12.2. The Einstein-Hilbert action

In our derivation of the Einstein field equations in Section 12.1 we assumed that there is an action with a local Lagrangian that gives rise to the dynamics of the metric field. By exploiting Lovelock's theorem, we managed to derive the equation of motion without ever specifying the Lagrangian explicitly. Since the EFEs are conceptually simple and "inevitable", we should expect the action that gives rise to these equations to be simple and "inevitable" as well:

1 | ≺ Generally covariant action for metric field $g_{\mu\nu}$:

(We omit prefactors for the sake of clarity.)

$$S[g] = \int \overbrace{\mathrm{d}^4x\sqrt{g}}^{\text{Scalar}} \Big[\overbrace{\underbrace{1}_{\sim\partial^0 g} + \underbrace{R}_{\sim\partial^2 g} + \underbrace{R^2 + R_{\mu\nu}R^{\mu\nu} + R_{\mu\nu\sigma\pi}R^{\mu\nu\sigma\pi} + \dots}_{\sim(\partial^2 g)^2}}^{\text{Scalars}} \Big] \tag{12.52}$$

GENERAL RELATIVITY — *Modifications* of GENERAL RELATIVITY

Goal: We want a *second*-order differential equation for $g_{\mu\nu}$ (← 2ND in Section 12.1)

Facts:

- If the Lagrangian only depends on *first* derivatives, the Euler-Lagrange equations include at most *second* derivatives. ☺
- Problem: There is *no* scalar including only first derivatives of the metric! ☹

  This is easy to see: At every point we can choose locally geodesic coordinates where all first derivatives of the metric vanish. Since a scalar is independent of coordinates, this leaves only the trivial possibility that the scalar does not depend on the first derivatives at all, and is therefore constant.
- The next best (and only!) scalar *linear* in *second* derivatives is the Ricci scalar.

  This is our last hope to obtain a (non-trivial) second-order differential equation, since the *linearity* in $\partial^2 g$ makes all terms of potentially third and higher derivatives vanish. ☺

  To understand why, remember (well, probably you don't remember because this is rarely covered in basic courses on classical mechanics) that for a Lagrangian $L(t, q, q', q'')$ that depends on *second* derivatives q'', the Euler-Lagrange equation reads

$$\frac{\partial L}{\partial q} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial q'} + \frac{\mathrm{d}^2}{\mathrm{d}t^2}\frac{\partial L}{\partial q''} = 0\,. \tag{12.53}$$

  If $L(t, q, q', q'') = f(t)q''(t) + \tilde{L}(t, q, q')$ only depends linearly on the second derivative, $\frac{\partial L}{\partial q''}$ does not contain the function $q(t)$, so that there are no derivatives beyond q'' that can show up in the Euler-Lagrange equation.

  That the Ricci scalar R is the *only* (non-trivial) scalar that can be constructed from the metric and its first and second derivatives, and is linear in the latter, has been shown by HERMANN VERMEIL in 1917 [167]; this statement is known as ↑ *Vermeil's theorem*.

2 | The ⁂ *Einstein-Hilbert action*:

These arguments lead us to propose following the simple action:

$$S[g] := \underbrace{\frac{c^3}{16\pi G}}_{(2\kappa c)^{-1}} \int \mathrm{d}^4x \sqrt{g}\,(R - 2\Lambda) \tag{12.54}$$

- The factor 2 in front of the cosmological constant is chosen such that the Euler-Lagrange equations take the conventional form of the Einstein field equations. The global prefactor $\frac{1}{2\kappa c}$ is irrelevant for our current purpose because, first, we are interested in pure gravity (= no matter action $S_g[\phi]$), and second, we are only interested in *classical* equations of motion, i.e., we don't use the action to define a ↓ *path integral* (which would be needed for a theory of *quantum* gravity). The strange additional c in the prefactor $\frac{1}{2\kappa c}$ is necessary for dimensional reasons (→ *below*) and due to our choice to measure time coordinates in units of length: $x^0 = ct$.
- If you want to use an action for a path integral, it must have the *dimension* of an action $E \cdot T = M \cdot L^2 \cdot T^{-1}$, so that the exponent of $\exp\left[\frac{i}{\hbar} S\right] = \exp\left[2\pi i \frac{S}{h}\right]$ is a dimensionless number [$\hbar$ is the reduced Planck's constant, the "quantum of action" (*Wirkungsquantum*) that quantifies the strength of quantum fluctuations].

  A bit of dimensional analysis yields for Eq. (12.54)

$$[S] = [\kappa]^{-1} \underbrace{[c^{-1}\mathrm{d}^4x]}_{L^3 \cdot T} \underbrace{[g]^{\frac{1}{2}}}_{1} \underbrace{[R]}_{L^{-2}} \overset{!}{=} M \cdot L^2 \cdot T^{-1} \;\Rightarrow\; [\kappa] = T^2 \cdot M^{-1} \cdot L^{-1} \tag{12.55}$$

  which is conveniently the dimension of $\kappa = \frac{8\pi G}{c^4}$.
- The Einstein-Hilbert action was introduced by mathematician David Hilbert in 1915 in Ref. [151] – where he independently found the Einstein field equations (essentially at the same time as Einstein, give or take a few days). For a historical account on the race between Einstein and Hilbert see Ref. [152].

3 | Euler-Lagrange equations:

One can now check (→ *below*) that the stationary solutions satisfy the EFE in vacuum Eq. (12.10):

$$\delta S[g] \overset{!}{=} 0 \quad \overset{*}{\Longleftrightarrow} \quad R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} \overset{!}{=} 0 \tag{12.56}$$

¡! Please appreciate this almost magical result: You start by writing down the *simplest non-trivial covariant action* that can be constructed from the metric, *and the Einstein field equations follow*.

This underpins our previous statement that the EFEs are "inevitable" under quite general assumptions. It also illustrates in which sense General Relativity is *simple*: Without cosmological constant (and dropping the unnecessary prefactor), the action of General Relativity in vacuum is

$$S[g] = \int \mathrm{d}^4x \sqrt{g} R\,. \tag{12.57}$$

If this isn't simple and elegant, what is?

The proof of Eq. (12.56) is straightforward but a bit tedious (➲ Problemset 5):

i | With Eq. (11.112) it follows

$$\delta(\sqrt{g}) = -\frac{1}{2}\sqrt{g}g_{\mu\nu}\delta g^{\mu\nu} \tag{12.58}$$

which is the variation of the cosmological term in Eq. (12.56).

ii | The variation of the more complicated first term in Eq. (12.56) yields

$$\delta(\sqrt{g}R) = \delta(\sqrt{g}R_{\mu\nu}g^{\mu\nu}) = \sqrt{g}g^{\mu\nu}\delta R_{\mu\nu} + \left(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}\right)\delta g^{\mu\nu}\sqrt{g} \tag{12.59}$$

where we again made use of Eq. (11.112); the variation $\delta R_{\mu\nu}$ remains to be evaluated.

iii | To do so, we proceed in *locally geodesic coordinates* where the Christoffel symbols vanish and we can use Eq. (10.104) so that

$$R_{\mu\nu} = R^{\lambda}{}_{\mu\nu\lambda} = \Gamma^{\lambda}{}_{\mu\lambda,\nu} - \Gamma^{\lambda}{}_{\mu\nu,\lambda}\,, \tag{12.60}$$

and therefore

$$g^{\mu\nu}\delta R_{\mu\nu} = g^{\mu\nu}\left[\left(\delta\Gamma^{\lambda}{}_{\mu\lambda}\right)_{,\nu} - \left(\delta\Gamma^{\lambda}{}_{\mu\nu}\right)_{,\lambda}\right] \tag{12.61a}$$

$$= \left(g^{\mu\nu}\delta\Gamma^{\lambda}{}_{\mu\lambda} - g^{\mu\lambda}\delta\Gamma^{\nu}{}_{\mu\lambda}\right)_{,\nu} \tag{12.61b}$$

$$\equiv C^{\nu}{}_{,\nu}\,. \tag{12.61c}$$

Here we used $\delta(\Gamma_{,\nu}) = (\delta\Gamma)_{,\nu}$ and that $g^{\mu\nu}{}_{,\sigma} = 0$ in locally geodesic coordinates; we also renamed the indices $\lambda \leftrightarrow \nu$ in the second term.

iv | In locally geodesic coordinates, Eq. (12.61) is equivalent to $g^{\mu\nu}\delta R_{\mu\nu} = C^{\nu}{}_{;\nu}$. If C^{ν} is a *vector field* (→ *below*), then this equation is actually valid in *arbitrary* coordinates, and we can apply Gauss's theorem:

$$\int \mathrm{d}^4x\,\sqrt{g}g^{\mu\nu}R_{\mu\nu} = \int \mathrm{d}^4x\,\sqrt{g}\,C^{\nu}{}_{;\nu} \overset{10.95}{=} \int \mathrm{d}^4x\,\left(\sqrt{g}C^{\nu}\right)_{,\nu} = \oint \mathrm{d}\sigma_{\nu}\sqrt{g}C^{\nu} = 0\,. \tag{12.62}$$

The surface integral vanishes because $C \propto \delta\Gamma \propto \delta g^{\mu\nu} = 0$ on the surface, which is true if we consider local variations of the metric. We have thereby shown that the first summand in Eq. (12.59) does not contribute to the variation of the Einstein-Hilbert action and can therefore be dropped.

So why is C^{ν}, defined in Eq. (12.61), a vector field? This is not so obvious because it is defined in terms of connection coefficients $\Gamma^{\mu}{}_{\nu\rho}$ – which are not tensors! The crucial point is that the coordinate transformation law Eq. (10.39) of connection coefficients is tensorial up to a non-tensorial contribution that depends only on the coordinate transformation (but not the connection itself): Let $\tilde{g} = g + \delta g$ be an infinitesimal variation of the metric. Then the variation of the coefficients of the Levi-Civita connection is

$$\delta\Gamma = \Gamma(\tilde{g}) - \Gamma(g)\,, \tag{12.63}$$

where we omit indices for clarity and the dependence on the metric is given by Eq. (10.79). Under an arbitrary coordinate transformation, this difference transforms like a tensor because the problematic term in Eq. (10.39) is independent of the metric and therefore cancels in the difference.

v | We can now combine our results:

The variation of the Einstein-Hilbert action Eq. (12.54) evaluates to

$$\delta S[g] \overset{\substack{12.58\\12.59\\12.62}}{=} \frac{1}{2\kappa c}\int \mathrm{d}^4x\,\sqrt{g}\left[R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu}\right]\delta g^{\mu\nu} \overset{!}{=} 0\,, \tag{12.64}$$

since this variation must vanish for all $\delta g^{\mu\nu}$, this is equivalent to the Einstein field Eq. (12.10) in vacuum ($T_{\mu\nu} = 0$). ■

4 | Action with matter:

We can now insert the Einstein-Hilbert action into the "Action of Everything" introduced in Section 12.1:

Eq. (12.1) $\xrightarrow{\text{Eqs. (12.5b) and (12.54)}}$

$$S[g,\phi] = \frac{1}{c}\int \mathrm{d}^4x\,\sqrt{g}\left[\frac{1}{2\kappa}(R-2\Lambda) + L_{\text{Matter}}\right] \tag{12.65}$$

¡! Now the prefactor with the Einstein gravitational constant κ is important: it determines the coupling strength between gravity and matter.

The additional prefactor $\frac{1}{c}$ is only needed for dimensional reasons because we measure time in units of length: $x^0 = ct$; it does not affect the equations of motion.

→

$$\delta_g S[g,\phi] \overset{!}{=} 0 \quad \overset{\substack{12.5a\\12.6\\12.64}}{\Longleftrightarrow} \quad R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = -\kappa T_{\mu\nu} \tag{12.66}$$

5 | What is the energy-momentum tensor of the gravitational field?

i | Since the metric is now our dynamical "gravitational potential", it should be able to carry energy (and momentum) in some form. And surely it does: The first (indirect) observation of → *gravitational waves* was based on the observation of a neutron star circling a pulsar (the ↑ *Hulse-Taylor pulsar*, also known as PSR B1913+16). Over time their orbital period changes, indicating a decay of the orbit [168–170]. But this means that energy must be radiated away, and the only possible carrier is gravitational waves! (By the way, the observations match precisely the quantitative predictions of GENERAL RELATIVITY.) So clearly gravitational waves – which are excitations of the metric field – carry energy.

ii | It is therefore reasonable to expect that there is an energy-momentum tensor associated to the gravitational field, just as there is for any other field that carries energy and momentum.

Recall that the Hilbert energy-momentum tensor was defined in Eq. (11.106) as

$$T^{\text{Matter}}_{\mu\nu} = \frac{2}{\sqrt{g}}\frac{\delta(\sqrt{g}L_{\text{Matter}})}{\delta g^{\mu\nu}}\,, \tag{12.67}$$

and by comparison with previous results (e.g., electrodynamics) we verified that this quantity indeed captures the concepts of energy (currents) and momentum (currents) correctly [recall Eq. (6.110)].

iii | But in GENERAL RELATIVITY, the metric field is "just another dynamical field," described by an action (the Einstein-Hilbert action), which is given by the Lagrangian

$$L_{\text{Metric}}(g,\partial g,\partial^2 g) \overset{12.54}{=} \frac{1}{2\kappa}(R-2\Lambda)\,. \tag{12.68}$$

It is therefore reasonable to expect that the energy-momentum tensor of the gravitational field is given by

$$T^{\text{Metric}}_{\mu\nu} \overset{?}{=} \frac{2}{\sqrt{g}}\frac{\delta(\sqrt{g}L_{\text{Metric}})}{\delta g^{\mu\nu}} \overset{\substack{12.6\\12.64}}{=} \kappa^{-1}(G_{\mu\nu} + \Lambda g_{\mu\nu})\,. \tag{12.69}$$

This already looks strange: This is the left-hand side of the Einstein field equations!

iv | Let us assume that there is no matter present, so that the EFEs read $G_{\mu\nu} + \Lambda g_{\mu\nu} = 0$. Since the propagation of gravitational waves does not rely on the presence of matter, they should still be able to carry energy. However:

$$\text{Eq. (12.69)} \quad \overset{12.10}{\Longrightarrow} \quad T^{\text{Metric}}_{\mu\nu} \doteq 0\,. \tag{12.70}$$

So the HEMT $T^{\text{Metric}}_{\mu\nu}$ of pure gravity vanishes for all solutions of the field equations. This tells us that $T^{\text{Metric}}_{\mu\nu}$ is *not* a reasonable choice for the EMT of the gravitational field.

[Side note: What happened here is a consequence of the diffeomorphism invariance of the Einstein-Hilbert action (which is a gauge symmetry). A *global* continuous symmetry yields a conserved current via Noether's *first* theorem. A *local* continuous symmetry yields Noether identities via Noether's *second* theorem. The latter necessarily make the conserved quantity associated to "global gauge transformations" vanish on-shell [76]. Here this means $T^{\text{Metric}}_{\mu\nu} \doteq 0$. This is similar to the vanishing of the Hamiltonian of the reparametrization invariant theory in Section 5.4.]

v | But there is something even weirder going on: If this tensor would describe the energy density of the gravitational field, then (local) energy conservation means

$$(T^{\text{Metric}})^{\mu\nu}{}_{;\nu} = \kappa^{-1}(G^{\mu\nu}{}_{;\nu} + \Lambda g^{\mu\nu}{}_{;\nu}) \equiv 0\,. \tag{12.71}$$

This follows from the Bianchi identity Eq. (10.122) and metric compatibility Eq. (10.74).

Compare this to the "normal" (local) energy-momentum conservation Eq. (11.109) of proper matter fields, which is only true for fields that satisfy the equations of motion ($\doteq$), i.e., it provides a *constraint* on field evolutions that can be realized in nature (one often employs such constraints to solve complicated equations of motion). But since Eq. (12.71) is an *identity*, it does not constrain the field evolution $g_{\mu\nu}$ whatsoever! This makes the constrain of "energy-momentum conservation" rather vacuous, and the "energy" defined by Eq. (12.69) a quite useless quantity (independent of the fact that it vanishes on-shell for pure gravity).

At this point it should be clear that Eq. (12.69) is *not* a reasonable candidate for the energy-momentum tensor of the gravitational field.

But we can escalate the situation further by asking …

vi | What is the total energy-momentum tensor of *Everything*?

Well, the "Action of Everything" is Eq. (12.1), so that the "HEMT of Everything" should be

$$\begin{aligned} T^{\text{Everything}}_{\mu\nu} &\overset{?}{=} \frac{2}{\sqrt{g}} \frac{\delta(\sqrt{g}[L_{\text{Metric}} + L_{\text{Matter}}])}{\delta g^{\mu\nu}} \\ &= T^{\text{Metric}}_{\mu\nu} + T^{\text{Matter}}_{\mu\nu} \overset{12.69}{=} \frac{1}{\kappa}(G_{\mu\nu} + \Lambda g_{\mu\nu}) + T^{\text{Matter}}_{\mu\nu}\,. \end{aligned} \tag{12.72}$$

That's the pinnacle ob absurdity! This is simply what one obtains from the Einstein field Eq. (12.10) if one collects all terms on one side. Since every realizable configuration (g, ϕ) of all fields must satisfy the EFE, the "energy-momentum tensor of Everything" again vanishes on-shell:

$$T^{\text{Everything}}_{\mu\nu} \doteq 0\,. \tag{12.73}$$

We could boldly conclude that "the total energy of the universe is zero," but this misses the point because Eq. (12.73) has no operational meaning – it is simply the Einstein field equations in disguise.

[If you recall the general structure of the theory considered in Section 12.1 and take Eq. (11.106) as the definition of the energy-momentum tensor, we have just shown that *the energy-momentum tensor of any diffeomorphism invariant theory vanishes on-shell*. A theory is diffeomorphism invariant if (1) it is background independent (= g is a dynamical field) and (2) the action is generally covariant.]

vii | If $T^{\text{Metric}}_{\mu\nu}$ is not a reasonable choice for the energy-momentum tensor of the gravitational field, what is? The answer may be surprising:

> There is no energy-momentum tensor of the gravitational field.
> Gravitational energy is necessarily *non-local*.

Important: This does *not* mean that the gravitational field carries no energy. It only means that this energy cannot be associated to a *local energy density* in any reasonable way; gravitational energy is necessarily *delocalized*.

viii | Here is another (hand-waving) argument to reason why there cannot be a energy-momentum tensor for gravity (the argument is flawed [171], but demonstrates at least that the gravitational field is "different"):

The energy-momentum tensors of all field theories relevant to fundamental physics are quadratic in *first* derivatives of the field [Examples: ← Eq. (11.115) for electrodynamics and ← Eq. (11.118) for the Klein-Gordon field]. Since $g_{\mu\nu}$ is the field of GENERAL RELATIVITY, a conventional EMT should then be quadratic in $g_{\mu\nu,\rho}$. But we already argued previously that one cannot construct a tensor from first derivatives of the metric alone (because one can make these derivatives vanish in locally geodesic coordinates). Hence such a "conventional" EMT cannot exist for the gravitational field.

The flaw of this argument, pointed out in Ref. [171], is of course that we already know that gravity is very different from all other fundamental field theories (recall Section 8.2). Thus it is quite a leap of faith to exclude an energy-momentum tensor that depends on higher-than-first derivatives, based solely on our experience that other field theories behave differently.

ix | The problem concerning the energy of the gravitational field has a long-standing history; for more details see Ref. [171] and references therein. See also MISNER *et al.* [2] (§20.2, pp. 466–468). For a discussion of the so called ↑ *energy-momentum pseudotensor* that can be used to study the energy of gravitational waves see CARROLL [102] (§7.6, pp. 307–315).

12.3. ‡ Modifications of GENERAL RELATIVITY

- Our approach to come up with a covariant action already suggests modifications of GENERAL RELATIVITY by adding higher-order curvature terms, recall Eq. (12.52). This begs the question in which ways GENERAL RELATIVITY can be *modified* to obtain other relativistic theories of gravity.
- ¡! So far, GENERAL RELATIVITY has passed the test of time with flying colors:

  → *Applications* in Chapter 13

  Despite some unexplained phenomena (← *below*), there is currently no widely accepted evidence that GENERAL RELATIVITY *needs* to be modified (at least not on the classical level, i.e., in the "infrared limit").

1 | Two classes of modifications:

- Modifications in the UV-limit (= at high energies, small distances …)

  Motivation: GENERAL RELATIVITY is not a quantum theory → Quantum gravity?

  This is quite *uncontroversial*: It is widely believed that GENERAL RELATIVITY is the classical limit of a more fundamental theory that is most likely governed by the laws of quantum mechanics (that is, a theory of ↑ *quantum gravity*). The modifications due to quantum effects

will become important on the ↓ *Planck scale* at the latest; on a semi-classical level, this might manifest as additional curvature terms showing up in the action / field equations, which modify the predictions of GENERAL RELATIVITY on very small (high) length (energy) scales. Note that such modifications are not relevant for large-scale physics (such as the motion of galaxies or the expansion of the universe).

Not everyone agrees that gravity must be quantized; Roger Penrose, for example, advocates that "quantum mechanics must be gravitized." He denies that gravity has a quantum nature at all, and that the collapse of the quantum wavefunction is an objective dynamical process, induced by gravity, that makes a unique, classical, macroscopic world emerge out of a microscopic quantum world [172]. This is in direct contradiction to most other interpretations of quantum mechanics (collapse theories are not interpretations but *modifications* of quantum mechanics) like ↑ *Everett's many-worlds interpretation* or ↑ *decoherence theory*.

→ *Excursions* in Part III

- Modifications in the IR-limit (= at low energies, large distances …)

  Motivation: Unexplained gravitational phenomena on large scales.

  This is *controversial* and a stance not shared by many physicists: Modifications of GENERAL RELATIVITY in the IR-limit typically means messing with well-established classical observations such as Newtonian gravity and/or the equivalence principle (in its various incarnations, ← Section 9.1). While it is of course possible that these classical laws and principles are violated by some as of yet undiscovered physical process (which then would require a revision of GENERAL RELATIVITY as taught in this course), there is currently no hard evidence for such (regarding the problem of *dark matter*: → *next*).

  → *Below* we briefly discuss potential IR-modifications of GENERAL RELATIVITY.

2 | Dark matter:

i | Arguments for IR-modifications must be based on *classical*, *large-scale* observations that don not match the predictions of GENERAL RELATIVITY and/or its non-relativistic limit: Newtonian gravity.

ii | The most prominent and widely accepted discrepancy of this kind is based on the *rotation curves of galaxies*: The gravitational pull experienced by stars in the outskirts of galaxies is much larger than computed from the mass of gas and stars one observes (using GENERAL RELATIVITY in the non-relativistic limit, i.e., Newtonian mechanics). That this discrepancy is real is confirmed beyond any doubt.

If one plots the orbital velocities of stars in spiral galaxies over their distance from the galaxy center, one obtains curves that flatten out for large distances (Fig. 12.1 a). This is true for almost all galaxies (there are a few exceptions though). If one uses telescopes to infer the mass-energy distribution of these galaxies (i.e., gas, dust and stars), and then computes how the velocity profile according to Newtonian mechanics should look like (by equating centrifugal and gravitational force), one finds that the rotation curves should *drop off* for large distances (Fig. 12.1 b). In a nutshell: the stars in the outskirts of galaxies are *too fast*, if they were only pulled by the gravitational force due to matter we can see, the galaxies would be ripped apart by the centrifugal force (but they are not!).

iii | Potential solutions to the puzzle fall into two categories:

- GENERAL RELATIVITY is *correct*

  → There must be matter/energy in galaxies that we cannot detect ("see").

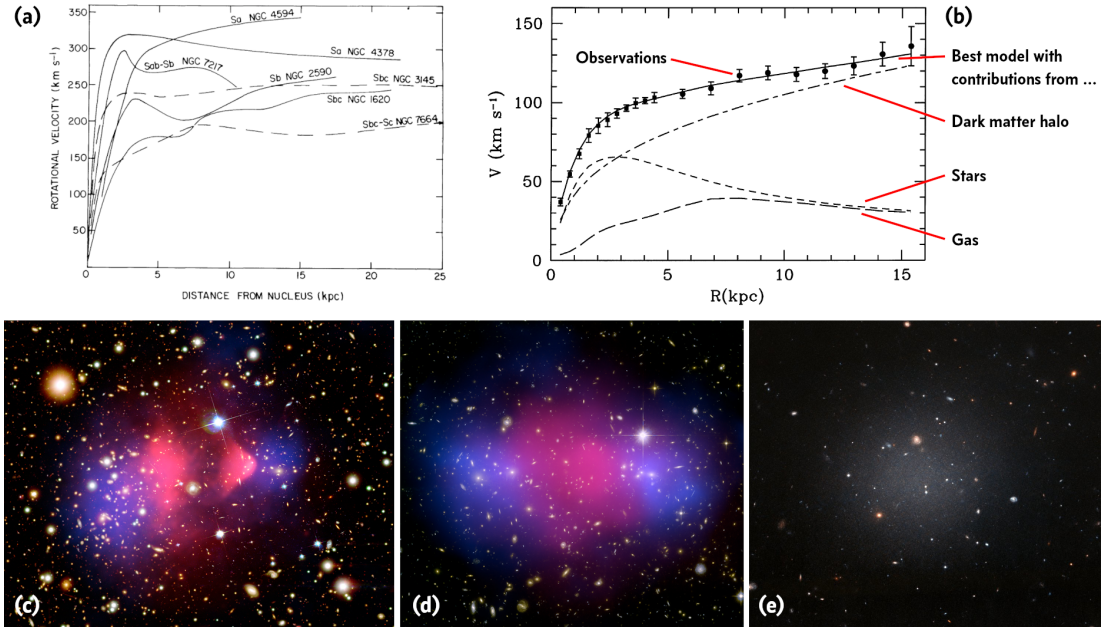


FIGURE 12.1. • Why dark matter? **(a)** That the rotation curves of galaxies do not match the visible matter distribution was first noticed in 1970 [173] and repeatedly confirmed over the years [174, 175]. For a review on the rotation curves of spiral galaxies see Ref. [176]. (Plot from Ref. [174].) **(b)** Without a modification of GENERAL RELATIVITY (↑ *MOND*, ↑ *TeVeS*, ...), there must be invisible matter ("dark matter") responsible for this phenomenon. The distribution of this hypothetical matter can then be mapped out by studying rotation curves [177]. (Plot from Ref. [177].) **(c)** Evidence for dark matter: Shown is a superimposed image of the galaxy cluster 1E 0657-56 (↑ *Bullet cluster*). Pink: X-ray (hot gas, baryonic matter); White/Orange: Optical (stars, baryonic matter); Blue: Lensing map (baryonic and dark matter). This direct observation of a spatial dislocation of baryonic and gravitating matter is believed to be a strong evidence for the existence of dark matter [178]. These measurements can even be used to constrain the properties of dark matter [179]. Photo: https://chandra.si.edu/photo/2006/1e0657/. **(d)** By now there are observations of other galaxy clusters with similar features, such as MACS J0025.4-1222 [180]. The color map is the same as for c). Photo: https://www.chandra.harvard.edu/photo/2008/macs/. **(e)** The observation of the ultra-diffuse galaxy NGC 1052-DF2 revealed a rotation curve consistent with the *absence* of dark matter [181, 182]. A similar galaxy without dark matter was found quickly after [183]. Note that the absence of dark matter in specific galaxies can be interpreted as evidence *for the existence* of dark matter. Photo: https://esahubble.org/images/heic1806a/.

→ ⁂ *Dark matter*?

Note that "dark matter" is a placeholder term. It is simply matter that we cannot detect for whatever reason. There is nothing magical about it (→ *below*).

- GENERAL RELATIVITY is *not correct*
  (on very large length-scales or at very low accelerations)

  → Modifications of GENERAL RELATIVITY?

  Because stars in the outskirts of galaxies are non-relativistic (low velocities, weak gravitational fields), they must be described by the non-relativistic limit of the correct relativistic theory of gravity; for GENERAL RELATIVITY, this is good old Newtonian gravity. So a modification of GENERAL RELATIVITY that makes sense of the flat rotation curves of galaxies *without* postulating additional "dark" matter must necessarily modify Newton's law of universal gravitation. But this law works perfectly well in our solar system (up to corrections that GENERAL RELATIVITY can explain). Thus any

reasonable IR-modification of GENERAL RELATIVITY must ensure that its hampering with Newton's law only affects extremely large distances (and therefore extremely low gravitational accelerations). The most prominent theory of this kind is called ↑ *Modified Newtonian Dynamics (MOND)*; in it original formulation by MILGROM [184–186] it is simply a non-relativistic, non-covariant modification of Newtonian gravity. Since (local) Lorentz covariance and the principle of general covariance are rather well-established cornerstones of physics that one shouldn't carelessly mess with, it is desirable to derive the modifications proposed by MOND from a generally covariant modification of GENERAL RELATIVITY; such a modification was proposed by BEKENSTEIN and dubbed ↑ *Tensor-Vector-Scalar gravity (TeVeS)* [187]. It is a rather contrived theory that is significantly more complex than GENERAL RELATIVITY.

Recent studies (using new data from space-borne observatories that piled up over the last few years) have shown that the modifications proposed by MOND-like theories do not match observations [188–191]. In short: the future for MOND-like modifications of GENERAL RELATIVITY looks bleak. (There is nothing wrong with this; that's how science works: there is a unexplained phenomenon; you propose a solution and derive its implications; as more observations pour in, you check whether they are compatible with your theory; if not, you modify or, if this doesn't help, abandon the theory.)

Here is the corresponding XKCD comic that sums up the situation quite well:

Source: https://xkcd.com/1758/

iv | The case for dark matter:

To be very clear: While the rotation curve/dark matter discrepancy has been the strongest case for potential IR-modifications of GENERAL RELATIVITY, this route has always been pursued only by a minority of physicists.

To laymen and students of physics alike, the alternative "solution" to postulate "dark matter" to patch up the discrepancy between observations and GENERAL RELATIVITY often looks like a cheap cop-out (Fig. 12.1 b). However, there are reasons why the majority of physicists believe that this is the most promising route to solve the puzzle:

1. Postulating unseen particles has been successful in the past. For example, Wolfang Pauli postulated 1930 the neutrino to explain missing momentum in radioactive beta decay (the neutrino was found in 1956). Peter Higgs (and others) postulated 1964 the Higgs boson to explain the mass of the weak gauge bosons (which was then discovered at the LHC in 2012). In 1973 the third generation of quarks (later called *top* and *bottom*) was predicted to explain CP violations in the decay of kaons (the bottom quark was discovered in 1977, the top quark in 1995).

   [To be fair: Postulating the existence of things that have not yet been observed has also failed in the past. For example, to explain the anomalous perihelion precession of

Mercury, a new planet named ↑ *Vulcan* was postulated (orbiting between Sun and Mercury). The planet does not exist, and today we know that corrections due to GENERAL RELATIVITY are responsible for the precession.]

2. The Standard Model of particle physics, our best theory of the very small, describes the properties of fundamental particles. While the model is restricted by symmetries (one of them being Lorentz invariance), it still can be modified and extended in many ways; for example, it is quite natural to add right-handed ↑ *sterile neutrinos* without breaking the math. Thus, from the viewpoint of a particle physicist, it is general practice to extend theories by new particles (= fields) and study the consequences. "Dark matter" could just be one or more fields the excitations of which evaded our detectors so far (sterile neutrinos are such a candidate for dark matter).

3. By now, there is strong *indirect* evidence for dark matter (whatever it is made of) from astronomical observations (Fig. 12.1):

   - Fig. 12.1 c and Fig. 12.1 d show images of the galaxy clusters `1E 0657-56` (↑ *Bullet cluster*) [178, 179] and `MACS J0025.4-1222` [180], respectively. The images superimpose data from different instruments: Pink denotes X-rays that indicate where the hot interstellar gas is located. The white/orange structures on the black background are the optical signatures of galaxies coming from stars. The most interesting is the blue cloud: it encodes the distribution of gravitating matter inferred from a so called ↑ *lensing map*. The idea is to use → *gravitational lenses* to map out the mass distribution of a region of space. Essentially you look how the light coming from the stars in the background is disturbed by masses in the foreground. In these pictures, only the blue density map is sensitive to dark matter (because dark matter does not emit light, but distorts light from the background stars).

     The situation in both Fig. 12.1 c and Fig. 12.1 d is similar: we see the aftermath of two clusters of galaxies that collided. This sounds more exciting than it actually is, because galaxies (and even more so clusters of galaxies) consist mostly of empty space with a bit of dust and gas. This means that in such collisions there are almost no collisions of *stars*; they all miss each other! By contrast, the low-density gas between the stars behaves like a fluid; the two "blobs" of interstellar gas hit each other and slow down. This is what the two pink clouds in Fig. 12.1 c show: the X-ray emitting gas is lagging behind the actual stars (mostly in the blue region) that missed each other and are flying to the left and right.

     So far, there is no hint of dark matter: the blue (gravitating) mass is on top of the stars and the gas is lagging behind. The twist is that almost all of the (visible) mass of a galaxy (cluster) comes from the gas between the stars – and not the stars themselves! This might sound strange, but there is a lot of space between stars, and even if this space is *almost* vacuum, the total mass still outweighs the stars significantly. But now we have a problem: If most of the *visible* mass is gas (pink), why is most of the *gravitating* mass where the stars are (blue)? Well, because the blue cloud is mostly caused by the *dark matter halo* of the two galaxies, and not by the stars! And this fits exactly the properties expected for dark matter: The particles making up dark matter cannot interact in any significant way, otherwise we would have already detected them. But this means that a cloud of dark matter does not behave like "normal gas" would; in particular, two colliding clouds of dark matter cannot slow each other down. Thus it is perfectly consistent that the two dark matter clouds (blue) passed each other, just as the visible stars did (but for very different reasons). It is this observable separation between *visible* mass (pink) and *gravitating* mass (blue) that makes a strong case for dark matter.

   - Another recent observation supporting the existence of dark matter is, quite surprisingly, the observation of the ultra-diffuse galaxy `NGC 1052-DF2` (Fig. 12.1 e) with a rotation curve that is consistent with the *absence* of dark matter [181, 182]

(by now a second of these rare galaxies has been found [183]). The argument is quite simple:

If you want to avoid dark matter, you must mess with GENERAL RELATIVITY (like MOND and TeVeS do) and thereby Newton's law of universal gravitation. But now that we have examples of galaxies *where Newtonian gravity works without postulating dark matter*, you have a problem: why is your modification not valid for these galaxies? You cannot go around and modify the laws of physics from place to place! But if dark matter exists (and is responsible for the flat rotation curves), it is at least plausible that a few galaxies with an extravagant history somehow got their cloud of dark matter stripped away (perhaps by the gravitational interaction with another galaxy), and therefore have rotation curves that fall off, without the need for additional mass.

In summary, it seems likely that dark matter exists and is responsible for the rotation curve problem. Conversely, it seems more and more unlikely that GENERAL RELATIVITY must be modified anytime soon. But until we identify and measure what dark matter actually is, we don't know for sure.

3 | Potential modifications:

Arguments for or against modifications of GENERAL RELATIVITY aside, which possibilities do we have to construct alternatives to GENERAL RELATIVITY?

For more details on alternative theories of gravity see Ref. [110] and CARROLL [102] (§4.8, pp. 181–190).

← *Lovelock's theorem* [134, 135, 171]:

$$\left.\begin{array}{r}\text{Only a metric field}\\ \text{Second-order field equations}\\ \text{Four-dimensional spacetime}\\ \text{Local action}\end{array}\right\} \quad \Rightarrow \quad \text{GENERAL RELATIVITY} \tag{12.74}$$

→ Options for modifications of GENERAL RELATIVITY:

- ◁ *Other fields* in addition to (or replacing) the metric
  - ◁ Scalar fields

    Theories that augment the metric tensor field $g_{\mu\nu}$ by an additional scalar field ϕ are known as ↑ *scalar-tensor theories* of gravity. You may wonder how ϕ differs from any other matter field? The reason why ϕ cannot be simply identified as another matter field is that its coupling to the metric is *non-minimal*. Note that this suggests a definition which fields describe *matter* and which describe *gravity*: Matter fields are minimally coupled to the metric, additional gravitational fields are non-minimally coupled. Since non-minimally coupled fields tend to violate the equivalence principle, such theories often violate its *strong* version SEP.

    Example: One of the first and most famous scalar-tensor theories is ↑ *Brans-Dicke theory* [192]. It is defined by the gravitational action (here for $c = 1$)

$$S_{\text{BD}}[g,\phi] = \frac{1}{16\pi}\int \mathrm{d}^4x\,\sqrt{g}\left[\phi R - \frac{\omega}{\phi}g^{\mu\nu}(\partial_\mu\phi)(\partial_\nu\phi)\right] \tag{12.75}$$

with massless scalar field $\phi(x)$ and ⁂ *Dicke coupling constant* ω. For $\omega \to \infty$ one recovers GENERAL RELATIVITY. In theories of this kind, the scalar field $\phi(x)$ can be interpreted as a position and time dependent replacement of the gravitational coupling constant $1/\kappa \propto 1/G$.

- ◁ Connections *other than* the Levi-Civita connection

  As discussed in Section 9.4, the concepts of *connection* and *metric* are independent in principle. Only when demanding a *metric-compatible* and *torsion-free* connection does one obtain the unique Levi-Civita connection and everything is determined by the metric alone.

  Example: One could start with the Einstein-Hilbert action, but treat connection Γ and metric g as independent fields:

$$S[g,\Gamma] := \frac{1}{2\kappa c}\int \mathrm{d}^4x\,\sqrt{g}\,g^{\mu\nu}R_{\mu\nu}(\Gamma)\,. \tag{12.76}$$

  Here, the curvature is directly computed from the connection via Eq. (10.70); this is known as the ↑ *Palatini action* (➲ Problemset 5).

  Quite surprisingly, if one starts from Eq. (12.76) and assumes either …

  * Γ is metric-compatible, or …
  * Γ is torsion-free,

  the variation $\delta_\Gamma S$ of the action wrt. the connection coefficients ${\Gamma^\mu}_{\nu\rho}$ vanishes only for the Levi-Civita connection (which brings us back to GENERAL RELATIVITY). Only if one drops all restrictions, and allows for arbitrary connections, does one find a modification of GENERAL RELATIVITY.

  However, note that the *difference* of two arbitrary connections is a *tensor* [this follows from Eq. (10.39)]. But this means that *w.l.o.g.* you can write any connection in the form

$${\Gamma^\mu}_{\nu\rho} = \underbrace{\left\{{\mu \atop \nu\rho}\right\}}_{\text{Levi-Civita}} + \underbrace{{T^\mu}_{\nu\rho}}_{\text{Tensor}}\,, \tag{12.77}$$

  so that the decoupling of metric and connection boils down to the extension of GENERAL RELATIVITY by some additional tensor field ${T^\mu}_{\nu\rho}$.

  For example, see Refs. [131, 132] for potential extensions of GENERAL RELATIVITY by allowing connections with torsion.

- …

  There is of course no limit to your imagination. One can consider any combination of arbitrary-rank tensor fields to augment the metric. One example is the previously mentioned ↑ *Tensor-Vector-Scalar gravity (TeVeS)* by Bekenstein [187] which, as the name suggests, comprises a metric tensor field, a vector field, and a scalar field.

- ◁ *Higher than second derivatives* of the metric in the field equations

  Example: ↑ $f(R)$*-gravity theories* [193] are defined by the generalized Einstein-Hilbert action

$$S_f[g] := \frac{1}{2\kappa c}\int \mathrm{d}^4x\,\sqrt{g}\,f(R) \tag{12.78}$$

  with some differentiable function $f : \mathbb{R} \to \mathbb{R}$ that specifies the theory. For $f(R) = R$ one recovers GENERAL RELATIVITY (without cosmological constant), but for $f(R) \neq R$ the field equations differ from the EFEs (and typically contain higher than second derivatives of the metric).

- ◁ Spacetime dimensions $D \neq 4$

  It is straightforward to generalize GENERAL RELATIVITY (e.g., using the Einstein-Hilbert action) to arbitrary spacetime dimensions D. For an example, in Problemset 4 we study $D = 2+1$-dimensional GENERAL RELATIVITY. These theories can behave very differently from $D = 3+1$-dimensional GENERAL RELATIVITY. However, they obviously do not describe reality correctly as our spacetime has undeniable $D = 3+1$ dimensions. The only viable route is then to postulate additional spatial dimensions and "curl them up" (↑ *compactification*) so that one cannot see them on the large scales accessible to us. Such theories are known as ↑ *Kaluza-Klein theories* because THEODOR KALUZA introduced a $D = 4+1$-dimensional version in 1921 [194], which was later extended by OSKAR KLEIN in 1926 [88, 195].

  Example: A simple metric of a $D = 4+d$-dimensional spacetime could have the form

  $$\mathrm{d}s^2 = G_{ab}\mathrm{d}X^a\mathrm{d}X^b \equiv \underbrace{g_{\mu\nu}(x)\mathrm{d}x^\mu\mathrm{d}x^\nu}_{\text{Observable 4D spacetime}} + \underbrace{b^2(x)\gamma_{ij}(y)\mathrm{d}y^i\mathrm{d}y^j}_{\text{Compact extra dimensions}} \tag{12.79}$$

  with $a, b = 0, \ldots, d+3$ for $d > 0$, $\mu, \nu = 0, 1, 2, 3$, and $i, j = 1, \ldots, d$ are the coordinates of the compactified additional d dimensions.

  As action one could postulate the Einstein-Hilbert action,

  $$S[G] := \frac{1}{16\pi G_{4+d}} \int \mathrm{d}^{4+d}X \sqrt{G}\, R[G]\,, \tag{12.80}$$

  where G_{4+d} denotes the "gravitational constant" of this hypothetical $4+d$-dimensional theory and $R[G]$ is the Ricci scalar computed from the $4+d$-dimensional metric G_{ab}. Note that this is not equivalent to GENERAL RELATIVITY in $4+d$ dimensions because (1) we constrain the form of the metric to Eq. (12.79), and (2) the additional d space dimensions are compact and not extended.

  Remarkably, if one integrates out these compact extra dimensions (↑ *dimensional reduction*), one finds theories equivalent to GENERAL RELATIVITY *with extra fields* [under additional constraints on γ_{ij}, the simple metric (12.79) yields a ← *scalar-tensor theory*, see CARROLL [102] (§4.8, pp. 186–189) for details]. The intuition behind this is that the geometric degrees of freedom of the curled-up dimensions manifest in the extended 4D spacetime as *additional fields* that can couple non-minimally to the metric.

  In a nutshell: Extending spacetime by compact extra dimensions is often equivalent to adding new fields on a 4D spacetime (without extra dimensions).

- ◁ Field equations that *cannot* be derived from the metric variation of an action.

  The field equations of such theories are not necessarily rank-2 tensor equations; and even if so, they are not necessarily symmetric in the indices and/or divergence-free (recall our derivation in Section 12.1 of these properties starting from a covariant action).

- ◁ *Non-local* theories

  Physicists don't like non-local theories very much. Whenever you work with a continuum theory that can be described by (a set of) differential equations on some manifold, the theory is local. Non-local theories therefore must be described by other equations (for example: ↑ *integro-differential equations*). Such theories and equations are often hard to work with. Fortunately, nature seems to be rather local, which explains the prevalence of local theories in physics (though this might be an illusion of sorts, → ??).

12.4. ‡ Diffeomorphism invariance and the Hole argument

Now that GENERAL RELATIVITY has been fully developed as a relativistic theory of gravity, there are a few conceptual issues that need to be clarified.

1 | The Hole argument:

This discussion is based on Ref. [196]. For reviews of the hole argument see Refs. [197, 198].

i | ≺ Fields (g, ϕ) & Action of Everything (AoE) $S[g, \phi] = S[g] + S_g[\phi]$

Recall Eq. (12.1) in Section 12.1.

≺ Diffeomorphism $\varphi \in \mathrm{Diff}(M) \to$ Transformed fields $(\bar{g}, \bar{\phi})$

Recall Eq. (11.85) in Section 11.4.

¡! Remember that we interpret $(\bar{g}, \bar{\phi})$ as new/different fields (in the same coordinates).

AoE is *generally covariant* $\xrightarrow{\text{Eq. (11.89)}}$ AoE is *diffeomorphism invariant*:

$$S[\bar{g}, \bar{\phi}] = S[g, \phi] \tag{12.81}$$

This implies for the EOMs →

$$\delta S[\bar{g}, \bar{\phi}] = 0 \quad \Leftrightarrow \quad \delta S[g, \phi] = 0 \tag{12.82}$$

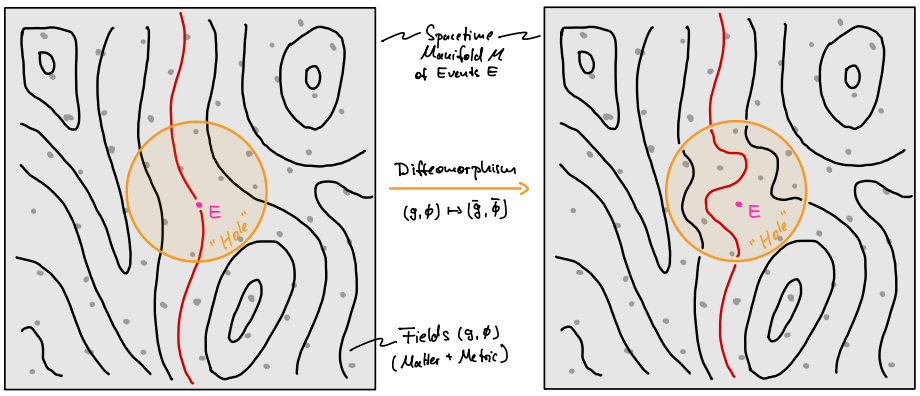
In words: If (g, ϕ) is a solution of the equations of motion (the Einstein field equations and the matter EOMs), then the new fields $(\bar{g}, \bar{\phi})$ obtained by any diffeomorphism φ are another solution. The group of diffeomorphisms $\mathrm{Diff}(M)$ on the spacetime manifold is therefore a *symmetry/invariance group* of GENERAL RELATIVITY.

Note that this hinges on the fact that both the metric g and the matter fields ϕ are transformed by the same diffeomorphism.

ii | At this point, it is unclear why the fact that $\mathrm{Diff}(M)$ is a symmetry group of GENERAL RELATIVITY poses a problem. To understand the issue, recall our current interpretation:

$$\text{Spacetime} = \Big(\underbrace{\text{Manifold } M}_{\substack{\text{Coincidence classes} \\ \text{of events (?)}}} , \underbrace{\text{Metric } g}_{\substack{\text{Gravitational} \\ \text{field}}} \Big) \tag{12.83}$$

Diffeomorphism invariance then implies that the following two field configurations (sketched here on a 2D spacetime manifold for simplicity) both satisfy the EOMs if one of them does:



The crucial point is that diffeomorphisms φ can act *locally* on compact regions of spacetime (here the "hole"), leaving the fields everywhere else unchanged.

Here the conincidence classes of events E that make up the spacetime manifold M are denoted by gray dots, the fields (both metric g and matter ϕ) are indicated by contour lines. Note that the two diffeomorphic field configurations (g, ϕ) and $(\bar{g}, \bar{\phi})$ differ only in a compact region denotes as "hole".

→ Problem:

The problem that arises from such a construction can be phrased in various ways:

- Assume that time runs upwards in the 2D patch of spacetime above. The two field configurations (g, ϕ) and $(\bar{g}, \bar{\phi})$ are then identical in the "past" (= lower boundary of the patch), but differ in the "hole". But this is a problem for *determinism*: A useful physical model should make unambiguous predictions for the future evolution of a system, based on a set of initial data. The diffeomorphism invariance of GENERAL RELATIVITY thwarts this, for it cannot distinguish between the two field evolutions above that coincide in the past.

  → Is GENERAL RELATIVITY *indeterministic*?

- In our current reading, the points of the spacetime manifold are (coincidence classes of) *events* E. The fields (metric and matter alike) are functions on M. If we interpret the red contour line in the sketch as the trajectory of a particle (given by the excitation of some field), diffeomorphisms can be used to deform this trajectory arbitrarily. But the statement "the red particle passes through event E" is not invariant under such transformations. This seems to be problematic because it is a statement about ← *coincidences*, and as such should be objective (as argued in Section 1.1). Put differently: The EOMs of GENERAL RELATIVITY cannot decide wether the particle meets the event E or not!

  → What is the relation between *fields* and (coincidence classes of) *events*?

- Einstein originally considered a spacetime filled with matter – except for a "hole" that was assumed to be free of matter (that's were the term "hole" comes from). He then asked whether the metric in the hole was determined by the distribution of matter (and the metric) outside the hole. Diffeomorphism invariance said *no*. This implies that knowledge of the distribution of matter *outside* the hole, together with the initial geometry of spacetime, is not enough to predict the metric *inside* the hole. Einstein called this a "violation of the law of causlity" – which is essentially the problem of *indeterminism* identified above.

Einstein introduced the argument in late 1913 to rationalize his failure to find a generally covariant field equations that were consistent with Newtonian gravity in the non-relativistic limit. He used the hole argument to convince himself that a generally covariant theory of gravity was *impossible* (← *last point above*). The argument then coaxed him into a (misguided) search for *non-covariant* field equations that, in hindsight, delayed the genesis of GENERAL RELATIVITY by two years. Einstein found the flaw in his argument in late 1915 (→ *next*); freed from this conceptual roadblock, he published the correct field Eq. (12.10) shortly after.

iii | Solution:

To solve the problems above, we have no choice but to concede the following:

- If we want GENERAL RELATIVITY to be a deterministic (= predictive) theory, we must *identify* diffeomorphic solutions as *physically indistinguishable*.

  [Similar to gauge fields A_μ and $\tilde{A}_\mu$ that are related by $\tilde{A}_\mu = A_\mu + \partial_\mu \lambda$ are physically indistinguishable in electrodanymics.]

- We cannot interpret the points $E \in M$ of the manifold as observable entities that exist (in some physical sense) independent of the fields.

→ ⁂ *Leibniz equivalence*:

- Diffeomorphic solutions $(g, \phi) \overset{\varphi}{\sim} (\bar{g}, \bar{\phi})$ describe *the same* physics.
- GENERAL RELATIVITY is a *gauge theory*; Diff(M) is its *gauge group*.
- The spacetime manifold M itself does *not* exist as physical entity.
- The *fields* (g, ϕ) on the spacetime manifold M exist as physical entities.

- The points $E \in M$ *of* the spacetime manifold cannot exist in the same way the fields *on* the manifold do. The stance that M exists as a physical entity is known as ↑ *manifold substantivalism*; the hole argument is therefore an argument *against* this philosophical reading of GENERAL RELATIVITY. (See also my perspective → *below*.) Note that this does not affect the independent existence of the *metric field*, which is responsible for the elevation of "spacetime" from a static background to a dynamical participant in the evolution of the universe. This view is known as ↑ *substantivalism* (without the prefix "manifold") and remains unaffected by the hole argument.
- *Disclaimer:* No matter what I claim here, you will always find a paper by a philosopher of science who disagrees. That's fine; the whole purpose of philosophy is to disagree about stuff that we cannot (yet) pin down by experiments.
- Historical note:

  Einstein finally discarded the hole argument and embraced Leibniz equivalence (which led him to his field equations in November 1915). On January 3. 1916, Einstein writes in a letter to his friend Michele Besso [199]:

  > *An der Lochbetrachtung war alles richtig bis auf den letzten Schluss. Es hat keinen physikalischen Inhalt, wenn inbezug auf dasselbe Koordinatensystem K zwei verschiedene Lösungen $G(x)$ und $G'(x)$ existieren. Gleichzeitig zwei Lösungen in dieselbe Mannigfaltigkeit hineinzudenken, hat keinen Sinn und das System K hat ja keine physikalische Realität. Anstelle der Lochbatrachtung tritt folgende Überlegung. Real ist physikalisch nichts als die Gesamtheit der raumzeitlichen Punktkoinzidenzen. Wäre z. B. das physikalische Geschehen aufzubauen aus Bewegungen materieller Punkte allein, so wären die Begegnungen der Punkte, d. h. die Schnittpunkte ihrer Weltlinien das einzig Reale, d. h. prinzipiell beobachtbare. Diese Schnittpunkte bleiben natürlich bei allen Transformationen erhalten (und es kommen keine neuen hinzu), wenn nur gewisse Eindeutigkeitsbedingungen gewahrt bleiben. Es ist also das natürlichste, von den Gesetzen zu verlangen, dass sie nicht mehr bestimmen als die Gesamtheit der zeiträumlichen Koinzidenzen. Dies wird nach dem Gesagten bereits durch allgemein kovariante Gleichungen erreicht.*

  For your entertainment, the letter also contains the following (unrelated) statement:

  > *Das Studium von Minkowski würde Dir nichts helfen.*
  > *Seine Arbeiten sind unnütz kompliziert.*

iv | ‡ Another perspective:

What follows is my own take on what diffeomorphism invariance might tell us about reality. The purpose of the following arguments is to demonstrate that the "hole issue", diffeomorphism invariance, and general covariance, are all "symptoms" of mathematical surplus structure that – while being useful for our description of GENERAL RELATIVITY– cannot (and should not) be identified with real physical entities.

a | Let me start by pointing out an intrinsic flaw of our previous interpretation of the mathematical objects we are working with. So far, our reading was as follows:

$$\begin{aligned}\text{Spacetime manifold } M &= \text{Set of } \textit{coincidence classes } E = \{e_1, e_2, \ldots\} \text{ of } \textit{events } e_i\\ \text{Gravitational field } g &= \text{Tensor field on } M\colon g : M \to T^*M \otimes T^*M\\ \text{Electromagnetic field } A &= \text{Vector field on } M\colon A : M \to TM\\ \text{Klein-Gordon field } \phi &= \text{Scalar field on } M\colon \phi : M \to \mathbb{R}\end{aligned}$$

Now think of a coincidence class including the events "particle here" and "photon here". We can think of this as a combined event where the photon is absorbed or emitted by the particle. But particle physics tells us that there are no particles (funny, I know), just fields. In the modern reading of quantum field theory, "particles" are simply localized (and quantized) excitations of fields. For simplicity, let us say that the event "photon here" simply means $A(E) \neq 0$ for some point $E \in M$, and the coincidental presence of the particle is similarly described by $\phi(E) \neq 0$ (since ϕ is a scalar, this would have to be a scalar particle like the Higgs boson [which is not electrically charged]; but this is not important here).

This shows that elementary events of the type "particle of type X is here" are associated with specific *values* of fields of type X – not with their *arguments* (= points of the manifold)! But if these points are supposed to be coincidence classes of events, we arrive at a strange circular construction where fields are defined on points that contain (and are characterized by) the value of the field at that very point.

In a nutshell:

All observable features of physical systems are determined by the values of fields, their coincidences and causal relations.

This suggests that the points E of the spacetime manifold M *cannot* be events themselves (nor can they be classes of events). However, we were also not far off identifying spacetime points with coincidence classes of events. Let me explain:

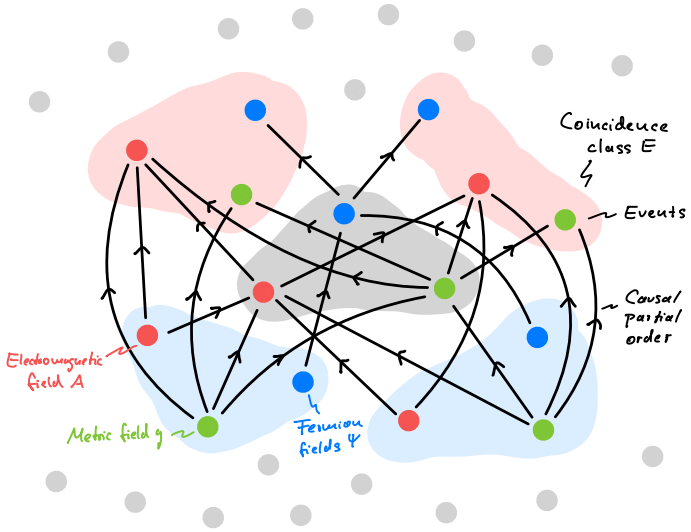
b | Let us put forward the following Postulate:

Only *events* (e), related by *coincidences* ($\sim$) and a *causal partial order* ($\prec$) exist.

Recall our discussion in SPECIAL RELATIVITY of events and their relations in Section 1.6.

That we identify these events with realizations of values of fields it not important.

→ Reality is a causal network of events, grouped by coincidences:



This looks rather messy! It is certainly hard to formulate a workable model (= theory) for this reality without putting in a bit more effort into the layout of the causal graph:

c | Unexplained fact: The causality graph of our universe is *4D-embeddable*.

- "4D-embeddable" means that you can lay out the graph on a 4D manifold such that the edges (= causal relations) only connect "nearby" nodes (= events) of the graph ("nearby" being defined by the ↑ *topology* of the manifold).
- If you randomly construct a causality graph, there is absolutely no reason to expect that it has this property. Hence this is a feature of reality that must be explained (→ *below*). Note that the embeddability is a *local* feature; we do not claim that the graph can be embedded in a topologically trivial manifold like $\mathbb{R}^4$.

This suggests the following procedure to lay out the causality graph:

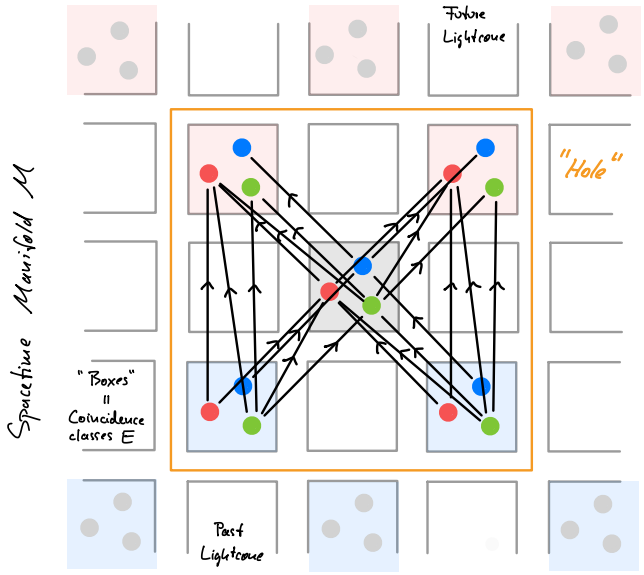
(1) Construct a 4D manifold using *empty boxes* (these are the points of the manifold) by arranging them in a 4D hypercubic lattice (for simplicity). This is our new spacetime manifold M. Note that it doesn't contain any events yet; its a completely artificial structure without physical existence.

(2) Place events of the causality graph into the boxes of the manifold such that …
   - … events that coincide ($\sim$) are placed in the *same* box.
   - … events connected by an edge ($\prec$) are placed in *nearby* boxes.

   This procedure succeeds because the graph is 4D-embeddable.

   (Since empty boxes don't exist, you can think of them as not being there at all.)

→ For example:

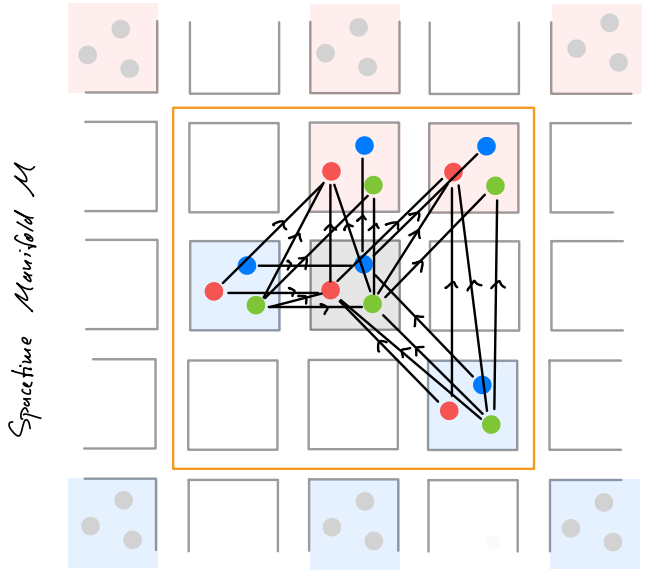


There you have it. This is the structure we called "spacetime manifold". You can even see the light/null-cone structure of a Lorentzian manifold emerge from the causal relations (recall Section 11.1).

What changed is our interpretation: The points $E \in M$ are the *boxes themselves* (not the sets of events collected in them). Nothing of this construction has to do with reality (we don't change the causality graph); this *layout* is merely a convenient way to *represent* reality.

d | The diffeomorphism invariance of GENERAL RELATIVITY (and thus the hole argument) are now trivial consequences of the fact that the above description to lay out the causality graph on a 4D grid of boxes *is not unique*: It is obvious that there is a quite lot of freedom in placing the causally related events in nearby boxes.

For example, an alternative layout that differs from the previous one only in the orange "hole" is the following:

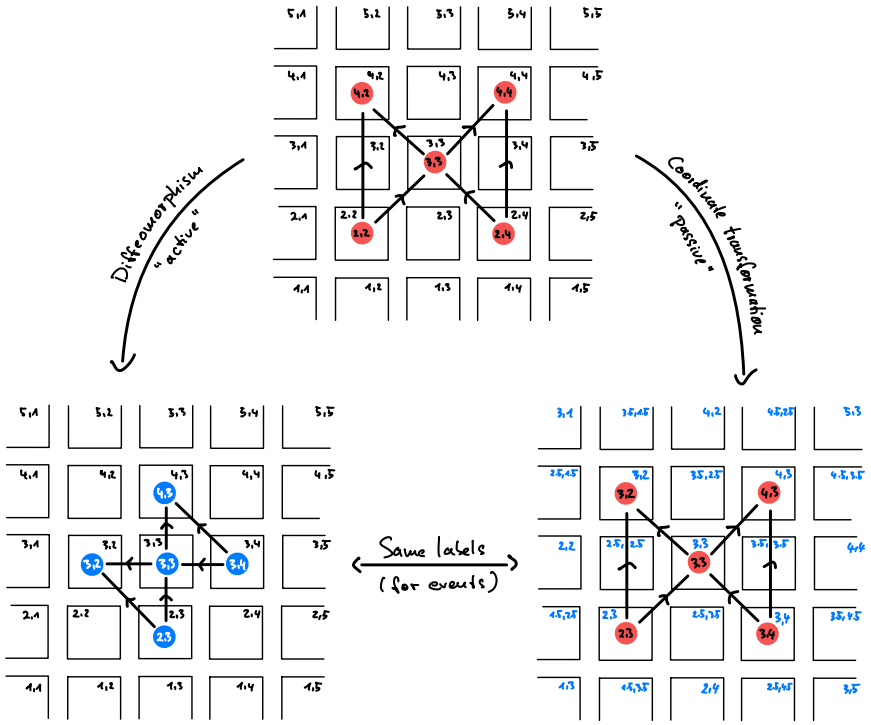


This is the discrete version of Einstein's "hole diffeomorphisms". From this perspective, it is trivial that such a transformation must be *gauge* because all physics is encoded in the causality graph of events (which remains the same). Note how the emerging light/null-cone texture is "warped", as expected from an (active) diffeomorphism that affects the metric.

e | We now understand diffeomorphism invariance and the hole argument. But what about *general covariance*? Up to now we didn't even mention *coordinates*!

Coordinates are simply *labels* that we print onto the boxes to refer to them in our equations. This is what we mean by a chart that assigns the coordinates x^μ to a point (= box) $E \in M$. It is also convenient to assign the label of a box to all events placed *in* the box; this is what we mean by expressions like $A^\mu(x)$ if $A(E) = A^\mu(x)\partial_\mu$ is the value of the field at $E \in M$.

For example, here is a systematic way to label the boxes and 5 exemplary events (e.g., the values of the EM field):



Now that we have boxes (that don't exist) with labels (that don't exist either), the duality of (active) diffeomorphisms and (passive) coordinate transformations becomes evident:

(A) *Diffeomorphism (active view):*

Keep the labels of the boxes, but move the events around (thereby assigning new labels to the events).

(B) *Coordinate transformation (passive view):*

Keep the events in their boxes, but change the labels of all boxes (thereby assigning new labels to the events).

Note that both transformation can lead to the same labeling of events (if the coordinate transformation is chosen "inverse" to the diffeomorphism)!

- From this perspective, the statement that GENERAL RELATIVITY is gauged by (active) diffeomorphisms is dual to the statement that its equations are generally covariant, i.e., form invariant under (passive) coordinate transformations.
- This equivalence hinges on the fact that *all physical content is encoded in the causality graph* which implies that *the boxes are not physical entities*; this is ← *background independence*. By contrast, a physical theory that is background *dependent* assigns physical reality to the boxes themselves (but not the labels) by associating them with some events (= field) that are not moved around with the other events. With such "static" events in place, the duality between diffeomorphisms and coordinate transformations is lost!

  This is why SPECIAL RELATIVITY can be formulated generally covariant without being diffeomorphism invariant. In this case, the "static" background structure is the Minkowski metric and the boxes make up Minkowski space.

v | Comment on ↑ *Scientific realism*:

When we deny the manifold M existence (and relegate it to a useful auxiliary structure of "empty boxes"), we must find an answer to the following question (otherwise the 4D embeddability of the causal network of events is a "miracle"):

> Why does the causal graph have the topology of a 4D manifold?

- I have no answer to this question ☹ (and there is certainly no consensus among scientists, let alone philosophers). However, it seems that any reasonable theory beyond GENERAL RELATIVITY (quantum gravity …) must answer this question.
- A potential solution to the question is the line of arguments discussed in Section 4.4.
- Scientific realism is the epistemological stance that *there exists physical entities out there* that we descibe by our theories – independent of whether (and how) we observe them. In philosophy, scientific realism is an attempt to explain "why science works."

  [For example: To understand the effectiveness of Maxwell's equation in describing electromagnetic phenomena, it is certainly useful to assume that the electromagnetic field $F_{\mu\nu}$ (or, to some extent, the gauge field A_μ) *really exists* – despite the fact that nobody has ever directly observed these fields.]

2 | Where is SPECIAL RELATIVITY?

i | Here is a riddle:

1. In SPECIAL RELATIVITY we were proud of our discovery that Maxwell's equations were forminvariant under Lorentz transformations (Lorentz covariant) but *not* under Galilei transformations.
2. In GENERAL RELATIVITY we were proud of our discovery that coordinate systems don't exist, and all fundamental physical theories must be expressible in a generally covariant form. We achieved this for Maxwell's equations.
3. But then these generally covariant Maxwell equations must be forminvariant under both Lorentz and Galilei transformations (among others). The distinguished status of Lorentz transformations seems to be lost.

What is going on?

ii | We use the massless Klein-Gordon field Eq. (11.36) for its simplicity to resolve the puzzle:

You can of course use the (more complicated) Maxwell equations to make the same points.

a | The Klein-Gordon field theory is defined in SPECIAL RELATIVITY as follows:

$$\eta^{\mu\nu}\partial_\mu\partial_\nu\phi(x) = 0 \tag{12.84}$$

- Black: Equation (= definition of the theory/model)
- Red: Solution (= possible evolution)

b | $\prec$ Arbitrary diffeomorphism $\bar{x} = \varphi(x)$

$\rightarrow$ Define new field $\bar{\phi}(\bar{x}) := \phi(\varphi^{-1}(\bar{x})) = \phi(x)$

$\rightarrow$ φ is *symmetry* of Eq. (12.84) iff

$$\left\{\begin{array}{l}\eta^{\mu\nu}\partial_\mu\partial_\nu\phi(x) = 0\\ \eta^{\alpha\beta}\bar{\partial}_\alpha\bar{\partial}_\beta\phi(\bar{x}) = 0\end{array}\right\} \quad \underset{\text{sym.}}{\overset{\varphi}{\Longleftrightarrow}} \quad \left\{\begin{array}{l}\eta^{\mu\nu}\partial_\mu\partial_\nu\bar{\phi}(x) = 0\\ \eta^{\alpha\beta}\bar{\partial}_\alpha\bar{\partial}_\beta\bar{\phi}(\bar{x}) = 0\end{array}\right\} \tag{12.85}$$

The differential equations in braces are trivially equivalent because x and $\bar{x}$ are dummy variables (α and β are dummy indices) and the equations are assumed to be satisfied for all coordinates x / points on the manifold.

Let us check under which conditions on φ we can get from the left-hand side of Eq. (12.85) to the righ-hand side (and vice versa):

$$\eta^{\mu\nu}\partial_\mu\partial_\nu\phi(x) = 0 \tag{12.86a}$$

$$\Leftrightarrow \quad \eta^{\mu\nu}\partial_\mu\partial_\nu\bar{\phi}(\bar{x}) = 0 \tag{12.86b}$$

$$\Leftrightarrow \quad \eta^{\mu\nu}\partial_\mu\left[\left(\bar{\partial}_\beta\bar{\phi}(\bar{x})\right)\frac{\partial\bar{x}^\beta}{\partial x^\nu}\right] = 0 \tag{12.86c}$$

$$\Leftrightarrow \quad \eta^{\mu\nu}\left[\left(\bar{\partial}_\alpha\bar{\partial}_\beta\bar{\phi}(\bar{x})\right)\frac{\partial\bar{x}^\beta}{\partial x^\nu}\frac{\partial\bar{x}^\alpha}{\partial x^\mu} + \left(\bar{\partial}_\beta\bar{\phi}(\bar{x})\right)\frac{\partial^2\bar{x}^\beta}{\partial x^\mu\partial x^\nu}\right] = 0 \tag{12.86d}$$

When is this expression equivalent to the right-hand side of Eq. (12.85)?

First, the linear order term must vanish. This implies

$$\frac{\partial^2\bar{x}^\beta}{\partial x^\mu\partial x^\nu} \overset{!}{=} 0 \quad \Leftrightarrow \quad \bar{x}^\beta = M^\beta{}_\alpha\, x^\alpha + b^\beta\,, \tag{12.87}$$

which means that the diffeomorphism φ must be an *affine map*.

With this constraint, Eq. (12.86d) simplifies to

$$\left(M^\alpha{}_\mu\,\eta^{\mu\nu}M^\beta{}_\nu\right)\bar{\partial}_\alpha\bar{\partial}_\beta\bar{\phi}(\bar{x}) = 0\,. \tag{12.88}$$

This is equivalent to the right-hand side of Eq. (12.85) if

$$\left(M^\alpha{}_\mu\,\eta^{\mu\nu}M^\beta{}_\nu\right) \overset{!}{=} \eta^{\alpha\beta}\,. \tag{12.89}$$

But this is the defining relation for *isometries* of Minkowski space [← Eq. (4.21)], and we already know that this defines the Lorentz group $\mathrm{O}(1,3)$ (recall Section 4.2). Hence we can conclude:

$\overset{\circ}{\rightarrow}$

The symmetries (← *invariance group*, Section 1.2) of the Klein-Gordon equation include ← *Poincaré transformations*.

¡! This is a statement about *active* transformations of fields: Poincaré transformations are a "machine" to construct new solutions of the Klein-Gordon equation. First, this is a useful mathematical tool, and second, it is physically significant as it implies that if the field evolution ϕ can be observed, then so can $\bar{\phi}$. Nothing of this has to do with coordinates!

This suggests the following definition:

A theory is *relativistic* (in the sense of SPECIAL RELATIVITY) if its invariance group contains the *Poincaré group*.

Note that this definition makes no reference to coordinate transformations and how equations transform under such!

c | But in SPECIAL RELATIVITY we always talked about "Lorentz covariant equations" that do not change under Lorentz/Poincaré transformations, now interpreted as *coordinate transformations*.

To understand how this relate to the previous discussion, let us once again focus on the Klein-Gordon equation, but now we perform a

$\prec$ Arbitrary coordinate transformation $\bar{x} = \varphi(x)$

It is convenient to interpret Eq. (12.84) as a generally covariant equation:

$$g^{\mu\nu}\nabla_\mu\nabla_\nu\phi = 0 \tag{12.90}$$

with $g^{\mu\nu}(x) = \eta^{\mu\nu}$, $\nabla_\nu\phi = \partial_\nu\phi$, and

$$\nabla_\mu\Box_\nu = \partial_\mu\Box_\nu - \Gamma^\alpha{}_{\mu\nu}\Box_\alpha \quad \text{with} \quad \Gamma^\alpha{}_{\mu\nu} = 0\,. \tag{12.91}$$

Under $\bar{x} = \varphi(x)$ the equation remains *forminvariant* in the sense that:

$$g^{\alpha\beta}(x)\nabla_\alpha\nabla_\beta\phi(x) = 0 \quad \overset{\bar{x}=\varphi(x)}{\Longleftrightarrow} \quad \bar{g}^{\mu\nu}(\bar{x})\bar{\nabla}_\mu\bar{\nabla}_\nu\bar{\phi}(\bar{x}) = 0 \tag{12.92}$$

with

$$\bar{g}^{\mu\nu}(\bar{x}) = \frac{\partial\bar{x}^\mu}{\partial x^\alpha}\frac{\partial\bar{x}^\nu}{\partial x^\beta}g^{\alpha\beta}(x) = \frac{\partial\bar{x}^\mu}{\partial x^\alpha}\frac{\partial\bar{x}^\nu}{\partial x^\beta}\eta^{\alpha\beta} \tag{12.93}$$

and

$$\bar{\nabla}_\mu\bar{\nabla}_\nu\bar{\phi}(\bar{x}) \overset{\text{def}}{=} \bar{\partial}_\mu\bar{\partial}_\nu\bar{\phi}(\bar{x}) - \bar{\Gamma}^\alpha{}_{\mu\nu}\bar{\partial}_\alpha\bar{\phi}(\bar{x}) \tag{12.94a}$$

$$\overset{10.39}{=} \big(\partial_\alpha\partial_\beta\phi(x)\big)\frac{\partial x^\beta}{\partial\bar{x}^\nu}\frac{\partial x^\alpha}{\partial\bar{x}^\mu} + \cancel{\big(\partial_\beta\phi(x)\big)\frac{\partial^2 x^\beta}{\partial\bar{x}^\mu\partial\bar{x}^\nu}} \tag{12.94b}$$
$$\cancel{-\left(\frac{\partial\bar{x}^\alpha}{\partial x^\rho}\frac{\partial^2 x^\rho}{\partial\bar{x}^\mu\partial\bar{x}^\nu}\right)\big(\partial_\beta\phi(x)\big)\frac{\partial x^\beta}{\partial\bar{x}^\alpha}}$$

$$\overset{12.91}{=} \frac{\partial x^\alpha}{\partial\bar{x}^\mu}\frac{\partial x^\beta}{\partial\bar{x}^\nu}\nabla_\alpha\nabla_\beta\phi(x) \tag{12.94c}$$

Of course you don't have to do this step-by-step calculation; the whole point of introducing covariant derivatives was that the object transforms like a tensor!

d | Now comes the punchline:

- General covariance:

  The property Eq. (12.92) is what we call *general covariance*; it is valid for *arbitrary* coordinate transformations φ, including Lorentz *and* Galilei transformations:

$$g^{\mu\nu}\nabla_\mu\nabla_\nu\phi(x) = 0 \quad \overset{\varphi=\left\{\begin{array}{l}\text{Lorentz}\\ \text{Galilei}\\ \dots\end{array}\right.}{\Longleftrightarrow} \quad \bar{g}^{\mu\nu}\bar{\nabla}_\mu\bar{\nabla}_\nu\bar{\phi}(\bar{x}) = 0 \tag{12.95}$$

  But this does not imply that φ is a *symmetry* of the equation *because the transformed equation on the right is not* functionally equivalent *to the equation on the left!* This means: If you relabel the dummy variable $\bar{x} \mapsto x$ in the right equation, you don't end up with original equation on the left because in general:

$$g^{\mu\nu}(x) \neq \bar{g}^{\mu\nu}(x) \quad \text{and} \quad \nabla_\mu \neq \bar{\nabla}_\mu \tag{12.96}$$

  → The transformed solution $\bar{\phi}(x)$ solves a *functionally different* equation!

- Let use show explicitly that the two equations are not functionally identical. To this end, introduce the explicit notation $\Gamma^\alpha{}_{\mu\nu}[g](x)$, which tells us to use the metric $g_{\mu\nu}$ to compute the Christoffel symbols from their definition Eq. (10.79), and interpret the result as a function of the spacetime coordinates x.

  With this notation, the left equation of (12.95) reads explicitly

  $$g^{\mu\nu}(x)\left[\partial_\mu\partial_\nu\phi(x) - \Gamma^\alpha{}_{\mu\nu}[g](x)\partial_\alpha\phi(x)\right] = 0\,, \tag{12.97}$$

  whereas the right equation reads

  $$\bar{g}^{\mu\nu}(\bar{x})\left[\bar{\partial}_\mu\bar{\partial}_\nu\bar{\phi}(\bar{x}) - \Gamma^\alpha{}_{\mu\nu}[\bar{g}](\bar{x})\bar{\partial}_\alpha\bar{\phi}(\bar{x})\right] = 0\,. \tag{12.98}$$

  These two equations have the same form – they are *forminvariant*; and Eq. (12.98) is equivalent to Eq. (12.97) if both $g_{\mu\nu}$ and ϕ transform as usual for a tensor and a scalar. But the variable $\bar{x}$ in Eq. (12.98) is a dummy variable (ignoring potential domain issues); thus let us rename it $\bar{x} \mapsto x$ so that the differential equation reads

  $$\bar{g}^{\mu\nu}(x)\left[\partial_\mu\partial_\nu\bar{\phi}(x) - \Gamma^\alpha{}_{\mu\nu}[\bar{g}](x)\partial_\alpha\bar{\phi}(x)\right] = 0\,. \tag{12.99}$$

  But this differential equation is not the same as Eq. (12.97) because $\bar{g}^{\mu\nu} \neq g^{\mu\nu}$ for arbitrary transformations φ. This is why the new function $\bar{\phi}(x)$ solves a *different* equation in general – and not the old one. But then φ does not automatically lead to a symmetry that can be used to construct new solutions from old ones.

- Some might complain: Wait, wasn't the point of general covariance that equations are *forminvariant* under arbitrary coordinate transformations? Well, yes, but with "forminvariance" one means exactly the above transformation; and not that the equation remains functionally *identical*.

  Recall (Chapter 3) that the whole point of introducing tensors and "generally covariant equations" (= tensor equations) was to characterize coordinate-*dependent* (!) equations that encode coordinate-*independent* equations (= relations between geometric objects). The transformation Eq. (12.95) guarantees that the equation can be written coordinate-*free* as (→ *below*)

  $$g^{ab}\nabla_a\nabla_b\phi = 0\,, \tag{12.100}$$

  and that's the whole point.

- Symmetry:

  For a scalar field, a transformation φ is a symmetry if, for a solution $\phi(x)$, the new function $\bar{\phi}(x) := \phi(\varphi^{-1}(x))$ is another solution of the *old* equation:

  $$g^{\alpha\beta}\nabla_\alpha\nabla_\beta\phi(x) = 0 \quad \overset{\varphi = \left\{\begin{array}{l}\text{Lorentz ?}\\ \text{Galilei ?}\\ \dots\ ?\end{array}\right.}{\Longleftrightarrow} \quad g^{\mu\nu}\nabla_\mu\nabla_\nu\bar{\phi}(x) = 0 \tag{12.101}$$

  This is clearly not the same equivalence as in Eq. (12.95).

  Eq. (12.101) & Eqs. (12.97) and (12.99) →

  $$\Gamma^\alpha{}_{\mu\nu}[\bar{g}] \overset{!}{=} \Gamma^\alpha{}_{\mu\nu}[g] = 0 \quad \text{and} \quad \bar{g}^{\mu\nu} \overset{!}{=} g^{\mu\nu} = \eta^{\mu\nu} \tag{12.102}$$

  Using the transformation of connection coefficients Eq. (10.39), one immediately derives Eq. (12.87) from the first condition; this again implies an affine

form Eq. (12.87) for the transformation φ. The second condition is equivalent to Eq. (12.89) and restricts φ to the *isometry group* of Minkowski space, that is: ← *Poincaré transformations*.

$\overset{\circ}{\to}$ Symmetries of the Klein-Gordon equation on Minkowski space:

$$g^{\alpha\beta}\nabla_\alpha\nabla_\beta\phi(x) = 0 \quad \overset{\varphi=\{\text{ Poincaré}}{\Longleftrightarrow} \quad g^{\mu\nu}\nabla_\mu\nabla_\nu\bar{\phi}(x) = 0 \tag{12.103}$$

Brief round-up:

- We started by writing the (special) relativistic Klein-Gordon equation in tensorial form. The equation then becomes generally covariant, i.e., forminvariant under arbitrary coordinate transformations (in particular: Galilei transformations). But these do not translate to (active) symmetries: The transformations of fields that map solutions onto new solutions are still only Poincaré transformations. The takeaway is that Galilei transformations (or any other non-Poincaré transformations) *are not isometries of Minkowski space*, and this spoils their use for constructing new solutions from old ones.
- A nice benefit of the generally covariant form Eq. (12.90) is that it can be used to define the Klein-Gordon field on arbitrary curved spacetimes, not only on Minkowski space. If $g_{\mu\nu}$ is the metric of some generic spacetime, the equation remains of course forminvariant under coordinate transformations. But which of these passive transformations can be reinterpreted as active *symmetries*? Our argument above still goes trough and we are tasked with finding the *isometries* of the new spacetime. But, as mentioned in Section 11.5, a generic spacetime doesn't have any Killing fields, and therefore also no (continuous group of) symmetries. Thus, on a generic spacetime, the Klein-Gordon equation does not have the Poincaré group as (part of) its symmetry group, because the spacetime on which it is formulate doesn't have this symmetry either.

iii | *Question:*

Is it possible to construct a generally covariant theory for which every *(passive) coordinate transformation can be interpreted as an (active)* symmetry*?*

Compare Eq. (12.97) and Eq. (12.99):

$$\text{Solve for } \phi\text{:} \quad \begin{cases} g^{\mu\nu}(x)\left[\partial_\mu\partial_\nu\phi(x) - \Gamma^\alpha{}_{\mu\nu}[g](x)\partial_\alpha\phi(x)\right] = 0 \\ \bar{g}^{\mu\nu}(x)\left[\partial_\mu\partial_\nu\bar{\phi}(x) - \Gamma^\alpha{}_{\mu\nu}[\bar{g}](x)\partial_\alpha\bar{\phi}(x)\right] = 0 \end{cases} \tag{12.104}$$

Problem: $\phi(x)$ and $\bar{\phi}(x)$ solve *different* equations (compare the black equations).

Idea: Interpret the metric as *solution* and not as background (= part of the equation).

→

$$\text{Solve for } (g,\phi)\text{:} \quad \begin{cases} g^{\mu\nu}(x)\left[\partial_\mu\partial_\nu\phi(x) - \Gamma^\alpha{}_{\mu\nu}[g](x)\partial_\alpha\phi(x)\right] = 0 \\ \bar{g}^{\mu\nu}(x)\left[\partial_\mu\partial_\nu\bar{\phi}(x) - \Gamma^\alpha{}_{\mu\nu}[\bar{g}](x)\partial_\alpha\bar{\phi}(x)\right] = 0 \end{cases} \tag{12.105}$$

(g,ϕ) and $(\bar{g},\bar{\phi})$ solve *the same* equation ☺.

→ We just prepared the theory for ← *background independence*.

Whether the theory really is background independent depends on the presence and type of additional equations of motion that constrain the metric field (→ *below*). However, what we can say is that the theory has no longer any *absolute objects* (= non-dynamical tensor fields).

We conclude:

$$\underbrace{\left.\begin{array}{l}\text{No absolute objects}\\ \text{General covariance}\end{array}\right\}}_{\textit{Passive}\text{ transformations}} \quad\Rightarrow\quad \underbrace{\text{Diffeomorphism invariance}}_{\textit{Active}\text{ transformations}}$$

- As argued above, diffeomorphism invariance must be interpreted as a *gauge symmetry* that relates physically equivalent solutions. This means that our "trick" to declare the metric as a dynamical field to lift all coordinate transformations to (active) symmetries (now also Galilei transformations are symmetries!) didn't really work out as intended. It is as if we wanted "too much": Now that *every* diffeomorphism is a symmetry, *none* of them is *physical* anymore – all of them are *gauge*! But the good old (physical) Poincaré symmetry can of course be resurrected for metric *solutions* that have the appropriate Killing fields.
- Interpreting Eq. (12.90) as an equation for (g, ϕ) makes the theory diffeomorphism invariant, i.e., every coordinate transformation can be interpreted as an active symmetry transformation. Without restricting the new dynamical field $g_{\mu\nu}$ by an additional equation of motion [e.g., the Einstein field equation Eq. (12.10) (→ *below*)], this is a rather useless construction because the theory has solutions for *every* metric you ask for. Hence it cannot *predict* anything about the metric, only about the evolution of the Klein-Gordon field in relation to the metric.

iv | Conclusion:

We can sum up our findings as follows:

- Global Lorentz/Poincaré transformations $\phi \overset{\Lambda}{\mapsto} \bar{\phi}$ of matter fields are *not* symmetries of GENERAL RELATIVITY, because the metric typically lacks the necessary Killing fields.
- Global Lorentz/Poincaré transformations $(g, \phi) \overset{\Lambda}{\mapsto} (\bar{g}, \bar{\phi})$ of *both* matter fields and metric are *gauge* symmetries of GENERAL RELATIVITY; they have *no* physical significance.

- So is SPECIAL RELATIVITY gone? Well, yes, if we identify the theory with "global Lorentz symmetry" the answer must be affirmative: GENERAL RELATIVITY does *not* contain SPECIAL RELATIVITY in its pure form because spacetime is a dynamical field – and solving for it usually does *not* produce flat Minkowski space. Only in the situations where it does, GENERAL RELATIVITY reduces to SPECIAL RELATIVITY (which is approximately true in interstellar space far away from matter, and with appropriate boundary conditions).
- *Is this a problem?* The answer is of course *no*, but it is instructive understand why:

  Minkowski space is the defining entity of SPECIAL RELATIVITY and has two characteristic features: it is *flat* and it is *Lorentzian* [it has metric signature $(1, 3)$]. The crucial

insight is that its *flatness* is not a characteristic feature of reality, it is a simplicity assumption that makes SPECIAL RELATIVITY "unnaturally symmetric" (10 Killing fields!). Hence the global Lorentz symmetry of relativistic theories – which is imprinted by the symmetry of spacetime due to general covariance – is "unnatural" *mutatis mutandis*. The core feature of reality, that SPECIAL RELATIVITY actually brought to the table, is the *locality of causality*, which is ensured by the *Lorentz signature* $(1,3)$ alone. The realization that this core feature is not tied to the flatness of Minkowski space leads directly to GENERAL RELATIVITY.

→ *Essence* of SPECIAL RELATIVITY that survives in GENERAL RELATIVITY:

$$\left.\begin{array}{r}\text{General covariance}\\ \text{Lorentzian metric}\end{array}\right\} \quad \Rightarrow \quad \left\{\begin{array}{l}\textit{Local}\text{ Lorentz symmetry}\\ \textit{Locality}\text{ of causality}\\ \textit{Local}\text{ constancy of the speed of light}\end{array}\right\}$$

- Recall the "spoiler" in Section 0.6.
- *Aside:* You know now three theories to describe classical mechanics: Good old Newtonian mechanics, SPECIAL RELATIVITY, and GENERAL RELATIVITY. It is established practice to teach these subjects in this very order:

$$\underbrace{\text{Newtonian mechanics}}_{\text{2nd term}} \xrightarrow{\text{then}} \underbrace{\text{SPECIAL RELATIVITY}}_{\text{5th term}} \xrightarrow{\text{then}} \underbrace{\text{GENERAL RELATIVITY}}_{\text{6th term}}$$

It seems to be consensus that the reason for this is how hard or easy the subjects are. While there certainly is some (pedagogic) truth to this assessment, I would like to point out that the "complexity" of a theory can be gauged in (at least) two different ways. For lack of better terms I will refer to them as *operational complexity* and *conceptual complexity*.

Operational complexity captures how hard it is to work out actual problems in the respective theory. This leads to the following grading:

$$\vec{F} = m\vec{a} \xrightarrow[\text{than}]{\text{simpler}} K^\mu = m\frac{\mathrm{d}u^\mu}{\mathrm{d}\tau} \xrightarrow[\text{than}]{\text{simpler}} K^\mu = m\frac{\mathrm{d}u^\mu}{\mathrm{d}\tau} + m\Gamma^\mu{}_{\alpha\beta}\,u^\alpha u^\beta$$

Since students must solve problems to internalize a theory, it is this order, from the operationally simple Newtonian mechanics to the operationally hard GENERAL RELATIVITY, that supports the conventional approach to teach these subjects.

By contrast, *conceptual complexity* captures how much "conceptual scaffolding" is needed to formulate the theory precisely. Based on the discussions above, my claim is that the order is exactly opposite:

$$\text{GENERAL RELATIVITY} \xrightarrow[\text{than}]{\text{simpler}} \text{SPECIAL RELATIVITY} \xrightarrow[\text{than}]{\text{simpler}} \text{Newton}$$

The argument is simple: While more symmetries (or structures) make *solving* problems easier (thereby *lowering* the *operational* complexity), they tend to clutter the conceptual framework of a theory (thereby *increasing* the *conceptual* complexity). In addition, they often obfuscate the actually important structures of the theory (recall the discussion of the flatness of Minkowski space above).

One might counter that surely the conceptual framework of Newtonian mechanics is not harder than that of SPECIAL RELATIVITY! I beg to disagree: If one carves out the mathematics of Newtonian spacetime *properly*, one has to deal with ↑ *affine*

manifolds, ↑ *fiber bundles*, etc…[for a proper definition of Newtonian spacetime, see STRAUMANN [8] (pp. 10–16)]. Put bluntly: Newtonian mechanics looks conceptually "simple" because it is usually not done rigorously! (At least compared to how rigorously we do GENERAL RELATIVITY.)

3 | Interlude: ⁂ *Abstract index notation*:

It is often convenient to write equations in a coordinate-*free* form, without loosing information about the types of tensors involved, and how they act on each other. This is achieved by abstract index notation:

$$\begin{aligned}
&\text{Scalar:} && \phi := \phi && \in \mathbb{R} && (12.106a)\\
&\text{Vector:} && A^a := A^\mu \partial_\mu && \in TM && (12.106b)\\
&\text{Covector:} && B_a := B_\mu \mathrm{d}x^\mu && \in T^*M && (12.106c)\\
&\text{Mixed tensor:} && T^a{}_b := T^\mu{}_\nu \, \partial_\mu \otimes \mathrm{d}x^\nu && \in TM \otimes T^*M && (12.106d)\\
&\dots
\end{aligned}$$

¡! The roman indices $a, b, \dots$ are *not* numerical indices, they are *labels* that indicate "slots" of tensors and how they are applied to each other. For example:

$$B_a A^a := B_a(A^a) = B_\mu \mathrm{d}x^\mu (A^\nu \partial_\nu) = B_\mu A^\nu \mathrm{d}x^\mu(\partial_\nu) = B_\mu A^\nu \delta^\mu_\nu = B_\mu A^\mu \,. \quad (12.107)$$

Example: The Klein-Gordon equation in coordinate-free notation reads:

$$\eta^{ab} \nabla_a \nabla_b \phi = 0 \quad (12.108)$$

where $\eta_{ab} = g_{\mu\nu}(x) \mathrm{d}x^\mu \mathrm{d}x^\nu$ denotes the Minkowski metric. [This is not the *matrix* $\eta_{\mu\nu} = \mathrm{diag}\,(1, -1, -1, -1)$! Furthermore, the *components* $g_{\mu\nu}(x)$ only equal $\eta_{\mu\nu}$ in inertial coordinates.]

4 | Background independence:

How does the concept of ← *background independence* (Section 9.2) mesh with these concepts?

This discussion is based on Ref. [118].

i | We have found the following implication:

$$\text{Background independence} \Rightarrow \left\{ \begin{array}{l} \text{No absolute objects} \\ \text{General covariance} \end{array} \right\} \Rightarrow \text{Diffeomorphism invariance} \quad (12.109)$$

This suggests the following identification:

(?) A theory is background independent *iff* it is diffeomorphism invariant.

ii | Problem:

◁ The following theory (in abstract index notation):

$$\texttt{KG-SRT}: \left\{ \begin{array}{rl} g^{ab} \nabla_a \nabla_b \phi = 0 & \text{(Matter EOM)} \\ R^a{}_{bcd} = 0 & \text{(Metric EOM)} \end{array} \right. \quad (12.110)$$

This theory …

- … is coordinate-free (in components: generally covariant).
- … has no absolute objects [solutions are tuples (g_{ab}, ϕ)].

- … is equivalent to the Klein-Gordon field in SPECIAL RELATIVITY.

  (The only flat metric on $M \simeq \mathbb{R}^4$ is the Minkowski metric $g_{ab} = \eta_{ab}$.)

→ KG-SRT is a *diffeomorphism invariant* formulation of SPECIAL RELATIVITY!

But clearly SPECIAL RELATIVITY should count as a background-*dependent* theory, for it is defined on non-dynamical Minkowski space! How can we extract this characteristic feature from artificially diffeomorphism invariant formulations like KG-SRT?

iii | To this end, let us compare KG-SRT with the Klein-Gordon field in GENERAL RELATIVITY:

$$\text{KG-GRT}: \begin{cases} g^{ab}\nabla_a\nabla_b\phi = 0 & \text{(Matter EOM)} \\ R_{ab} - \frac{1}{2}Rg_{ab} = -\kappa T_{ab} & \text{(Metric EOM)} \end{cases} \tag{12.111}$$

Here T_{ab} depends on the KG-field with components given in Eq. (11.118) for $m = 0$.

→ Compare solutions:

$$\text{KG-SRT} \quad \Rightarrow \quad (\eta_{ab}, \phi^1),\ (\eta_{ab}, \phi^2),\ (\eta_{ab}, \phi^3), \ldots \tag{12.112}$$

$$\text{KG-GRT} \quad \Rightarrow \quad (g^1_{ab}, \phi^1),\ (g^2_{ab}, \phi^2),\ (g^3_{ab}, \phi^3), \ldots \tag{12.113}$$

→ KG-SRT has a "hidden" *absolute object* (η_{ab}) shared by all solutions!

iv | The fact that all solutions of a theory like SPECIAL RELATIVITY share some invariant objects allows for a cascade of "specializations" of the formulation of the theory:

Formulation	Solve for …	Diff. inv.	Coord. free	Gen. cov.
$g^{ab}\nabla_a\nabla_b\phi = 0$ $R^a{}_{bcd} = 0$	g_{ab}, ϕ	✓	✓	✓
Fix shared metric $g_{ab} = \eta_{ab} = \tilde{\eta}_{\mu\nu}(x)\mathrm{d}x^\mu\mathrm{d}x^\nu$ as absolute element →				
$\eta^{ab}\nabla_a\nabla_b\phi = 0$	ϕ	✗	✓	✓
Write in components wrt. some coordinates →				
$\tilde{\eta}^{\mu\nu}(x)\nabla_\mu\nabla_\nu\phi = 0$	ϕ	✗	✗	✓
Choose inertial coordinates to exploit symmetry of Minkowski space →				
$\eta^{\mu\nu}\partial_\mu\partial_\nu\phi = 0$	ϕ	✗	✗	✗

- All formulations above are equivalent in the sense that they describe the same physics.
- Thus, the very fact that we could formulate SPECIAL RELATIVITY in a *non*-diffeomorphism invariant form (and even a *non*-generally covariant form) characterizes it as a background-*dependent* theory.
- For GENERAL RELATIVITY, the first step (were one fixes the metric as an absolute element of the theory) fails and the cascade cannot take off. This prevents non-diffeomorphism invariant formulations of the theory. Since there is no distinguished metric, there cannot be a distinguished coordinate system, and thereby no non-generally covariant formulation either.

v | This leads us to the following refined definition of background independence:

> Background independent theories (like GENERAL RELATIVITY) are characterized by their *lack* of a formulation that is *not* diffeomorphism invariant.

This explains why we didn't encounter a formulation of GENERAL RELATIVITY that is *not* generally covariant, while we did use such formulations when discussing SPECIAL RELATIVITY.