On the Category-Theoretic Approach to Differential Geometry

Federico G. Lastaria

Abstract

The purpose of this paper is to outline the program - formulated by F.W. Lawvere in 1967 - of giving mathematical foundations for continuum mechanics, differential geometry and calculus of variations in category-theoretic terms. Among the remarkable features of this approach are the unification of the finite and the infinite-dimensional cases, and the rigorous systematic use of nilpotent infinitesimal quantities, widely used in mathematics at least until the end of the 19th century and extensively used still now, at least in heuristic form, by physicists and engineers. Suitable axioms allow to reduce differential calculus to simple algebra, fully combining rigor with intuition. The importance of such an approach in the teaching of calculus deserves the utmost attention. The construction of models for categories of smooth spaces is finally sketched.

Introduction

The category-theoretic approach to continuum mechanics, differential geometry and the calculus of variations was first formulated by F.W. Lawvere in a series of lectures in Chicago in 1967. In the sequel, such a theory will be referred to with the well-established phrase *Synthetic Differential Geometry*, though *Categorical Dynamics* - the title of Lawvere's 1967 Chicago lectures - might be more appropriate.

In general terms, the issue is to give an axiomatic theory of continuous bodies and continuous variations: the very notion of smoothness and the decisive conceptual constructions that can be performed on continuous bodies are described as ways in which bodies are transformed or mapped into each other. Otherwise stated, fundamental concepts and constructions are described mathematically through consideration of morphisms in suitable categories, namely cartesian closed categories satisfying appropriate axioms, which are chosen to model objective properties of the physical world. Category theory provides the indispensable conceptual tools for such an approach, as well for the construction of suitable models.

The paper is organized as follows. In the first section it is explained what cartesian closed categories are and why they are required, along with infinitesimal nilpotent quantities, for a mathematical description of the physical world. The second section briefly shows

[0]Proceedings of the Third International Palestinian Conference on Mathematics and Mathematics Education, Bethlehem University, 9-12 August 2000.

how elementary calculus may be developed in a purely algebraic way, by a systematic use, on a rigorous basis, of nilpotent infinitesimal quantities. Finally section 3 accounts for consistency of the whole theory, sketching the construction of models for Synthetic Differential Geometry.

This paper is an invitation to the topic, meant to provide the reader with some motivations to go through the specialized literature cited in the bibliography. In particular Lawvere's paper [7], his Introduction in [8], Kock's fundamental book [4], Bell's booklet [1], [6] and [15] by Moerdijk and Reyes, from each of which the present paper has largely drawn, are suggested as further readings. For an outline of more recent developments of Synthetic Differential Geometry, see [11] and the references therein.

It must be finally remarked explicitly, to avoid confusion, that Lawvere's categorical approach to calculus, outlined in this note, completely differs from A. Robinson's *'Non-standard analysis'* (1966), both for conceptual origins and for technical developments.

Acknowledgments. My thanks go to Anders Kock for his corrections and for his constructive criticisms, and to Bill Lawvere for his interest in this work and for his kindness and patience in explaining me the fundamental philosophical ideas underlying the theory. Of course, I insist on retaining sole responsibility, at least for any errors or misconceptions should still remain.

1 Why Cartesian Closed Categories and Infinitesimal Quantities?

The category **Man** of finite-dimensional smooth manifolds is not a convenient setting for the study of continuum physics, differential geometry and the calculus of variations - more generally, for the study of quantities which undergo smooth variations - because it lacks good categorical properties, such as the existence of all finite limits and the existence of exponentials. To begin with, the category **Man** lacks the property of being closed with respect to the fundamental construction of function spaces, or exponentials. Roughly speaking, the space of smooth maps between two smooth manifolds is not a smooth manifold. More precisely, **Man** is not cartesian closed, according to the terminology that we explain below. To illustrate the importance of such a concept, consider the following constructions, of fundamental importance in continuum physics ([7],[9]).

In order to describe the motion of bodies in space, we need to consider a category **E**. A time interval, a body and the ordinary flat space are modelled as objects of **E**, say T, B and E respectively. For any pair of objects X, Y, denote by Y^X the set of morphisms from X to Y. A particular motion of the body may be described in any of the following three ways:

- as a map $B \times T \longrightarrow E$, that to each pair (*particle*, *instant*) associates a position in space;
- as a path $T \longrightarrow E^B$ in the map space E^B, where the latter is the space of all possible

placements of B in the space E;

- as a placement $B \longrightarrow E^T$ of B in the path space E^T, the map space of all paths of E. Here one assigns to every particle of B its path through E.

For the purposes of continuum physics, it is necessary that these three descriptions be all available and equivalent, because each of them has its own range of applications. Here are some examples.

1. Let $E \longrightarrow R$ be a numerical function giving, for instance, the distance from a fixed reference point P_0 of E. Composition with the map $B \times T \longrightarrow E$ gives the distance from P_0 of each particle at each instant during the motion.

2. The *center of mass* of a body depends only on the way the body is placed in space, and is therefore a map $E^B \longrightarrow E$. By composing this map with the motion described as a map $T \longrightarrow E^B$, one describes the motion $T \longrightarrow E$ of the center of mass of the body.

3. Let V be the object of velocities of E, i.e. the vector space of translations of the flat space E. Then differentiation gives a morphism $E^T \longrightarrow V^T$ which associates a velocity path to every space path. Composing the latter with the motion of a body in the form $B \longrightarrow E^T$, one obtains the velocity path of the particles of the body B as a map $B \longrightarrow V^T$.

The existence and equivalence of the different ways of representing a motion described above (for any three objects in **E**) are ensured by the property of cartesian closedness, which defines the map-space construction as an adjoint functor.

A category **E** is called *cartesian closed* if it has finite cartesian products (including an empty product 1, the final object) and has exponential, right adjoint to product. This means that for any two objects A, Y there is a an object Y^A such that for any object X there is a natural bijection

$$\frac{A \times X \longrightarrow Y}{X \longrightarrow Y^A}$$

between the indicated maps. For every pair of objects A, Y, the points of the map object Y^A are (identified with) the maps from A to Y.

Another limitation of the category of manifolds is the absence of ways of dealing with infinitesimal quantities. Reasoning with infinitesimals - more generally, synthetic reasoning - has been a fundamental issue in studying and teaching mathematics, physics and engineering at least until the end of the nineteenth century, when it was discredited and slandered as non-rigorous by the process of arithmetization. As A. Weil points out: *"Depuis les premieres tâtonnements du calcul infinitésimal et de la géométrie différentielle, les géomètres ont pensé "points infinement voisins" (ou "points proches"; le terme latin est "proximus"), quitte a rédiger tout autrement lorsque cette notion fut tombée en discrédit: il y a là un chapitre d'histoire qui mériterait d'être écrit."* ([17].)

Lawvere's proposal is to work in cartesian closed categories in which the real line R is (no longer a field but) a commutative ring with non-zero nilpotents. Through an appropriate axiom, that will be stated in the next section, the powerful intuitive ways of thinking and computing in terms of infinitesimals are thus recovered and systematically used in a rigorous setting.

The consideration of cartesian closed categories finally allows the unification of the finite and the infinite-dimensional cases. For instance, along with any pair of (finite-dimensional) spaces X, Y, the (infinite-dimensional) map-space Y^X is available inside the category. This issue is already explicit in 1854 B. Riemann's lecture "*On the hypotheses which lie at the foundations of geometry*", where it is pointed out that there are also "*manifolds in which the determination of position requires not a finite number, but either an infinite sequence or a continuous manifold of determinations of quantities.*" [1]

2 Calculus in a Cartesian Closed Category of Smooth Objects

Differential and integral calculus, in one or several variables, may be developed in category-theoretic terms. We refer the reader to [1] and [4] for a thorough treatment and limit ourselves to some basic principles and examples.

We assume to work in a cartesian closed category $\mathbf{E}$. We further assume that in $\mathbf{E}$ there is a commutative ring object R, referred to as the *real line*. The object D of nilpotent infinitesimals of first order is definable in terms of the smooth line R as [2]

$$D = \{d \in R \mid d^2 = 0\}$$

Geometrically, in the plane $R \times R$, D may be thought of as the intersection of the circle $x^2 + (y-1)^2 = 1$ with the tangent line $y = 0$.

The fundamental axiom which allows to develop calculus is the following.

Axiom 1(Kock-Lawvere) *For any $f : D \longrightarrow R$ there exist unique $a, b \in R$ such that*

$$\forall d \in D \qquad f(d) = a + b \cdot d$$

We can phrase the axiom by saying that:

D is so small that all functions $f : D \longrightarrow R$ are affine,

[1] *"Es gibt indess auch Mannigfaltigkeiten, in welchen die Ortsbestimmung nicht eine endliche Zahl, sondern entweder eine unendliche Reihe oder eine stetige Mannigfaltigkeit von Grössenbestimmungen erfordert."* [16]. (Quotation in [15].)

[2] Actually D is not a set, but an object of $\mathbf{E}$, namely the equalizer of the pair of maps $(-)^2 : R \longrightarrow R$ and $0 : R \longrightarrow R$. Such an equalizer exists, because $\mathbf{E}$ has all finite limits. One can prove that a coherent semantics ([4],[14]) allows to work in $\mathbf{E}$ as if it were the category of sets, once some cautions about the use of the logical principle of the excluded middle are taken, as explained below.

while, on the other hand,

D is large enough to determine the slope of any $f : D \longrightarrow R$ uniquely.

Thus every function $R \longrightarrow R$, restricted to D, is an infinitesimal element of a straight line, and this straight line is uniquely determined by such infinitesimal element.

Axiom 1 implies that $D \neq \{0\}$ (otherwise the uniqueness claim would fail). Therefore the ring of line type R is required not to be a field. Furthermore, notice that Axiom 1 implies:

$$(\forall h \in D \quad h \cdot b_1 = h \cdot b_2) \Longrightarrow b_1 = b_2$$

i.e. 'universally quantified first order nilpotents may be cancelled'.

Now there is a subtle logical point to acknowledge: it turns out that if we insist on an axiomatic approach to smoothness, as we do by Axiom 1, then the law of excluded middle (roughly speaking: a statement is either true or false) has to be rejected, as shown by the following argument, due to Steve Schanuel. By using the law of excluded middle, define a function $f : D \longrightarrow R$ by

$$f(h) = \begin{cases} 1 & \text{if } h \neq 0, \\ 0 & \text{if } h = 0. \end{cases}$$

If $h_0 \in D$, $h_0 \neq 0$, then, by Kock-Lawvere axiom,

$$1 = f(h_0) = f(0) + h_0 \cdot b = h_0 \cdot b$$

By squaring, $1 = 0$. Thus, when working in cartesian closed categories satisfying Kock-Lawvere axiom, one needs to leave classical logic. The weaker logic that must be used in cartesian closed categories is the Heyting logic.

"The moral to be learnt here, and that we believe is new in mathematics, is that weaker logics have their virtue, not by being more 'secure' or 'trustworthy' than full classical logic, but by their ability to carry theories (like Axiom 1) which are so strong as to be inconsistent in classical logic." (See Kock's paper in [8].)

Now define the derivative $R \xrightarrow{f'} R$ of a function $R \xrightarrow{f} R$ as the function (unique, by Axiom 1) defined by

$$f(x + h) = f(x) + f'(x) \cdot h$$

for all $x \in R$ and for all $h \in D$.

The process of forming derivatives may be indefinitely iterated. Thus every function $R \xrightarrow{f} R$ has higher derivatives $f', f'', \ldots$ of any order. The assumption of enough nilpotents ensured by the Kock-Lawvere axiom, allows to prove the basic formulas of calculus in a rigorous way by means of straightforward algebraic manipulations. Here are some instances.

Theorem *For any $f, g : R \longrightarrow R$ and for any $r \in R$:*

1. $(f + g)' = f' + g'$
2. $(r \cdot f)' = r \cdot f'$

3. $(f \cdot g)' = f' \cdot g + f \cdot g'$

4. $(g \circ f)' = (g' \circ f) \cdot f'$

Proof. As a matter of example, we prove 3 and 4. For all $d \in D$,

$$\begin{aligned}(f \cdot g)(x+d) &= f(x+d) \cdot g(x+d) \\ &= (f(x) + f'(x) \cdot d) \cdot (g(x) + g'(x) \cdot d) \\ &= (f \cdot g)(x) + (f'(x) \cdot g(x) + f(x) \cdot g'(x)) \cdot d\end{aligned}$$

(We used $d^2 = 0$.) Analogously, as far as equality 4. is concerned, for any $x \in R$ and for any $d \in D$, we have:

$$\begin{aligned}(g \circ f)(x+d) &= g(f(x+d)) \\ &= g(f(x) + f'(x) \cdot d) \\ &= g(f(x)) + g'(f(x)) \cdot f'(x) \cdot d\end{aligned}$$

(Notice that $f'(x) \cdot d \in D$.) By definition of derivative, it follows

$$(g \circ f)'(x) = g'(f(x)) \cdot f'(x)$$

for all $x \in R$. ∎

Notice that that proofs like the one reported above are qualitatively simpler than the usual ones and, on the other hand, are as rigorous as the latter.

Algebraic manipulations of nilpotent quantities have long been known as effective techniques to argue in the 'infinitely small'. To give an idea of synthetic reasoning, here are two other instances of proofs, taken from Lawvere's paper [7], based on the use of nilpotent infinitesimals. Numerous other examples from geometry and physics may be found in [1].

- *The length $C(r)$ of the boundary of a disk of radius r.* Define π to be the area of a unit disk. By homogeneity, the area of a disk of radius r is $A(r) = \pi r^2$. If the radius is perturbed by a small quantity h, the increment of area is

  $$A(r+h) - A(r) = \pi((r+h)^2 - r^2) = \pi(2rh + h^2) = 2\pi rh$$

  if 'small' is intended to mean $h^2 = 0$. On the other hand, we may assume that

  $$A(r+h) - A(r) = C(r)h$$

  Therefore $C(r)h = 2\pi rh$, for all h such that $h^2 = 0$. Since universally quantified nilpotents can be cancelled, it follows that $C(r) = 2\pi r$.

- *The electric field of a dipole.* Two charged particles $+q$ and $-q$ are placed at a small $2h$ distance apart ($1 \gg h$). The electric field E at a substantial distance x from

the mid-point between the particles, on the line containing the particles, is given by Coulomb's law:

$$\begin{aligned} E &= \frac{1}{(x-h)^2} - \frac{1}{(x+h)^2} \\ &= \frac{1}{x^2 - 2xh} - \frac{1}{x^2 + 2xh} \\ &= \frac{1}{x^2}\left(\frac{1}{1-\frac{2h}{x}} - \frac{1}{1+\frac{2h}{x}}\right) \\ &= \frac{1}{x^2}\left(1 + \frac{2h}{x} - \left(1 - \frac{2h}{x}\right)\right) \\ &= \frac{4h}{x^3} \end{aligned}$$

Algebraic calculations based on heuristic arguments of this sort, often found in physics books, become rigorous in our context, if assumptions like $1 \gg h$ are interpreted as h being nilpotent - i.e. satisfying $h^{n+1} = 0$ for some $n = 1, 2, ..$ - while 'substantial' quantities are considered to be invertible elements.

By cartesian closedness, for every object M in $\mathbf{E}$ one can consider the space M^D of infinitesimal paths in M, the *tangent bundle* of M. Thus the tangent bundle of an object M is representable as the object of maps from D to M. The representability of the tangent bundle functors greatly simplifies constructions and computations in the theory of vector fields, flows and differential equations. For a recent account, see ([11]).

3 A Sketch of the Construction of Models

Up to now, we did not consider the problem of proving the existence of *models* that may be used as a universe of discourse for Synthetic Differential Geometry. Good models do exist, and the basic principles for their construction as (dual) sheaves on finitely-generated C^∞-algebras were given by Lawvere in 1967 ([3]). We now roughly sketch the ideas underlying their construction. As emphasized in section 1, such models are required to be cartesian closed categories, in order to guarantee the existence of function spaces. Actually, they will turn out to be *toposes*, namely cartesian closed categories in which the notion of subobject is representable by a truth-value object Ω. This means that, for any object X, there is a natural bijection

$$\frac{S \hookrightarrow X}{X \longrightarrow \Omega}$$

between subobjects $S \hookrightarrow X$ of X and maps $X \longrightarrow \Omega$, the characteristic functions of the subobjects. (In the topos **Set** of sets, for example, the truth-value object is the two-element set $\Omega = \{0, 1\}$. For other examples see [9].)

Furthermore, these toposes are required to contain infinitesimal objects (like the space D of nilpotent infinitesimals of square zero) and are subjected to suitable axioms (such as Kock-Lawvere's axiom or the integration axiom), chosen to suit different purposes. To

construct such models, one starts from the category **Man** of manifolds, whose objects are finite dimensional smooth manifolds and whose morphisms are infinitely differentiable maps. This category is not cartesian closed and it obviously does not contain infinitesimal objects like D.

Now the basic idea is to specify a category of spaces as the dual of a category of commutative rings. An instance of such a line of thought was the introduction of the category of affine schemes by Grothendieck (around 1957) as a specification of the dual of the category **Rings** of commutative rings. The use of non-reduced commutative rings, i.e. rings with nilpotent non-zero elements, allowed to apply differential geometric concepts in algebraic geometry, by making it possible to render infinitesimal variations of points on an algebraic variety ([17]).

In the same vein, the basic guideline in Synthetic Differential Geometry is to construct categories of smooth spaces as a specification of the dual of suitable subcategories of the category of commutative rings. A space will be re-constructed as the spectrum of a ring of functions. In this way, for example, the infinitesimal object D will be recovered as the spectrum of its coordinate ring R^D (the ring of smooth maps from D to the reals) where the latter, by the Kock-Lawvere Axiom, is isomorphic to the ring of dual numbers $R[X]/(X^2)$. To be more precise, as a first step for the construction of models, we need to select a category $\mathbf{A}$ satisfying:

- its objects include all the coordinate rings of (ordinary) manifolds as well as all the coordinate rings of infinitesimal spaces. The rings whose spectrums are infinitesimal objects are the *local algebras* or *Weil algebras* ([17]);
- the morphisms of $\mathbf{A}$ are required to be exactly the ring homomorphisms induced by smooth maps.

One such candidate is the category $\mathbf{A}$ of *finitely generated C^∞-rings* ([15]). The functor $\mathbf{Man} \xrightarrow{C} \mathbf{A^{op}}$, taking each manifold M to its coordinate ring $C(M)$, is a full and faithful embedding of **Man** in the category $\mathbf{A^{op}}$. Now the category $\mathbf{A^{op}}$ has all finite limits and contains all ordinary manifolds as well as all infinitesimal objects, but it lacks function spaces in general, i.e. it is not cartesian closed. A natural step to achieve cartesian closedness is considering the functor category $\mathbf{Set^A}$ of presheaves of sets on $\mathbf{A^{op}}$. The composition of the functors

$$\mathbf{Man} \xrightarrow{C} \mathbf{A^{op}} \xrightarrow{Y} \mathbf{Set^A}$$

(where Y is the Yoneda embedding) gives an embedding $\mathbf{Man} \longrightarrow \mathbf{Set^A}$. Now the category of presheaves $\mathbf{Set^A}$ is cartesian closed, contains infinitesimal objects and can be easily seen to satisfy Kock-Lawvere's axiom. Nevertheless, the embedding $\mathbf{Man} \longrightarrow \mathbf{Set^A}$ is still unsatisfactory, because it does not preserve such colimits of **Man** as open covers. Preservation of open covers is needed in order to ensure, for example, that compact manifolds are mapped to compact spaces in the model or to allow the consideration of local questions. In order to overcome this difficulty, one defines a Grothendieck topology, or covering system, on $\mathbf{A^{op}}$ and considers the category (Grothendieck topos) $\mathbf{Sh}(\mathbf{A^{op}}) = \mathbf{E}$ of sheaves with

respect to the fixed covering system: only those presheaves are considered that 'perceive' that open covers in **Man** are still open covers in $\mathbf{Set}^{\mathbf{A}}$.

We finally have functors

$$\mathbf{Man} \xrightarrow{C} \mathbf{A}^{\mathbf{op}} \xrightarrow{Y} \mathbf{Set}^{\mathbf{A}} \xrightarrow{L} \mathbf{E}$$

whose composition $J : \mathbf{Man} \longrightarrow \mathbf{E}$ is a full and faithful embedding satisfying:

1. $J(\mathbb{R}) = R$, where $\mathbb{R}$ denotes the real numbers and R is the real line in $\mathbf{E}$;
2. $J(\mathbb{R} \setminus \{0\}) =$ the set of invertible elements of R;
3. for every $f : \mathbb{R} \longrightarrow \mathbb{R}$ with derivative $f' : \mathbb{R} \longrightarrow \mathbb{R}$, $J(f') = (J(f))'$;
4. for every manifold M, $J(TM) = (JM)^D$, where TM is the tangent bundle of M;
5. J preserves transversal pullbacks.

The category $\mathbf{E}$ is said to be a *well-adapted* model, in the sense that the embedding $J : \mathbf{Man} \longrightarrow \mathbf{E}$ is suitable for comparison between properties and constructions already available in the category of manifolds **Man** and those available in the enlarged universe $\mathbf{E}$. The notion and the construction of well adapted models of Synthetic Differential Geometry, indispensable for for the applications to classical differential geometry, are due to Eduardo Dubuc ([2]).

References

[1] Bell, J.L., *A Primer of Infinitesimal Analysis*, Cambridge Univesity Press, 1998.

[2] Dubuc, E., Sur les Modeles de la Géométrie Différentielle Synthetique, *Cahiers de Topologie et Géométrie Différentielle*, Vol. XX-3, 1979.

[3] Kock, A. (ed.), *Topos Theoretic Methods in Geometry*, Various Publications Series No. 30, Aarhus Universitet, Matematisk Institut, 1979.

[4] Kock, A. *Synthetic Differential Geometry*, London Math. Soc. Lecture Note Series, 51, Cambridge University Press, 1981.

[5] Kock, A. Synthetic differential calculus, and nilpotent real numbers, "Meeting on Existence in Mathematics", University of Roskilde, Nov. 24-25 2000, available at http://home.imf.au.dk/kock/

[6] Lavendhomme, R., *Basic Concepts of Synthetic Differential Geometry*, Kluwer, 1996.

[7] Lawvere, F.W., Toward the description in a smooth topos of the dynamically possible motions and deformations of a continuous body, *Cahiers de Topologie et Géométrie Différentielle*, Vol. XXI-4, 1980.

[8] Lawvere, F.W. and Schanuel, S.H. (ed.) *Categories in Continuum Physics*, Lecture Notes in Mathematics N. 1174, Springer-Verlag, 1982.

[9] Lawvere, F.W. and Schanuel, S.H., *Conceptual Mathematics: A First Introduction to Categories*, Cambridge University Press, 1997.

[10] Lawvere, F. William, Toposes of Laws of Motion, Montreal 1997, available at http:// www.acsu.buffalo.edu/~ wlawvere/

[11] Lawvere, F. William, Outline of Synthetic Differential Geometry, 1998, available at http:// www.acsu.buffalo.edu/~ wlawvere/

[12] Lawvere, F.W., Everyday Physics of Extended Bodies or Why Functionals Need Analyzing, 1998, available at http://www.acsu.buffalo.edu/~ wlawvere/

[13] Lawvere, F.W., *Categorie e spazio: un profilo,* Lettera Matematica Pristem, Springer, 31, 1999.

[14] McLarty, C., *Elementary Categories, Elementary Toposes*, Clarendon Press, Oxford, 1992.

[15] Moerdijk, I. and Reyes, G., *Models for Smooth Infinitesimal Analysis*, Springer-Verlag, 1991.

[16] Riemann, B., *Gesammelte Mathematische Werke*, Dover, New York, 1953.

[17] Weil, A., *Théorie des points proches sur les variétés différentiables,* Colloque Géom. Diff., Strasbourg, 1953. Oeuvres Scientifiques, Collected Papers, Vol. II, Springer-Verlag, 1979., p.103-109. Commentaire, p.534-536.

Federico G. Lastaria
Dipartimento di Matematica
Politecnico di Milano
Piazza Leonardo da Vinci 32
20133 Milano, Italy
fedlas@mate.polimi.it