8.1 Minimal coupling

We will now discuss the influence of the gravitational field on matter. We have learned how to write equations on a manifold in a coordinate-independent way. By construction, when written out in components, such equations have the same form in any coordinate system and therefore they satisfy the principle of General Covariance. However, as mentioned before, not all generally covariant equations have to do with gravitation. In order to write the equations that govern physical phenomena in a gravitational field, we have to resort to the Principle of Equivalence, which tells us that the laws of physics in a gravitational field must reduce to those of Special Relativity in the infinitesimally inertial frames. These frames are characterized by the property that the metric can be approximated by the Minkowski metric and the connection coefficients can be neglected. In a flat spacetime such properties hold exactly throughout spacetime. In the presence of a gravitational field they can be satisfied to any desired degree of accuracy by choosing a reference frame which is in free fall and restricting attention to a sufficiently small subset of spacetime.

The condition of reducing to the laws of special relativity in the infinitesimally inertial frames puts strong restrictions on the form of the coupling of matter to gravity. In practice, there is a simple heuristic principle that allows us to write equations for matter in the presence of gravity. We note that if an equation is written for an arbitrary metric g and connection ∇, when restricted to Minkowski spacetime it still holds with $g_{\mu\nu}$ replaced by $\eta_{\mu\nu}$ and ∇_μ replaced by ∂_μ. Conversely, given any special relativistic equation written in tensorial form, we will obtain a generally covariant equation replacing $\eta_{\mu\nu}$ by $g_{\mu\nu}$ and ∂_μ by ∇_μ and the volume element d^4x by $\sqrt{|g|}d^4x$. This resulting equation will automatically satisfy the Principle of General Covariance and the Equivalence Principle. This procedure is called "minimal coupling". While useful, it does not yield the only possible equations. In general, there may exist many (in fact, infinitely many) generally covariant equations that reduce to the same special relativistic equation. We shall discuss some examples of these ambiguities in this chapter.

8.2 Particle in a gravitational field

In Special Relativity, the world line of a free particle in an inertial frame is a straight line. If we denote $x^\mu(\tau)$ the coordinates of the worldline of the particle, where τ is proper time, and $u^\mu = \frac{dx^\mu}{d\tau}$ the four-velocity of the particle (the vector tangent to the curve), such a free motion obeys the equation

$$\frac{du^\mu}{d\tau} = \frac{d^2x^\mu}{d\tau^2} = 0 \ . \tag{8.2.1}$$

Such worldlines have two special properties: they are self parallel (in the sense that the four-velocity u^μ is constant and hence always parallel to itself), and it is also an extremal for the length. We will now generalize both properties to a curved spacetime M.

Let $c(\lambda)$ be the worldline of a particle, where λ is a parameter along the curve. We denote $x^\mu(\lambda)$ the coordinates of $c(\lambda)$, $u^\mu = \frac{dx^\mu}{d\lambda}$ the vector tangent to the curve and ∇_μ the covariant derivative associated to any linear connection in TM. A curve c is said to be self-parallel with respect to ∇, or a geodesic for the connection ∇, if

$$\nabla_u u = \alpha u \ , \tag{8.2.2}$$

or, in coordinates,

$$u^\mu \nabla_\mu u^\nu = \alpha u^\nu \ , \tag{8.2.3}$$

where α is a function of λ. Without loss of generality one can assume that $\alpha = 0$. In fact, consider another parameter τ such that the four-velocity $u'^\mu = \frac{dx^\mu}{d\tau}$ is related to u by $u = fu'$, for some $f = \frac{d\tau}{d\lambda}$, a function of λ. Then Eq.(8.2.3) implies that

$$u'^\mu \nabla_\mu u'^\nu = \frac{\alpha - u'^\mu \nabla_\mu f}{f} u'^\nu \ , \tag{8.2.4}$$

The right-hand side vanishes provided we choose f such that

$$\frac{1}{f} u^\mu \nabla_\mu f \equiv u(\log f) \equiv \frac{d \log f}{d\lambda} = \alpha \tag{8.2.5}$$

which can be integrated twice to give $\tau(\lambda)$. A parameter τ with the property that the r.h.s. of (8.2.2) vanishes is called an affine parameter. If τ is an affine parameter, any other affine parameter can differ from

τ at most by a transformation of the form $\tau' = a\tau + b$. Using an affine parameter, the equation for the geodesics in the connection ∇ becomes simply

$$\nabla_u u = 0 \ . \tag{8.2.6}$$

Let us write this equation more explicitly in a coordinate system. Using the formula (6.4.2) we have

$$0 = \frac{dx^\mu}{d\tau}\left(\partial_\mu \frac{dx^\nu}{d\tau} + A_\mu{}^\nu{}_\lambda \frac{dx^\lambda}{d\tau}\right) = \frac{d^2x^\nu}{d\tau^2} + A_\mu{}^\nu{}_\lambda \frac{dx^\mu}{d\tau}\frac{dx^\lambda}{d\tau} \ . \tag{8.2.7}$$

Note that the notion of geodesic depends only on the connection and not on a metric. However, in a space with a metric there is also an alternative characterization of geodesics as stationary points of the length functional. In more physical terms, we will interpret $(-m)$ times the length as the action functional of a particle of mass m (the minus sign is such that the action is positive for timelike curves in a metric of signature $-+++$).

Let τ be the proper time along the worldline of the particle:

$$d\tau^2 = -ds^2 = -g_{\mu\nu}dx^\mu dx^\nu \ . \tag{8.2.8}$$

We can take τ as the parameter along the curve. We define the action functional

$$S[c] = -m\int d\tau \tag{8.2.9}$$

and we apply the standard variational procedure with fixed endpoints:

$$\begin{aligned}\delta S =& -m\int \delta\sqrt{d\tau^2}\\ =&\frac{1}{2}m\int\left[2g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{d\delta x^\nu}{d\tau} + \partial_\lambda g_{\mu\nu}\delta x^\lambda \frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}\right]d\tau\\ =&m\int \delta x^\lambda\left[-\frac{d}{d\tau}\left(g_{\mu\lambda}\frac{dx^\mu}{d\tau}\right) + \frac{1}{2}\partial_\lambda g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}\right]d\tau\end{aligned}$$

where in the last step we have performed an integration by parts dropping the boundary terms due to the assumption that $\delta x^\lambda = 0$ at the endpoints. This can be rewritten as

$$\delta S = m\int \delta x^\lambda\left[-\partial_\alpha g_{\mu\lambda}\frac{dx^\alpha}{d\tau}\frac{dx^\mu}{d\tau} - g_{\mu\lambda}\frac{d^2x^\mu}{d\tau^2} + \frac{1}{2}\partial_\lambda g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}\right]d\tau$$

The combination of the first term, suitably symmetrized in μ and α, with the third term reconstructs the Christoffel coefficients $\Gamma_\alpha{}^\mu{}_\beta$ of the metric g, so

$$\delta S = -m\int \delta x^\lambda g_{\lambda\mu}\left[\frac{d^2x^\mu}{d\tau^2} + \Gamma_\alpha{}^\mu{}_\beta\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau}\right]d\tau \ .$$

Requiring stationarity of the action yields the equation for the geodesics of the Levi-Civita connection of g:

$$0 = \frac{d^2x^\nu}{d\tau^2} + \Gamma_\mu{}^\nu{}_\lambda\frac{dx^\mu}{d\tau}\frac{dx^\lambda}{d\tau} = 0 \ . \tag{8.2.10}$$

We thus discover that the geodesics of the Levi-Civita connection of a metric are also the extremal curves of the arc-length for that metric. From a physical point of view, we have rederived the geodesic equation from a variational principle.

In the derivation we also learned that the proper time along a curve is an affine parameter, and furthermore that using this affine parameter, the tangent vector to the curve (*i.e.* the four–velocity) is normalized:

$$g(u,u) = g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = -1 \ . \tag{8.2.11}$$

Let us discuss the physical interpretation of the geodesic equation. First we note that it reduces to (8.2.1) in the case of Minkowski space, in an inertial coordinate system. In the case of Minkowski space with a noninertial coordinate system it becomes identical to (1.2.11), as it should. We have seen that in equation (1.2.11) the coefficients ${\Gamma_\mu}^\nu{}_\lambda$ represent the effect on the particle of the inertial forces. In equation (8.2.10) the connection coefficients contain, in addition to possible inertial forces, the effect on the particle of the gravitational force. As we have seen in section 7.7, it is possible to choose the coordinate system in such a way that $\Gamma = 0$ and the metric becomes Minkowskian at a point. This is then an infinitesimally inertial coordinate system: the effects of gravity vanish to zeroth order in the distance from the chosen point. Whether a genuine gravitational field is present in some region of spacetime or not must be established by looking at the Riemann tensor: its vanishing is a necessary and sufficient condition for the existence of inertial coordinate systems, and hence for the absence of a gravitational field in that region. We argued in section 1.2 that the physical effect of the gravitational field is embodied in the tidal forces. In the next section we will see in detail how the Riemann tensor relates to these forces.

For the time being we observe that in General Relativity gravity is not described as a force but rather as a deformation of the metric structure of spacetime. A particle which moves only under the influence of gravity follows a geodesics of the Levi–Civita connection determined by the metric. It is said to be in "free fall". The possibility of interpreting gravity in this geometrical way hinges crucially on the experimental observation that different particles follow identical trajectories under the influence of gravity. If a deviation from this behavior was ever observed, the geometrical interpretation would collapse.

8.3 Geodesic deviation

In this section M will be a (pseudo)-riemannian manifold with a metric g and ∇ the Levi-Civita connection of g. Consider a one-parameter family of geodesics, parametrized by a parameter s. The affine parameter along the geodesics will be called t. We call $u = \frac{\partial}{\partial t}$ the tangent to the geodesic; they satisfy Eq.(8.2.6) and (8.2.11). The vector $X = \frac{\partial}{\partial s}$ gives the separation between two infinitely close geodesics; it is called the geodesic deviation. We shall now derive an equation obeyed by this vector.

As the parameters t and s vary, the point $c_s(t)$ moves on a two-dimensional surface in M. Because t and s are coordinates on this surface, the vectorfields u and X commute. If ∇ is any torsionfree connection (in particular if it is the Levi-Civita connection of the metric g) we have from (7.4.7)

$$u^\mu\nabla_\mu X^\nu = X^\mu\nabla_\mu u^\nu \ . \tag{8.3.1}$$

Using that u is self-parallel, then the fact that u and X commute, and then using (8.2.11) we have

$$u^\lambda\nabla_\lambda(u^\mu X_\mu) = u^\mu u^\lambda\nabla_\lambda X_\mu = u^\mu X^\lambda\nabla_\lambda u_\mu = \frac{1}{2}X^\lambda\nabla_\lambda(u^\mu u_\mu) = 0 \ .$$

Without loss of generality we can choose the origin of t in each geodesic in such a way that $g(u,X) = u^\mu X_\mu = 0$ at $t = 0$. Using (8.3.2) there follows that

$$u^\mu X_\mu = 0 \tag{8.3.2}$$

everywhere. The vector $\nabla_u X$ is the rate of change of the separation between neighboring geodesics, *i.e.* the relative velocity of neighboring geodesics. The vector $\nabla_u\nabla_u X$ is the relative acceleration of neighboring geodesics. The components of this vector can be computed as follows. Using (8.3.1) one finds first that

$$u^\mu\nabla_\mu(u^\nu\nabla_\nu X^\alpha) = u^\mu\nabla_\mu(X^\nu\nabla_\nu u^\alpha) = u^\mu\nabla_\mu X^\nu\nabla_\nu u^\alpha + u^\mu X^\nu\nabla_\mu\nabla_\nu u^\alpha$$

Using again (8.3.1) on the first term and (6.6.8) on the second term we find

$$u^\mu\nabla_\mu(u^\nu\nabla_\nu X^\alpha) = X^\mu\nabla_\mu u^\nu\nabla_\nu u^\alpha + u^\mu X^\nu\nabla_\nu\nabla_\mu u^\alpha + u^\mu X^\nu {R_{\mu\nu}}^\alpha{}_\beta u^\beta$$

The first and second terms sum up to

$$X^\mu(\nabla_\mu u^\nu \nabla_\nu u^\alpha + u^\nu \nabla_\mu \nabla_\nu u^\alpha) = X^\mu \nabla_\mu (u^\nu \nabla_\nu u^\alpha) = 0$$

so

$$u^\mu \nabla_\mu (u^\nu \nabla_\nu X^\alpha) = u^\mu X^\nu R_{\mu\nu}{}^\alpha{}_\beta \, u^\beta \tag{8.3.3}$$

which can also be written in a coordinate-independent way as

$$\nabla_u \nabla_u X = R(u, X)u \tag{8.3.4}$$

Equation (8.3.4) or (8.3.3) is called the equation of geodesic deviation.

Using this equation one can measure the components of the curvature tensor by performing local experiments with freely falling particles. For example, recall the discussion in section 1.3 about the equivalence principle. Particles in an accelerating rocket in empty space undergo exactly the same acceleration, so if they are initially at rest the relative distances will remain unchanged. The vectors X are constant and the curvature is zero. If we are near the surface of the Earth, we feel a gravitational acceleration g. Most of this acceleration is due to the fact that the coordinate system is not inertial. It can be almost entirely cancelled by going to a system of coordinates in free fall. In this system the Christoffel symbols are zero at a chosen spacetime point x_0 and are of order $x - x_0$ away from it, as discussed in section 9.4. The residual effect is given by the tidal forces and is described by the equation of geodesic deviation.

Exercise 8.3.1. The curvature tensor at the surface of the Earth. In a freely falling elevator, choose local orthonormal coordinates so that the z axis is in the vertical direction, while x and y axes are in the horizontal plane. Place two particles at positions $(t, x, y, z) = (0, 0, 0, 0)$ and $(0, d, 0, 0)$, so that the separation four-vector X has components $(0, d, 0, 0)$, and set them free at the time $t = 0$. To a very good approximation, their four-velocities have components $(1, 0, 0, 0)$. From the point of view of a coordinate system that is fixed to the earth, the acceleration of the particles is in the direction of the center of the Earth, which is nearly exactly the z direction, and is equal to $g \approx 10^{-16} m^{-1}$ in units where $c = 1$. The acceleration of the particles towards each other is in the x direction and its magnitude is equal to $g \sin \delta\theta \approx gd/r_\oplus$ where $\delta\theta$ is the angle between the two particles, seen from the center of the earth. The only nonvanishing component in equation (8.3.3) is the $\alpha = x$ component and using that the radius of the Earth is $r_\oplus \approx 6 \cdot 10^6 m$ we read off that $R_{0x}{}^x{}_0 = -g/r_\oplus \approx -1.6 \cdot 10^{-21} m^{-2}$ (the minus sign is because the acceleration is in the direction of negative x). For $R_{0y}{}^y{}_0$ one finds the same result. If the particle separation is in the vertical direction: $X = (0, 0, 0, d)$, the tidal acceleration is the difference in the gravitational accelerations at the positions of the particles, and it gives $R_{0z}{}^z{}_0 = 2g/r_\oplus \approx 3.2 \cdot 10^{-21} m^{-2}$. We thus learn that the radius of curvature of spacetime at the surface of the Earth is of the order of $10^{11} m$, thus much larger than the radius of the Earth. Furthermore, summing these three components we find for the Ricci tensor $R_{00} = 0$.

This example implies that the components of the curvature tensor can be measured and therefore are observable. This point can give rise to misunderstandings and therefore requires a comment. Since gravity is a gauge theory, only gauge invariant quantities can be observed. Let us restrict ourselves to the metric formulation, so that the gauge group consists of diffeomorphisms (for this discussion it is convenient to adopt the active viewpoint). Since the components of a tensor are not invariant under diffeomorphisms, they are not observable. Even the value of a scalar at a point is not an observable, because under a diffeomorphism f the value of the transformed scalar at a point x is equal to the value of the original scalar at the point $f(x)$ (see equation (3.3.1)). Hence, a scalar is not invariant.

In pure gravity only global quantities such as $\int d^4x \sqrt{|g|} R_{\mu\nu} R^{\mu\nu}$ are observable. Such quantities are not particularly interesting for local physics, however. In fact, since the position of points changes under coordinate transformations, the whole concept of locality becomes questionable. But the world does not consist of gravity alone, and matter can be used to pinpoint positions. If $\bar{x}(t)$ denotes the trajectory of a particle, and R is, for example, the Ricci scalar, consider the value of the Ricci scalar at the position of the particle, $R(\bar{x})$. Under a diffeomorphism f the position of the particle is changed to $\bar{x}'(t) = f^{-1}(\bar{x}(t))$; then the value of the transformed scalar $R' = f^*R$ at $\bar{x}'$ is equal to

$$R'(\bar{x}') = f^*R(f^{-1}(\bar{x})) = R(\bar{x}) \ .$$

Thus, $R(\bar{x})$ is invariant, and therefore observable.

Similarly, suppose that the particle at $\bar{x}$ is surrounded by three other infinitesimally close particles, located along the axes of an orthonormal reference frame. (By infinitesimally close we mean that the distance between the particles is much less that the curvature radii, so that we can regard the displacements between the particles as elements of the tangent space at $\bar{x}$.) As discussed above, with such a setup we can measure the components of the Riemann tensor at $\bar{x}$. It should then be clear that to make local measurements we need a cloud of particles with typical interparticle distances that are much smaller than the curvature radii. This cloud can be described phenomenologically as a fluid. A concrete example of this is used in cosmology, where galaxies can be treated as a perfect pressureless fluid.

8.4 Scalar and electromagnetic field

Applying the prescription of minimal coupling to the Klein-Gordon equation for a scalar field ϕ coupled to gravity we obtain

$$-g^{\mu\nu}\nabla_\mu\nabla_\nu\phi + m^2\phi = 0 \ . \tag{8.4.1}$$

Using the expression for the Christoffel symbols (7.5.5), one finds

$$g^{\mu\nu}\Gamma_\mu{}^\lambda{}_\nu \ = -\frac{1}{\sqrt{|g|}}\partial_\tau(\sqrt{|g|}g^{\tau\lambda}) \tag{8.4.2}$$

and therefore (8.4.1) can be rewritten in the form

$$-\frac{1}{\sqrt{|g|}}\partial_\tau(\sqrt{|g|}g^{\tau\lambda}\partial_\lambda\phi) + m^2\phi = 0 \ . \tag{8.4.3}$$

The first term is seen to coincide with the Laplacian on zero-forms defined in (8.4.9). Equation (8.4.3) can be derived from the action functional

$$S(\phi) = \int d^4x\sqrt{|g|}\left[-\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - \frac{1}{2}m^2\phi^2\right] \ , \tag{8.4.4}$$

where the metric is treated as a fixed background. (Recall that we use a metric with signature $-+++$ so the integrand in (8.4.4) is the correct Lagrangian density $\frac{1}{2}(\dot{\phi})^2 - \frac{1}{2}(\partial_i\phi)^2$.)

Equation (8.4.1) is not the unique one that reduces to the ordinary Klein–Gordon equation in Minkowski space. For example, a common modification is to add a nonminimal term $\frac{1}{2}\xi R\phi^2$ to the Lagrangian, where ξ is a coupling constant. This leads to an additional term $\xi R\phi$ in (8.4.1); in the limit when space becomes flat, this term vanishes. The addition of this term can be motivated by considerations of conformal invariance. The action of a massless scalar field in four dimensions is invariant under scale transformations $g_{\mu\nu} \mapsto \Omega^2 g_{\mu\nu}$, $\phi \mapsto \Omega^{-1}\phi$ with constant Ω, but not under Weyl transformations, when Ω is a positive function. However, the nonminimally coupled scalar theory with $\xi = \frac{1}{6}$ is Weyl invariant.

The electromagnetic field is described in flat spacetime by the four-potential A_μ and the field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. In coordinate-independent language the field A is a one-form and F a two–form and their relation $F = dA$ makes sense on any manifold, with or without metric. If we applied the principle of minimal coupling we would be led to define, in a gravitational field $F_{\mu\nu} = \nabla_\mu A_\nu - \nabla_\nu A_\mu$, but since the Christoffel symbols are symmetric in the lower indices, they cancel out, so this alternative definition of F coincides with the original one.

The prescription of minimal coupling leads us to the following action for an electromagnetic field coupled to gravity:

$$S(A) = -\frac{1}{4}\int d^4x\sqrt{|g|}\, g^{\mu\nu}g^{\rho\sigma}F_{\mu\rho}F_{\nu\sigma} \tag{8.4.5}$$

for an electromagnetic field in a gravitational field. Varying with respect to A_μ leads to the equation

$$\frac{1}{\sqrt{|g|}}\partial_\mu(\sqrt{|g|}F^{\mu\nu}) = 0 \ , \tag{8.4.6}$$

where indices on F have been raised with the metric g. The operator on the l.h.s. is recognized as the covariant divergence δ defined in (8.4.9). Using the formula

$$\Gamma_\mu{}^\mu{}_\nu = \frac{1}{\sqrt{|g|}}\partial_\nu\sqrt{|g|} \tag{8.4.7}$$

which follows from (7.5.5), equation (8.4.6) can also be written in the form

$$\nabla_\mu F^{\mu\nu} = 0 \ , \tag{8.4.8}$$

which would follow from Maxwell's equations in Minkowski space applying the principle of minimal coupling.

Exercise 8.4.2. Divergence and integration by parts. For an antisymmetric p-tensor β, the formula (7.2.12) for the divergence operator can be rewritten in the manifestly covariant form

$$\delta\beta^{\mu_1\ldots\mu_{p-1}} = -\nabla_\nu\beta^{\nu\mu_1\ldots\nu_{p-1}} = -\frac{1}{\sqrt{|g|}}\partial_\nu(\sqrt{|g|}\beta^{\nu\mu_1\ldots\nu_{p-1}}) \ . \tag{8.4.2.1}$$

In the derivation of (8.4.8) we have used the general rule for integrations by parts in a coordinate system. One can rewrite that rule in an alternative and convenient form using the covariant derivatives with respect to the Levi–Civita connection. Using (8.4.2.1),

$$\int dx\sqrt{|g|}\,\alpha_{\nu_1\ldots\nu_{p-1}}\nabla_\mu\beta^{\mu\nu_1\ldots\nu_{p-1}} = -\int dx\sqrt{|g|}\,\nabla_\mu\alpha_{\nu_1\ldots\nu_{p-1}}\beta^{\mu\nu_1\ldots\nu_{p-1}} \ . \tag{8.4.2.2}$$

In flat spacetime and in the Lorentz gauge $\partial_\mu A^\mu = 0$, each component of the four-potential satisfies a massless Klein-Gordon equation. In curved spacetime the result is not what one would guess by applying minimal coupling to the equation for the potential. Writing $F_{\mu\nu} = \nabla_\mu A_\nu - \nabla_\nu A_\mu$ and using the fact that the connection is metric, (8.4.8) can be rewritten in the form $\nabla_\mu(\nabla^\mu A_\rho - \nabla_\rho A^\mu) = 0$. Commuting the covariant derivatives in the second term by means of (6.6.8) one finds in the Lorentz gauge $\nabla_\mu A^\mu = 0$

$$\nabla_\mu\nabla^\mu A_\rho - R_\rho{}^\nu A_\nu = 0 \tag{8.4.9}$$

The operator appearing in this equation is (minus) the covariant Laplacian on one-forms defined in (8.4.9). We note that this equation could not be obtained applying the procedure of minimal coupling to the corresponding flat space equation. This is a manifestation of the intrinsic ambiguities of the procedure: since partial derivatives in flat space commute but covariant derivatives in curved space don't, there are ordering ambiguities when transforming a flat space equation into a curved space one. Such ambiguities can only be resolved by independent theoretical arguments or, better still, experimentally. Unfortunately, experiments are not sufficiently precise to allow the detection of such effects. In particular, most experimental tests of general relativity are performed in the gravitational field of the Earth or the Sun, where, as we shall see in the sequel, the Ricci tensor vanishes.

Exercise 8.4.3. Using the Levi-Civita connection on M, the explicit coordinate expressions for the Laplacian on p-forms $d\delta + \delta d$, for $p = 0, 1, 2$, is

$$\Delta\omega_\mu = -\nabla_\lambda\nabla^\lambda\omega_\mu + R_\mu{}^\nu\,\omega_\nu \ ,$$

$$\Delta\omega_{\mu\nu} = -\nabla_\lambda\nabla^\lambda\omega_{\mu\nu} + R_\mu{}^\rho\,\omega_{\rho\nu} - R_\nu{}^\rho\,\omega_{\rho\mu} - 2R_\mu{}^\rho{}_\nu{}^\sigma\omega_{\rho\sigma} \ .$$

8.5 Spinors and gravity

The treatment of spinors requires a different approach from the one that we used for scalar and electromagnetic fields. At a very basic level, scalar and electromagnetic fields are tensors and therefore representations of the linear group, while spinors are not. Thus, the machinery of tensor calculus developed in chapters 2, 3 and 4 cannot be applied. However, we can use the more general machinery of vector bundles developed in chapter 6.

A spinor is an element of a vectorspace K^m, where K is either the real or the complex field, carrying a representation of a double covering of the (pseudo)–orthogonal group $SO(p,q)$, called $Spin(p,q)$. (We will keep the signature of the metric generic as far as possible.) There are well–known isomorphisms in low dimensions (e.g. $Spin(3) = SU(2)$, $Spin(4) = SU(2) \times SU(2)$, $Spin(1,3) = SL(2,C)$ etc.). We call $\mu : Spin(p,q) \to SO(p,q)$ the double covering map; the nontrivial element of $Spin(p,q)$ that is mapped to the identity by μ is called -1. A spinor field on flat space is a section of the trivial bundle $M \times K^m$. We want to generalize this notion to an arbitrary manifold M.

As in section 9.1, it is convenient to think of a vectorbundle V which is isomorphic to TM but not necessarily identical to TM. The isomorphism is given by the soldering form θ. Let γ be a fiber metric in V with signature (p,q) and let $SO(V)$ be the bundle of oriented orthonormal frames in V. It is a principal $SO(p,q)$–bundle. If K^m carried a true representation of $SO(p,q)$ we could immediately construct the bundle associated to $SO(V)$, $(SO(V) \times K^m)/SO(p,q)$, as described in section 7.1. Since they are only "representations up to a sign", we need first to define a so-called "spin structure". This is a principal $Spin(p,q)$-bundle $Spin(V)$ and a bundle M-morphism $\eta : Spin(V) \to SO(V)$ which is a double covering. The elements of $Spin(V)$ will be called the spin frames, and to each orthonormal frame e there correspond exactly two spin frames $\tilde{e}$ and $-\tilde{e}$.

Two spin structures $(Spin(V), \eta)$ and $(Spin(V)', \eta')$ are said to be equivalent if there is a M-isomorphism $u : Spin(V)' \to Spin(V)$ such that

$$\eta \circ u = \eta' \ , \quad \eta(\tilde{e}\tilde{g}) = \eta(\tilde{e})\mu(\tilde{g}) \tag{8.5.1}$$

for any spin frame $\tilde{e}$ and $\tilde{g} \in Spin(p,q)$. In gauge theory language, this means that the two spin structures differ by a $Spin(p,q)$ gauge transformation.

For example if $SO(V)$ is trivial, $SO(V) = M \times SO(p,q)$, a spin structure is given by $Spin(V) = M \times Spin(p,q)$ and a double covering map

$$\eta(x, \tilde{r}) = (x, \mu(\tilde{r})) \ . \tag{8.5.2}$$

where $\tilde{r} \in Spin(p,q)$. The map η' defined by $\eta'(\tilde{e}) = \eta(-\tilde{e})$ is an equivalent spin structure according to the definition given above.

Since every bundle is locally trivial, we can try to construct a spin structure for V by patching together local double coverings. This procedure is not unambiguous. Let $\{(U_A, \Psi_A)\}$ be an atlas for $SO(V)$ with transition functions $g_{AB} : U_A \cap U_B \to SO(p,q)$. We seek an atlas for a $Spin(p,q)$–bundle with transition functions $\tilde{g}_{AB} : U_A \cap U_B \to Spin(n)$ such that $\mu \circ \tilde{g}_{AB} = g_{AB}$. Without loss of generality we can assume that the domains of the transition functions are connected and simply connected. Therefore, on each intersection there are precisely two such maps $\tilde{g}_{AB}$, and we have to make a choice between the two alternatives on each intersection. Depending on the topology of M, it may or may not be possible to make these choices respecting the conditions (6.1.2), and when it is possible the choice may not be unique The simplest example of a manifold that does not admit a spin structure is the complex projective space CP^2 ([1]). If there are no obstructions, we have then defined a principal $Spin(n)$-bundle $Spin(V)$. The double covering η can be defined by assuming that (8.5.2) holds in each chart, namely, for any $\tilde{r} \in Spin(p,q)$

$$\eta(\tilde{\psi}^{-1}(x, \tilde{r})) = \psi^{-1}(x, \mu(\tilde{r})) \ . \tag{8.5.3}$$

This definition is independent of the local trivialization. To see this, let $\tilde{\psi}'$ be another local trivialization; if $\tilde{\psi}(\tilde{e}) = (x, \tilde{r})$, there is a (locally defined) function $\tilde{g}$ such that $\tilde{\psi}'(\tilde{e}) = (x, \tilde{g}(x)^{-1}\tilde{r})$. To this there corresponds

([1]) S.W. Hawking and C.N. Pope, "Generalized Spin Structures in Quantum Gravity", Phys.Lett.**B73**:42-44 (1978).

another local trivialization ψ' of $SO(V)$ such that if $\psi(e) = (x, r)$, $\psi'(e) = (x, g(x)^{-1}r)$, with $g = \mu \circ \tilde{g}$. If we define η by

$$\eta(\tilde{\psi}'^{-1}(x, \tilde{r}')) = \psi'^{-1}(x, \mu(\tilde{r}')) \ ,$$

for any $\tilde{r}'$, then putting $\tilde{r}' = \tilde{g}(x)^{-1}\tilde{r}$ this is seen to be identical to (8.5.3). Thus, η is globally well defined.

The spinor space K^m carries a representation of $Spin(n)$ and therefore one can now define the associated vectorbundle $S = (Spin(V) \times K^m)/Spin(p, q)$. A spinor field is a section of S. Its local representative in a certain field of spin frames $\{\tilde{e}_a\}$ (covering a field of orthonormal frames $\{e_a\}$) is a map ψ with values in K^m.

When a spin structure exists, it may not be unique. Consider for example a torus $T^2 = S^1 \times S^1$ with coordinates (x, y), periodic with period 2π, and with a flat metric $g_{\mu\nu} = \delta_{\mu\nu}$. The groups $SO(2)$ and $Spin(2)$ are both isomorphic to $U(1)$, and the double covering is given by $e^{i\alpha} \mapsto e^{2i\alpha}$. Since T^2 is parallelizable we can choose a global field of orthonormal frames $e = \{e_1 = \partial_x, e_2 = \partial_y\}$. There exists a spin structure with total space $Spin(T^2) = T^2 \times Spin(2)$, and η the obvious double covering, defined as in (8.5.2): $\eta((x, y), e^{i\alpha}) = ((x, y), e^{2i\alpha})$. To the field of frames e there corresponds a field of spin frames $\tilde{e}$. There exist other, inequivalent, spin structures that differ in the map η. Consider for example the function $R(x, y) = e^{ix}$; since $R(2\pi, y) = R(0, y) = 1$, it is a continuous map of T^2 into the rotation group $SO(2)$, and can be regarded as a globally defined $SO(2)$ gauge transformation. The frames $e' = eR$ define another global trivialization of the bundle of orthonormal frames. However, there is no continuous $Spin(2)$ gauge transformation $\tilde{R}$ covering R: if we define $\tilde{R}(0, y) = 1$, then $\tilde{R}(2\pi, y) = -1$. Therefore, the map $\eta'((x, y), e^{i\alpha}) = ((x, y), e^{2i\alpha} \cdot R(x, y))$ defines an inequivalent spin structure.

Because the bundle $Spin(T^2)$ is globally trivialized (by $\tilde{e}$), a spinor field relative to the second spin structure can be described by a map $\psi' : T^2 \to S$, giving the components of the section relative to the frames $e' = eR$. This function is continuous, in particular

$$\psi'(2\pi, y) = \psi'(0, y) \ . \tag{8.5.4}$$

It is often preferable to work with the frames e. The corresponding transformation of the spinor field is $\psi = \tilde{R}\psi'$, There follows that the local representative of a spinor field belonging to the second spin structure, relative to the standard field of frames e, is antiperiodic in y:

$$\psi(2\pi, y) = -\psi(0, y) \ . \tag{8.5.5}$$

If in this discussion we replace the function $R(x, y) = e^{ix}$ by the function $R(x, y) = e^{2ix}$, then the resulting spin structure is equivalent to the first one, because there exists a smooth, globally defined $Spin(2)$-gauge transformation $\tilde{R}(x, y) = e^{ix}$ covering R (the map u of equation (6.3.1) is then given by $u(x, \tilde{e}) = (x, \tilde{e}\tilde{R})$). One can repeat the same arguments in the y direction, leading to the conclusion that there are altogether four spin structures on the torus, corresponding to spinor fields that, relative to the field of frames e, are either periodic or antiperiodic in x and y. Similar considerations apply whenever M is not simply connected ([2]).

Let us return to general issues. In General Relativity the status of spinor and tensor fields is profoundly different. In the case of the scalar or electromagnetic field, for example, one can define the kinematical variables ϕ, A, F etc. on any manifold. To write the Klein-Gordon or the Maxwell equation one needs a metric, but even if a metric did not exist a scalar or an electromagnetic field could still exist. This is not so in the case of spinors. The construction of the vectorbundle V_S and of the space of spinor fields, was based on a metric g. One cannot even define what a spinor field is, until a globally defined, smooth, nondegenerate metric is given. Furthermore, a different metric g' will lead to a different vectorbundle of spinors and therefore a different space of spinor fields. There is no canonical way of identifying these different spaces, so each metric gives rise to a different space of spinors. This gives rise to further questions. Under a diffeomorphism f, the metric g is transformed into $g' = f^*g$. Therefore, a spinor defined with the metric g must be transformed into a spinor defined with the metric g'. The transformation of a spinor is defined by means of the trasformation of the orthonormal frames. A diffeomorphism f maps a frame e to a frame $Tf(e)$.

([2]) S.J. Avis, C.J. Isham, "Generalized Spin Structures On Four-Dimensional Space-Times", Commun. Math. Phys. **72** 103 (1980).

This lifts to an action on spin frames. If the spinor field is represented in a spin frame $\tilde{e}$ by a spinor ψ, the transformed spinor is represented by the same spinor ψ in the transformed spin frame. This is commonly expressed by the statement that "a spinor field transforms as a scalar under coordinate transformations". This statement is not incorrect, but it is somehow tautological. Exactly the same can be said of any tensor field: the components of the transformed tensor with respect to the transformed frame are identical to the components of the original tensor with respect to the original frame. The transformation properties of tensors (2.3.10) or (3.3.14) give instead the components of the transformed tensor in the old frame (active point of view) or the components of the old tensor in the transformed frame (passive point of view) ([3]).

We now choose to work in the vierbein gauge (7.4.4), so that the metric is given by

$$g_{\mu\nu} = \theta^a_\mu \theta^b_\nu \eta_{ab} \ , \tag{8.5.6}$$

In this gauge the soldering form is usually called a vierbein or tetrad. In order to define a dynamics for spinors, we need a connection. Since $Spin(n)$ and $SO(n)$ are locally isomorphic, a connection in $SO(V)$ automatically defines a unique connection in $Spin(V)$. For example, if the connection is the Levi-Civita connection, its components are given by equation (9.2.6). The covariant derivative of a spinor field is given by (7.2.8):

$$\nabla_\mu \psi = \partial_\mu \psi + A_\mu{}^{ab} \Sigma_{ab} \psi \ . \tag{8.5.7}$$

Here $\Sigma_{ab} = \frac{1}{8}[\gamma_a, \gamma_b]$ are the Lorentz generators in the given spinor representation and spinor indices are not written explicitly. With these preparations, the Dirac equation for a massless spinor field in a background metric g can be written

$$\gamma^a \theta_a{}^\mu \nabla_\mu \psi = 0 \ . \tag{8.5.8}$$

Exercise 8.5.4. Using the algebra of the Dirac matrices compute the square of the Dirac operator $D = \gamma^a e_a{}^\mu \nabla_\mu$. Show that the Dirac spinors obey the equation (in any dimension):

$$\left(-\nabla_\mu \nabla^\mu + \frac{R}{4} \right) \psi = 0 \ . \tag{8.5.1.1}$$

8.6 The limit of geometric optics

The fields described in the previous sections obey wave equations of the form

$$\nabla_\mu \nabla^\mu \psi + cR\psi = 0 \ , \tag{8.6.1}$$

where c is some constant and R is a linear operator acting on ψ constructed with the curvature tensor. In general the solutions of these equations will be propagating waves. If the wavelength is much smaller that the typical curvature radii, the propagation is similar to that of plane waves in flat space.

Let us write the solution of (8.6.1) in the form

$$\psi = Ce^{iS} \ , \tag{8.6.2}$$

where C is a slowly varying amplitude and S is the phase. Both are real functions on M. The condition that S be constant defines three dimensional surfaces called wave fronts. We assume that $\nabla_\mu C \ll \nabla_\mu S$ and $RC \ll \nabla_\mu S$. This is called the geometric optics limit. Inserting (8.6.2) in (8.6.1) and keeping only the leading terms we obtain

$$0 = iC\nabla_\mu \nabla^\mu S e^{iS} - C\nabla_\mu S \nabla^\mu S e^{iS}$$

and therefore $\nabla_\mu \nabla^\mu S = 0$, $\nabla_\mu S \nabla^\mu S = 0$. Let $k_\mu = \nabla_\mu S$ be the vector normal to the wave fronts. It is a null vector: $k^\mu k_\mu = 0$. We have

$$0 = \nabla_\lambda (\nabla_\mu S \nabla^\mu S) = 2k^\mu \nabla_\mu k_\lambda \ , \tag{8.6.3}$$

therefore the integral lines of k are null geodesics. These geodesics are called "light rays" and can also be thought of as the world lines of massless particles. The angular frequency of a wave measured by an observer with four–velocity u is

$$\omega = u^\mu \partial_\mu S = u^\mu k_\mu \ . \tag{8.6.4}$$

([3]) A precise definition of the transformation of spinors under diffeomorphisms was given in L. Dabrowski, R. Percacci, "Spinors And Diffeomorphisms", Commun. in Math. Phys. **106**, 691 (1986)

Solution to Exercise 8.5.4.

When squaring the Dirac operator one has to pay attention to what type of covariant derivative one is dealing with. It is therefore convenient to spell out also the spinor indices, or at least to keep mentally track of them. When we write

$$\gamma^a \theta_a{}^\mu \nabla_\mu (\gamma^b \theta_b{}^\nu \nabla_\nu \psi)$$

the covariant derivative on the left acts on the object in parenthesis, which carries only a spinor index. Therefore this covariant derivative is again of the form (8.5.7). However, when we apply the Leibnitz rule we break up the object inside the parenthesis into pieces that also carry other types of indices. When the spinor covariant derivative (8.5.7) acts on such objects, it is not a covariant derivative. Only the sum of all terms would be covariant. It is therefore better to think that the covariant derivative on the left is a "fully covariant" derivative, i.e. it is covariant also with respect to the indices $a, b \dots$ and $\mu\nu \dots$. Since in the parenthesis all such indices are contracted, the outer ∇_μ is just (8.5.7). However when we apply the Leibnitz rule

$$\nabla_\mu (\gamma^b \theta_b{}^\nu \nabla_\nu \psi) = (\nabla_\mu \gamma^b) \theta_b{}^\nu \nabla_\nu \psi + \gamma^b (\nabla_\mu \theta_b{}^\nu) \nabla_\nu \psi + \gamma^b \theta_b{}^\nu \nabla_\mu \nabla_\nu \psi$$

each term is now covariant. Let us consider the three terms on the r.h.s. The gamma matrix has the Lorentz index b as well as two spinor indices: $\gamma^{bA}{}_B$. Its fully covariant derivative is

$$\begin{aligned}
\nabla_\mu \gamma^b &= \partial_\mu \gamma^b + A_\mu{}^b{}_c \gamma^c + A_\mu{}^{cd} [\Sigma_{cd}, \gamma^a] \\
&= A_\mu{}^b{}_c \gamma^c + \frac{1}{4} A_\mu{}^{cd} (\gamma_c \gamma_d \gamma^a - \gamma^a \gamma_c \gamma_d) \\
&= 0
\end{aligned}$$

where in the first step we used that γ^b is constant and in the last we used $\{\gamma^a, \gamma^b\} = 2\eta^{ab}$. In the second term we have

$$\nabla_\mu \theta_b{}^\nu = \partial_\mu \theta_b{}^\nu + \Gamma_\mu{}^\nu{}_\rho \theta_b{}^\rho - \Gamma_\mu{}^c{}_b \theta_c{}^\nu \ .$$

This is zero because of (7.3.4). The last term can be written $\gamma^b \theta_b{}^\nu \nabla_\mu \nabla_\nu \psi$.

Then we have

$$\begin{aligned}
\gamma^a \theta_a{}^\mu \nabla_\mu (\gamma^b \theta_b{}^\nu \nabla_\nu \psi) &= \gamma^a \gamma^b \theta_a{}^\mu \theta_b{}^\nu \nabla_\mu \nabla_\nu \psi \\
&= \left(\frac{1}{2} \{\gamma^a, \gamma^b\} + \frac{1}{2} [\gamma^a, \gamma^b] \right) \theta_a{}^\mu \theta_b{}^\nu \nabla_\mu \nabla_\nu \psi \\
&= \left(\eta^{ab} + 4\Sigma^{ab} \right) \theta_a{}^\mu \theta_b{}^\nu \nabla_\mu \nabla_\nu \psi \\
&= g^{\mu\nu} \nabla_\mu \nabla_\nu \psi + 2\Sigma^{ab} \theta_a{}^\mu \theta_b{}^\nu [\nabla_\mu, \nabla_\nu] \psi \ .
\end{aligned}$$

We have $[\nabla_\mu, \nabla_\nu]\psi = R_{\mu\nu}{}^{cd} \Sigma_{cd} \psi$ with $R_{\mu\nu}{}^{ab} = R_{\mu\nu}{}^{\rho\sigma} \theta^a{}_\rho \theta^b{}_\sigma$. Thus the last term is equal to

$$2\Sigma^{ab} R_{abcd} \Sigma^{cd} = \frac{1}{8} R_{abcd} \gamma^a \gamma^b \gamma^c \gamma^d \ .$$

Now we can use the gamma matrix identity

$$\gamma^a \gamma^b \gamma^c = \eta^{ab} \gamma^c + \eta^{bc} \gamma^a - \eta^{ac} \gamma^b + \gamma^{abc}$$

where γ^{abc} is totally antisymmetric. When contracted with R_{abcd} the first term gives zero because of symmetry, the last gives zero because of the Bianchi identity (7.6.3). The rest is $-2R_{ab}\gamma^a\gamma^b = -2R$. Thus finally

$$\gamma^a \theta_a{}^\mu \nabla_\mu (\gamma^b \theta_b{}^\nu \nabla_\nu \psi) = \left(g^{\mu\nu} \nabla_\mu \nabla_\nu - \frac{R}{4} \right) \psi \ .$$