The Incredible Shrinking Manifold

Basic Differential Geometry from the Synthetic Standpoint, with some Remarks on Spacetime

John L. Bell

Traditionally, there have been two methods of deriving the theorems of geometry: the *analytic* and the *synthetic.* While the analytical method is based on the introduction of numerical coordinates, and so on the theory of real numbers, the idea behind the synthetic approach is to furnish the subject of geometry with a purely geometric foundation in which the theorems are then deduced by purely logical means from an initial body of postulates.

The most familiar examples of the synthetic geometry are classical Euclidean geometry and the synthetic projective geometry introduced by Desargues in the 17th century and revived and developed by Carnot, Poncelet, Steiner and others during the 19th century.

The power of analytic geometry derives very largely from the fact that it permits the methods of the calculus, and, more generally, of mathematical analysis, to be introduced into geometry, leading in particular to *differential geometry* (a term, by the way, introduced in 1894 by the Italian geometer *Luigi Bianchi*). That being the case, the idea of a "synthetic" differential geometry seems elusive: how can differential geometry be placed on a "purely geometric" or "axiomatic" foundation when the apparatus of the calculus seems inextricably involved?

To my knowledge there have been two attempts to develop a synthetic differential geometry. The first was initiated by Herbert Busemann in the 1940s, building on earlier work of Paul Finsler. Here the idea was to build a differential geometry that, in its author's words, "requires no derivatives": the basic objects in Busemann's approach are not differentiable manifolds, but metric spaces of a certain type in which the notion of a geodesic can be defined in an intrinsic manner. I shall not have anything more to say about this approach.

The second approach, that with which I shall be concerned here, was originally proposed in the 1960s by F. W. Lawvere, who was in fact striving to fashion a decisive axiomatic framework for continuum mechanics. His ideas have led to what I shall simply call *synthetic differential geometry* (SDG) (sometimes called *smooth infinitesimal analysis*). SDG is formulated within *category theory,* the branch of mathematics created in 1945 by Eilenberg and Mac Lane which deals with mathematical form and structure in its most general manifestations. As in biology, the viewpoint of category theory is that mathematical structures fall naturally into species or *categories*. But a category is specified not just by identifying the species of mathematical structure which constitute its *objects*; one must also specify the transformations or *maps* linking these objects. Thus one has, for example, the category **Set** with objects all sets and maps all functions between sets; the category **Grp** with objects all groups and maps all group homomorphisms; the category **Top** with objects all topological spaces and maps all continuous functions; and **Man**, with objects all (Hausdorff, second countable) smooth manifolds and maps all smooth functions. Since differential geometry "lives" in **Man**, it might be supposed that in formulating a "synthetic differential geometry" the category-theorist's goal would be to find an axiomatic description of **Man** itself.

But in fact the category **Man** has a couple of "deficiencies" which make it unsuitable as the object of axiomatic description:

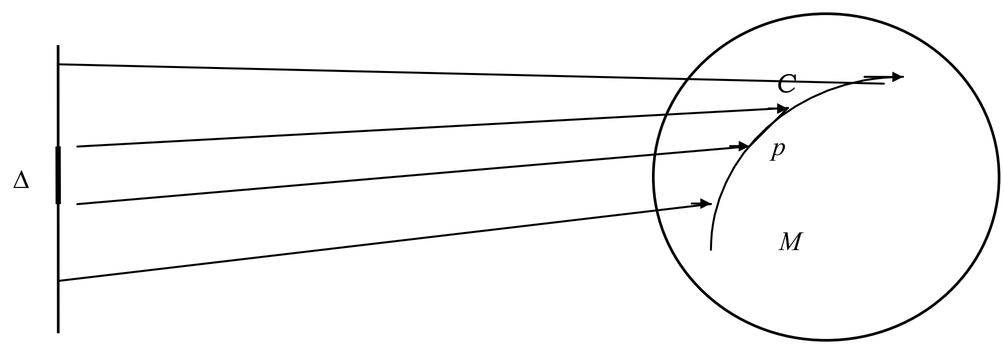
1. It lacks exponentials: that is, the "space of all smooth maps" from one manifold to another in general fails to be a manifold. And even if it did—
2. It also lacks "infinitesimal objects"; in particular, there is no "infinitesimal" or *incredible shrinking manifold* Δ for which the tangent bundle *TM* of an arbitrary manifold *M* can be identified as the exponential "manifold" M^Δ of all "infinitesimal paths" in *M.* (It may be remarked parenthetically that it is this deficiency that makes the construction of the tangent bundle in **Man** something of a headache.)

Lawvere's idea was to enlarge **Man** to a category **S**—a category of so-called *smooth spaces* or a *smooth category*—which avoids these two deficiencies, admits a simple axiomatic description, and at the same time is sufficiently similar to **Set** for mathematical construction and calculation to take place in the familiar way.

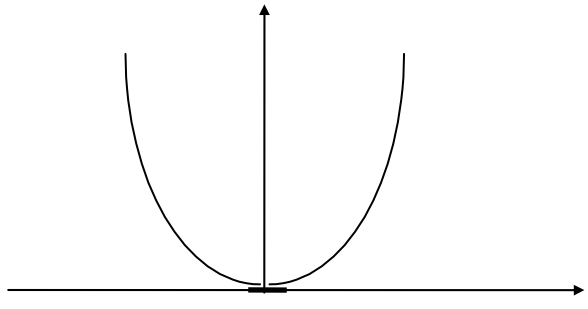
The essential features of a smooth category **S** are these:

- In enlarging **Man** to **S** no "new" maps between manifolds are added, that is, all maps in **S** between objects of **Man** are smooth. (Notice that this is not the case when **Man** is enlarged to **Set.**)
- **S** is *Cartesian closed*, that is, contains products and exponentials of its objects in the appropriate sense.
- **S** satisfies the *principle of microstraightness.* Let $\mathbb{R}$ be the real line considered as a object of **Man,** and hence also of **S**. Then there is a nondegenerate segment Δ of $\mathbb{R}$ around 0 which remains *straight* and *unbroken* under any map in **S.** In other words, Δ is subject in **S** to *Euclidean motions only.*

Δ may be thought of as a *generic tangent vector.* For consider any curve *C* in a space *M*—that is, the image of a segment of $\mathbb{R}$ (containing Δ) under a map *f* into *M.* Then the image of Δ under *f* may considered as a short straight line segment lying along *C* around the point $p = f(0)$ of *C.*



In fact, by considering the curve in $\mathbb{R} \times \mathbb{R}$ given by $f(x) = x^2$, we see that Δ is the intersection of the curve $y = x^2$ with the *x*-axis:



That is,

$$\Delta = \{x \in \mathbb{R} : x^2 = 0\}.$$

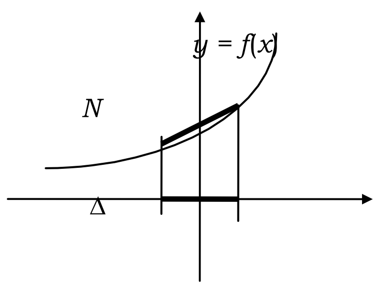
Thus Δ consists of *nilsquare infinitesimals,* or *microquantities.* We use the letter ε to denote an arbitrary microquantity.

Now classically Δ coincides with {0}, but a precise version of the principle of microstraightness—the *Principle of Microtaffineness*—ensures that this is not the case in **S.** The principle states that

- in **S**, any map $f: \Delta \to \mathbb{R}$ is (uniquely) *affine*, that is, for some *unique* $b \in \mathbb{R}$, we have, for all ε,

$$f(\varepsilon) = f(0) + b\varepsilon.$$

Here b is the *slope* of the segment N in the diagram:



Thus the principle of microaffineness asserts that each map $\Delta \to \mathbb{R}$ has a *unique* slope. This reduces the development of the *differential calculus* to simple algebra.

The principle of microaffineness asserts also that the map $\mathbb{R}^\Delta \to \mathbb{R} \times \mathbb{R}$ which assigns to each $f \in \mathbb{R}^\Delta$ the pair ($f(0)$, slope of f) is an isomorphism:

$$\mathbb{R}^\Delta \cong \mathbb{R} \times \mathbb{R}.$$

Since $\mathbb{R}^\Delta$ is the *tangent bundle* of $\mathbb{R}$, so is $\mathbb{R}^\Delta$.

This suggests that, for any space M in **S**, we take the tangent bundle TM of M to be the exponential M^Δ. Elements of M^Δ are called *tangent vectors* to M. Thus a *tangent vector* to M at a point $p \in M$ is just a map $t: \Delta \to M$ with $t(0) = x$ That is, a tangent vector at p is a *micropath in M with base point p*. The *base point map* $\pi: TM \to M$ is defined by $\pi(t) = t(0)$. For $p \in M$, the fibre $\pi^{-1}(p) = T_pM$ is the *tangent space to M at p.*

Observe that, if we identify each tangent vector with its image in *M,* then *each tangent space to M may be regarded as lying in M.* In this sense each space in **S** is "infinitesimally flat".

We check the compatibility of this definition of *TM* with the usual one in the case of Euclidean spaces:

$$T(R^n) = (R^n)^\Delta \cong (R^\Delta)^n \cong (R \times R)^n \cong R^n \times R^n.$$

The assignment $M \mapsto TM$ can be turned into a functor in the natural way—the *tangent bundle functor.* (For $f: M \to N$, $Tf: TM \to TN$ is defined by $(Tf)t = f \circ t$ for $t \in TM$.)

The whole point of synthetic differential geometry is to render *the tangent bundle functor representable*: *TM* becomes identified with the space of all maps from some fixed object—in this case Δ)—to *M.* (Classically, this is impossible.) This in turn simplifies a number of fundamental definitions in differential geometry.

For instance, a *vector field* on a space *M* is an assignment of a tangent vector to *M* at each point in it, that is, a map $\xi: M \to TM = M^\Delta$ such that $\xi(x)(0) = x$ for all $x \in M$. This means that $\pi \circ \xi$ is the identity on *M*, so that a *vector field is a section of the base point map*.

A *differential k- form* ((0, *k*) tensor field) on *M* may be considered as a map $M^{\Delta^n} \to \mathbb{R}$.

Recall the condition that **S** be Cartesian closed. This means that for any pair *S,T* of spaces in **S**, **S** also contains their *product* $S \times T$ and their *exponential* T^S, the space of all (smooth) maps $S \to T$. These are connected in the following way: for any spaces *S, T, U*, there is a natural bijection of maps

$$\frac{S \to T^U}{S \times U \to T}$$

In the usual function-argument notation, this bijection is given by:

$$(f: S \times U \to T) \mapsto (f^\wedge: S \to T^U) \quad \text{with} \quad f^\wedge(s)(u) = f(s, u) \text{ for } s \in S,\ u \in U.$$

This gives rise to a bijective correspondence between vector fields on *M* and what we shall call *microflows* on *M:*

$$\frac{\xi: M \to M^\Delta}{\xi^\wedge: M \times \Delta \to M} \quad \begin{matrix}\text{(vector fields on } M)\\ \text{(microflows on } M),\end{matrix}$$

with

$$\xi^\wedge(x,\varepsilon) = \xi(x)(\varepsilon).$$

Notice that then $\xi^\wedge(x,0) = x$.

We also get, in turn, a bijective correspondence between microflows on *M* and *micropaths in M^M with the identity map as base point:*

$$\frac{\xi^\wedge: M \times \Delta \to M}{\xi^*: \Delta \to M^M} \quad \begin{matrix}\text{(microflows on } M)\\ \text{(micropaths in } M^M),\end{matrix}$$

with

$$\xi^*(\varepsilon)(x) = \xi^\wedge(x,\varepsilon) = \xi(x)(\varepsilon).$$

Thus, in particular,

$$\xi^*(0)(x) = \xi(x)(0) = x,$$

so that $\xi^*(0)$ is the identity map on M. Each $\xi^*(\varepsilon)$ is a microtransformation of M into itself which is "very close" to the identity map.

Accordingly, in **S**, *vector fields, microflows, and micropaths are equivalent.* Classically, this is a metaphor at best.

The notions of affine connection, geodesic, and the whole apparatus of Riemannian geometry can also be developed within *SDG*, as has been shown by Bunge, Kock and Reyes. Guts and Grinkevich have shown how Einstein's field equations can be formulated within *SDG*, resulting in a synthetic theory of relativity.

In a spacetime the metric can be written in the form

$$(*) \qquad ds^2 = \Sigma g_{\mu\nu} dx_\mu dx_\nu \qquad \mu,\nu = 1,2,3,4.$$

In the classical setting (*) is in fact an abbreviation for an equation involving derivatives and the "differentials" ds and dx_μ are not really quantities at all. What form does this equation take in SDG? Notice that the "differentials" cannot be taken as nilsquare infinitesimals since all the squared terms would vanish. But the equation does have a very natural form in terms of nilsquare infinitesimals. Here is an informal way of obtaining it.

We think of the dx_μ as being multiples $k_\mu e$ of some small quantity e. Then (*) becomes

$$ds^2 = e^2\Sigma g_{\mu\nu} k_\mu k_\nu,$$

so that

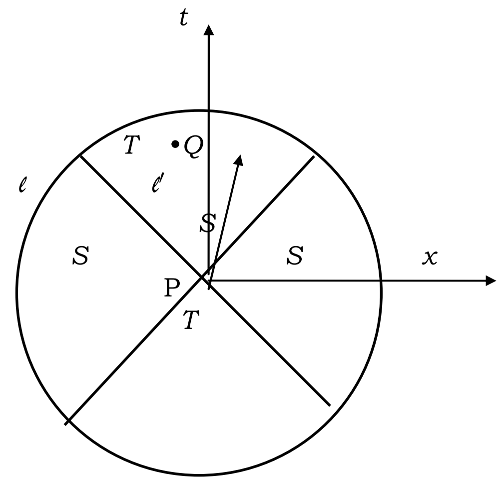
$$ds = e[\Sigma g_{\mu\nu} k_\mu k_\nu]^{1/2}$$

Now replace e by a nilsquare infinitesimal ε. Then we obtain the metric relation in SDG:

$$ds = \varepsilon[\Sigma g_{\mu\nu} k_\mu k_\nu]^{1/2}.$$

This tells us that the "infinitesimal distance" ds between a point P with coordinates (x_1, x_2, x_3, x_4) and an infinitesimally near point Q with coordinates $(x_1 + k_1\varepsilon, x_2 + k_2\varepsilon, x_3 + k_3\varepsilon, x_4 + k_4\varepsilon)$ is $\varepsilon[\Sigma g_{\mu\nu} k_\mu k_\nu]^{1/2}$. Here a curious situation arises. For when the "infinitesimal interval" ds between P and Q is timelike (or lightlike), the quantity $\Sigma g_{\mu\nu} k_\mu k_\nu$ is nonnegative, so that its square root is a real number. In this case ds may be written as εd, where d is a real number. On the other hand, if ds is spacelike, then $\Sigma g_{\mu\nu} k_\mu k_\nu$ is negative, so that its square root is imaginary. In this case, then, ds assumes the form $i\varepsilon d$, where d is a real number (and, of course $i = \sqrt{-1}$). On comparing these we see that, if we take ε as the "infinitesimal unit" for measuring infinitesimal timelike distances, then $i\varepsilon$ serves as the "imaginary infinitesimal unit" for measuring infinitesimal spacelike distances.

For purposes of illustration, let us restrict the spacetime to two dimensions (x, t), and assume that the metric takes the simple form $ds^2 = dt^2 - dx^2$. The infinitesimal light cone at a point P divides the infinitesimal neighbourhood at P into a timelike region T and a spacelike region S,



bounded by the null lines ℓ and ℓ′ respectively. If we take P as origin of coordinates, a typical point Q in this neighbourhood will have coordinates ($a\varepsilon$, $b\varepsilon$) with a and b real numbers: if $|b| > |a|$, Q lies in T; if $a = b$, P lies on ℓ or ℓ′; if $|a| < |b|$, P lies in S. If we write $d = |a^2 - b^2|^{1/2}$, then in the first case, the infinitesimal distance between P and Q is εd, in the second, it is 0, and in the third it is $i\varepsilon d$.

Minkowski introduced "i*ct*" to replace the "*t*" coordinate so as to make the metric of relativistic spacetime positive definite. This was, despite its daring, purely a matter of formal convenience, and was later rejected by (general) relativists (see, for example Box 2.1, *Farewell to* "i*ct*", of Misner, Thorne and Wheeler *Gravitation* [1973]). In conventional physics one never works with nilpotent quantities so it is always possible to replace formal imaginaries by their (negative) squares. But spacetime theory in SDG *forces* one to use imaginary units, since, infinitesimally, one can't "square oneself out of trouble". This being the case, it would seem that, infinitesimally, Wheeler *et al.'s* dictum needs to be replaced by

Vale "i*c*(*t*)", *ave* "iε" !

To quote once again from Misner, Thorne and Wheeler's massive work,

> *Another danger in curved spacetime is the temptation to regard ... the tangent space as lying in spacetime itself. This practice can be useful for heuristic purposes, but is incompatible with complete mathematical precision.*

The consistency of synthetic differential geometry shows that, on the contrary, yielding to this temptation *is* compatible with complete mathematical precision: there tangent spaces may indeed be regarded as lying in spacetime itself. If (as Hilbert said) set theory is "Cantor's paradise" then I would submit that SDG is nothing less than "Riemann's paradise"!