantity $K(\mathbf{r}, t\,;\, \mathbf{r}_0, t_0)$ is in fact left unspecified for $t < t_0$ But then this is irrelevant for the physical, causal solution we seek here. In fact, causality imposes an even stronger constraint, as we shall see. Since the disturbance propagates with a finite speed c, we expect that the signal cannot reach a point $\mathbf{r}$ before the instant $t_0 + |\mathbf{r} - \mathbf{r}_0|/c$, because $|\mathbf{r} - \mathbf{r}_0|/c$ is the time it takes to propagate from $\mathbf{r}_0$ to $\mathbf{r}$. This, too, will emerge automatically in the solution.

For completeness, let us mention that the specific solution $u(\mathbf{r}, t)$ in which we are interested here is called the *retarded* or *causal Green function* corresponding to the wave operator (the differential operator in brackets on the LHS of (1)) together with the natural boundary condition just stated; while the function K is the corresponding *propagator*.

We now have a well-defined mathematical problem. Although its solution is a standard exercise, it does involve a rather surprising number of subtleties – especially in the incorporation of the conditions that enable us to arrive at a physically acceptable solution. These finer (but important) points are so often slurred over or misrepresented in otherwise respectable texts, that it seems to be worthwhile to spell them out with some care, at the risk of appearing to dwell on technicalities, not to mention increasing the length of this article.

The solution u ($\mathbf{r}$, t) is called the *retarded* or *causal Green function* corresponding to the wave operator.

Fourier Transform

To proceed, we could substitute in (2) the form given in (3) for $u(\mathbf{r}, t)$, and work with the resulting partial differential equation for K But it is just as convenient to work directly with (2). We note that the variables $t - t_0$ and $\mathbf{r} - \mathbf{r}_0$ merely represent shifts of t and $\mathbf{r}$ by constant amounts, for any given t_0 and $\mathbf{r}_0$. This suggests immediately that we change variables from $\mathbf{r}$ and t to $\mathbf{R} \equiv \mathbf{r} - \mathbf{r}_0$ and $\tau \equiv t - t_0$ respectively. Equation (2) becomes

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial \tau^2} - \nabla^2_{\mathbf{R}}\right) u = \delta^{(D)}(\mathbf{R})\,\delta(\tau) \qquad (4)$$

We have used the obvious notation $\nabla^2_{\mathbf{R}}$ for the Laplacian operator with respect to $\mathbf{R}$. The causality condition is $u = 0$ and $\partial u/\partial \tau = 0$ for all $\tau < 0^-$, while the boundary condition is $u \to 0$ for $R \to \infty$, where $R \equiv |\mathbf{R}|$. It is evident that, under these circumstances, $u(\mathbf{r}, t)$ is in fact a function of $\mathbf{R}$ and τ. In anticipation of this, we have retained the symbol u for the unknown function in (4). Note, in passing, that in a region of *finite* extent, in the presence of boundary conditions at finite values of r, the dependence of u on $\mathbf{r}$ and $\mathbf{r}_0$ cannot be reduced in general to a dependence on the difference $\mathbf{r} - \mathbf{r}_0 = \mathbf{R}$ alone.

Now, the Fourier transform of $u(\mathbf{R}, \tau)$ with respect to both $\mathbf{R}$ and τ is defined as

$$\widetilde{u}(\mathbf{k}, \omega) = \int d^D\mathbf{R} \int_{-\infty}^{\infty} d\tau\, e^{-i(\mathbf{k}\cdot\mathbf{R} - \omega\tau)}\; u(\mathbf{R}, \tau) \qquad (5)$$

The inverse relation that yields $u(\mathbf{R}, \tau)$ in terms of $\widetilde{u}(\mathbf{k}, \omega)$ is

$$u(\mathbf{R}, \tau) = \int \frac{d^D\mathbf{k}}{(2\pi)^D} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, e^{i(\mathbf{k}\cdot\mathbf{R} - \omega\tau)}\; \widetilde{u}(\mathbf{k}, \omega) \qquad (6)$$

Here $d^D\mathbf{k}$ denotes the volume element $dk_1\, dk_2 \quad dk_D$ in the space $\mathbb{R}^D$ of the D-dimensional vector $\mathbf{k}$. Similarly, exploiting the fact that the Fourier transform of the delta function is just unity, we have the representation

$$\delta^{(D)}(\mathbf{R})\,\delta(\tau) = \int \frac{d^D\mathbf{k}}{(2\pi)^D} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, e^{i(\mathbf{k}\cdot\mathbf{R}-\omega\tau)} \qquad (7)$$

Substitution of (6) and (7) in (4) gives

$$\int \frac{d^D\mathbf{k}}{(2\pi)^D} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, e^{i(\mathbf{k}\cdot\mathbf{R}-\omega\tau)} \Big((\omega^2 - c^2k^2)\,\widetilde{u}(\mathbf{k},\omega) + c^2 \Big) = 0, \qquad (8)$$

where k stands for $|\mathbf{k}|$ as usual. But the set of functions $\exp\,[i(\mathbf{k} \cdot \mathbf{R} - \omega\tau)]$, where each Cartesian component of $\mathbf{k}$ and τ can take on all real values, forms a 'complete orthonormal basis'[2] in the space of integrable functions of $\mathbf{R}$ and τ. Therefore the LHS in (8) can only vanish if the coefficient of $\exp[i(\mathbf{k} \cdot \mathbf{R} - \omega\tau)]$ for *each* value of $\mathbf{k}$ and ω itself vanishes. We must therefore have $(\omega^2 - c^2k^2)\,\widetilde{u}(\mathbf{k},\omega) + c^2 = 0$ i.e.,

$$\widetilde{u}(\mathbf{k},\omega) = -c^2/(\omega^2 - c^2k^2) \qquad (9)$$

for all $\mathbf{k}$ and ω. The idea behind the introduction of the Fourier transform should now be obvious – namely, *to convert the partial differential equation for $u(\mathbf{R},\tau)$ into a trivially-solved algebraic equation for $\widetilde{u}(\mathbf{k},\omega)$.*

The idea behind the introduction of the Fourier transform is to convert the partial differential equation for $u(\mathbf{R},\tau)$ into a trivially-solved algebraic equation for $\widetilde{u}\,(\mathbf{k},\omega)$.

Putting this expression for $\widetilde{u}(\mathbf{k},\omega)$ back in (6), we have the *formal* solution

$$u(\mathbf{R},\tau) = -c^2 \int \frac{d^D\mathbf{k}}{(2\pi)^D}\, e^{i\mathbf{k}\cdot\mathbf{R}} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, \frac{e^{-i\omega\tau}}{(\omega^2 - c^2k^2)}. \qquad (10)$$

But this does not make sense as it stands, because the integral over ω diverges owing to the vanishing of the denominator of the integrand at $\omega = -c|k|$ and again at $\omega = c|k|$. The dilemma is resolved by invoking the physical requirement of causality, as we shall see now.

[2] A brief account of basis sets in a linear vector space may be found in *Resonance*, Vol.8, No.8, p.57, 2003.

The trick is to carry out the integration by adding a well-chosen zero to the integral, so as to convert it to an integral over a closed contour in the complex plane.

Simplification of the Formal Solution

The trick is to carry out the integration over ω in (10) by adding a well-chosen zero to the integral, so as to convert it to an integral over a closed contour in the complex ω-plane. The latter is then evaluated by applying Cauchy's Residue Theorem.

Let Ω be a large positive constant. Consider a closed contour comprising a straight line from $-\Omega$ to $+\Omega$ along the real axis in the ω-plane, and a semicircle of radius Ω that takes us back from $+\Omega$ to $-\Omega$ in either the upper or lower half-plane. The limit $\Omega \to \infty$ is to be taken after the contour integral is evaluated. If the contribution from the semicircle vanishes in the limit $\Omega \to \infty$, the original line integral from $-\infty$ to $+\infty$ over ω is guaranteed to be precisely equal to the integral over the closed contour.

Now, for $\tau < 0$, this semicircle *must* lie in the *upper* half-plane in ω, because it is only in this region that the factor $\exp(-i\omega\tau)$ in the integrand vanishes exponentially as $\Omega \to \infty$. The addition of the semicircle to the contour would then simply add a vanishing contribution to the original line integral that we want to evaluate. Therefore, if no singularities of the integrand lie on the real axis or in the upper half-plane in ω, the contour integral is guaranteed to vanish identically for $\tau < 0$. But this is precisely what causality requires: namely, that $u(\mathbf{r}, t)$ remain equal to zero for all $t < t_0$, that is, for all $\tau < 0$.

On the other hand, for $\tau > 0$, i.e., for $t > t_0$ we do expect to have a signal that does not vanish identically. But now the semicircle closing the contour *must* lie in the *lower* half-plane, because it is only then that the contribution from the semicircle to the contour integral vanishes as $\Omega \to \infty$. Therefore, if all the singularities of the integrand are in the lower half-plane, all our requirements are satisfied.

This is ensured by displacing each of the poles of the integrand at $\omega = -ck$ and $\omega = +ck$ by an infinitesimal *negative* imaginary quantity $-i\epsilon$ where $\epsilon > 0$, and then passing to the limit $\epsilon \to 0$ after the integral is evaluated. In general, of course, each of the two poles of the integrand can be displaced so as to lie in the upper or lower half-plane. This leads to four possible ways of making the divergent integral finite. It is easy to see that any two of these are linearly independent solutions, the other two being linear combinations of the former pair. The particular '$i\epsilon$ prescription' we have used above is tailored to ensure that the correct causal solution is picked up from among the set of possible solutions. In general, such '$i\epsilon$ prescriptions' are a way of incorporating boundary conditions (here, initial conditions) into the solutions of differential equations. The causal solution to (2) or (4) is therefore given by the modified version of (10) that reads

'$i\varepsilon$ prescriptions' are a way of incorporating boundary conditions into the solutions of differential equations.

$$u(\mathbf{R},\tau) = -c^2 \lim_{\epsilon\downarrow 0} \int \frac{d^D\mathbf{k}}{(2\pi)^D}\, e^{i\mathbf{k}\cdot\mathbf{R}} \times$$

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{e^{-i\omega\tau}}{(\omega + ck + i\epsilon)(\omega - ck + i\epsilon)} \qquad (11)$$

But, as discussed above,

$$\begin{aligned} &\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{e^{-i\omega\tau}}{(\omega + ck + i\epsilon)(\omega - ck + i\epsilon)} \\ &= \lim_{\Omega\to\infty} \int_{-\Omega}^{\Omega} \frac{d\omega}{2\pi} \frac{e^{-i\omega\tau}}{(\omega + ck + i\epsilon)(\omega - ck + i\epsilon)} \\ &= \lim_{\Omega\to\infty} \int_{C_\pm} \frac{d\omega}{2\pi} \frac{e^{-i\omega\tau}}{(\omega + ck + i\epsilon)(\omega - ck + i\epsilon)} \end{aligned} \qquad (12)$$

where $C_\pm$ are the closed contours shown in *Figure* 1. As explained above, we must use C_+ for $\tau < 0$ and C_- for $\tau > 0$. However, C_+ does not enclose any singularity of the integrand, and so the corresponding integral vanishes, as we want it to. In the case of C_-, the integral

The integral is $-(2\pi i)$ times the sum of the residues of the integrand at the two poles enclosed.

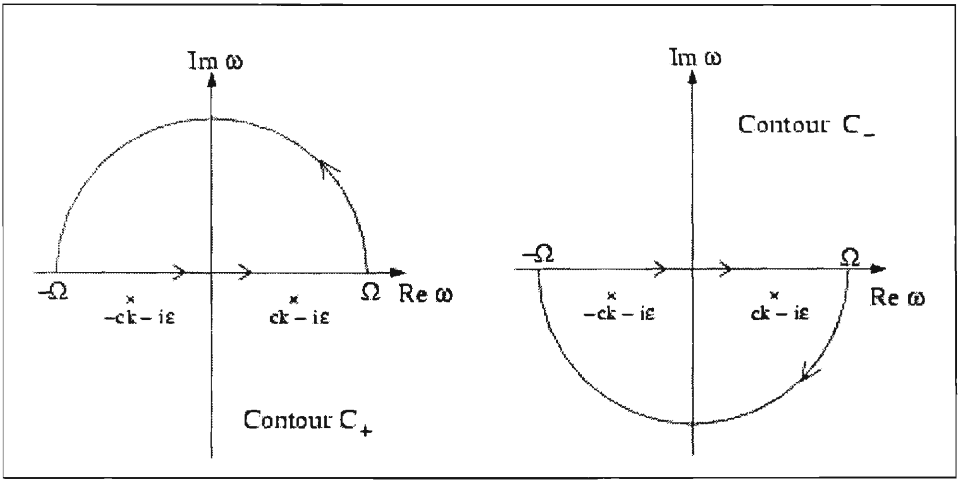

Figure 1. Contours C_+ and C_-.

is $-(2\pi i)$ times the sum of the residues of the integrand at the two poles enclosed: the extra minus sign arises because C_- is traversed in the clockwise sense. We thus obtain, after simplification,

$$u(\mathbf{R},\tau) = c\,\theta(\tau)\int \frac{d^D\mathbf{k}}{(2\pi)^D}\,\frac{\sin c\tau k}{k}\,e^{i\mathbf{k}\cdot\mathbf{R}} \qquad (13)$$

Note how the factor $\theta(\tau)$ required by causality has emerged automatically in the solution for $u(\mathbf{R},\tau)$.

For completeness, we mention in passing that solutions corresponding to the other possible '$i\epsilon$ prescriptions' described above do play a role in physics – for instance, in the so-called *Feynman propagator* in quantum field theory.

Returning to the causal solution of interest to us, we would like to distinguish between solutions obtained for different values of D. We shall therefore write the solution henceforth as $u^{(D)}(\mathbf{R},\tau)$ instead of $u(\mathbf{R},\tau)$. In Part 2 of this article, we shall deduce and analyse the explicit form of $u^{(D)}(\mathbf{R},\tau)$ for individual values of D, to bring out the special features of the solution in each case.

Suggested Reading

[1] G Barton, *Elements of Green's Functions and Propagation – Potentials, Diffusion and Waves*, Clarendon Press, Oxford, 1989.

[2] R Courant and D Hilbert, *Methods of Mathematical Physics*, Vol. 2, Interscience, New York, 1962.

[3] P R Garabedian, *Partial Differential Equations*, Wiley, New York, 1964.

[4] G B Whitham, *Linear and Nonlinear Waves*, Wiley, New York, 1999.

Address for Correspondence
V Balakrishnan
Department of Physics
Indian Institute of Technology
Madras
Chennai 600 036, India.
Email: vbalki@iitm.ac.in