

Who's afraid of coordinate systems? An essay on representation of spacetime structure

David Wallace

March 22, 2016

Abstract

Coordinate-based approaches to physical theories remain standard in mainstream physics but are largely eschewed in foundational discussion in favour of coordinate-free differential-geometric approaches. I defend the conceptual and mathematical legitimacy of the coordinate-based approach for foundational work. In doing so, I provide an account of the Kleinian conception of geometry as a theory of invariance under symmetry groups; I argue that this conception continues to play a very substantial role in contemporary mathematical physics and indeed that supposedly “coordinate-free” differential geometry relies centrally on this conception of geometry. I discuss some foundational and pedagogical advantages of the coordinate-based formulation and briefly connect it to some remarks of Norton on the historical development of geometry in physics during the establishment of the general theory of relativity.

1 Introduction

There is a rather sharp difference between elementary theories of mechanics as presented in the mainstream physics literature on the one hand, and in the foundational and philosophical literature on the other. A scalar field in special relativity, for instance, is typically presented in the physics literature as a smooth (i. e. , infinitely-often differentiable) function $\varphi : \mathbb{R}^4 \rightarrow \mathbb{R}$, satisfying the equation

$$\eta^{\mu\nu} \frac{\partial^2 \varphi}{\partial x^\mu \partial x^\nu} + m^2 \varphi + \frac{\lambda}{3!} \varphi^3 = 0 \quad (1)$$

where(as always in this paper) the indices μ, ν take values 0, 1, 2, 3, $\eta^{\mu\nu}$ denotes the components of the diagonal 4x4 matrix with diagonal elements $-1, 1, 1, 1$, and the Einstein summation components applies.

Of course, the choice of coordinates here has to be understood as to some degree arbitrary, and this is captured by the fact that certain coordinate transformations leave the form of this equation invariant. Specifically, it is invariant under any transformation of form

$$x^\mu \rightarrow \Lambda^\mu_\nu x^\nu + a^\mu \quad (2)$$

where Λ^μ_ν are the components of a matrix satisfying $\Lambda\eta\Lambda^{-1} = \eta$ and the a^μ are arbitrary: that is, it is invariant under the standard action of the Poincaré group on \mathbb{R}^4 .

In foundational discussions, the theory is characterised instead as something like the ordered quadruple $\langle \mathcal{M}, g, \nabla, \varphi \rangle$, where

- \mathcal{M} is a differentiable manifold diffeomorphic to \mathbb{R}^4 ;
- g is a smooth symmetric tensor field of rank $(0, 2)$ and signature $(-1, 1, 1, 1)$;
- ∇ is a flat symmetric affine connection satisfying $\nabla g = 0$;
- $\varphi : \mathcal{M} \rightarrow \mathbb{R}$ is a smooth real function on the manifold satisfying

$$g(\nabla, \nabla)\varphi + m^2\varphi + \frac{\lambda}{3!}\varphi^3 = 0. \quad (3)$$

(Use of an abstract-index notation (as per Wald 1984) makes (3) look somewhat more like (1) but the equations remain completely different sorts of mathematical object.)

The former version is usually called ‘coordinate-based’ or ‘coordinate-dependent’; the latter, ‘coordinate-free’. In the case of general relativity in particular, coordinate-free formulations based upon differential geometry have been widely used in physics for decades (Misner, Thorne and Wheeler’s classic *Gravitation* (1973) was written with the explicit aim of making them standard), but a wide range of pre-general-relativistic theories, including the special-relativistic field theory above, but also other special-relativistic field and particle theories, and pretty much the whole of non-relativistic dynamics, can also be formulated in the framework of differential geometry. The work of, in particular, Anderson (1964, 1967, 1971), Trautman (1966), Earman (Earman 1970; Earman 1974) and Friedman (Earman and Friedman 1973) brought this coordinate-free formulation onto widespread use in foundational discussions; this was solidified by hugely influential books by Friedman (1983) and Earman (1989). In contemporary philosophy of physics it is close to an orthodoxy that spacetimes, and theories written on spacetimes, should be formulated in this way (see, for instance, Malament (2012), Pooley (2013), Pitts (2006) and Weatherall (2014) for recent examples).

The use of differential geometry as the standard language for spacetime theorising is by now so uniform that it is scarcely defended explicitly. But those defences which are made (in particular, in the work of Earman, Friedman, Trautman and Anderson, *ibid*) generally cite two reasons. Firstly, an important foundational topic in spacetime theory is the comparisons of the different structures posited by different theories of spacetime; such comparisons require, or at any rate are greatly facilitated, by phrasing all such theories in a common language. And since general relativity (it is said) is most naturally expressed in the coordinate-free differential-geometric language, the clearly preferred strategy is to formulate all spacetime theories in this language.

Secondly, the traditional physics approach (it is claimed) is based upon an untenable understanding of coordinates. Expressions like (1) are not invariant under arbitrary changes of coordinates: in this particular case, the form of the equation is preserved *only* under coordinate changes of form (2); under more general coordinate changes, the equation will change form.

So it is tempting to say that the theory can *only* be expressed in certain special coordinate systems. But of course this is not correct: coordinates are just ways of labelling points in a space, and we can, if we choose to, describe any theory we like with respect to any coordinates we like. This suggests that if we want to get out the physical content of a theory, we need to move away from a characterisation in terms of particular coordinate systems and find some intrinsic way of describing the structure of the theory. This suggests a move to a framework which does not intrinsically privilege one coordinate system above another, and in due course away from coordinate systems altogether: in other words, to differential geometry.

This objection is generally first attributed to Kretschmann, and was influential in the development of general relativity and in particular in the understanding of general covariance (see Norton (1993) for a detailed discussion and a more nuanced account of Kretschmann's actual view). In the standard (and reasonably accurate) account, Einstein originally saw the move from special to general relativity as a move from a theory which could be formulated only with respect to a certain collection of coordinate systems (any two of which being related via a Poincaré transformation) to a 'generally covariant' theory which could be formulated with respect to any coordinate system. Kretschmann objected that (plausibly) *any* theory can be so formulated, since coordinates per se are simply labels, and that any genuinely substantive sense of general covariance had to be understood some other way.

In contemporary discussions, the most common understanding has been that general relativity — unlike other spacetime theories — contains no 'absolute objects', understood intuitively to be those differential-geometric objects in a theory which represent background structure rather than anything dynamical: in the case of the scalar field theory above, for instance, the absolute objects are taken to be g and ∇ . Anderson and Friedman (ibid.) were influential both in developing this no-absolute-object interpretation of substantive general covariance and in proposing a concrete definition of absolute objects; the task of refining this definition continues today (see Pitts (2006), Pooley (2015), and references therein for a contemporary summary; see also section 6).

This perspective also informs the consensus view in foundational work on how the coordinate-based approach is to be understood. Firstly, the *spacetime symmetry group* of the theory is declared to be the subgroup of the diffeomorphism group which leaves the absolute structures invariant. In the case of the scalar field theory, this group turns out to be the Poincaré group.

It is then noted that, if the equations defining the theory are expressed with respect to two coordinate systems (that is, if global issues are put aside, two smooth injective maps from the manifold to \mathbb{R}^4) related by an element of the spacetime symmetry group, then the numerical values of the components of the

absolute objects will be the same in each. If in particular we can find some coordinate system in which those components take a particularly simple form, then we will actually have found a family of such coordinate systems — and the standard coordinate-based way of writing a theory is to be understood simply as the differential-geometric theory described with respect to one of these simple coordinate systems.

For instance, given that ∇ is flat, in the scalar field theory we can find coordinates in which the components of g are just the components of the matrix η and in which the covariant derivative reduces to the ordinary derivative. In those coordinates, equation (3) reduces to (1). Physics orthodoxy is recovered, but only derivatively on the more sophisticated differential-geometric framework.

The purpose of this paper is to attempt a rehabilitation of the coordinate-based approach to physical theories as a legitimate and informative way to formulate them, and not simply as a crude first stab at a more sophisticated approach. My intention is not to deny the usefulness for many purposes of the differential-geometry, coordinate-free approach: that approach is indisputably useful for a great many purposes. It *is* to deny its primacy: it is (I will argue) perfectly possible to make sense of the coordinate-based approach in its own terms, and doing so, in some circumstances, has advantages pedagogical, technical and conceptual. Furthermore, when the coordinate-based approach is understood properly, it will be realised that the ‘coordinate-free’ approach isn’t as coordinate-free as all that.

2 Kleinian geometry and the concept of a structured space

Coordinates are rules that assign to each point in a space some N -tuple of real numbers representing that point. Formally, that makes a coordinate system a 1:1 map $f : \mathcal{S} \rightarrow \mathbb{R}^N$ from a space \mathcal{S} into the N -fold product of the real numbers. (In due course we will consider *local* coordinate systems, which coordinatise only patches of a space; for the moment, we treat all coordinate systems as global.)

\mathcal{S} may have some intrinsically characterised structure, in which case we can use that structure to pick out certain coordinate systems as preferred. For instance, suppose that \mathcal{S} is an N -dimensional vector space. Then certain 1:1 maps from \mathcal{S} to \mathbb{R}^N will be linear, and we can restrict our attention to that collection of coordinate systems.

Given two coordinate systems f, f' , we can consider the *coordinate transformation map* $f \cdot f'^{-1}$ which maps us from one coordinatisation scheme to another. In our vector space example, any transformation map between preferred coordinates will have the form

$$x^i \rightarrow M_j^i x^j \tag{4}$$

where M_j^i are the components of an invertible matrix; conversely, given any preferred (i. e. , linear) coordinate system f and any map g of this form, $g \cdot f$ will also be preferred.

A vector space is defined by the operations of addition of vectors and multiplication of vectors by scalars. And if x^i are the coordinates of some vector x with respect to a given system of coordinates, these operations look like

$$(ax + by)^i = ax^i + by^i. \quad (5)$$

Now suppose that we are given \mathcal{S} , with its collection of preferred coordinatisations but without the intrinsic vector-space structure on \mathcal{S} . We can recover that structure just by *defining* addition and multiplication via (5), with respect to a quite arbitrary coordinatisation in the preferred set. How can it be arbitrary? Because the requirement that transformations between coordinatisations have the form (4) guarantees that if that definition holds with respect to one coordinate system, it holds with respect to any coordinate system.

This means that we have an alternative definition of a (finite-dimensional) vector space available. Instead of defining it in the usual way, define it as an unstructured set, plus a collection of bijections from the set to \mathbb{R}^N , such that given any one such bijection f in the set, any other bijection f' is in the set iff $f \cdot f'^{-1}$ has form (4).¹

In what sense are these coordinatisations ‘preferred’? Not in the sense that the space can only be coordinatised by them. (No law prohibits us coordinatising a vector space any way we like, after all — for all that is usually unprofitable to coordinatise it any way except via linear coordinate functions.) Rather, in the austere mathematical sense that a set equipped with a family of mappings in this way is a perfectly respectable mathematical object whose properties can be studied.

The framework can be generalised. Suppose that \mathcal{G} is a group of bijections of \mathbb{R}^N . Then a \mathcal{G} -structured space is a set \mathcal{S} together with a nonempty collection \mathcal{C} of bijections from \mathcal{S} to \mathbb{R}^N (the ‘coordinatisations’ of \mathcal{S}), such that if $f \in \mathcal{C}$, then $f' \in \mathcal{C}$ iff $f \cdot f'^{-1} \in \mathcal{G}$.

This is, in effect, a form of the definition of geometry in Klein’s famous *Erlangen* program (Klein 1892). A space coordinatised by only a single map to \mathbb{R}^N (that is, a space structured by the trivial group) is effectively just a copy of \mathbb{R}^N , with all the associated structure. As the group is made larger, the space becomes less structured, because any operation defined on the space will have to be defined in terms of the coordinate systems (there is no other structure available) and will have to be defined in a way that works equally well for any such coordinate system (because no one coordinatisation is preferred).

For instance, consider these groups of transformations of \mathbb{R} :

1. $x \rightarrow x$
2. $x \rightarrow \lambda x, \lambda > 0$
3. $x \rightarrow \lambda x, \lambda \neq 0$

¹Formally speaking: the category whose objects are sets equipped with coordinates in this way, and whose morphisms are maps whose coordinatisations are linear, is equivalent to the category of vector spaces and linear maps.

4. $x \rightarrow \pm(x + a)$
5. $x \rightarrow \lambda x + a, \lambda \neq 0$
6. $x \rightarrow f(x), f$ smooth and with smooth inverse
7. $x \rightarrow f(x), f$ continuous and with continuous inverse
8. $x \rightarrow f(x), f$ invertible.

These groups each define a structured space (call them \mathcal{S}_1 through \mathcal{S}_8). To discuss their structure, I use unprimed letters x, y, z to discuss the coordinates of points of the space with respect to one coordinate system and primed letters x', y', z' to discuss the coordinates of those same points with respect to another system. So $x' = f(x)$ where f is some element of the appropriate group.

So: \mathcal{S}_1 is just a copy of \mathbb{R} : any of its features (multiplicative structure, additive structure, ordering) can be carried over to \mathcal{S}_1 via the single coordinatisation. \mathcal{S}_2 has discarded the multiplicative structure of \mathbb{R} , for $(xy)' \neq x'y'$ and so the definition of multiplication is not coordinate-independent. But it maintains the linear structure $((ax + by)' = ax' + by')$ and the ordering (if $x > y, x' > y'$). \mathcal{S}_3 discards the ordering but keeps the linear structure. \mathcal{S}_4 discards the linear structure but preserves the distances between points (discarded in \mathcal{S}_2 and \mathcal{S}_3): $|x' - y'| = |x - y|$. \mathcal{S}_5 discards both distance and linear structure, but retains the latter for the differences of pairs of points: $(a(x - y) + b(z - w))' = a(x - y)' + b(z - w)'$. \mathcal{S}_6 discards even this, but maintains enough structure that we can say of a function from \mathcal{S}_6 to \mathbb{R} whether it is smooth. In \mathcal{S}_7 , all we can do is say if such a function is continuous. And \mathcal{S}_8 has no structure at all beyond the bare cardinality of the real line: it is has as many elements as \mathbb{R} but that is all that can be said. (This by no means exhausts the range of structures that can be so defined; taking $f(x)$ to be monotonic, for instance, gives a structured space with the same order structure as the real line but no other structure.)

This Kleinian framework is extremely powerful, as the real-line examples demonstrate: all that is necessary to define a space is to give a group of transformations of \mathbb{R}^N . To illustrate further:

- As already noted, a *vector space* can be defined as having structure group

$$x^i \rightarrow M_j^i x^j \tag{6}$$

for M_j^i invertible. Putting further restrictions on the form of the matrix \mathbf{M} whose components are M_j^i defines further spaces:

- Requiring \mathbf{M} has positive determinant gives an *oriented vector space*.
- Requiring $\mathbf{M}^{-1} = \mathbf{M}^T$ gives an inner product space. (Requiring only $\mathbf{M}^{-1} = \lambda \mathbf{M}^T$ for nonzero λ gives a weakening of an inner product space in which angle but not length is defined.)

- Requiring M to have determinant ± 1 gives a vector space with volume.

In each case, the structure-defining group is smaller than (i. e., a subgroup of) the general linear group that defines a vector space, so more concepts can be consistently defined. In an oriented vector space, it's consistent to define whether two N -tuples of vectors have the same orientation; in an inner product space the quantity $\sum_i v^i w^i$ is well-defined — i. e. coordinate-independent — and so can be taken to define a bilinear map from vectors to real numbers; in a vector space with volume the volume enclosed by an N -tuple is well-defined. (And any of the concepts can be combined to yield, for instance, an oriented inner-product space).

Alternatively, the second requirement can be replaced by $M\eta M^{-1} = \eta$. This makes the space into a Minkowski vector space; its structure group can be recognised as the Lorentz group.

- An *affine space* can be defined as having structure group

$$x^i \rightarrow M_j^i x^j + a^i \quad (7)$$

for M_j^i invertible. This *affine group* is larger than (i. e. contains) the general linear group, so fewer concepts are definable within it: in particular, addition of points is no longer defined. However, we have

$$(y^i - x^i) \rightarrow (M_j^i y^j + a^i - M_j^i x^j - a^i) = M_j^i (y^j - x^j)$$

so the space of differences between points inherits a vector-space structure. We can then consistently define, given two pairs of points A, B and C, D , whether the vector from A to B is the same as, or a multiple of, the vector from C to D .

Any of the various restrictions on M_j^i imposed in the case of vector spaces can be imposed here too. Imposing $M^{-1} = M^T$, for instance, restricts the affine group to the *Euclidean group*, and defines Euclidean space, in which the lengths and angles of the vectors joining points are defined. Imposing $M\eta M^{-1} = \eta$ — that is, taking the structure group to be the Poincaré group — defines Minkowski spacetime.

- The various spacetimes of pre-relativistic dynamics can also be defined as structured spaces, just as Minkowski spacetime can. Such spacetimes are coordinatised by points in \mathbb{R}^4 , and in accordance with normal usage I label such points (t, x^1, x^2, x^3) . They are defined by these groups of maps (here I follow the terminology of Earman (1989, ch.2)):

1. $t \rightarrow \pm t + \tau, x^i \rightarrow R_j^i x^j + a^i$: *Newtonian spacetime*.
2. $t \rightarrow \pm t + \tau, x^i \rightarrow R_j^i x^j + v^i t + a^i$: *Galilean spacetime*.
3. $t \rightarrow \pm t + \tau, x^i \rightarrow R_j^i x^j + a^i(t)$: *Maxwellian spacetime* (called *Newton-Huygens spacetime* by Saunders (2013)).

4. $t \rightarrow \pm t + \tau, x^i \rightarrow R_j^i(t)x^j + a^i(t)$: *Leibnizian spacetime*
5. $t \rightarrow f(t)$ for monotonic $f, x^i \rightarrow R_j^i(t)x^j + a^i(t)$: *Machian spacetime*.

In each case R_j^i is orthogonal and any functions of t are smooth.

The first two of these are subgroups of the affine group, so that these spacetimes can be regarded as affine spaces with extra structure. The last three cannot be so regarded, due to the arbitrary functions of t that they contain. Each group is a subgroup of those below it, so that we can see the move from one spacetime to the next as a successive discarding of structure.

- Finally, for some structured spaces with even larger structure groups, we can simply require that the transition functions between coordinatisations are diffeomorphisms (that is, smooth and with smooth inverse) or homeomorphisms (continuous with continuous inverse). These two structure groups give rise, respectively, to a *globally trivial differentiable manifold* or to a *globally trivial topological manifold*, two familiar examples to which I return in the next section.

I close this section with three general observations about structured spaces. Firstly, any two such spaces $\mathcal{S}_1, \mathcal{S}_2$ for given \mathcal{G} (with coordinatisations $\mathcal{C}_1, \mathcal{C}_2$), are mathematically equivalent, in the sense that there is a bijection $\varphi : \mathcal{S}_1 \rightarrow \mathcal{S}_2$ such that $f \in \mathcal{C}_2$ iff $f \cdot \varphi \in \mathcal{C}_1$. (Pick arbitrary $f_1 \in \mathcal{C}_1, f_2 \in \mathcal{C}_2$, and define $\varphi = f_2^{-1} \cdot f_1$.) So we can speak (in the same way that we speak of “the torus” or “the complex field”) of “the \mathcal{G} -structured space”: there are no mathematically salient differences between any two such spaces.

Secondly, if \mathcal{H} is a subgroup of \mathcal{G} , then any \mathcal{H} -structured space can be turned into a \mathcal{G} -structured space just by closing the set of coordinatisations under the action of \mathcal{G} . This amounts to ‘forgetting’ those features of the structure that are given by \mathcal{H} in particular. (Hence, we can forget the inner-product structure of an inner-product space by widening its structure group from the orthogonal to the general linear group, and thus obtain a vector space; we can widen it further to the affine group, obtaining an affine space.)

Finally, the requirement that coordinatisations take values in \mathbb{R}^N does no formal work in the definitions above (though it does in each of the specific examples). We can perfectly well ‘coordinatise’ spaces by complex numbers, or even by points in more abstract spaces like a vector space. An affine space, for instance, can be understood as a space structured by the transformation group $x \rightarrow x + a$ on a vector space, reflecting the heuristic idea that an affine space is a vector space whose origin has been discarded. In this characterisation, the interpretation of the coordinate maps as providing coordinates in the ordinary sense has fallen away (unlike tuples of real numbers, points in an abstract vector space are not uniquely individuated and so make poor coordinates) but we retain the idea of characterising the structure of a space in terms of a class of identifications of that space with a more richly structured space.

3 Local coordinate systems and locally structured spaces

The structured spaces we have discussed are all *globally* coordinatised by points in \mathbb{R}^N . But of course many spaces — the surface of the sphere, for instance — can be coordinatised only in patches. A coordinatisation of such a space is a 1:1 map, not from the space to \mathbb{R}^N , but from some subset of the space to some subset of \mathbb{R}^N . And the compatibility condition between two coordinatisations can be imposed only on their overlap.

The framework of structured spaces can be extended to cover this case. Firstly, recall that an *injective partial function* from X to Y is a function from some subset of X to Y such that if $f(x) = f(y)$, then $x = y$. Local coordinatisations of a space \mathcal{S} will then be injective partial functions from \mathcal{S} to \mathbb{R}^N . We can define the *domain* $\text{Dom}(f) \subset X$ of the partial function f as the set of points x for which $f(x)$ is defined, and the *range* $\text{Ran}(X) = f(\text{Dom}(X))$ as the set of points y for which $f(x) = y$ for some x . The composite $g \cdot f$ of two partial functions is defined by $g \cdot f(x) = g(f(x))$ whenever that is defined; it is a partial function with domain

$$\text{Dom}(g \cdot f) = \text{Dom}(f) \cap f^{-1}(\text{Dom}(g)); \quad \text{Ran}(g \cdot f) = \text{Ran}(g) \cap g(\text{Ran}(f)). \quad (8)$$

(If these are empty, the partial function is the trivial *empty function*, uniquely defined by the fact that its domain is the empty set.) And the *domain set* $\text{Dom}(\mathcal{G})$ of a collection \mathcal{G} of partial functions is the collection of all domains of functions in \mathcal{G} ; the *range set* $\text{Ran}(\mathcal{G})$ is defined similarly.

Given two coordinatisations f and f' , we can consider the partial function $f' \cdot f^{-1}$, which is a partial function from \mathbb{R}^N to itself and which represents the transformation between the two coordinate systems for those points coordinatised by both. Imposing a structure on the allowable coordinatisations is now a matter of restricting these partial compositions. The relevant notion is the *pseudogroup*, a generalisation of a transformation group introduced by Cartan (1904): a pseudogroup \mathcal{G} is a collection of injective partial functions on some space $\mathcal{X}_{\mathcal{G}}$ (normally \mathbb{R}^N) satisfying these two properties:

Closure: \mathcal{G} is closed under composition and taking the inverse.

Local definedness: If f is an injective partial function on $\mathcal{X}_{\mathcal{G}}$, and if $\{X_i\}$ is a collection of sets in $\text{Dom}(\mathcal{G})$ whose union contains $\text{Dom}(f)$, then f is in \mathcal{G} whenever its restriction to each of the X_i is.

(Unlike a group, a pseudogroup is an essentially geometric rather than algebraic object, defined directly as a collection of maps. Various definitions of pseudogroup are in use (cf, e.g., Veblen and Whitehead (1932, p.37), Kobayashi and Nomizu (1963, pp.1–2); it is common, though not universal, to include various topological assumptions about the domains of the mappings, which I here omit.)

It follows from these definitions that if $X \in \text{Dom}(\mathcal{G})$, then $\text{id}_X \in \mathcal{G}$. Since for any f , $\text{id}_X \cdot f$ is the restriction of f to X , it follows that \mathcal{G} is also closed

under restriction to any set in $\text{Dom}(\mathcal{G})$, and hence that $\text{Dom}(\mathcal{G})$ is closed under intersections.

Any nonempty collection of *total* functions from some \mathcal{X} to itself closed under composition and inversion is a pseudogroup, so transformation groups are special cases of pseudogroups. A less trivial example is the collection $\text{Diff}_{loc}(\mathbb{R}^N)$ of *local diffeomorphisms* of \mathbb{R}^N : the 1:1 partial functions on \mathbb{R}^N which (i) have open sets as their domains; (ii) are smooth; (iii) have smooth inverses. Similarly, we can define the group of local homeomorphisms (maps which are homeomorphisms between open sets in \mathbb{R}^N (continuous 1:1 maps between open sets in \mathbb{R}^N with continuous inverses), local affine maps (1:1 maps between open sets in \mathbb{R}^N that are affine — $x^i \rightarrow R_j^i x^j + a^i$ with R_j^i invertible — on any connected subset), local Euclidean maps (1:1 maps between open sets in \mathbb{R}^N that are Euclidean — $x^i \rightarrow R_j^i x^j + a^i$ with R_j^i self-adjoint — on any connected subset), and so forth. If we use C^N in place of \mathbb{R}^N as a coordinate space, the group of locally analytic maps is also a pseudogroup.

Now, given a pseudogroup \mathcal{G} , a *locally \mathcal{G} -structured space* is then defined by:

- A set \mathcal{S} .
- A collection \mathcal{C} of 1:1 partial functions from \mathcal{S} to $\mathcal{X}_{\mathcal{G}}$ — the *local coordinatisations* (or *charts*)— such that
 1. Every point in \mathcal{S} lies in the domain of some element of \mathcal{C} ;
 2. $\text{Ran}(\mathcal{C}) \subset \text{Dom}(\mathcal{G})$, i. e. the range of every local coordinatisation is the domain of one of the partial functions in \mathcal{G} .
 3. For any partial function $f : \mathcal{S} \rightarrow \mathbb{R}^N$ whose range is in $\text{Dom}(\mathcal{G})$, $f \in \mathcal{C}$ iff for any $f' \in \mathcal{C}$ the compatibility condition

$$f \cdot f'^{-1} \in \mathcal{G} \tag{9}$$

is satisfied.

It should be apparent that this generalises our previous concept of structured space. More interestingly, it contains the standard case of differentiable manifolds as a special case: on this definition (I claim) a differentiable manifold is a space locally structured by $\text{Diff}_{loc}(\mathbb{R}^N)$.

This is worth dwelling on. A manifold \mathcal{M} as typically defined (see, e. g. , (Kobayashi and Nomizu 1963, pp.2–3)) is a *topological* space equipped with an *atlas* of local homeomorphisms between open subsets of \mathcal{M} and open subsets of \mathbb{R}^N , such that (i) any two charts in the atlas are compatible in the above sense, and (ii) any point in \mathcal{M} lies in the domain of at least one chart. An atlas is *maximal* if additional charts cannot be added to it without breaking the compatibility condition, but it is not usually required of a manifold that its atlas be maximal.

However, any atlas has a unique extension to a maximal atlas; conversely, it is tacit in differentiable topology that any notion defined on a manifold must

remain well-defined if the manifold's atlas is extended to be maximal. (A manifold globally diffeomorphic to \mathbb{R}^N may be coordinatised by a single chart, but it would not be legitimate in differentiable topology to define a function on the manifold as *linear* simply because it is linear with respect to that one chart!) Or, more precisely: sometimes the atlas *is* restricted so as not to be maximal, but such restrictions are always used to define extra structure (as in the case of analytic manifolds or fiber bundles, both of which I return to below).

Furthermore, the requirement that a manifold's maps be local homeomorphisms is dispensable: the topology, as well as the differentiable structure, of the manifold may be inherited from the charts of a maximal atlas. Specifically, the set of all domains of charts in the atlas is closed under intersection and its union is the whole manifold; this set forms a base, in the topological sense, for the manifold, and with respect to this topology it may readily be verified that all charts are in fact homeomorphisms.²

Just as $\text{Diff}_{loc}(\mathbb{R}^N)$ defines the differentiable manifolds, each of our other examples of pseudogroups defines a locally structured space. The local homeomorphisms define the topological manifolds; the locally affine maps, the locally affine manifolds; the locally Euclidean maps, the locally Euclidean manifolds; the locally analytic maps, the analytic manifolds. In each case, a *class* of spaces is defined, whereas for the global notion of structure, each group of transformations defined only a single space. However, the notion of atlas generalises from the manifold case (an atlas is a collection of coordinatisations obeying the compatibility condition and jointly coordinatising the space), as does the result that any atlas determines a unique maximal atlas. For suppose that the charts in the atlas are h_1, \dots, h_i, \dots with $\text{Dom}(h_i) = X_i$. and that charts f, f' are both compatible with each h_i . Then

$$(f' \cdot h_i^{-1}) \cdot (h_i \cdot f^{-1}) = f' \cdot \text{id}_X \cdot f^{-1} \in \mathcal{G}, \quad (10)$$

$$\text{Dom} f^{-1} \subseteq \cup_i f(X_i) \quad (11)$$

and

$$f' \cdot \text{id}_X \cdot f^{-1} = f' \cdot f^{-1}|_{X_i}. \quad (12)$$

By the local definedness condition, $f^{-1} \cdot f' \in \mathcal{G}$ and so the two charts are compatible with each other; the set of all charts compatible with the atlas is therefore a maximal atlas and is unique. (This is just a generalisation of the well-known proof of the same result in manifold theory.)

For instance, the circle can be coordinatised by two charts f_1, f_2 , with compatibility conditions

$$\begin{aligned} f_2 \cdot f_1^{-1}(x) &= \pi - x & 0 < x < \pi \\ f_2 \cdot f_1^{-1}(x) &= -\pi + x & -\pi < x < 0. \end{aligned} \quad (13)$$

The partial function $f_2 \cdot f_1^{-1}$ is smooth, and so this pair of charts may be taken as an atlas and the circle can be regarded as a differentiable manifold. However,

²See (Lang 1999, pp.22-23) for an author who defines manifold topology this way.

the partial function also lies in the pseudogroups of local homeomorphisms, local affine maps, and local Euclidean maps. Constructing a maximal atlas with respect to each of these pseudogroups allows us to realise the circle as a topological manifold, a locally affine manifold, or a locally Euclidean manifold, including different amounts of the circle's structure in each case. For instance, on the circle *qua* locally Euclidean space we can say what the distance is between any two points; we cannot do this for the circle *qua* locally affine space but we can say of two pairs of points x, y and x', y' what the ratio is between the two pairs of distances.

An important class of additional examples are the *fiber bundles*, also familiar from mathematical physics. A fiber bundle is normally³ defined (cf. e.g., (Nakahara 1991, pp.305–312) as a tuple $\langle E, B, F, \mathcal{G}, \Phi \rangle$, where:

- E (the *total space*), B (the *base space*), and F (the *typical fiber*) are manifolds;
- \mathcal{G} is a group of 1:1 diffeomorphisms of F ;
- Φ is a collection of diffeomorphisms f_i between open subsets of F and sets $U_i \times F$, where $U \subseteq B$ is open, where every $x \in E$ lies in the domain of some f_i , where the f_i satisfy the compatibility condition

$$f_i \cdot f_j^{-1}(x, y) = (x, \varphi(x) \cdot y) \quad (14)$$

whenever it is defined for φ a smooth map from $U_i \times U_j$ to \mathcal{G} . (Φ is typically not explicitly required to be maximal under this condition, but as with manifolds, well-defined notions on a fiber bundle need to remain well-defined if it is extended to be maximal.)

This clearly has the flavour of our local-structure definitions: the bundle is defined as being locally isomorphic to a product space, and there is no single preferred local product structure but rather a family related by some transformation group. We can shift it still closer by combining the charts that define B itself with the charts that define the bundle. In this case, Φ is replaced with a collection of partial functions from E to $\mathbb{R}^N \times F$, where each partial function has domain $U_i \times F$ for some open $U_i \subset \mathbb{R}^N$. The compatibility condition is then

$$f_i \cdot f_j^{-1}(x, y) = (h(x), \varphi(x) \cdot y) \quad (15)$$

where h is a local diffeomorphism of \mathbb{R}^N .

The family of partial functions satisfying this condition is readily identified as a pseudogroup, and a fiber bundle can be redefined simply as a space locally structured by this pseudogroup. Indeed, we can readily generalise to bundles over structured spaces that are not manifolds: simply define a *bundle pseudogroup* as a collection of injective partial functions on some space $T \times F$, where

³Specifically, this is the normal way to define a fiber bundle when the structure group is intrinsic in the definition. An alternative approach characterises a fiber bundle without specific mention of structure group, and adds the latter later.

1. each such function has domain $D \times F$ and range $R \times F$ for some $D, R \subset T$;
2. there is a group \mathcal{G}_F , the *structure group*, of (total) 1:1 functions on F , and a pseudogroup \mathcal{G}_T of partial functions on T , such that elements of the bundle pseudogroup are exactly the functions of form

$$g(x, y) = (g_T \cdot, \varphi(x) \cdot y) \quad (16)$$

where $g_T \in \mathcal{G}_T$ and φ is a smooth map⁴ from the domain of g_T to \mathcal{G}_F .

A bundle can then be defined as a space locally structured by a bundle pseudogroup. The equivalence relation $u \sim v$ if $f(u) = (x, y)$ and $f(v) = (x, y')$, defined locally on the bundle by an arbitrary chart f , is easily verified to be chart-independent; the equivalence classes of this relation are the *fibers*, each of which is structured by the structure group, and the quotient with respect to the relation is the *base space*, coordinatised locally by the quotients of the charts for the bundle and hence locally structured by the pseudogroup \mathcal{G}_T .

4 Physical theories on structured spaces

Let us return to the scalar field. The Kleinian language of structured spaces allows us to define that theory as follows: it is a real function on a structured space, satisfying equation (1) with respect to any one of the charts of that structured space, and where the structure group of that space is the dynamical symmetry group of that equation. (If we wanted to leave the global topology of the theory open, we could stipulate simply that the function satisfies this equation with respect to any one of the charts of a *locally* structured space.)

I want to claim that this presentation of a theory is on just as solid mathematical and physical grounds as the differential-geometric presentations more normally used in foundational discussions. Firstly (and counter-intuitively) it is a *coordinate-independent* specification of the theory. For to say that a statement is coordinate-dependent is to say that its truth value depends on the coordinate system in which it is evaluated. If I say that I am standing at coordinates (0,0,1), without saying which coordinate system I have in mind, or that I have velocity zero, I say something coordinate-dependent. But a statement does not become coordinate-dependent simply because it refers to a coordinate system. If I say that I am standing at latitude 51.7519 N, longitude 1.2578 W in standard longitude-latitude coordinates, I communicate my location as coordinate-independently as if I tell you that I am in Oxford. Similarly, to say that I have velocity zero in a reference frame comoving with respect to Oxford, my statement is as coordinate-independent as if I say that my world line and Oxford's are parallel.

To stress this point, note that there is no prohibition on describing the scalar field in a coordinate system that *isn't* in the preferred set of coordinatisations

⁴This definition requires, therefore, that notions of smoothness are defined in the spaces F and T , and so that each has (at least) the structure of a differentiable manifold.

that structures the space. A coordinate system is just a map from the space to \mathbb{R}^4 ; the scalar field can be described with respect to any such map. Indeed, if we use a set of coordinates X^0, X^1, X^2, X^3 expressible as twice-differentiable functions of some set x^0, x^1, x^2, x^3 of preferred coordinates, elementary calculus tells us that the equation of motion with respect to these coordinates is

$$\eta^{\mu\nu} \left\{ \left(\frac{\partial X^\sigma}{\partial x^\mu} \right) \left(\frac{\partial X^\tau}{\partial x^\nu} \right) \frac{\partial^2 \varphi}{\partial X^\sigma \partial X^\tau} + \frac{\partial^2 X^\sigma}{\partial x^\mu \partial x^\nu} \frac{\partial \varphi}{\partial X^\sigma} \right\} + m^2 \varphi + \frac{\lambda}{3!} \varphi^3 = 0 \quad (17)$$

More elementary calculus tells us that if (and only if) the new coordinates are actually in the original set, this reduces to the original equation.

This is not of the same form as the original equation, of course. But what of it? Kretschmann may have had a point that any theory *can be* cast in a form where its equations of motion are form-invariant, but that is far from saying that a theory *must* be so cast. But in any case, the same moral applies in generally covariant theories. They have preferred coordinate systems too: all the elements of a maximal atlas, which is to say all the coordinate systems where the coordinate functions are smooth. But nothing prohibits us from using a set of coordinates that are continuous but not smooth. Indeed, nothing (beyond the fact that life is short) prohibits us from “coordinatising” a manifold via a wildly discontinuous map from the manifold into \mathbb{R} , if we so choose, say by using an ordinary coordinate system and then applying one of the standard maps from \mathbb{R}^4 onto \mathbb{R} . The equations of motion of the theory will not be simple in such a coordinatisation (to put it mildly) but it’s a free country: we *can* use it if we so choose.

A critic might object: your account is coordinate-*independent*, but the differential-geometry account is coordinate-*free*, in that it does not refer to coordinates at all. And if a “coordinate-free” account of a theory is one that makes no use of functions from spacetime to \mathbb{R}^N in characterising the structure of that spacetime, fair enough: the structured-space account is not coordinate-free. But in the first place, it is not clear why this should matter. The “coordinate systems” to which we refer here — for better or for worse — are not definitionally things to be understood operationally, as ways in which humans label events. They are pieces of mathematical structure, as abstract as any metric or affine connection, and — prior to physical interpretation — as devoid of operational content.

More importantly, if this is what “coordinate-free” means, the differential-geometry account is not coordinate-free either. For as we have seen, the manifolds on which differential geometry is defined are themselves just structured spaces in the Kleinian sense, structured by the diffeomorphism group (or the local diffeomorphism pseudogroup in the case of globally nontrivial manifolds). So if there is something wrong with specifying the structure of a space *a la* Klein, differential geometry does not right this wrong.

Nor is this unique to the case of manifolds. We have also seen that fiber bundles, the workhorses of contemporary mathematical physics, are also locally structured spaces, characterised by classes of coordinatisations just as manifolds are.

The notation of coordinate-free differential geometry hides this dependence, building a hierarchy of geometrically defined entities on the manifold and characterising those entities intrinsically in ways that make no explicit reference to any coordinate system. But reference is there nonetheless. A vector field, for instance, characterised in about the most abstract way available as a derivation of smooth functions, is nonetheless a functional on a space of functions defined as exactly those functions which, in coordinates, are infinitely-often differentiable.

(As a further reminder that differentiable structure is not innocent, it is worth noting that there are uncountably many non-diffeomorphic manifolds all topologically identical to \mathbb{R}^4 (the so-called “exotic \mathbb{R}^4 s— Freedman and Taylor 1986). We are saying something real and substantive — at least mathematically speaking — about a space when we give its differentiable structure.)

Ironically, genuinely coordinate-free theories are somewhat easier to come by in simpler systems. Galilean and Minkowski spacetime, from a structured-space perspective, are affine spaces with additional structure, and affine spaces can be characterised intrinsically as sets on which a vector space acts freely and transitively as an additive group. (Stachel (1993) has stressed this alternative conception of pre-general-relativistic spacetimes and its foundational significance.) But whether or not characterising a space via the action of a group on that space is somehow preferable to characterising it via a group of maps of that space to another space, there is no elementary coordinate-free way to define a differentiable manifold.⁵

5 Beyond real scalar fields

The appeal to symmetries of equations to determine the structure of the space on which they are defined goes beyond the case of the real scalar fields we have discussed so far. In the case of electromagnetism, for instance, the realisation that the Maxwell equations had the Poincaré group as a symmetry was a central part of the discovery of special relativity. But the action of that group is only a symmetry if the field components as well as the spacetime coordinates are transformed in the appropriate way, and ‘the appropriate way’ — mixing electric and magnetic fields in a subtle combination — did not follow purely from the geometry but required the equations themselves. (That is: *why* do the fields transform this way? Because if they do, the form of the equations is invariant.)

The general framework of structured spaces can be extended straightforwardly to more general fields. To begin with, suppose that we have a theory which, in coordinates, is concerned with a field represented by some

⁵It can be defined as a topological space together with a ring of real functions on that space, taken definitionally to be the smooth functions (see, e.g., Nestruev (2003) for details, and Rosenstock, Barrett, and Weatherall (2015) for conceptual discussion in the context of general relativity); but (i) that ring, definitionally, must be locally isomorphic to the algebra of smooth functions on \mathbb{R}^N , so use of \mathbb{R}^N as a coordinatisation of the space is not thereby abandoned; (ii) it’s not clear to me that there is really a substantive conceptual advantage to characterising a space via a preferred family of functions to \mathbb{R}^N rather than via a preferred family of functions to \mathbb{R} .

function

$$f : \mathbb{R}^N \rightarrow \mathbb{R}^K. \quad (18)$$

(In electromagnetism, for instance, $N = 4$ and $K = 6$: fields assign to each quadruple of spacetime coordinates a triple of E -field coordinates and a triple of B -field coordinates.) A *field symmetry* (here, roughly, I follow Olver 2000) is then a bijection of $\mathbb{R}^N \times \mathbb{R}^K$ of the particular form

$$(x, u) \rightarrow (\varphi(x), \psi(x, u)) \quad (19)$$

— that is, the transformation of the field values can depend on the spacetime coordinate but not vice versa. (In the physics literature, the common terminology of *dependent* and *independent* variables reflects this.) Equivalently, for each $x \in \mathbb{R}^N$, $\psi(x, \cdot)$ is a bijection of \mathbb{R}^K , and I write this bijection as ψ^x .⁶

This induces a transformation on fields:

$$f \rightarrow f'; \quad f'(\varphi(x)) = \psi^x \cdot f(x) \quad (20)$$

More succinctly (and in closer conformity with the physics literature), writing $x' = \varphi(x)$ we have

$$f'(x') = \psi^x \cdot f(x). \quad (21)$$

We can now consider a group of such transformations; or, equivalently, the action of an abstractly-characterised group \mathcal{G} *via* such transformations:

$$g \rightarrow (\varphi_g, \psi_g). \quad (22)$$

The group composition rule becomes

$$\varphi_{hg} = \varphi_h \varphi_g \quad \psi_{hg}^x = \psi_h^{\varphi_g(x)} \psi_g^x. \quad (23)$$

We can define the *internal subgroup* \mathcal{I} of \mathcal{G} as the group of elements which keep the independent variable fixed:

$$g \in \mathcal{I} \text{ iff } \varphi_g = \text{id}. \quad (24)$$

The *spacetime group* \mathcal{G}_S defined by the action of \mathcal{G} is then just the quotient of \mathcal{G} by \mathcal{I} ; equivalently, it is the quotient of \mathcal{G} by the equivalence relation

$$g \sim g' \text{ iff } \varphi_g = \varphi_{g'}. \quad (25)$$

And of course, we can consider a structured space modelled on $\mathbb{R}^N \times \mathbb{R}^K$, with this action of \mathcal{G} as its structure group. Any such space will have associated with it another structured space, modelled on \mathbb{R}^N , with \mathcal{G}_S (acting via $g \rightarrow \varphi_g$) as its structure group; we can consider this space, the *base space* \mathcal{S} , as the spacetime on which the fields are defined.

⁶Readers familiar with category theory and/or functional programming will recognise the operation of *currying*.

For the moment, consider the case where the internal symmetry group is trivial, so that the action of an element of \mathcal{G} on spacetime points determines its action on fields. We can distinguish two cases: the *global case*, where ψ_g^x is independent of x , and the *local case*, where ψ_g^x generically depends on x . In the former case, we can think of fields as (represented by) a kind of *function* to some other space $\mathcal{T} \simeq \mathbb{R}^K$; in the latter case, as some kind of *section*, where $f(x)$ lives in its own space $T_x \simeq \mathbb{R}^K$. But in each case, the space in which $f(x)$ lives cannot be thought of as a structured space independent of \mathcal{S} : any coordinatisation of \mathcal{S} determines a unique coordinatisation of T , or of the T_x .

For instance, taking $K = N$, a (contravariant) *vector field* is, by definition, one that transforms like

$$V'^{\mu}(x') = \frac{\partial x'^{\mu}}{\partial x^{\nu}} V^{\nu}. \quad (26)$$

If the spacetime symmetry group is trivial, a vector field is just an arbitrary N -tuple of real functions on R^N . If — as in the cases of Euclidean space, and of Newtonian, Galilean and Minkowski spacetime — it is nontrivial but linear (so that the partial derivatives do not themselves depend on x), then the transformation rule does not depend on the point of spacetime at which it is evaluated, and the vector field can be thought of as a function to a single space, which can be identified with the associated vector space of the spacetime regarded as an affine space. If the spacetime structure group is the full diffeomorphism group, the transformation rule is itself spacetime-dependent, and the vector field at a point must be thought of as defined in a space associated to that point, and which can be identified as the tangent space to that point (so that the whole structured space is the tangent *bundle*). And there are intermediate cases: in Leibnizian or Maxwellian spacetime, for instance, the partial derivatives depend on time but not on space, so that a vector at a point can be thought of as living in a space defined for the entire spatial slice at that point (and which can be identified as the vector space associated to the 3-dimensional affine space of spacetime points at that fixed time).

Tensor fields of more general form can be defined in a way familiar from the mainstream physics literature (see, e.g., Weinberg (1972) or Zee (2013)): for instance, a (1,2) tensor transforms, by definition, like

$$T'_{\alpha\beta}{}^{\mu} = \frac{\partial x'^{\mu}}{\partial x^{\nu}} T_{\sigma\tau}{}^{\nu} \frac{\partial x^{\sigma}}{\partial x'^{\alpha}} \frac{\partial x^{\tau}}{\partial x'^{\beta}} \quad (27)$$

As a slightly different example, taking $K = N^3$, an *affine connection* can be defined by the transformation law

$$T'_{\alpha\beta}{}^{\mu} = \frac{\partial x'^{\mu}}{\partial x^{\nu}} T_{\sigma\tau}{}^{\nu} \frac{\partial x^{\sigma}}{\partial x'^{\alpha}} \frac{\partial x^{\tau}}{\partial x'^{\beta}} - \frac{\partial x^{\mu}}{\partial x'^{\tau}} \frac{\partial^2 x'^{\tau}}{\partial x^{\alpha} \partial x^{\beta}}. \quad (28)$$

If the spacetime coordinate transformations are linear, then the second term vanishes identically, and so the transformation rule for a connection is the same as for a (1,2) tensor — and so in this case, a connection *is* a (1,2) tensor. But in the more general case (for instance, for a differentiable manifold), this term

is nonvanishing and so, by definition, a connection is not a tensor in these cases. A connection can be used to define a derivative operator on vector fields:

$$\nabla_\nu V^\mu = \partial_\nu V^\mu + T_{\nu\alpha}^\mu V^\alpha, \quad (29)$$

and indeed the transformation rule (28) can be derived by requiring $\nabla_\nu V^\mu$ to transform as a (1,1) tensor.

Now consider the case of nontrivial internal symmetries. For a given point x , the *local internal symmetry group* \mathcal{I}_x of x is $\mathcal{I}/\mathcal{K}_x$, where

$$\mathcal{K}_x = \{g \in \mathcal{I} : \psi_g^x = \text{id}^x\}. \quad (30)$$

Given $x, y \in \mathbb{R}^N$, suppose there is some $\xi \in \mathcal{G}$ such that $\varphi_\xi(x) = y$. Then a quick calculation gives $g \in \mathcal{K}_y$ iff $\xi^{-1}g\xi \in \mathcal{K}_x$. It follows that \mathcal{I}_x and \mathcal{I}_y are isomorphic as groups, related by the automorphism $g \rightarrow \xi^{-1}g\xi$. If we assume that the spacetime group is transitive, i. e. for *any* x, y there is such a ξ (something that holds for all the spacetimes we have considered so far, and which just codifies the fact that any point in \mathbb{R}^N can be used as a coordinate), then the local internal symmetry groups of any two points are isomorphic and we can speak unambiguously of ‘the’ local internal symmetry group \mathcal{I}_{loc} .

We can now distinguish two important possibilities (these are mathematically non-exhaustive, but cover all the physical cases of which I am aware):

Global internal symmetries: $\mathcal{I} \simeq \mathcal{I}_{loc}$, so that the action ψ^x of the internal symmetry group at any point determines its action everywhere.

Local internal symmetries: elements of \mathcal{I} are smooth functions from \mathbb{R}^N to \mathcal{I}_{loc} , with the group rule defined pointwise: $hg(x) = h(x)g(x)$.

To visualise the geometry here, first consider the *scalar* case: here $\mathcal{G} = \mathcal{G}_S \times \mathcal{I}$, and the action of (g, h) is

$$(g, h) \rightarrow (\varphi_g, \psi_h^x). \quad (31)$$

In this case, the space of internal degrees of freedom of the field transform trivially under spacetime transformations. If the internal symmetry group is global, fields are then maps from spacetime (structured by \mathcal{G}_S) to an internal space structured by \mathcal{I} . If it is local, fields are instead sections of a fibre bundle over spacetime, with \mathcal{I}_{loc} as structure group. The nonscalar case, in which the various field components transform under both internal and spacetime transformations, is more complicated but still describable in the general framework.

6 Virtues of the structured-space approach

I have argued that the structured space approach to defining a spacetime theory — give the field’s equations in coordinates, and state that the theory is defined on a structured space and that the structure group is the symmetry group of the equations — is conceptually and mathematically well-motivated. But does it have any actual advantages? Here I give several: my purpose is not to argue for its preeminence in all circumstances but simply for its usefulness in some.

Cohesion with physical practice

In physics — and in particular in particle physics — it is absolutely standard to define theories via coordinate systems. While any such theory can usually be cast into a coordinate-free form, doing so can often be complicated and can render obscure pieces of physical reasoning that are fairly transparent from the coordinate-based perspective. This in turn can make it unnecessarily difficult to engage with arguments in the physics literature.

As an illustrative example, consider the case of a four-vector field on Minkowski spacetime. It is normally claimed (e.g., Lifshitz *et al* (1982, Ch.II)) that the only viable equation of motion for such a field (in the absence of interactions) is given (in coordinates) by

$$\partial^\nu \partial_\nu A^\mu - \partial^\mu \partial_\nu A^\nu + m^2 A^\mu = 0. \quad (32)$$

(Here, as usual, $\partial^\mu = \eta^{\mu\nu} \partial_\nu$.) This equation is readily cast into differential-geometric form, on a flat pseudo-Riemannian manifold with metric g and covariant derivative ∇ : since the connection is flat, ∇^2 is a well-defined operator on any vector or tensor field and we have

$$\nabla^2 A - \nabla(\nabla \cdot A) + m^2 A = 0 \quad (33)$$

where A is a vector field, i.e. a section of the tangent bundle.

Prima facie, there is at least one other equation of motion available. We can equally well consider

$$\partial^\nu \partial_\nu A^\mu + m^2 A^\mu = 0, \quad (34)$$

which certainly looks like a dynamical equation for a vector field, and which goes over perfectly happily into differential geometry as

$$\nabla^2 A + m^2 A = 0. \quad (35)$$

So in what sense is this not a legitimate equation for a vector field? In the coordinate-based approach, we need to look at the symmetry group of each equation. The symmetry group of the first is simply the Poincaré group, acting simultaneously on the coordinates and field values:

$$A'^\mu(\Lambda_\beta^\alpha x^\beta + a^\alpha) = \Lambda_\nu^\mu A^\nu(x^\alpha) \quad (36)$$

for arbitrary Lorentz transformation Λ . In the coordinate-based approach this is what it *means* for A^μ to be a vector field. On the other hand, the symmetry group of the second is the product of the Poincaré group with the general linear group $GL(4, \mathbb{R})$, acting like

$$A'^\mu(\Lambda_\beta^\alpha x^\beta + a^\alpha) = R_\nu^\mu A^\nu(x^\alpha) \quad (37)$$

where now R is an arbitrary invertible matrix, independent of Λ . So (again in the coordinate-based approach) this means that A^μ is not a vector field at

all, but rather a scalar field with a 4-dimensional space of internal degrees of freedom (a Higgs scalar, that is).

From a coordinate-free perspective, the “right” way to represent the second theory is not as a section of the tangent bundle, but rather as a function from the manifold to a vector space of internal degrees of freedom. But it is not so trivial to say just what was ‘wrong’ about the representation of the theory as a vector field. After all, it is perfectly meaningful to say, in a coordinate-free way, that the dynamically allowed fields are just those satisfying $(\nabla^2 + m^2)A = 0$. The right way to understand the problem is probably to observe that the spacetime and dynamical symmetries of the theory do not match correctly (cf Earman (1989, pp.46–7)). But the puzzle only arises to begin with because we are determining the geometric representation of the theory before rather than after looking at the dynamical symmetries of the equations of motion.

Identification of absolute structure, and the search for substantive general covariance

In the Anderson-Friedman absolute-objects program, the presence of preferred coordinate systems (preferred, that is, beyond the family of smooth coordinate systems preferred in the definition of a manifold) is taken to indicate the presence of absolute structure (Newtonian spacetime being the paradigm of such structure) and the absolute structure can be identified by casting the theory into generally covariant form and looking for those objects in the theory that are invariant across models (in an appropriate sense, albeit one that has proven awkward to make precise). “Substantive general covariance”, the elusive property which Einstein sought in general relativity, is then to be characterised as the absence of any such absolute objects. From this perspective, leaving a theory in a formulation that makes use of a preferred coordinatisation simply hides the absolute structure from view.

I have already suggested that differential-geometric frameworks are not as coordinate-free as they look, so pushing the Anderson-Friedman program to its logical conclusion might suggest that even the differentiable and topological structures on manifolds ought to be regarded as absolute structure in some sense (a suggestion that periodically arises, usually informally, in conversations about quantum gravity, though not one that has led to concrete results to the best of my knowledge). But leaving that aside, the Anderson-Friedman program seems to build in an *a priori* assumption that the absolute structures to be identified can be naturally characterised in differential-geometric terms.

To expand: given a theory whose equations of motion, given in coordinates, are invariant under smooth transformations, that theory will naturally end up geometrised via differential geometry. But a theory whose invariance group is much smaller might well have a very different geometry. We have seen that the abstract geometries of vector and affine spaces can be characterised via their structure groups; conversely, theories with such structure groups might most naturally be thought of as vector or affine spaces and not as manifolds at all (a point stressed by (Stachel 1993)). While the latter spaces *can* in turn be

realised as certain manifolds, it is less clear that this is the most perspicuous way of realising them.

Now, there may be concrete reasons why (a given version of) the Anderson-Friedman program can justify a particular focus on differential geometry. Perhaps we need to place all the spacetime theories under consideration in a single framework; perhaps we are moved by Kretschmann's point-coincidence argument to privilege topological (or perhaps differential-topological) features of a theory above all others. But the issue is not much discussed and a reader of the literature can be left with the impression that the differential-geometry characterisation of a spacetime theory is used in this program not because of its advantages over other approaches but because it is the only game in town.

Furthermore, even if we are concerned with comparison of general relativity (where the differential-geometry framework is more or less compulsory) with other spacetime theories, differential geometry is not the only way to make the comparison, and hides features as well as revealing them. In particular, from a differential-geometry perspective, the difference between spacetime geometry in general relativity and in its predecessors comes down to whether that geometry is flat or curved. From the alternative perspective, it is more a question of whether it is global or local: that is, whether or not the geometry is *differential*. Describing the global geometries of Newtonian or special-relativistic physics via a locally-defined connection that just happens to be flat has its advantages but obscures this global/local distinction.

Similarly, *one* way to work out what is distinctive about general relativity is to cast it and its predecessors into differential-geometric form and look at which objects are absolute. But another is to cast it and its predecessors into structured-space form and compare the structure groups. From this perspective, what is distinctive about general relativity in this framework is not that it *can* be formulated on a space with as little (Kleinian) structure as the local diffeomorphism group, but that apparently it *must* be so formulated, whereas other theories have straightforward formulations on much more structured spaces. This is not a new idea in the substantive-general-covariance dispute (Norton's extensive review (1993) of that dispute attributes it first (p.828) to Bergmann (1942, p.129)) and has recently been defended by Pooley (2015) ("The idea is that it is the lack of a non-dieomorphism-invariant formulation, rather than the existence of a dieomorphism-invariant formalism, that is the mark of a background-independent theory") but it seems to have received relatively little attention in recent discussions.

Brown's dynamical characterisation of geometry

Harvey Brown (most sustainedly in Brown 2005) has long advocated an approach to spacetime structure whereby the structure of spacetime is derivative on the dynamical symmetries of the equations of motion, in contrast to the (Brown claims) more usual view in philosophy of physics that we specify a dynamical theory by first specifying the spacetime on which it is defined and then setting out equations of motion specified with respect to the structures present

on that spacetime. (See Earman *ibid*, Friedman *ibid*, and Maudlin (2012) for explicit advocacy of this latter view.)

As long as spacetime theories are automatically cast in differential-geometric form, it is hard even to make sense of Brown's program. (The best attempt I know is Pooley 2013, s.6.3.2.) But it is clear that this is not what Brown has in mind: indeed, his discussion makes extensive use of coordinate systems and coordinate transformations and little use of differential geometry. The structured-space way of formulating a dynamical theory offers a theoretical framework in which Brown's basic proposal is easier to state and thus to assess fairly. (I am unclear whether my discussion should count as a development of Brown's position or simply as Brown exegesis; certainly Brown is aware of, and makes use of, the Kleinian conception of spaces in terms of their symmetry groups.)

As a first attempt at realising Brown's position in the structured-space framework, we might try the following (for field theories): there is some coordinatisation of spacetime (or perhaps, as Pooley suggests, of the various parts of the field regarded as an extended body) and of the space of field values, such that, with respect to that coordinatisation, the equations of motion are XYZ. (Mathematically, this is perfectly definable: the theory is given by a family of maps from unstructured space \mathcal{S} to unstructured space \mathcal{F} , each representing a dynamically possible state of affairs. *Which* family of maps? One such that there exist coordinate systems on \mathcal{S} and \mathcal{T} with respect to which the equations of motion are XYZ.) The structure group can then be defined in terms of the equations of motion, and need not be postulated prior to the stipulation of these equations.

That is not to say that Brown's approach is *compulsory* in a presentation of spacetime theories based on structure (pseudo-)groups. I characterised such theories as having structure groups tacitly determined by the equations of motion, but it's mathematically coherent (if perhaps physically obscure) to start with a structured space with group \mathcal{G} and then write down on that space a theory whose dynamical symmetry group is larger than \mathcal{G} . (This is the realisation in the structured-space format of Earman's distinction between spacetime and dynamical symmetries.) It is to say that the structure-group framework at least provides a starting point for discussion of the approach.

Pedagogy

Last but not least, the structured-space way of defining spaces is very often much simpler to explain and work with than the differential-geometry approach. To characterise Newtonian spacetime — a conceptually fairly simple and intuitive spacetime — even in qualitative differential-geometric terms takes Friedman (1983, pp.71–78) some eight pages. To do it properly requires familiarity with mathematics generally studied at graduate level. In contrast, it takes about a line to define the spacetime in terms of its symmetry group, and that line needs no mathematics beyond a little linear algebra.

Similarly, the five prerelativistic spacetimes discussed in section 2 can be defined in the structured-space approach simply by giving their structure groups,

and the relations between them can be seen simply in the fact that each structure group contains its predecessors as subgroups. To define these spaces through differential geometry is in each case quite nontrivial and unintuitive.⁷ (The most elegant differential-geometric definition I know of Maxwellian spacetime, for instance (Weatherall 2014) makes use of two degenerate metrics and an infinite equivalence class of flat affine connections.)

The same can also be seen in the more abstract mathematical spaces discussed in section 2. It is rather striking that the move from a vector space, to an oriented vector space, to a Euclidean vector space (oriented or not), to a vector space with a notion of relative but not absolute distance, is achieved in each case by a simple modification of the structure group. To define these various notions in a coordinate-free way takes work, and often the work can be feel as if it is just a reconstruction of something antecedently understood perfectly well in terms of coordinate systems.

For one more example, consider the case of phase space (understood in its own right and not as derivative on a configuration space). Foundational definitions of phase space (e.g. Arnol'd 1989; Abraham and Marsden 1978) typically define phase space as a symplectic manifold: a differentiable manifold of even dimension equipped with a closed nondegenerate two-form ω . The dynamics on that phase space is then defined as associating to any smooth function F on the manifold a vector field X_F , defined by

$$\omega(X_F, \cdot) = dF. \quad (38)$$

The Poisson bracket of two functions is defined by

$$X_{\{F,G\}} = [X_F, X_G]_{LB} \quad (39)$$

where $[\cdot, \cdot]_{LB}$ denotes the Lie bracket.

This is undeniably concise and elegant, especially for those familiar with differential forms. It is also strikingly unfamiliar even to those who have studied abstract classical mechanics in more familiar forms. Fairly significant work is required to relate the above definition to the usual talk of q and p .

From a structured-space starting point, we instead define the Poisson bracket on R^{2N} in an utterly familiar way: using q^1, \dots, q^N to label the first N coordinates and p_1, \dots, p_N the second N , we define

$$\{F, G\} = \frac{\partial F}{\partial q^i} \frac{\partial G}{\partial p_i} - \frac{\partial F}{\partial p_i} \frac{\partial G}{\partial q^i}. \quad (40)$$

A canonical transformation is then any smooth map of R^{2N} to itself that preserves the Poisson bracket, i.e. f is canonical iff $\{F \cdot f, G \cdot f\} = \{F, G\} \cdot f$ for all F, G . The pseudogroup of local canonical transformations consists of all local diffeomorphisms that are canonical transformations on their connected

⁷I concede that simplicity and intuitiveness in matters of pedagogy can be somewhat subjective; not all share my intuitions here!

components. And a phase space is a space locally structured by canonical transformations. (This definition also has the advantage that it makes explicit that any two phase spaces of the same dimension are locally isomorphic. This result (Darboux's theorem) is far from obvious in the symplectic manifold approach.)

This is not to say that the abstract symplectic definition does not sometimes do important foundational work (it is crucial, for instance, in Wald's (1994) discussion of quantization). But adopting that definition simply out of a misplaced conviction that mathematical rigor demands it just seems to serve to make one's arguments less accessible than they might be.

Of course, true mastery of general relativity, or classical mechanics, or differential geometry requires a proper understanding of coordinate-free definitions of concepts. But then true mastery also requires easy familiarity with coordinate-based approaches. In exegesis, meanwhile, one ought to use whichever approach is clearer and easier to understand. Often (by no means always) that is a coordinate-transformation-based and not coordinate-free approach.

7 Kleinian vs. Riemannian conceptions of geometry

I have noted that much of the contemporary foundational literature on spacetime uncritically assumes that differential geometry is the only sensible way to characterise spacetime structure. An important exception is Norton (2001, 1999, 1993), who explicitly discusses Klein's Erlangen program and indeed attributes much of the historical controversy over substantive general covariance to confusion between transformation-based and differential-geometric ways of characterising geometry.

Norton, however, regards the Kleinian approach to geometry as essentially superseded in contemporary spacetime physics. Specifically, he draws a contrast between two rival programs for the characterisation of geometry (as of the early 20th century): Klein's, in which geometry is characterised via the invariance groups of the geometry under transformations, and Riemann's, in which geometry is characterised via metric tensors and similar differential-geometric objects. As Norton sees it, the move from special to general relativity is really a move from a Kleinian to a Riemannian conception of spacetime geometry. (He does not claim originality for this observation, and attributes it (Norton 1993, p.832) first to Cartan (1927).)

But it seems too strong to say that geometry *simpliciter* in modern physics is Riemannian in character. For one thing, we have seen that even a general-relativistic spacetime still has a Kleinian geometry, given by its atlas, and its Riemannian geometry is not even definable until that Kleinian geometry is in place. Furthermore, other spaces used in contemporary physics have significant Kleinian geometry, as we have seen from our consideration of fiber bundles. Even some contemporary formulations of general relativity (such as those on spinor or $SO(3,1)$ vector bundles; cf Rovelli (2004) and references therein) have

a Klein geometry significantly richer than that defined by smooth maps. And of course, many of the more abstract spaces used in physics — vector spaces, in particular — may have manifold structure but are not defined *as* manifolds with additional structure; their geometries are Kleinian if they are either.

Rather than one conception of geometry having won out in modern physics, we actually have peaceful coexistence. Spaces used in physics pretty much always have a Kleinian geometry. Many also have a Riemann-type geometry. Sometimes, indeed, the same geometry can be characterised in Kleinian or Riemann-type ways: symplectic or canonical-transformation definitions of phase space; differential-geometric or structured-space characterisation of classical spacetimes.

But which is the *true* geometry? In a particular physical context this question may be well posed: Knox 2013, 2014, for instance, building on Brown (*ibid*, characterises spacetime geometry functionally in terms of inertial structure, and from this point of view spacetime geometry is realised *a la Klein* in non-gravitational theories and *a la Riemann* in general relativity, just as Norton (and Cartan) observed. But abstracted from such particular contexts, the question of which is the ‘true’ geometry is largely a question of semantics. Both Riemann’s and Klein’s programs continue to play a major part in contemporary mathematics, and *a fortiori* in those mathematical tools used in physics.

8 Conclusion

It is generally fine, and often actively useful, to characterise mathematical spaces via classes of preferred coordinatisations of these spaces. Indeed, differentiable manifolds themselves are standardly so characterised. It is, equally, generally fine, and often actively useful, to specify dynamical equations in physics via coordinate expressions.

Differential geometry is an extraordinarily powerful tool, and the coordinate-free discipline in differential geometry has achieved remarkable results. This paper advocates a move away from an insistence that all geometric structure in foundational work is described via differential geometry, and from the eschewal of coordinate systems, but it does not advocate an abandonment of these tools. It is, more modestly, a call for pluralism. There are other tools, too, and they are on a perfectly sound conceptual and mathematical footing. Aspects of modern physical theories, and their relations to older theories, may be best understood through coordinate-free differential geometry; other aspects may best be understood through explicit use of coordinates. Or, to put it another way: coordinate-free and coordinate-transformation-based approaches are both valid coordinatisations of the underlying conceptual questions that we seek to answer, and which coordinatisation is best depends on the particular question being asked.

Acknowledgements

This paper has benefitted greatly from conversations with Erik Curiel, Neil Dewar, Eleanor Knox, Oliver Pooley, Simon Saunders, and in particular, and over many years, Harvey Brown.

References

- Abraham, R. and J. E. Marsden (1978). *Foundations of Mechanics*. Cambridge, Mass.: Perseus Books.
- Anderson, J. L. (1964). Relativity principles and the role of coordinates in physics. In *Gravitation and Relativity*, pp. 175–194. New York: W.A. Benjamin.
- Anderson, J. L. (1967). *Principles of Relativity Physics*. New York: Academic Press.
- Anderson, J. L. (1971). Covariance, invariance and equivalence: a viewpoint. *General Relativity and Gravitation 2*, 161–172.
- Arnol'd, V. (1989). *Mathematical Methods of Classical Mechanics* (2nd ed.). Berlin: Springer.
- Berestetskii, V., E. Lifshitz, and L. Pitaevskii (1982). *Landau and Lifshitz Course of Theoretical Physics Volume 4: Quantum Electrodynamics* (2nd ed.). Oxford: Butterworth-Heinemann.
- Bergmann, P. G. (1942). *Introduction to the Theory of Relativity*. New York: Dover.
- Brown, H. R. (2005). *Physical Relativity*. Oxford: Oxford University Press.
- Cartan, E. (1904). Sur la structure des groupes infinis de transformations. *Annales Scientifiques de l'Ecole Normale Supérieure 21*, 153–206.
- Cartan, E. (1927). *L'enseignement Mathématique 26*, 200–225.
- Earman, J. (1970). Who's afraid of absolute space? *Australasian Journal of Philosophy 48*, 287–319.
- Earman, J. (1974). Covariance, invariance, and the equivalence of frames. *Foundations of Physics 4*, 267–289.
- Earman, J. (1989). *World Enough and Space-Time*. Cambridge, Massachusetts: MIT Press.
- Earman, J. and M. Friedman (1973). The meaning and status of Newton's law of inertia and the nature of gravitational forces. *Philosophy of Science 40*, 329–359.
- Freedman, M. H. and L. R. Taylor (1986). A universal smoothing of four-space. *Journal of Differential Geometry 24*, 69–78.
- Friedman, M. (1983). *Foundations of Space-Time Theories*. Princeton: Princeton University Press.

- Klein, F. (1892). A comparative review of recent researches in geometry. *Bulletin of the New York Mathematical Society* 2, 215–249. Translation by M. W. Haskell. Online version available at <http://arxiv.org/abs/0807.3161>.
- Knox, E. (2013). Effective spacetime geometry. *Studies in the History and Philosophy of Modern Physics* 44, 346–356.
- Knox, E. (2014). Newtonian spacetime structure in light of the equivalence principle. *British Journal for the Philosophy of Science* 65, 863–880.
- Kobayashi, S. and K. Nomizu (1963). *Foundations of Differential Geometry*. New York: John Wiley and Sons.
- Lang, S. (1999). *Fundamentals of Differential Geometry*. Springer-Verlag.
- Malament, D. (2012). *Topics in the Foundations of General Relativity and Newtonian Gravitation*. Chicago University Press.
- Maudlin, T. (2012). *Philosophy of Physics: Space and Time*. Princeton: Princeton University Press.
- Misner, C. W., K. S. Thorne, and J. A. Wheeler (1973). *Gravitation*. New York: W.H. Freeman and Company.
- Nakahara, M. (1991). *Geometry, Topology and Physics*. Bristol: Institute of Physics Publishing.
- Nestruev, J. (2003). *Smooth manifolds and observables*. Berlin: Springer.
- Norton, J. (1993). General covariance and the foundations of general relativity: eight decades of dispute. *Reports on Progress in Physics* 56.
- Norton, J. (1999). Geometries in collision: Einstein, Klein and Riemann. In J. J. Gray (Ed.), *The Symbolic Universe: Geometry and Physics 1890-1930*, pp. 128–144. Oxford: Oxford University Press.
- Norton, J. (2001). Einstein’s triumph over the spacetime coordinate system: A paper presented in honor of Roberto Toretti.
- Olver, P. J. (2000). *Applications of Lie Groups to Differential Equations* (2nd ed.). Berlin: Springer.
- Pitts, J. B. (2006). The Anderson-Friedman absolute objects program: Several successes, one difficulty. Available online at <http://philsci-archive.pitt.edu/3005>.
- Pooley, O. (2013). Substantialist and relationalist approaches to spacetime. In R. Batterman (Ed.), *Oxford Handbook of the Philosophy of Physics*. Oxford University Press.
- Pooley, O. (2015). Background independence, diffeomorphism invariance, and the meaning of coordinates. In D. Lehmkuhl (Ed.), *Towards a Theory of Spacetime Theories*. Birkhauser.
- Rosenstock, S., T. W. Barrett, and J. O. Weatherall (2015). On Einstein algebras and relativistic spacetimes. *Studies in the History and Philosophy of Modern Physics* 52, 309–316.

- Rovelli, C. (2004). *Quantum Gravity*. Cambridge: Cambridge University Press.
- Saunders, S. (2013). Rethinking Newton's *Principia*. *Philosophy of Science* 80, 22–48.
- Stachel, J. (1993). The meaning of general covariance. In J. Earman, A. Janis, G. Massey, and N. Rescher (Eds.), *Philosophical Problems of the Internal and External Worlds: Essays on the philosophy of Adolf Grunbaum*, pp. 129–160. Pittsburgh: University of Pittsburgh Press.
- Trautman, A. (1966). Comparison of Newtonian and relativistic versions of space-time. In B. Hoffman (Ed.), *Perspectives in Geometry and Relativity: essays in honour of V. Hlavat'y*, pp. 413–425. Indiana University Press.
- Veblen, O. and J. H. C. Whitehead (1932). *The Foundations of Differential Geometry*. Cambridge: Cambridge University Press.
- Wald, R. M. (1984). *General Relativity*. Chicago: University of Chicago Press.
- Wald, R. M. (1994). *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*. Chicago: University of Chicago Press.
- Weatherall, J. O. (2014). Maxwell-Huygens, Newton-Cartan, and Saunders-Knox spacetimes. Forthcoming in *Philosophy of Science*. Available online at <http://philsci-archive.pitt.edu/11526/>.
- Weinberg, S. (1972). *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. New York: John Wiley and sons.
- Zee, A. (2013). *Einstein Gravity in a Nutshell*. Princeton, NJ: Princeton University Press.