Nonlinearity, Conservation Law and Shocks

Part I : Genuine Nonlinearity and Discontinuous Solutions

Phoolan Prasad

Phoolan Prasad is with the Department of Mathematics, Indian Institute of Science and has been working in the area of nonlinear waves and hyperbolic partial differential equations. He is deeply interested in mathematics education at all levels : he has written popular articles for students in schools and colleges, a text book on partial differential equations and written/ edited research monographs. He played an important role in planning and organising the Mathematics Olympiad in India.

We present in two parts, a mathematical theory of conservation laws using the language of physics. In Part I we explain the concept of a special type of nonlinearity which appears in an important class of evolutionary processes governed by hyperbolic partial differential equations. For simplicity, we develop the theory using a simple model equation. We show that it is possible to extend the concept of solutions with discontinuities with the help of a conservation form of the equation.

Introduction

Almost all natural phenomena, and social and economic changes, are governed by nonlinear equations and attempts to understand them using linearised equations turn out to be futile. Analysis and solution of nonlinear equations is more difficult than linear equations, as most of the scientists, students and mathematicians are aware. However, they may not be aware of the fact that there is a special type of nonlinearity, called *genuine nonlinearity* which distinguishes itself from other nonlinearities due to very special properties. This is because genuine nonlinearity is a subtle concept that appears in a type of partial differential equations — *quasilinear hyperbolic systems*. Only a specialist can understand this concept and appreciate its properties. In this article, an attempt has been made for the first time to explain this concept and its important properties to readers having almost no knowledge of partial differential equations.

Genuine Nonlinearity

Almost all physical phenomena, biological evolutions, population growth and social and economic changes are nonlinear. Therefore, is it not unfair that the word 'nonlinearity' is defined as a negation of linearity ? However, this term has come from mathematics, where it is much simpler to define a linear mapping or operator. As an example of a linear process without a source term i.e. linear homogeneous system, consider the statement

$$\left\{\begin{array}{l}\text{rate of change of}\\ \text{a quantity } u\end{array}\right\} \text{is} \left\{\begin{array}{l}\text{proportional to the}\\ \text{quantity } u\end{array}\right\} \qquad (1a)$$

or

$$\frac{du}{dt} = \alpha\, u, \ \ \alpha = \text{constant of proportionality} \qquad (1b)$$

where we take the rate of change to be with respect to time t. This gives a law of growth or decay of the quantity u. According to this law $u = u_0 e^{\alpha t}$, u_0 = constant; which implies that if u is initially zero, it always remains zero and if u is nonzero at any time it remains finite and nonzero for all time. An important consequence of the linear evolution is that the quantity u can neither become zero nor can tend to infinity at a finite time. It always takes infinite time to attain zero or infinity. In mathematics, equation (1) is called *linear* because if u_1 and u_2 satisfy (1) then their linear combination $a\,u_1 + b\,u_2$, where a, b are constants, also satisfies (1).

Genuine nonlinearity is a subtle concept that appears in a type of partial differential equations — *quasilinear hyperbolic systems.*

In contrast to the above law, let us consider a process in which

$$\left\{\begin{array}{l}\text{rate of change of}\\ \text{a quantity } u\end{array}\right\} \text{is} \left\{\begin{array}{l}\text{proportional to the}\\ \text{quantity } u^2\end{array}\right\} \qquad (2a)$$

or

$$\frac{du}{dt} = u^2 \qquad (2b)$$

where we have taken the constant of proportionality to be 1.This process does not satisfy a linearity criterion, so if u_1 and u_2 evolve according to the above rule then $a\,u_1 + b\,u_2$ does not evolve according to the same rule. The rule (2) gives $u = u_0 / \{1 - u_0 t\}$ which implies that starting with a finite positive value u_0 at

An important consequence of the linear evolution of a quantity is that it can neither become zero nor can tend to infinity at a finite time.

$t = 0$, the quantity u tends to infinity in just finite time $T = 1/u_0$. Had we taken another nonlinear law of evolution

$$\left\{ \begin{array}{l} \text{rate of change of} \\ \text{of } u \end{array} \right\} \text{is} \left\{ \begin{array}{l} \text{proportional to} \\ u^{2/3} \end{array} \right\} \tag{3}$$

then taking the constant of proportionality to be one, we find a solution as per this rule to be $u = (1/27)(t^3)$. Now u can attain a non-zero value at a finite time starting even from zero at $t=0$.

These are all examples of a nonlinear ordinary differential equation of the form

$$\frac{du}{dt} = f(u) \tag{4}$$

We can easily write the simplest example of a partial differential equation having nonlinearity of the type present in (4):

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = f(u), \quad c = \text{constant} \tag{5a}$$

A physical interpretation of this equation is (see next paragraph for explanation)

$$\left\{ \begin{array}{l} \text{time rate of change of} \\ u \text{ as we move with a} \\ \text{constant velocity } c \end{array} \right\} = \left\{ f(u) \right\} \tag{5b}$$

Any nonlinear effect we observe at time t in (4) will also manifest in (5) at the same time but a distance ct from its source.

Almost all results in the theory of hyperbolic partial differential equations can be interpreted in the language of wave propagation.

The topic of discussion is a different type of nonlinearity. It is observed only in hyperbolic systems of quasilinear partial differential equations. Almost all results in the theory of hyperbolic systems of partial differential equations can be interpreted in the language of wave propagation and it is quite instructive to use this language to explain the concept of genuine nonlinearity. Let us consider a point $X(t)$ moving along the x-axis and assume that its velocity of propagation (dX/dt) is equal to c. Then the rate of change of a quantity $u(x,t)$ as we move with velocity c is $(d/dt)\,u\,(X(t),t) = u_t + (dX/dt)\,u_x = u_t + cu_x$

Before we discuss the special type of nonlinearity, we consider the partial differential equation (5) with $f(u) = 0$ i.e.

$$u_t + cu_x = 0, \quad c = \text{constant} \tag{6a}$$

for which the rule (5b) can be restated in the form

$$\left\{\begin{array}{l}\text{velocity of propagation of}\\ \text{a point on a pulse having}\\ \text{amplitude } u\end{array}\right\} = \left\{\begin{array}{l}\text{the same constant } c \text{ for}\\ \text{all points of the pulse}\end{array}\right\}. \tag{6b}$$

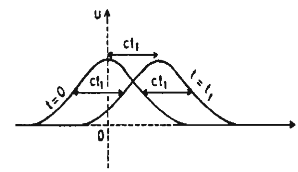

Figure 1a: Linear wave propagation

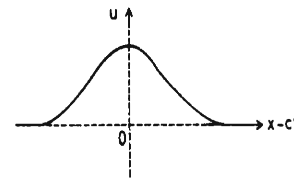

Figure 1b: Stationary pulse in a moving frame

Figure 1 Linear wave propagation in a homogeneous medium in one space dimension assuming that the wave motion is free from dispersion and diffusion. In the frame of reference moving with velocity c, a pulse becomes stationary.

We can easily verify that if $u(x,t)$ has the form $u(x-ct)$ then it satisfies the equation (6). We also notice that if u_1 and u_2 are two solutions of (6), then

$$\frac{\partial}{\partial t}(u_1 + u_2) + c\frac{\partial}{\partial x}(u_1 + u_2) = \left\{\frac{\partial u_1}{\partial t} + c\frac{\partial u_1}{\partial x}\right\} + \left\{\frac{\partial u_2}{\partial t} + c\frac{\partial u_2}{\partial x}\right\} = 0$$

i.e. the superposition $u_1 + u_2$ is also a solution of the same equation. This is an example of linear wave propagation and a solution is graphically represented in *Figure 1*.

In a system governed by the equation (6), the velocity of propagation of a point on the pulse is independent of the amplitude u there. However we almost always find in nature that the amplitude of the wave does influence the position of a wavefront. When we translate this statement into the language of mathematics, we find that the equations governing the wave are no longer linear which implies that superposition of two solutions is no longer a solution of the system. We can verify this for the equation (7) below. For this reason, the physical system is called a nonlinear system and the wave is called a nonlinear wave. The dependence of the velocity of propagation on the amplitude brings in what we define to be *genuine nonlinearity* (see *Box*). Genuine nonlinearity is a very subtle concept. It may appear in certain modes but not in others in the same system. In a small amplitude genuinely nonlinear wave in a homogeneous medium, the propagation velocity of a point on the pulseexceeds the constant velocity c by a quantity proportional to the amplitude of the wave. Denoting the distance in the frame of reference moving with velocity c also by the same symbol x and rescaling the amplitude we get the following rule of propagation

$$\left\{\begin{array}{l}\text{the velocity of propagation}\\ \text{of a point on the pulse}\end{array}\right\} = \left\{\begin{array}{l}\text{amplitude } u \text{ of the wave}\\ \text{at the point of the pulse}\end{array}\right\} \tag{7a}$$

The partial differential equation giving such a wave is

$$u_t + u\,u_x = 0. \tag{7b}$$

S D Poisson (1808) was the first to show that the solution of a problem containing this type of nonlinearity can be obtained from an implicit relation. J Challis (1848) observed that the implicit relation may not always be solved uniquely. E E Stokes suggested (in 1848) introduction of discontinuities in solutions of such equations. S Earnshaw (1858) developed the theory of simple waves and B Riemann (1860) introduced the Riemann invariants and laid the foundation of a general theory. S Chandrashekhar worked out in 1942 solution of problems containing weak nonlinearities and it was soon followed by K O Friedrichs. O A Oleinik in Russia and P D Lax in USA developed the modern theory in the early fifties. P D Lax first defined the term *genuine nonlinearity* precisely in 1957. The subject has also become quite abstract as can be seen from the publication of the book *Shock Waves and Reaction Diffusion Equation* in 1983 by J A Smoller. *Shock wave*, an important type of discontinuity in a solution is now a purely abstract mathematical term but has not lost it relevance to application. In the last one and half centuries some of the greatest mathematicians have contributed to this subject. Inspite of all these, there is no end to the open questions and unsolved problems in the subject. There is only one snag, the open questions are difficult to answer and problems are very difficult to solve. Moreover, the solution requires not only a very specialized training in abstract theory but also a feeling for the correct answers from long association with phenomena having genuine nonlinearity.

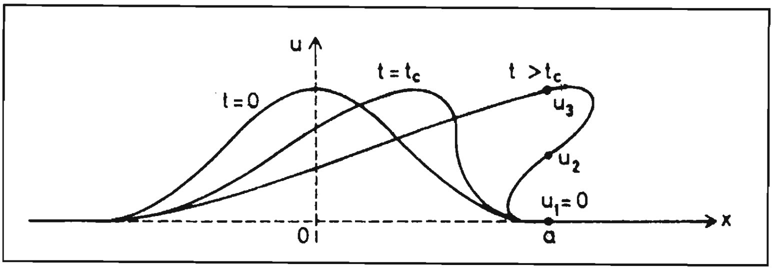

Figure 2 As t increases, the pulse shape of a non-linear wave changes.

A solution of (7b) has been graphically represented in *Figure 2*, where we note that

- since different points of the pulse move with different velocity the pulse now deforms,
- at a critical time t_c the pulse has a vertical tangent at some point,
- after $t > t_c$, the pulse represents the graph of a multivalued function (for example, at $x = a$ it has three values $u_1 = 0, u_2$, u_3) and the physical interpretation fails due to the fact that a physical variable cannot have three values at a point at the same time. In nature, it has been observed that a discontinuity appears in the quantity u immediately after the time t_c, this moving discontinuity at $x = X(t)$ cuts off two lobes of the pulse on either side in such a way that the pulse gives a single valued function (see *Figure 3*).

Our aim in this article is to discuss solutions containing special type of discontinuities, called *shocks*. The solution will belong to a class of functions: a function u from this class is assumed to be

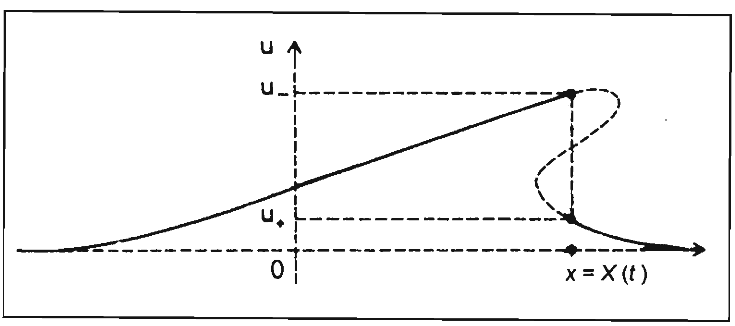

Figure 3 A pulse at a time t with a discontinuity at x=X(t).

smooth everywhere except at these discontinuities with an additional property that the limiting values of u as we approach a discontinuity at $x = X(t)$ from smaller and larger values of x are finite. We denote these limiting values by u_- and u_+ respectively i.e.

$$u_- = \lim_{x \to X(t)-0} u(x,t), \; u_+ = \lim_{x \to X(t)+0} u(x,t).$$

While representing such a piecewise smooth function u, we need not prescribe the value of the function at the point of a discontinuity. In the graph of such a function, the limiting values u_- and u_+ are shown by dots. The two dots are joined by a dotted vertical line showing that the value of the function u jumps from one dot to the next. The dotted vertical line, representing infinitely many values of u at $x = X(t)$ does not form a part of the graph of the function u.

Conservation Law

Consider now a solution containing a discontinuity at a point $X(t)$ in u as shown in *Figure 3*. The law of propagation (7) governing the motion of the different points of the pulse becomes meaningless at the point of discontinuity $x = X(t)$ since u tends to two different values u_- and u_+ as we approach this point from the left and right sides respectively. Hence, there is a need for a new formulation or more precisely a more general formulation which would give the law of propagation of a discontinuity also. It is more than just a mere coincidence that the basic laws of nature are not expressed locally as that in (7) but for a total quantity contained in an interval (or a domain in three dimensions). Such laws are fundamental conservation laws from which local statements are obtained by limiting processes. To give an example of such a conservation law, let u denote the line density of a quantity which evolves due to flux $F(u)$ at point x in the positive x–direction (*Figure 4*). Then the conservation law is

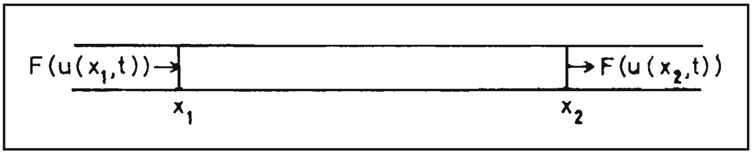

Figure 4 x_1 and x_2 are fixed points where the flux (in positive x–direction) of u are $F(u(x_1, t))$ and $F(u(x_2, t))$ respectively.

$$\left\{\begin{array}{l}\text{Time rate of change of the}\\ \text{total quantity } u \text{ contained in}\\ \text{a spatial interval from } x_1 \text{ to } x_2\end{array}\right\} = \left\{\begin{array}{l}\text{Difference } F(u(x_1,t)) - F(u(x_2,t))\\ \text{of the flux at the two ends of the}\\ \text{interval}\end{array}\right\}.$$

If the flux F vanishes at two fixed points x_1 and x_2 for all time t i.e. $F(u(x_1, t)) = 0 = F(u(x_2, t))$, then the total quantity u contained in the interval from x_1 to x_2 is conserved.

To derive the velocity of propagation S of the discontinuity from the conservation law, let

$u_- =$ a constant value of u behind the discontinuity i.e. for $x < X(t)$

$u_+ =$ another constant value of u ahead of the discontinuity i.e. for $x > X(t)$.

We consider two fixed points x_1 and x_2 $(x_1 < x_2)$ on the two sides of the discontinuity as shown in *Figure 5*.

Then

$$\left\{\begin{array}{l}\text{increase of } u \text{ in the interval}\\ (x_1, x_2) \text{ in time } t_2 - t_1\end{array}\right\} = (u_- - u_+).\{S(t_2 - t_1)\}$$

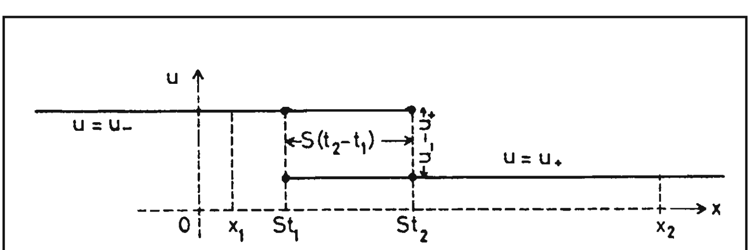

Figure 5 A discontinuity at X=St separates two uniform states u_- and u_+. An increase in u in the interval (x_1, x_2) in time $t_2 - t_1$ is due to change in u from u_+ to u_- in the spatial interval from St_1 to St_2.

which must be equal to

increase of u due to the flux $F=\{F(u_-)-F(u_+)\}\ \ (t_2-t_1)$.

Therefore

$$S=\frac{F(u_-)-F(u_+)}{u_- - u_+} \tag{9}$$

Different conservation laws i.e. different choices of the line density and the flux lead to different laws of propagation of a discontinuity. For a special choice of $F(u)$ = $(1/2)\,u^2$, the conservation law becomes

Different conservation laws i.e. different choices of the line density and the flux lead to different laws of propagation of a discontinuity.

$$\left\{\begin{array}{l}\text{The time rate of change of}\\ \text{the total quantity } u \text{ in the}\\ \text{interval from } x_1 \text{ to } x_2\end{array}\right\}=\left\{\begin{array}{l}\text{Difference of the fluxes}\\ \text{at } x_1 \text{ and } x_2 \text{ i.e.}\\ \frac{1}{2}u^2(x_1,t)-\frac{1}{2}u^2(x_2,t)\end{array}\right\} \tag{10a}$$

and (9) becomes

$$S=\frac{1}{2}(u_+ + u_-). \tag{10b}$$

We can use (9) to deduce the local law of propagation c of a point of continuity on a pulse. This is obtained by taking the strength $u_- - u_+$ of the discontinuity to tend to zero i.e. by taking $u_- \to u$ and $u_+ \to u$

$$c=\lim_{\substack{u_+\to u\\ u_-\to u}} S=\frac{dF}{du} \tag{11}$$

For a special choice of F = $(1/2)\,u^2$

$$c=\lim_{u_+,u_-\to u}\frac{1}{2}(u_+ + u_-)=u$$

which is the law of propagation stated in (7). Therefore, we have proved a theorem.

Theorem 1. A continuous solution of the conservation law (10 a) is a solution as per the rule (7).

The rule (7a) and the equation (7b) are not exactly the same. In order that a pulse satisfies (7a), it is enough if it is continuous but in order that a function, whose graph represents a pulse, satisfies (7b) it is necessary that the two partial derivations u_t and u_x exist and are continuous. We shall not go into such mathematical rigour. Instead, we concentrate on a smooth part of the graph and take fixed points x_1 and x_2. Then the time rate of change of the total quantity u contained in spatial interval from x_1 to x_2

$$= \frac{d}{dt}\int_{x_1}^{x_2} u(\xi,t)\,d\xi = \int_{x_1}^{x_2} u_t(\xi,t)\,d\xi$$

$$= -\frac{1}{2}\int_{x_1}^{x_2} \{u^2(\xi,t)\}_\xi\,d\xi \quad (\text{since } u_t = -uu_x = -\frac{1}{2}(u^2)_x)$$

$$= \frac{1}{2}u^2(x_1,t) - \frac{1}{2}u^2(x_2,t)$$

which shows that the conservation law (10a) is valid. Thus we have shown that the converse of the theorem 1 is also true.

Theorem 2 Every smooth solution of the equation (7b) satisfies the conservation law (10a).

It requires only a little more mathematical argument to show that this theorem remains true if *smooth solution* is replaced by smooth solution except for a finite number of discontinuities of u_t and u_x in the interval x_1 to x_2 provided u is continuous at these points of discontinuities.

In general, there is more than one conservation law (i.e. there is more than one choice of the line density and the corresponding flux F) which give the same law of propagation of a point on a continuous wave profile. For example we choose the line density to be $\rho(u) = u^2$ and the flux to be $F(u) = (2/3)\,u^3$. Then

$$S = \frac{\frac{2}{3}u_-^3 - \frac{2}{3}u_+^3}{u_-^2 - u_+^2} = \frac{2}{3}\,\frac{u_-^2 + u_-u_+ + u_+^2}{u_+ + u_-} \tag{12}$$

which leads to $c = u$ as $u_+, u_- \to u$.

As we have pointed out earlier, conservation laws are more fundamental, that is logically primitive and corresponding local statements like (7) and (12) can be derived from these. Choice of an appropriate conservation law for any system comes from physical considerations.

In the next part we will consider an example of a continuous solution with discontinuous initial data. We will also discuss stability considerations and some interesting examples.

Address for Correspondence
Phoolan Prasad
Department of Mathematics
Indian Institute of Science
Bangalore 560 012, India.
email:prasad@math.iisc.ernet.in
Fax: (080) 3341683

Suggested Reading

- **R Courant and K O Friedrichs. *Supersonic Flow and Shock Waves*. Interscience. reprinted by Springer-Verlag, 1948.**
- **J Smoller. *Shock Waves and Reaction-Diffusion Equations*. Springer-Verlag, 1983.**
- **Phoolan Prasad and Renuka Ravindran. *Partial Differential Equations*. John Wiley and Wiley Eastern, 1985.**

Robert H Goddard (1882–1945)

Wernher von Braun (1912–1977)

The first liquid fuel rocket was tested by Goddard in 1926 and V–2 rocket by Braun in 1942.

From the stamp collection of R G Sangoram