WAVE MAPS

TERENCE TAO

1. Wave maps

Let $c > 0$ be a parameter (the "speed of light"). The wave equation

$$\frac{\partial^2 u}{\partial t^2}(t,x) = c^2 \sum_{j=1}^n \frac{\partial^2 u}{\partial x_j^2}(t,x)$$

in $\mathbf{R}^n$ describes (at least as a first approximation) the free motion of an n-dimensional surface in an ambient Euclidean space (a particle, a string, the surface of a drum, or the atmosphere are examples of the $n = 0, 1, 2, 3$ cases); meanwhile, the geodesic flow equation

$$\nabla_t \frac{du}{dt}(t) = 0 \tag{1}$$

(the notation will be explained later), describes the free motion of a point in a space that is not necessarily Euclidean (more technically, it describes motion of a point in a Riemannian manifold; this manifold may be itself contained in a higher-dimensional Euclidean space, but may also be considered abstractly as an independent object). The wave map equation

$$\nabla_t \frac{\partial u}{\partial t}(t,x) = c^2 \sum_{j=1}^n \nabla_{x_j} \frac{\partial u}{\partial x_j}(t,x) \tag{2}$$

is the natural combination of these equations, describing the free motion of an n-dimensional surface in a non-Euclidean space; for instance, the motion of a string that is constrained to lie on a sphere would be given (at least to first approximation) as a wave map. As such, wave maps are one of the fundamental equations used to describe geometric motion, although they are not as well understood (and have fewer applications to geometry at present) than other geometric flows such as the Ricci flow or mean curvature flow. One contrast between the wave maps equations and these other flows is that the wave maps equation is time-reversible and non-dissipative; a surface never loses or gains any "energy" under this equation. In contrast, the Ricci and mean curvature flows are dissipative (or *parabolic*), which means that as time progresses, the surfaces should become smoother and less energetic (although singularities are still possible, and of tremendous importance to geometry). A major difficulty in understanding these equations is that they are *non-linear*; this may not be so apparent in the formulation (2) but is hidden in the notation for the covariant derivatives ∇_t, ∇_{x_j}, which we will come to later. Wave maps are also closely related to *harmonic maps*, which have played an important role in geometry, topology, and even integrable systems; it is thus conceivable that wave maps may eventually also be a useful tool for these fields. Wave maps also arise as a simplified model for Einstein's equations of general relativity.

Before we describe the wave maps equation further, let us first review the geodesic flow equation (1). This equation not only describes the free motion of particles in a Riemannian manifold M, but also describes (or more precisely, it parametrizes) *geodesics*. The precise notion of a Riemannian manifold is not necessary for our discussion here, but suffice to say that it comes with an object called a *Riemannian metric* g, which among other things allows us to assign a length[1] $|v|_{g(x)} \geq 0$ to any tangent vector $v \in T_x M$ at a point $x \in M$. As a consequence, given any continuously differentiable parameterized curve $\gamma : [0,1] \to M$ in the manifold, it is possible to define a notion of *length* $|\gamma|_g$ of that curve as

$$|\gamma|_g = \int_0^1 |\gamma'(t)|_{g(\gamma(t))}\, dt$$

since for each time t, $\gamma'(t)$ is a tangent vector to M at $\gamma(t)$. If M is a connected manifold, we can then define the distance $d(x,y)$ between any two points $x, y \in M$ to be the shortest length required to connect x and y by a curve γ, or more precisely

$$d(x,y) := \inf\{|\gamma|_g : \gamma : [0,1] \to M \text{ is continuously differentiable}, \gamma(0) = x, \gamma(1) = y\}.$$

It can be shown that given any two points $x, y \in M$ whose separation $d(x,y)$ is sufficiently small[2], there is a unique curve γ connecting x and y which actually attains the minimum length, thus $|\gamma|_g = d(x,y)$. Actually, this is technically incorrect, since one can reparameterize this curve γ by any (sufficiently regular) change of variables $t = f(t')$ and still obtain a curve with minimum length. The more precise statement is that there is a unique *equivalence class* of parameterized curves $\gamma : [0,1] \to M$ connecting x to y which attains the minimum length $d(x,y)$, with all curves in this class being equivalent to each other after change of variables. All these parameterized curves thus trace out the same set in M (but at different speeds); this set is then called the *geodesic* between x and y.

[1]For instance, in Euclidean space $\mathbf{R}^3$, the length $|v|_g$ of the tangent vector would be given by Pythagoras's theorem $|v|_{g(x)} = \sqrt{v_x^2 + v_y^2 + v_z^2}$, where v_x, v_y, v_z were the components of v in the three co-ordinate axes. For a more general Riemmanian manifold, the non-negative quadratic expression $v_x^2 + v_y^2 + v_z^2$ would be replaced by another non-negative quadratic form, with coefficients that could depend on what location $x \in M$ the tangent vector is based at.

[2]The situation is more complicated when x and y are far apart, as then there can be multiple geodesics connecting x and y; think of the longitudes connecting the north and south poles on the earth, for instance. Even worse, for x and y very far apart, there exist curves which are locally length-minimizing but not globally length-minimizing. For instance, if one imagines a rope which stretches three-quarters of the way around the equator of the earth, then this is clearly not the shortest way to connect the two endpoints, and one can pull the rope across the northern hemisphere, for instance, to continously decrease the total length. On the other hand, if one takes any segment of this rope which extends less than halfway around the equator, then this segment will lie on the shortest path on the earth that connects the two endpoints of the segment. Because of this, we still classify the curve traced out by this rope as a geodesic, even though it is not globally length-minimizing. We shall avoid these subtleties by only working locally, in situations where the curves are always too short to encounter these issues.

Thus we have given a "variational" description of the geodesic connecting two nearby points x and y, by asking that one minimize the length $|\gamma|_g = \int_0^1 |\gamma'(t)|_{g(\gamma(t)}\, dt$ of the curve. While this is a very intuitive and natural definition of a geodesic, as we saw above it only describes the *set* that the geodesic traces out, and does not determine the choice of parameterization of the curve. As such, it is insufficient to describe the motion of free particles on the manifold, since to describe such motion one has to not only specify the locus of this motion (the set that the particle traverses), but also where the particle is at each specific time t. However, this problem can be fixed by replacing the notion of length $|\gamma|_g$ of a parameterized curve γ with the notion of its *energy* $E(\gamma)$, defined as

$$E(\gamma) := \int_0^1 |\gamma'(t)|^2_{g(\gamma(t))}\, dt.$$

This quantity looks very similar to the length $|\gamma|_g$, differing only by the presence of the square, but is now no longer invariant under change of variables $t = f(t')$. Informally, what the energy does is "penalize" the curve for moving too fast at some times and moving too slow at others, as the presence of the square in the integral will then make that parameterization consume more energy than if the speed $|\gamma'(t)|_{g(\gamma(t))}$ remained constant throughout. This can be formalized by a simple application of the Cauchy-Schwarz inequality, which then also shows that in order to minimize the energy $E(\gamma)$ of a curve connecting two nearby points $x, y \in M$, the curve γ must trace out a geodesic *and* have constant speed throughout (in fact one can easily show that $|\gamma'(t)|_{g(\gamma(t))} = d(x, y)$ for all $0 \leq t \leq 1$).

The theory of calculus of variations tells us that if a curve γ minimizes the energy $E(\gamma)$, then it must obey a certain differential equation (the *Euler-Lagrange equation* for that energy); this equation is then called the *geodesic flow equation.* There are a number of ways to describe this equation. If one was in a Euclidean space $M = \mathbf{R}^n$ (with the standard Euclidean metric given by Pythagoras's theorem), then as is well known, geodesics are just straight lines, and the geodesic flow equation becomes *Newton's first law*

$$\frac{d}{dt}\gamma'(t) = 0.$$

It is tempting to then conjecture that this is also the equation for more general geodesic flows. Unfortunately, the expression $\frac{d}{dt}\gamma'(t)$ does not actually make any sense for general manifolds, for the following reason. Let us write this quantity out more explicitly as

$$\frac{d}{dt}\gamma'(t) = \lim_{h\to 0} \frac{\gamma'(t+h) - \gamma'(t)}{h}.$$

Now $\gamma'(t)$ is the velocity of γ at time t, and is a tangent vector to the manifold at the position $\gamma(t)$, thus $\gamma'(t) \in T_{\gamma(t)}M$. Similarly, $\gamma'(t+h) \in T_{\gamma(t+h)}(M)$. But $\gamma(t)$ and $\gamma(t+h)$ are different points, and so the two tangent spaces $T_{\gamma(t)}M$ and $T_{\gamma(t+h)}M$ are thus almost always *different* vector spaces, and so it is not possible to compute the difference $\gamma'(t+h) - \gamma'(t)$ in a meaningful manner. However, this problem can be rectified with the aid of the metric g. It can be shown that the metric g generates a natural (and uniquely defined, up to errors of at most $O(h^2)$) procedure of *parallel transport*, which allows one to transport (i.e. to identify) vectors in one tangent space such as $T_{\gamma(t)}M$ with vectors in a nearby tangent space $T_{\gamma(t+h)}M$. The formal name for this transport procedure is the *Levi-Civita connection*; it is the unique procedure of this type which is linear, which preserves the lengths of vectors, and is "torsion free" (a concept which is a bit technical to describe here, but roughly speaking asserts that one can use parallel transport to produce "parallelograms" in M of sidelengths h, whose sides only differ from being parallel by at most $O(h^3)$). Using such a procedure, it is then possible to define a notion of a derivative of $\gamma'(t)$, which is referred to as the *covariant derivative* of $\gamma'(t)$ with respect to t and denoted $\nabla_t \gamma'(t)$ instead of $\frac{d}{dt}\gamma'(t)$. One can then show that the Euler-Lagrange equation for minimizing the energy $E(\gamma)$ is then given[3] by the equation

$$\nabla_t \gamma'(t) = 0,$$

which is (1) with u replaced by γ. In other words, Newton's first law still survives when working in Riemannian manifolds instead of Euclidean space, but the catch is that rather than the velocity staying constant, the velocity vector is instead transported by the Levi-Civita parallel transport procedure. Note that the geodesic flow equation is now non-linear because the definition of parallel transport ∇_t depends on the location $\gamma(t)$ of the tangent vector that one is transporting, as well as the direction $\gamma'(t)$ that one is transporting it in.

Now we return to wave maps. Whereas a geodesic is a map $\gamma : [0,1] \to M$ from a one-dimensional space (the time interval $[0,1]$) to a Riemannian manifold M, a wave map is a map $u : \mathbf{R} \times \mathbf{R}^n \to M$ from an $n+1$-dimensional spacetime $\mathbf{R} \times \mathbf{R}^n := \{(t,x) : t \in \mathbf{R}, x \in \mathbf{R}^n\}$ to a Riemannian manifold M; one can thus think of a wave map as describing the time evolution of some n-dimensional object in M. Instead of working with the energy $E(\gamma) = \int_0^1 |\gamma'(t)|^2_{g(\gamma(t))}\, dt$, we form the *action*[4]

$$\mathcal{L}(u) := \int_{\mathbf{R}\times\mathbf{R}^n} |\frac{\partial u}{\partial t}(t,x)|^2_{g(u(t,x))} - c^2 \sum_{j=1}^n |\frac{\partial u}{\partial x_j}(t,x)|^2_{g(u(t,x))}\, dtdx;$$

the first term in this action resembles the energy of the geodesic and roughly resembles a kind of "kinetic energy" of this wave map, while the second term is a

[3] It is possible to avoid the covariant derivative, and provide a more traditional differential equation for geodesic flow by parameterizing the manifold M (locally) by some co-ordinates $x^1, \dots, x^n$, and then expressing the metric in these co-ordinates as $|v|^2_{g(x)} = \sum_{j=1}^n \sum_{k=1}^n g_{jk}(x) x^j x^k$ for some symmetric matrix g_{jk}. The geodesic flow equation then becomes

$$\frac{d^2}{dt^2}\gamma(t)^i = -\sum_{j=1}^n \sum_{k=1}^n \Gamma^i_{jk}(\gamma(t))(\frac{d}{dt}\gamma^j(t))(\frac{d}{dt}\gamma^k(t)) \text{ for } i = 1, \dots, n$$

where the *Christoffel symbols* Γ^i_{jk} are given by the formula

$$\Gamma^i_{jk} := \frac{1}{2} \sum_{l=1}^n g^{il}(\frac{\partial g_{jl}}{\partial x_k} + \frac{\partial g_{kl}}{\partial x_j} - \frac{\partial g_{jk}}{\partial x_l})$$

and g^{il} is the inverse matrix to g_{il}. As one can see, the equations certainly become more complicated (though more explicit) when written in concrete co-ordinates!

[4] Strictly speaking, this integral is likely to be divergent. However, just as we localized the time variable in a geodesic to a compact interval such as $[0,1]$, we can localize space and time to a bounded domain in order to make this action well-defined, so that we can derive the equation (2). But then (2) makes sense globally, so we can continue to adopt (2) as the definition of a wave map even when the action no longer makes sense; this is somewhat analogous to how one still retains a notion of geodesic over long distances even when these geodesics are no longer global length minimizers.

"potential energy". The two turn out to stay roughly in balance, which means that the wave map tends to propagate at the speed of light c. Instead of requiring that u be a minimizer of this action, as we did for geodesics, we ask instead for the slightly weaker condition that u is a *critical point* of the action, which means that

$$\frac{d}{d\varepsilon}\mathcal{L}(u_\varepsilon)|_{\varepsilon=0} = 0$$

whenever $u_\varepsilon : \mathbf{R} \to \mathbf{R}^n \to M$ is a continuously differentiable family of maps for each ε in a small neighborhood $[0, \varepsilon_0)$ of zero such that $u_0 = 0$. (The relationship between the two concepts is analogous to the relationship in single-variable calculus between a continuously differentiable function $f(x)$ attaining a minimum at some point x_0, versus merely having zero derivative $f'(x_0) = 0$ at that point.) It is possible to show (at least for u smooth, and if one localizes the integral defining the action to a bounded set) that the property of being a critical point is equivalent to solving the equation (2).

Wave maps share many features in common with the wave equation, which can be viewed as wave maps specialized to the case when the manifold M is Euclidean. For instance, wave maps have a conserved energy: the quantity

$$\int_{\mathbf{R}^n} \frac{1}{2}|\frac{\partial u}{\partial t}|^2(t,x) + c^2\frac{1}{2}\sum_{j=1}^n |\frac{\partial u}{\partial x_j}|^2(t,x)\ dx$$

is constant in time. Also, there is a finite speed of propagation property: the value of a wave map $u(t_0, x_0)$ at some future time $t_0 > 0$ is completely determined by the initial position $u(0,x)$ and initial velocity $\frac{\partial u}{\partial t}(0,x)$ in the ball $\{x \in \mathbf{R}^n : |x - x_0| \leq ct_0\}$; thus changing the initial conditions of the wave map at a distance further than t_0 from x_0 cannot influence the value of the wave map at t_0, x_0 (or more succinctly, wave maps cannot transmit information faster than the speed of light c). There are still however major gaps in our understanding of wave maps, notably in knowing when wave maps can form singularities, and if so what types of singularities can occur. In two dimensions, for instance, it is known that wave maps cannot form singularities from smooth initial data if the energy is sufficiently small, and for large energies it is conjectured that singularities can only occur if the wave map begins to converge (after renormalization) to a harmonic map; but this conjecture is still mostly unresolved at present.

Department of Mathematics, UCLA, Los Angeles CA 90095-1555

E-mail address: tao@math.ucla.edu