Remarks on the Notions of General Covariance and Background Independence

D. Giulini

Max-Planck-Institut für Gravitationsphysik, Am Mühlenberg 1,
14476 Golm/Potsdam
domenico.guiliniaei.mpg.de

1 Introduction

It is a widely shared opinion that *the* most outstanding and characteristic feature of general relativity is its manifest *background independence*. Accordingly, those pursuing the canonical quantization programme for general relativity see the fundamental virtue of their approach in precisely this preservation of 'background independence' (cf. Kiefer's and Thiemann's contributions). Indeed, there is no disagreement as to the background dependence of competing approaches, like the perturbative spacetime approach[1] (see the contribution by Lauscher and Reuter) or string theory (see the contribution by Louis, Mohaupt, and Theisen, in particular their Sect. 10). Accordingly, many string theorists would subscribe to the following research strategy:

> Seek to make progress by identifying the background structure in our theories and removing it, replacing it with relations which evolve subject to dynamical laws. ([18], p. 10).

But how can we reliably identify background structures?

There is another widely shared opinion according to which the principle of *general covariance* is devoid of any physical content. This was first forcefully argued for in 1917 by Erich Kretschmann [11] and almost immediately accepted by Einstein [20] (Vol. 7, Doc. 38, p. 39), who from then on seemed to have granted the principle of general covariance no more physical meaning than that of a formal heuristic concept.

From this it appears that it would not be a good idea to define 'background independence' via 'general covariance', for this would not result in a

[1] Usually referred to as the 'covariant approach', since perturbative expansions are made around a maximally symmetric spacetime, like Minkowski or DeSitter spacetime, and the theory is intended to manifestly keep covariance under this symmetry group (i.e. the Poincaré or the DeSitter group), not the diffeomorphism group!

D. Giulini: *Remarks on the Notions of General Covariance and Background Independence*, Lect. Notes Phys. **721**, 105–120 (2007)
DOI 10.1007/978-3-540-71117-9_6

physically meaningful selection principle that could effectively guide future research. What would be a better definition? 'Diffeomorphism invariance' is the most often quoted candidate. What precisely is the difference between general covariance and diffeomorphism invariance, and does the latter really improve on the situation? These are the questions to be discussed here. For related and partially complementary discussions, which also give more historical details, we refer to [14, 15] and [4] respectively.

As a historical remark we recall that Einstein quite clearly distinguished between the *principle of general relativity* (PGR) on one hand, and the *principle of general covariance* (PGC) on the other. He proposed that the formal PGC would imply (but not be equivalent to) the physical PGR. He therefore adopted the PGC as a heuristic principle, guiding our search for physically relevant equations. But how can this ever work if Kretschmann is right and hence PGC devoid of any physical content? Well, what Kretschmann precisely said was that *any* physical law can be rewritten in an equivalent but generally covariant form. Hence general covariance alone cannot rule out any physical law. Einstein maintained that it did if one considers the aspect of 'formal simplicity'. Only those expressions which are formally 'simple' after having been written in a generally covariant form should be considered as candidates for physical laws. Einstein clearly felt the lack for any good definition of formal 'simplicity', hence he recommended to experience it by comparing general relativity to a generally covariant formulation of Newtonian gravity (then not explicitly known to him), which was later given by Cartan [5, 6] and Friedrichs [9] and which did not turn out to be outrageously complicated, though perhaps somewhat unnatural. In any case, one undeniably feels that this state of affairs is not optimal.

2 Attempts to Define General Covariance and/or Background Independence

A serious attempt to clarify the situation was made by James Anderson [2, 3], who introduced the notion of *absolute structure* which here we propose to take synonymously with background independence. This attempt will be discussed in some detail below. Before doing this we need to clarify some other notions.

2.1 Laws of Motion: Covariance versus Invariance

We represent spacetime by a tuple (M, g), where M is a four-dimensional infinitely differentiable manifold and g a Lorentzian metric of signature $(+,-,-,-)$. The global topology of M is not restricted a priori, but for definiteness we shall assume a product-topology $\mathbb{R} \times S$ and think of the first factor as time and the second as space (meaning that g restricted to the tangent spaces of the submanifolds $S_t := \{t\} \times S$ is negative definite and positive definite along $\mathbb{R}_p := \mathbb{R} \times \{p\}$). Also, unless stated otherwise, the Lorentzian metric g is assumed to be at least twice continuously differentiable. We will generally not need to assume (M, g) to be geodesically complete.

Being a C^∞-manifold, M is endowed with a maximal atlas of coordinate functions on open domains in M with C^∞-transition functions on their mutual overlaps. Transition functions relabel the points that constitute M, which for the time being we think of as recognizable entities, as mathematicians do. (For physicists these points are mere 'potential events' and do not have an obvious individuality beyond an actual, yet unknown, event that realizes this potentiality.) Different from maps between coordinate charts are global diffeomorphisms on M, which are C^∞ maps $f : M \rightarrow M$ with C^∞ inverses $f^{-1} : M \rightarrow M$. Diffeomorphisms form a group (multiplication being composition) which we denote by Diff(M). Diffeomorphisms act (mostly, but not always, naturally) on geometric objects representing physical entities, like particles and fields.[2] The transformed geometric object has then to be considered a priori as a *different* object on the *same* manifold (which is not meant to imply that they are necessarily physically distinguishable in a specific theoretical context). This is sometimes called the 'active' interpretation of diffeomorphisms to which we will stick throughout.

Structures that obey equations of motion are, e.g., particles and fields. Classically, a structureless *particle* (no spin etc.) is mathematically represented by a map *into* spacetime:

$$\gamma : \mathbb{R} \rightarrow M \ , \tag{1}$$

such that the tangent vector-field $\dot{\gamma}$ is everywhere timelike, i.e. $g(\dot{\gamma}, \dot{\gamma}) > 0$. Other structures that are also represented by maps *into* spacetime are strings, membranes, etc.

A *field* is defined by a map *from* spacetime, i.e.

$$\Phi : M \rightarrow V \tag{2}$$

where V is some vector space (or, slightly more general, affine space, to include connections). To keep the main argument simple we neglect more general situations where fields are sections in non-trivial vector bundles or non-linear target spaces.

Let γ collectively represent all structures given by maps into spacetime and Φ collectively all structures represented by maps from spacetime. Equations of motions usually take the general symbolic form

$$\mathcal{F}[\gamma, \Phi, \Sigma] = 0 \tag{3}$$

which should be read as equation for γ, Φ *given* Σ.

[2] For example, diffeomorphisms of M lift naturally to any bundle associated to the bundle of linear frames and hence act naturally on spaces of sections in those bundles. In particular, these include bundles of tensors of arbitrary ranks and density weights. On the other hand, there is no natural lift to, e.g., spinor bundles, which are associated to the bundle of *orthonormal* frames (which are only naturally acted upon by isometries, but not by arbitrary diffeomorphisms).

Σ represents some *non-dynamical* structures on M. Only if the value of Σ is prescribed do we have definite equations of motions for (γ, Φ). This is usually how equations of motions are presented in physics: solve (3) for (γ, Φ), *given* Σ. Here only (γ, Φ) represent physical 'degrees of freedom' of the theory to which alone observables refer (or out of which observables are to be constructed). By 'theory' we shall always understand, amongst other things, a definite specification of degrees of freedom and observables.

The group Diff(M) acts on the objects (γ, Φ) (here we restrict the fields to tensor fields for simplicity) as follows:

$$(f, \gamma) \rightarrow f \cdot \gamma := f \circ \gamma \qquad \text{for particles etc.}\,, \tag{4a}$$

$$(f, \Phi) \rightarrow f \cdot \Phi := D(f_*) \circ \Phi \circ f^{-1} \qquad \text{for fields etc.}\,, \tag{4b}$$

where D is the representation of $GL(4,\mathbb{R})$ carried by the fields. In addition, we require that the non-dynamical quantities Σ to be geometric objects, i.e. to support an action of the diffeomorphism group.

Definition 1. *Equation (3) is said to be* ***covariant*** *under the subgroup* $G \subseteq$ Diff(M) *iff*[3] *for all* $f \in G$

$$F[\gamma, \Phi, \Sigma] = 0 \Leftrightarrow F[f \cdot \gamma\,,\, f \cdot \Phi\,,\, f \cdot \Sigma] = 0\,. \tag{5}$$

Definition 2. *Equation (3) is said to be* ***invariant*** *under the subgroup* $G \subseteq$ Diff(M) *iff for all* $f \in G$

$$F[\gamma, \Phi, \Sigma] = 0 \Leftrightarrow F[f \cdot \gamma\,,\, f \cdot \Phi\,,\, \Sigma] = 0\,. \tag{6}$$

Note the difference: in Definition 2 the non-dynamical structures Σ are the same on both sides of the equation, whereas in Definition 1 they are allowed to be also transformed by $f \in$ Diff(M). Covariance merely requires the equation to 'live on the manifold', i.e. to be well defined in a differential-geometric sense, whereas an invariance is required to transform solutions to the equations of motions to solutions of the *very same* equation,[4] which is a much more restrictive condition.

As a simple example, consider the vacuum Maxwell equations on a fixed spacetime (Lorentzian manifold (M, g)):

$$dF = 0\,, \tag{7a}$$

$$d \star F = 0\,, \tag{7b}$$

[3] I use 'iff' as an abbreviation for 'if and only if'.

[4] In the mathematical literature this is called a symmetry (of the equation). We wish to avoid the term 'symmetry' here altogether because that – in our terminology – is reserved for a further distinction of invariances into *symmetries*, which change the physical state, and *redundancies* (gauge transformations) which do not change the physical state. Here we will not need this otherwise very important distinction.

where F denotes the 2-form of the electromagnetic field and d the exterior differential. The $\star$ denotes the (linear) 'Hodge duality' map, which in components reads

$$\star F_{\mu\nu} = \tfrac{1}{2}\varepsilon_{\mu\nu\alpha\beta}F^{\alpha\beta}\,, \tag{8}$$

and which depends on the background metric g through ε and the operation of raising indices: $F^{\alpha\beta} := g^{\alpha\mu}g^{\beta\nu}\,F_{\mu\nu}$.[5] The system (7) is clearly Diff(M)-covariant since it is written purely in terms of geometric structures on M and makes perfect sense as equation on M. In particular, given any diffeomorphisms f of M, we have that $f \cdot F$ satisfies (7a) iff F does. But it is *not* likewise true that $d \star F = 0$ implies $d \star f \cdot F = 0$. In fact, it may be shown[6] that this is true iff f is a conformal isometry of the background metric g, i.e. $f \cdot g = \lambda\, g$ for some positive real-valued function λ on M. Hence the system (7) is not Diff(M)-invariant but only G-invariant, where G is the conformal group of (M, g).

2.2 Triviality Pursuit

Covariance Trivialized (Kretschmann's Point)

Consider the ordinary 'non-relativistic' diffusion equation for the $\mathbb{R}$-valued field ϕ (giving the concentration density):

$$\partial_t\phi = \kappa\Delta\phi\,. \tag{9}$$

This does not look Lorentz covariant, let alone covariant under diffeomorphisms. This changes if it is rewritten in the following form

$$\{n^\mu\nabla_\mu - \kappa(n^\mu n^\nu - g^{\mu\nu})\nabla_\mu\nabla_\nu\}\,\phi = 0\,. \tag{10}$$

Here $g^{\mu\nu}$ are the contravariant components of the spacetime metric (recall that we use the 'mostly minus' convention for its signature), ∇_μ is the associated Levi-Civita covariant derivative, and n^μ is a normalized covariant-constant timelike vector field which gives the preferred flow of time encoded in (9) (i.e. on scalar fields $\partial_t = n^\mu\nabla_\mu$). Equation (10) has the form (3) with no γ, $\Phi = \phi$, and $\Sigma = (g^{\mu\nu}, n^\mu)$ and is certainly diffeomorphism covariant in the sense of Definition 1. The largest invariance group – in the sense of Definition 2 – is given by that subgroup of Diff(M) whose elements stabilize the non-dynamical structures Σ. We write

$$\mathrm{Stab}_{\mathrm{Diff(M)}}(\Sigma) = \{f \in \mathrm{Diff(M)} \mid f \cdot \Sigma = \Sigma\} \tag{11}$$

[5] Note that in 3+1 dimensions this means that the $\star$ operation only depends on the conformal equivalence class of g, since $g^{\alpha\beta}g^{\gamma\delta}\sqrt{|\det\{g_{\mu\nu}\}|}$ is invariant under $g_{\mu\nu} \mapsto \Omega^2\,g_{\mu\nu}$. Accordingly, in this case, it is only the conformal equivalence class of g and not g itself that should be identified with Σ.

[6] This is true in 1+3 dimensions. In other dimensions higher than two, f must even be an isometry of g.

In our case, $\text{Stab}_{\text{Diff(M)}}(g)$ is the 10-parameter Poincaré group. In addition, f stabilizes n^μ if it is in the 7-parameter subgroup $\mathbb{R} \times E(3)$ of time translations and spatial Euclidean motions.

This example already shows (there will be more below) how to proceed in order to make any theory covariant under Diff(M). As already noted, Diff(M)-covariance merely requires the equation to be well defined in the sense of differential geometry, i.e. it should live on the manifold. It seems clear that any equation that has been written down in a special coordinate system on M (like (9)) can also be written in a Diff(M)-covariant way by introducing the coordinate system – or parts of it – as background geometric structure. This is, in more modern terms, the formal core of the critique put forward by Erich Kretschmann in 1917 [11].

Invariance Trivialized

Given that an equation of the form (3) is already G-covariant, we can equivalently express the condition of being G-invariant by

$$F[\gamma, \Phi, \Sigma] = 0 \Leftrightarrow F[\gamma, \Phi\,,\, f \cdot \Sigma] = 0\,, \quad \forall f \in G\,, \tag{12}$$

i.e. any solution of the equation parameterized by Σ is also a solution of the *different* equation parameterized by $f \cdot \Sigma$. Evidently, the more non-dynamical structures there are, the more difficult it is to satisfy (12). In generic situations it will only be satisfied if $G = \text{Stab}_{\text{Diff(M)}}(\Sigma)$. Hence, in distinction to the covariance group, increasing the amount of structures of the type Σ cannot enlarge the invariance group. The case of the largest possible invariance group deserves a special name:

Definition 3. *Equation (3) is called* ***diffeomorphism invariant*** *iff it allows* Diff(M) *as invariance group.*

In view of (12), the requirement of Diff(M)-invariance can be understood as a strong limit on the amount of non-dynamical structure Σ. Generically it seems to eliminate any Σ, i.e. the theory should contain no non-dynamical background fields whatsoever. Intuitively this is what background independence stands for. Conversely, any Diff(M)-covariant theory without non-dynamical fields is trivially Diff(M)-invariant. Hence it seems sensible to simply identify 'Diff(M)-invariance' and 'background independence', and this is what most working physicists seem to do.

But this turns out to be too simple. The origin of the difficulty lies in our distinction between dynamical and non-dynamical structures, which turns out not to be sufficiently sharp. Basically we just said that a structure (γ or Φ) was dynamical if it had no a priori prescribed values, but rather obeyed some equations of motion. We did not say what qualifies an equation as an 'equation of motion'. Can it just be *any* equation? If yes then we immediately object that there exists an obvious strategy to trivialize the requirement of Diff(M)-invariance: just let the values of Σ be determined by equations rather than by hand; in this way they formally become 'dynamical' variables and no non-dynamical quantities are left. Formally this corresponds to the replacement scheme

$$\Phi \mapsto \Phi' = (\Phi, \Sigma) \,, \tag{13a}$$

$$\Sigma \mapsto \Sigma' = \emptyset \,, \tag{13b}$$

so that invariance now becomes as trivial as the requirement of covariance.

More concretely, reconsider the examples (7) and (10) above. In the first case we now regard the spacetime metric g as 'dynamical' field for which we add the condition of flatness as 'equation of motion':

$$\mathbf{Riem}[g] = 0 \,, \tag{14}$$

where **Riem** denotes the Riemann tensor of (M, g). In the second case we regard g as well as the timelike vector field n as 'dynamical' and add (14) and the two equations

$$g(n, n) = c^2 \,, \tag{15a}$$

$$\nabla n = 0 \,. \tag{15b}$$

In this fashion we arrive at diffeomorphism invariant equations. But do they really represent the same theory as the one we originally started from? For example, are their solution spaces 'the same'? Naively the answer is clearly 'no', simply because the reformulated theory has – by construction – a much larger space of solutions. For any solution Φ of the original equations $F[\Phi, \Sigma] = 0$, where Σ is fixed, we now have the whole Diff(M)-orbit of solutions, $\{(f \cdot \Phi, f \cdot \Sigma) \mid f \in \mathrm{Diff(M)}\}$ of the new equations, which treat Σ as dynamical variable. A bijective correspondence can only be established if the transformations f that act non-trivially on Σ (i.e. $f \notin \mathrm{Stab}_{\mathrm{Diff(M)}}(\Sigma)$) are declared to be *gauge transformations*, so that any two field configurations related by such an f are considered to be physically identical.

If this is done, the simple strategy outlined here suffices to (formally) trivialize the requirement of diffeomorphism invariance. Hence defining background independence as being simple diffeomorphism invariance would also render it a trivial requirement. How could we improve its definition so as to make it a useful notion? This is precisely what Anderson attempted in [3]. He noted the following peculiarities of the reformulation just given:

1. The new fields g or (g, n) obey an autonomous set of equations which does not involve the proper dynamical fields F or ϕ respectively. In contrast, the equations for the latter *do* involve g or (g, n). Physically speaking, the system whose states are parameterized by the new variables acts upon the system whose states are parameterized by F or ϕ, but not vice versa. An agent which dynamically acts but is not acted upon may well be called 'absolute' – in generalization of Newton's absolute space. Such an absolute agent should be eliminated.

2. The sector of solution space parameterized by g or (g,n) consists of a single diffeomorphism orbit. For example, this means that for any two solutions (ϕ, g, n) and (ϕ', g', n') of (10), (14), and (15) there exists a diffeomorphism f such that $(g', n') = (f \cdot g\,,\ f \cdot n)$. So 'up to diffeomorphisms' there exists only one solution in the (g,n) sector. This is far from true for ϕ: the two solutions ϕ and ϕ' are generally not related by a diffeomorphism. This difference just highlights the fact that the added variables really did not correspond to new degrees of freedom (they were never supposed to) because the added equations were chosen strong enough to maximally fix their values (up to diffeomorphisms).

A closer analysis shows that the first criterion is really too much dependent on the presentation to be generally useful as a necessary condition. Absolute structures will not always reveal their nature by obeying autonomous equations. The second criterion is more promising and actually entered the literature with some refinements as criterion for absolute structures. Before going into this, we will discuss some attempts to disable the trivialization strategies just outlined.

2.3 Strategies Against Triviality

Involving the Principle of Equivalence

As diffeomorphism *covariance* is a rather trivial requirement to satisfy, we will from now on only be concerned with diffeomorphism *invariance*. As we explained, it could be achieved by letting the Σ's 'change sides', i.e. become dynamical structures (γ's and Φ's), as schematically written down in (13). We seek sensible criteria that will limit the number of such renegades. A physical criterion that suggests itself is to allow only those Σ to change sides which are known to correspond to dynamical variables in a wider context. For example, we may allow the spacetime metric g to become formally dynamical, since we know that it describes the gravitational field, even if in the context at hand the self-dynamics of the gravitational field is not relevant and therefore, as a matter of approximation, fixed to some value (e.g. the Minkowski metric). Doing this would render the Maxwell equations (7) (plus the equations for g) diffeomorphism invariant. But this alone would not work for the diffusion equation, where n would still act as a non-dynamical structure.

Hence we see that the requirement to achieve diffeomorphism invariance by at most adjoining g to the dynamical variables is rather non-trivial and connects to Einstein's principle of equivalence. Let us quote Wolfgang Pauli in this context:

> Einen physikalischen Inhalt bekommt die allgemeine kovariante Formulierung der Naturgesetze erst durch das Äquivalenzprinzip, welches zur Folge hat, daß die Gravitation durch die g_{ik} *allein* beschrieben

wird, und daß diese nicht unabhängig von der Materie gegeben, sondern selbst durch die Feldgleichungen bestimmt sind. Erst deshalb können die g_{ik} als *physikalische Zustandsgrößen* bezeichnet werden.[7] ([17], p. 181; the emphases are Pauli's)

Absolute Structures

As already remarked, another strategy to render the requirement of diffeomorphism invariance non-trivial was suggested by Anderson [3] by means of his notion of 'absolute structures'. However, most commentators share the opinion that Anderson did not succeed to give a proper definition of this term. Even worse, some feel that so far nobody has, in fact, succeeded in giving a fully satisfying definition.

To see what is behind this somewhat unhappy state of affairs, let us start with a tentative definition that suggests itself from the discussion given above:

Definition 4 (tentative). *Any field which is either not dynamical, or whose solution space consists of a single* Diff(M)*-orbit, is called an* ***absolute structure****.*

In general terms, let $\mathcal{S}$ denote the space of solutions to a given theory. If the theory is Diff(M)-invariant $\mathcal{S}$ carries an action of Diff(M). The fields can be thought of as parameterising on $\mathcal{S}$. An absolute structure is a parameter which takes the same range of values in each Diff(M) orbit and therefore cannot separate any two of them. If we regard Diff(M) as a gauge group, i.e. that Diff(M)-related configurations are physically indistinguishable, then absolute structures carry no observable content.

Following our general strategy we could now attempt to give a definition of 'background independence':

Definition 5 (tentative). *A theory is called* ***background independent*** *iff its equations are* Diff(M)*-invariant in the sense of Definition 3 and its fields do not include absolute structures in the sense of Definition 4.*

Before discussing these proposal, let us look at some more examples.

2.4 More Examples

Scalar Gravity a la Einstein–Fokker

In 1913, just before the advent of general relativity, Gunnar Nordsröm invented a formally consistent Poincaré-invariant scalar theory of gravity; see,

[7] 'The generally covariant formulation of the physical laws acquires a physical content only through the principle of equivalence, in consequence of which gravitation is described *solely* by the g_{ik} and these latter are not given independently from matter, but are themselves determined by field equations. Only for this reason can the g_{ik} be described as *physical quantities*' ([16], p. 150).

e.g., the survey by von Laue [22]. Shortly after its publication it was pointed out by Einstein and Fokker that Nordström's (second) theory can be presented in a 'covariant' way. Explicitly they said,

> Im folgenden soll dargetan werden, daß man zu einer in formaler Hinsicht vollkommen geschlossenen und befriedigenden Darstellung der Theorie [Nordströms] gelangen kann, wenn man, wie dies bei der Einstein-Grossmannschen Theorie bereits geschehen ist, das invarianten-theoretische Hilfsmittel benutzt, welches uns in dem absoluten Differentialkalkül gegeben ist.[8] ([20], Vol. 4, Doc. 28, p. 321)

The essential observation is this: consider conformally flat metrics:

$$g_{\mu\nu} = \phi^2 \, \eta_{\mu\nu} \, , \tag{16}$$

then the field equation is equivalent to

$$R[g] \; = \; 24\pi G \, g^{\mu\nu} T_{\mu\nu} \, , \tag{17a}$$

where $R[g]$ is the Ricci scalar for the metric g, whereas the equation of motion for the particle becomes the geodesic equation with respect to g:

$$\ddot{x}^\mu + \Gamma^\mu_{\alpha\beta} \dot{x}^\alpha \dot{x}^\beta = 0 \, . \tag{17b}$$

Now, the system (17), considered as equations for the metric g and the trajectory x, is clearly Diff(M)-invariant. But Nordström's theory is equivalent to (17) *plus* (16). Here η is a non-dynamical field so that (16, 17) is only Diff(M)-covariant. According to the general scheme outlined above this could be remedied by letting the metric η be a new dynamical variable whose equation of motion just asserts its flatness:

$$\mathbf{Riem}[\eta] = 0 \, . \tag{18}$$

But then η qualifies as an absolute structure according to Definition 4 and the theory (16, 17, 18) is not background independent. The subgroup $G \subset \mathrm{Diff(M)}$ that stabilizes η is – by definition – the inhomogeneous Lorentz group, which had already been the invariance group of Nordström's theory. So no additional invariance has, in fact, been gained in the transition from Nordström's to the Einstein–Fokker formulation.

Sometimes the absolute structures are not so easy to find because the theory is formulated in such a way that they are not yet isolated as separate field. For example, in the case at hand, (16) and (18) together are clearly equivalent to the single condition that g be conformally flat, which in turn

[8] 'In the following we wish to show that one can arrive at a formally complete and satisfying presentation of the theory [Nordström's] if one uses the methods from the theory of invariants given by the absolute differential calculus, as it was already done in the Einstein–Grossman theory.'

is equivalent to the vanishing of the conformal curvature tensor for g (Weyl tensor):

$$\mathbf{Weyl}[g] = 0 \; . \tag{19}$$

The field $\eta_{\mu\nu}$ has now disappeared from the description and the theory does not explicitly display any absolute structure anymore. But, of course, it is still there; it is now part of the field g. To bring it back to light, make a field redefinition $g_{\mu\nu} \mapsto (\phi, h_{\mu\nu})$ which isolates the part determined by (19); for example,

$$\phi \quad := [-\det\{g_{\mu\nu}\}]^{\frac{1}{8}} \; , \tag{20}$$

$$h_{\mu\nu} := g_{\mu\nu} \; [-\det\{g_{\mu\nu}\}]^{-\frac{1}{4}} \; . \tag{21}$$

Then any two solutions for the full set of equations are such that their component fields $h_{\mu\nu}$ and $h'_{\mu\nu}$ are related by a diffeomorphism. Hence $h_{\mu\nu}$ is an absolute structure.

Clearly there is a rather non-trivial mathematical theory behind the last statement of diffeomorphism equivalence of $h_{\mu\nu}$. We could not have made that statement had we not already been in possession of the full solution theory for (19) which, after all, is a complicated set of non-linear partial differential equations of second order.

A Massless Scalar Field from an Action Principle

Usually we require the equations of motion to be the Euler–Lagrange equations for some associated action principle. Would the somewhat bold strategy to render non-dynamical structures dynamical by adding *by hand* ‘equations of motion’ which fix them to their previous values also work if these added equations were required to be the Euler–Lagrange equations for some common action principle? The answer is by no means obvious, as the following simple example taken from [19] illustrates:

Consider a real massless[9] scalar field in Minkowski space:

$$\Box\phi := \eta^{\mu\nu}\nabla_\mu\nabla_\nu\phi = 0 \; . \tag{22}$$

According to standard strategy the non-dynamical Minkowski metric η is eliminated by introducing the dynamical variable g, replacing η in (22) by g, and adding the flatness condition

$$\mathbf{Riem}[g] = 0 \tag{23}$$

as new equation of motion. Is there an action principle whose Euler–Lagrange equations are (equivalent to) these equations? This seems impossible without

[9] This is just assumed for simplicity. The arguments work the same way if a mass term were included.

introducing yet another field λ (a Lagrange multiplier) whose variation just yields (23). The action would then be

$$S = \tfrac{1}{2}\int dV\ g^{\mu\nu}\nabla_\mu\phi\,\nabla_\nu\phi + \tfrac{1}{4}\int dV\ \lambda^{\alpha\beta\mu\nu}R_{\alpha\beta\mu\nu}\ , \tag{24}$$

where the symmetries of the tensor field λ are that of the Riemann tensor:

$$\lambda^{\alpha\beta\mu\nu} = \lambda^{[\alpha\beta][\mu\nu]} = \lambda^{\mu\nu\alpha\beta}\ . \tag{25}$$

Variation with respect to ϕ and λ yield (22) and (23) respectively, and variation with respect to g gives

$$\nabla_\mu\nabla_\nu\lambda^{\alpha\mu\beta\nu} = T^{\alpha\beta}\ , \tag{26}$$

where $T^{\alpha\beta}$ is the energy–momentum tensor for ϕ. These equations do not give a background independent theory for the fields (ϕ, g, λ) since g is an absolute structure. The solution manifold of the ϕ field is, in fact, the same as before. For this it is important to note that there is an integrability condition resulting from (23,26), namely $\nabla_\alpha T^{\alpha\beta} = 0$, which is however already implied by (22). Hence no extra constraints on ϕ result from (26).

However, the λ field seems to actually add more dimensions to the solution manifold and hence to the observable content of the theory. Indeed, using the Poincaré Lemma in flat space one shows that any divergenceless symmetric 2-tensor $T^{\mu\nu}$ can always be written as in (26), where λ has the symmetries (25). But this does not fix $\lambda^{\mu\alpha\nu\beta}$, so that the set of Diff(M)-equivalence classes of stationary points of (24) is strictly 'larger' than the set of solutions of (22). In other words, the (Diff(M) reduced) phase space for the theory described by (24) is 'larger' then that for (22).[10] A a result we conclude that the reformulation given here does *not* achieve an equivalent Diff(M)-invariant reformulation of (22) in terms of an action principle.

2.5 Problems with Absolute Structures

A first thing to realize from the examples above is that the notion of absolute structure should be slightly refined. More precisely, it should be made local in order to capture the idea that an absolute element in the theory does not represent local degrees of freedom. Rather than saying that a field corresponds to an absolute structure if its solution space consists of a single Diff(M)-orbit, we would like to make the latter condition local:

Definition 6. *Two fields T_1 and T_2 are said to be* ***locally diffeomorphism equivalent*** *iff for any point $p \in M$ there exits a neighbourhood U of p and a diffeomorphism $\phi_U : U \to U$ such that $\phi_U \cdot (T_1\big|_U) = T_2\big|_U$.*

[10] I am not aware of a reference where a Hamiltonian reduction of (24) is carried out.

Note that local diffeomorphism equivalence defines an equivalence relation on the set of fields. Accordingly, following a suggestion of Friedman [7], we should replace the tentative Definition 4 by the following:

Definition 7. *Any field which is either not dynamical or whose solutions are all locally diffeomorphism equivalent is called an* ***absolute structure****.*

In fact, this is what we implicitly used in the discussions above where we slightly oversimplified matters. For example, any two flat metrics g_1, g_2 (i.e. which satisfy $\mathbf{Riem}[g_{1,2}] = 0$) are generally only *locally* diffeomorphism equivalent. Likewise, a conformally flat metric g (i.e. which satisfy $\mathbf{Weyl[g]{=}0}$) is *locally* diffeomorphism equivalent to $f^2\eta$, where f is non-vanishing function and η is a fixed flat metric.

Having corrected this we should also adapt the tentative Definition 5:

Definition 8. *A theory is called* ***background independent*** *iff its equations are* Diff(M)*-invariant in the sense of Definition 3 and its fields do not include absolute structures in the sense of Definition 7.*

So far so good. Is this, then, the final answer? Unfortunately not! The standard argument against *this* notion of absolute structure is that it may render structures absolute that one would normally call dynamical. The canonical example, usually attributed to Robert Geroch [10], makes use of the well-known fact in differential geometry that nowhere vanishing vector fields are always locally diffeomorphism equivalent (see, e.g., Theorem 2.1.9 in [1]). Hence any diffeomorphism-invariant theory containing vector fields among their fundamental field variables cannot be background independent. For example, consider the coupled Einstein–Euler equations for a perfect fluid of density ρ and four-velocity u in spacetime with metric g. This system of equations is Diff(M)-invariant. By definition of a velocity field we have $g(u, u) = c^2$. This means that u cannot have zeros, even if for physical reasons we would usually assume the fluid to be present not everywhere in spacetime, i.e. the support of ρ is a proper subset of spacetime.[11] Then the four velocity of the fluid is an absolute structure, contrary to our physical intention.

I know of two suggestions how to avoid this conclusion in the present example. One is to use the 1-form $u_\mu \, dx^\mu$ rather than the vector field $u^\mu \partial_\mu$ as fundamental dynamical variable for the fluid. The point being that one-form fields are *not* locally diffeomorphism equivalent. For example, a closed (exact) one-form field will always be mapped into a closed (exact) one-form field, and hence cannot be locally diffeomorphism equivalent to a non-closed field. Another suggestion, in fact the only one that I have seen in the literature ([8] p. 59 footnote 9 and [21], p. 99, footnote 8) is to take the energy–momentum density Π rather than u as fundamental variable. To be sure, on

[11] It seems a little strange to be forced to consider velocity fields u in regions where $\rho = 0$, i.e. where there is no fluid matter. Velocity of what? one might ask. In a concrete application this means that we have to extend u beyond the support of ρ and that the physical prediction is independent of that extension.

the support of Π we can think of it as equal to ρu, but on the complement of its support there is no need to define a u. This avoids the unwanted conclusion whenever Π indeed has zeros; otherwise the argument given above for u just applies to Π.

An even simpler argument, which I have not seen in the physics literature, even applies to pure gravity. It rests on the following theorem from differential geometry, an elegant proof of which was given by Moser [12]: given two compact-oriented n-dimensional manifolds V_1 and V_2 with n-forms μ_1 and μ_2 respectively. There exists an orientation-preserving diffeomorphism $\phi : V_1 \rightarrow V_2$ such that $\phi^* \mu_2 = \mu_1$ iff the μ_1-volume of V_1 equals the μ_2-volume of V_2, i.e. iff

$$\int_{V_1} \mu_1 = \int_{V_2} \mu_2 \ . \tag{27}$$

If we take $V_1 = V_2$ to be the closure of an open neighbourhood U in the spacetime manifold M, this theorem implies that the metric volume forms, written in coordinates as

$$\mu = \sqrt{\left| \det[g(\partial_\mu, \partial_\nu)] \right|} \, dx^1 \wedge \cdots \wedge dx^n \ , \tag{28}$$

are locally diffeomorphism equivalent iff they assign the same volume to U. Hence it follows that the metric volume elements modulo constant factors are absolute elements in pure gravity. Note that this implies that for any metric g any point $p \in M$ there is always a local coordinate system $\{x^\mu\}$ in an open neighbourhood U of p such that $\sqrt{|\det[g(\partial_\mu, \partial_\nu)]|} = 1$.

3 Conclusion

Background independence is one of the central strategic issues in discussions on competing approaches to quantum gravity. This clearly emerges from the contributions of Kiefer, Thiemann, Nicolai and Peeters, Lauscher and Reuter, and Louis, Mohaupt, and Theisen to this book. Given the impressive amount of effort that is devoted to analyse the consequences of these different approaches, it seems a little strange to me that the very notion of background independence is tolerated to be in the state of relative elusiveness in which it appears to be. Clearly, in *specific* situations it is usually not difficult to associate a mathematically well-defined meaning to an 'intuitively obvious' interpretation of such a requirement of background independence. But when used as *general* strategic criterion one should, I think, come up with a generally valid and mathematically well-defined definition. I am not aware of such a definition. Attempts were made in the past, but they run into the problems outlined here. Hence the problem must be regarded as an outstanding one.

References

1. Ralph Abraham and Jerrold E. Marsden. *Foundations of Mechanics*. The Benjamin/Cummings Publishing Company, Reading, Massachusetts, second edition, 1978. 117
2. James L. Anderson. Relativity principles and the role of coordinates in physics. In Hong-Yee Chiu and William F. Hoffmann, editors, *Gravitation and Relativity*, pp. 175–194. W.A. Benjamin, Inc., New York and Amsterdam, 1964. 106
3. James L. Anderson. *Principles of Relativity Physics*. Academic Press, New York, 1967. [106, 111, 113]
4. Julian Barbour. On general covariance and best matching. In Craig Callender and Nick Huggett, editors, *Physics Meets Philosophy at the Planck Scale – Contemporary Theories in Quantum Gravity*, pp. 199–212. Cambridge University Press, Cambridge (England), 2001. 106
5. Elie Cartan. Sur les varietes a connexion affine et la théorie de la relativité généralisée. *Annales Scientifiques de l'École Normale Supérieure*, 40:325–412, 1923. 106
6. Elie Cartan. Sur les varietes a connexion affine et la théorie de la relativité généralisée. *Annales Scientifiques de l'École Normale Supérieure*, 41:1–15, 1924. 106
7. Michael Friedman. Relativity principles, absolute objects and symmetry groups. In Patrick Suppes, editor, *Space, Time and Geometry*, pp. 296–320, Dordrecht, Holland, 1973. D. Reidel Publishing Company. 117
8. Michael Friedman. *Foundations of Space-Time Theories*. Princeton University Press, Princeton, New Jersey, 1983. 117
9. Kurt Friedrichs. Eine invariante Formulierung des Newtonschen Gravitationsgesetzes und des Grenzüberganges vom Einsteinschen zum Newtonschen Gesetz. *Mathematische Annalen*, 98:566–575, 1927. 106
10. Roger Jones. The special and general principles of relativity. In P. Barker and C.G. Shugart, editors, *After Einstein*, pp. 159–173, Memphis, USA, 1981. Memphis State University Press. 117
11. Erich Kretschmann. Über den physikalischen Sinn der Relativitätspostulate, A. Einsteins neue und seine ursprüngliche Relativitätstheorie. *Annalen der Physik*, 53:575–614, 1917. [105, 110]
12. Jürgen Moser. On the volume elements on a manifold. *Transactions of the American Mathematical Society*, 120(2):286–294, 1965. 118
13. John Norton. Einstein, Nordström and the early demise of scalar Lorentz-covariant theories of gravitation. *Archive for the History of Exact Sciences*, 45:17–94, 1992.
14. John Norton. General covariance and the foundations of general relativity: Eight decades of dispute. *Reports of Progress in Physics*, 56:791–858, 1993. 106
15. John Norton. General covariance, gauge theories and the Kretschmann objection. In Katherine Brading and Elena Castellani, editors, *Symmetries in Physics: Philosophical Reflections*, pp. 110–123. Cambridge University Press, Cambridge, UK, 2003. 106
16. Wolfgang Pauli. *Theory of Relativity*. Dover Publications, Inc., New York, 1981. Unabridged and unaltered republication of the english translation by G. Field that was first published in 1958 by Pergamon Press. 113

17. Wolfgang Pauli. *Relativitätstheorie.* Springer Verlag, Berlin, 2000. Reprint of the original 'Encyclopädie-Artikel' from 1921 with various additions, including Pauli's own supplementary notes from 1956. Edited and annotated by Domenico Giulini. 113
18. Lee Smolin. The case for background independence. *www.arxiv.org/abs/ hep-th/0507235.* 105
19. Rafael Sorkin. An example relevant to the Kretschmann-Einstein debate. *Modern Physics Letters A*, 17(11):695–700, 2002. 115
20. John Stachel et al., editors. *The Collected Papers of Albert Einstein, Vols. 1–9.* Princeton University Press, Princeton, New Jersey, 1987–2005. [105, 114]
21. Norbert Straumann. *General Relativity.* Springer Verlag, Berlin, 2004. 117
22. Max von Laue. Die Nordströmsche Gravitationstheorie. *Jahrbuch der Radioaktivität und Elektronik*, 14(3):263–313, 1917. 114