Positive Energy Theorem & Spin Structure

Bryan, X. J. Xie, H. Q. Wu

Contents

1 Introduction

Relativity and quantum mechanics are the two cornerstones of modern physics. Attempted unions of these two theories have produced some highly non-trivial results in theoretical physics. Most surprisingly, by considering *classical limits* of quantum effects in general relativity, we can reveal some new structures of classical spacetime. Witten's proof of the positive energy theorem [1] is a major example of such treatment.

In general relativity, matter curves spacetime, and its effects are captured by Einstein's equations. The positive energy conjecture claims that "regular" matter[1] should give rise to spacetime with a positive energy, and the energy of spacetime is zero iff. it is flat [2].

[1] In theory, there *can* be "exotic" matter, however unlikely in reality. For example, matter with negative mass can be used to create traversable wormhole.

This is a statement about purely classical general relativity. It feels natural in physics, but is extremely hard to prove in a rigorous way. The first complete proof is given by Schon and Yau [3] using techniques from geometric analysis. Witten's proof, on the other hand, relies heavily on the existence of *spinor fields* on a given spacetime. Although it is inspired by ideas from quantum mechanics and even supersymmetry, the proof by itself is still fully contained in general relativity.

Spinor field is a natural consequence of quantum field theory, and can be treated semi-classically by introducing additional *spin structure* on the spacetime $\mathcal{M}$. This will be the focus of our following discussions.

As an introductory review, we try to illustrate the geometric aspects of spin structure and its role in Witten's proof of the positive energy theorem. To avoid technicalities of spinor analysis, we will only sketch some main ideas of the proof. For a more detailed treatment, see [1, 2] for physicists and [4, 5] for mathematicians.

1.1 Review of General Relativity and Conventions

General relativity is formulated on a *pseudo*-Riemannian (Lorentzian) manifold $(\mathcal{M}, g)$. $\mathcal{M}$ is locally identical to $\mathbb{R}^{3,1}$, here $\mathbb{R}^{3,1} \cong \mathbb{R}\times\mathbb{R}^3$ with the first component being the time coordinate. From now on we will use the symbol "$\cong$" to denote diffeomorphism, or *diffeo* for short. However, when there is extra structure involved (e.g. group product), "$\cong$" will stand for *isomorphism*, and we will use "$\cong$" for mere diffeo.

Metric tensor can then be pointwise diagonalized:

$$g_{\mu\nu} = \eta_{ab}\, e^a{}_\mu e^b{}_\nu + \mathcal{O}(x^2), \quad \eta_{ab} \sim \mathrm{diag}(-1,1,1,1) \sim \text{flat} \tag{1.1}$$

Here $\mu, \nu, \cdots = 0, 1, 2, 3$ denotes tensor components w.r.t. local coordinates, while $a, b, \cdots = 0, 1, 2, 3$ denotes components w.r.t. a *orthonormal frame field*[2] or *vierbein*:

$$\left\{\theta^a(x) = e^a{}_\mu(x)\, \mathrm{d}x^\mu\right\}_{a=0,1,2,3} \longleftrightarrow \left\{e_a(x) = e_a{}^\mu(x)\, \frac{\partial}{\partial x^\mu}\right\}_{a=0,1,2,3} \tag{1.2}$$

It is often helpful to think of $\mathcal{M}$ as the *foliation / time-evolution* of a spacelike (Riemannian) codimension-1 hypersurface Σ along a time direction[3].

Coordinates on Σ is indexed with $i, j, \cdots = 1, 2, 3$. In fact, we assume that:

$$\mathcal{M}_{(t_1,t_2)} \cong (t_1, t_2) \times \Sigma \cong \mathbb{R} \times \Sigma \tag{1.3}$$

Where $t \in (t_1, t_2)$ is some *global time* coordinate restricted in an interval.

Following the usual approach of Riemannian geometry, we can similarly define Riemann curvature $R^\lambda{}_{\rho\mu\nu}$, Ricci tensor $R_{\mu\nu}$ and scalar curvature R. Einstein's equations of relativity

[2] This is basically Cartan's *moving frame* formalism.

[3] In fact, global existence of such time slice Σ (so-called *Cauchy slice*) as an initial value surface is *not* automatically guaranteed, but relies on the causal structure of $\mathcal{M}$, see [6]; but locally we can always assume such structure.

is then neatly organized into the following covariant tensorial expression[4]:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi T_{\mu\nu} \tag{1.4}$$

The left-hand side (LHS) of the equation is purely geometric, while the right-hand side (RHS) is proportional to the symmetric *energy–momentum stress tensor* $T_{\mu\nu}$, which describes matter distribution in spacetime. In general, we expect natural matter on the RHS will lead to well-behaved spacetime on the LHS; positive energy theorem is one exact case of such observation.

2 Positive Energy Theorem

Although Witten's proof relies on insights from quantum mechanics, the statement of positive energy theorem is purely *classical* (in the sense that it does *not* depend on quantum mechanics)[5]. Here we summarize the main components of the theorem.

2.1 Constraints on Manifold and Total Energy

As is illustrated in (1.3), the main object to study can be reduced to a usual Riemannian manifold Σ. To make things easier, we shall further assume that Σ is *asymptotically flat*, i.e. Σ has finitely many non-compact asymptotic *ends*[6] $\Sigma_\ell \subset \Sigma$. Asymptotic flatness is then captured by the asymptotic behavior the metric: $g_{ij} \sim \delta_{ij}$.

Definition (Asymptotic Flatness) A Riemannian 3–fold $\Sigma \subset \mathcal{M}$ as a hyperspace in spacetime is *asymptotically flat* when it has the following decomposition:

$$\Sigma \cong \Sigma_0 \cup \coprod_{\ell \neq 0} \Sigma_\ell \tag{2.1}$$

[4] Here we have set Newton's constant $G_N = 1$ and cosmological constant $\Lambda = 0$. Proof of the Positive Energy Theorem with $\Lambda \neq 0$ is still incomplete to this day, and still an important field of research; see e.g. [7].

[5] There are some variations between statements in the literature; here we will follow Straumann's simplified version [2] and Witten's original work [1], with mathematical supplements from Parker *et al* [4].

[6] It might be helpful to visualize this as the spatial "end" of the world, 世界尽头。This poetic translation is due to X. J. Xie.

Where $\{\ell\}$: finite indices, and:

- Σ_0: compact;
- $\{\Sigma_\ell\}_\ell$: asymptotic ends, each diffeo to the complement of a contractible compact set in $\mathbb{R}^3$;
- Under such diffeo $\Sigma_\ell \to \mathbb{R}^3$, the metric g of every end $\Sigma_\ell \subset \mathcal{M}$ is asymptotically flat, i.e.

$$\begin{gathered} g_{ij} = \delta_{ij} + h_{ij}, \\ h_{ij} \sim \mathcal{O}\left(\tfrac{1}{r}\right), \quad \partial_k h_{ij} \sim \mathcal{O}\left(\tfrac{1}{r^2}\right), \quad \partial_l \partial_k h_{ij} \sim \mathcal{O}\left(\tfrac{1}{r^3}\right) \end{gathered} \tag{2.2}$$

- The second fundamental form K of $\Sigma_\ell \subset \mathcal{M}$ is also constrained:

$$K_{ij} \sim \mathcal{O}\left(\tfrac{1}{r^2}\right), \quad \partial_k K_{ij} \sim \mathcal{O}\left(\tfrac{1}{r^3}\right) \tag{2.3}$$

In general, it is quite difficult to define a total energy for some arbitrary spacetime $\mathcal{M}$. However, when asymptotic flatness is imposed, it can be treated as a *symmetry* of the ends Σ_ℓ. Energy of the spacetime[7] can then be defined as the *conserved charge* corresponding to the time translation Killing vector field $\xi = \partial_t$, or more explicitly, by integrating the *Noether current* of such symmetry on Σ_ℓ, as is done in *Noether's theorem*. See [8] for a comprehensive review; for a covariant treatment, see [9, 10].

In fact, the Noether current can be further shown to be exact: $\mathrm{d}Q_\xi$, then by Stokes' theorem, energy as a conserved charge can be further reduced to some integral along the "asymptotic boundaries" (plus some extra boundary terms, shown below in "$\cdots$") [10, 11]:

$$\begin{gathered} E_\ell = P^0(\Sigma_\ell) = \int_{\Sigma_\ell} \mathrm{d}Q_\xi + \int_{S^2_\ell} (\cdots) = \int_{S^2_\ell} \left(Q_\xi + \cdots\right), \\ S^2_\ell = \text{“}\,\partial\Sigma_\ell\text{”} \subset \Sigma_\ell \subset \mathbb{R}^3, \quad \xi = \partial_t \end{gathered} \tag{2.4}$$

Here S^2_ℓ is a sphere with radius $R \to \infty$[8]. This is the so-called *ADM energy* [12]. In fact, we can take $\xi = \partial_\mu, \mu = 0, 1, 2, 3$ and obtain the ADM 4–momentum [2]:

$$P^\mu_\ell \equiv P^\mu(\Sigma_\ell) \simeq -\frac{1}{16\pi} \int_{S^2_\ell} \sqrt{|g|}\, \omega_{\rho\sigma} \wedge \star \left(\mathrm{d}x^\rho \wedge \mathrm{d}x^\sigma \wedge \mathrm{d}x^\mu\right), \omega_{\rho\sigma} = \Gamma_{\rho\mu\sigma}\, \mathrm{d}x^\mu = g_{\rho\lambda}\Gamma^\lambda{}_{\mu\sigma}\, \mathrm{d}x^\mu \tag{2.5}$$

Where $\omega_{\rho\sigma}$ is the Levi-Civita connection form. This can be further expanded in induced metric (1^{st} fundamental form) g_{ij} and 2^{nd} fundamental form K_{ij} when $R \to \infty$, which gives the explicit expressions below.

---

[7] This includes contributions from gravitational field and matter field.

[8] Discussions here are clearly not mathematically exact (hence the quotation marks), but can be made rigorous by taking a cutoff geometry $\mathcal{M}'$ with (true) boundary $\partial\mathcal{M}' = \coprod_{\ell \neq 0} S_\ell$, and then prove that the end result is in fact cutoff independent.

Definition (ADM Energy–Momentum) Total ADM energy–momentum P_ℓ^μ of an asymptotically flat end $\Sigma_\ell \subset \Sigma \subset \mathcal{M}$ is defined in the following explicit form[a]:

$$\begin{aligned} E_\ell \equiv P_\ell^0 &= \lim_{R\to\infty} \frac{1}{16\pi} \int_{S_\ell^2} (\partial_j g_{ij} - \partial_i g_{jj})\, \mathrm{d}\Omega^i \,, \\ P_\ell^k &= \lim_{R\to\infty} \frac{1}{16\pi} \int_{S_\ell^2} 2\,(K_{ik} - \delta_{ik} K_{jj})\, \mathrm{d}\Omega^i \,, \end{aligned} \tag{2.6}$$

Where R is the radius of 2–sphere $S_\ell^2 \subset \Sigma_\ell \cong \mathbb{R}^3$,

- g_{ij} is the induced metric (1^{st} fundamental form),
- K_{ij} is the 2^{nd} fundamental form on $\Sigma_\ell \cong \mathbb{R}^3$,

and repeated indices are summed over.

[a] Note that the E_ℓ given in [4] has a typo. Check [1, 6] for the correct expression.

2.2 Energy Conditions

On the other hand, there are physical constraints on the stress tensor $T_{\mu\nu}$, which describes *local* energy–momentum density. Physical constraints on $T_{\mu\nu}$ are called *energy conditions* (EC), see *Wald* [6] or [13] for some detailed discussions. Here we assume the *dominant* EC (DEC), which requires that mass / energy can never be observed to be flowing faster than light.

Specifically, a non-rotating observer moving along some path $x(\tau) \in \mathcal{M}$ is characterized by a local velocity 4–vector $v^\mu = x'(\tau) = \mathrm{T}_x\mathcal{M}$. The observed energy–momentum flow p^μ and energy density ρ w.r.t. v^μ is obtained by the contraction[9]:

$$\begin{aligned} p^\mu &= -v_\nu T^{\mu\nu}, \\ \rho &= -v_\nu p^\nu, \end{aligned} \qquad v^\nu\text{: timelike, i.e. } v^\nu v_\nu = g_{\mu\nu} v^\mu v^\nu < 0 \tag{2.7}$$

p^μ should be a *timelike* or *lightlike* (non-spacelike) 4–vector, i.e. $p^\mu p_\mu \le 0$, $\forall\, v^\mu$: timelike. Also, matter density should always be non-negative[10], i.e. $\rho \ge 0$. These two constraints can be equivalently captured by simple restrictions of $T_{\mu\nu}$ components [2], shown below.

[9] Minus signs in (2.7) is due to the $\eta_{ab} \sim (-1,+1,+1,+1)$ metric convention, in this case $v^a \sim (1,0,0,0) \mapsto v_a \sim (-1,0,0,0)$, therefore we need an extra sign flip to get the correct sign of energy.

[10] The condition $\rho \ge 0$ by itself is also a common energy condition, called the weak EC (WEC).

Definition (Dominant Energy Condition) The symmetric stress tensor $T_{\mu\nu}$ satisfies the dominant energy condition (DEC), iff.

$$T^{00} \geq |T^{\mu\nu}|, \quad T^{00} \geq \sqrt{-T_{0i}T^{0i}} \tag{2.8}$$

Which is equivalent to $\rho \geq 0$, $p^\mu p_\mu \leq 0$, where ρ, p^μ is defined in (2.7).

2.3 Statement of the Theorem

With the above preparation, we can now state the positive energy theorem in a mathematically precise manner.

Theorem (Positive Energy Theorem) For a spacelike hypersurface $\Sigma \subset \mathcal{M}$ that:

- is asymptotically flat, $\Sigma \cong \Sigma_0 \cup \coprod_{\ell \neq 0} \Sigma_\ell$, and
- the dominant energy condition holds;

Then the total energy-momentum P^μ_ℓ is a future directed timelike or lightlike (non-spacelike) vector on each end Σ_ℓ, i.e.

$$E_\ell - |P_\ell| \equiv E_\ell - \sqrt{P_{\ell,i}P^i_\ell} \geq 0 \tag{2.9}$$

Furthermore, if $E_\ell = 0$ for some ℓ then Σ has only one end and $\mathcal{M}$ is flat along Σ, i.e. $\mathcal{M}$ is flat in a neighborhood of Σ, as is described in (1.3).

The proof of the positive energy theorem and its various generalizations has been the pursuit of many mathematicians and physicists, as is described in [1]. Witten's proof provides an attracting alternative to the first proof [3], and is much more approachable for physicists.

As is mentioned before, Witten's proof is based on the assumption that there is matter described by some *spinor field* ψ on the manifold $\mathcal{M}$. This is not so surprising to physicists, as spinors are the mathematical descriptions of *fermions* (incl. electrons, protons, and basically all "matter" in the usual sense).

The use of spinor is also suggested by considerations involving *supersymmetry* (SUSY) [11], as is discussed by Witten [1]. In supersymmetric gravity (*supergravity*, SUGRA), the energy is *formally* written as the sum of squares of Hermitian supercharges $H = \frac{1}{\hbar}\sum_\alpha Q^2_\alpha$, which is *formally* positive. Witten's proof is then found by designing a *classical limit* of SUGRA.

[11] Basically, SUSY is a new spacetime symmetry that relates spinor with their vector partners, and tries to describe them in a single framework. This simple idea has generated numerous developments in both math and physics.

Besides physical considerations, spinors are also fascinating objects in geometry. In the following part of our review we will focus the physical and geometric properties of spinor and spin structure, and give a very brief sketch of Witten's proof at the end of this review.

3 Spinor from Quantization[12]

This section serves as a non-technical introduction to spinor for a general audience. We follow *Weinberg* [14] for this informal discussion about quantization.

Particles have *spin*. Classically, spin is described by a local 3-*vector* $s^i \in \mathrm{T}_x\Sigma$. This vector can be naïvely promoted to a 4-*vector* in its rest frame[13]:

$$s^i \quad \longrightarrow \quad s^\mu \sim (0, s^i), \quad \text{rest frame} \tag{3.1}$$

In quantum mechanics, particle ceases to be a localized entity; rather, it is described by a *wave function* spread across the spacetime. Furthermore, in field theory, a particle is considered as an *excitation* of some *field*, just like a ripple in the pond. Therefore, instead of localized vectors, we should consider *vector fields*, i.e. sections of the tangent bundle:

$$s^\mu \in \mathrm{T}_x\mathcal{M} \quad \longrightarrow \quad \phi^\mu \in \Gamma(\mathcal{M}, \mathrm{T}\mathcal{M}) : \ \mathcal{M} \to \mathrm{T}\mathcal{M} \tag{3.2}$$

Another feature of the wave function description is that there is always a *Hilbert space* $\mathcal{H}$ of wave functions corresponding to the actual spacetime manifold $\mathcal{M}$. Quantum mechanical evolution of a system is in fact formulated in its Hilbert space, rather than its spacetime manifold. Symmetries of spacetime $\mathcal{M}$ lead to symmetries on the Hilbert space; in fact, the Hilbert space $\mathcal{H}$ can be seen as a *representation* of the spacetime symmetries.

For a flat spacetime, 4–momentum p_μ and (relativistic) angular momentum tensor $j_{\mu\nu}$ are symmetry generators, i.e. Killing vector fields of the spacetime; with the standard orthonormal bases of $\mathbb{R}^{3,1}$, we have:

$$p_\mu = \partial_\mu, \quad j_{\mu\nu} = x_\mu \partial_\nu - x_\nu \partial_\mu \tag{3.3}$$

Killing vector fields as derivations form a closed Lie algebra; for $\mathbb{R}^{3,1}$ this is the so called *Poincaré* algebra, here denoted as the Lie algebra $\mathfrak{iso}(3,1)$ of Lie group $\mathrm{ISO}^+(3,1)$, where " I " stands for inhomogeneous, i.e. with translations, and " $+$ " denotes its identity component. The finite transformation generated by $p_\mu, j_{\mu\nu}$ is exactly the Lorentz transformation. This shows that a (local) change of coordinates is interpreted in physics as switching between reference frames.

---

[12] This section is far too physical, therefore inappropriate in a report for differential manifold, and deserves to be dumped into an appendix. However, some necessary concepts are introduced in this section, so we will leave it here for now.

[13] In this review particles are always assumed to move slower than light, for simplicity. Therefore, it is possible to change the coordinates to follow a particle in its comoving rest frame. Most of our arguments here can be generalized to particles moving at the speed of light, but with considerable amount of subtleties.

Quantum mechanically, $p_\mu, j_{\mu\nu}$ is *represented* (literally and mathematically) as linear operators $P_\mu, J_{\mu\nu}$ on Hilbert space $\mathcal{H}$,

$$p_\mu \longrightarrow P_\mu, \quad j_{\mu\nu} \longrightarrow J_{\mu\nu} \tag{3.4}$$

Intuitively, particles are characterized with its energy[14] $H = P^0$, momentum $\mathbf{P}^i = P^i$ and their relations: $H = H(\mathbf{P})$. Conveniently, P_μ's commute with each other; therefore in field theory, particles are *defined* by simultaneous eigenstates of P_μ's.

For a given Lie algebra and its representation, we can construct *Casimir invariants* which commute with all generators, therefore are invariant (or *conserved*) under symmetry transformation. Naturally, we expect them to describe *intrinsic properties* of a particle, which is not affected by a change of coordinates.

For Poincaré algebra in $(3+1)$ dimensions, there are exactly two linearly independent Casimirs[15]:

$$M^2 = -P_\mu P^\mu, \quad W^2 = W_\mu W^\mu, \quad W_\lambda = -\frac{1}{2}\,\epsilon_{\lambda\mu\nu\rho} J^{\mu\nu} P^\rho \tag{3.5}$$

Here $\epsilon_{\lambda\mu\nu\rho}$ is the totally anti-symmetric Levi-Civita symbol, and the minus signs are merely conventional. W_λ is the *Pauli–Lubanski pseudovector*.

Physical states are required[16] to have $M^2 > 0$, and states of the same eigenvalue $M = m \geq 0$ form a sub-representation (*submodule*) of $\mathrm{ISO}^+(3,1)$ acting on $\mathcal{H}$. Different species of particles are therefore distinguished by different values of $M = m$. Physically, m is naturally interpreted as the *mass* of a particle. Similarly, W^2 should characterize another intrinsic property of a particle; let's further define sub-species of particle labeled by different eigenvalues of W^2. Assume $m > 0$ for simplicity, then in the rest frame of some particle with constant eigenvalue W^2, we have $P^\mu = p^\mu \sim (p^0, 0) = (m, 0)$,

$$W_\lambda = -\frac{1}{2}\,\epsilon_{\lambda\mu\nu\rho} J^{\mu\nu} P^\rho = -\frac{1}{2}\,\epsilon_{\lambda\mu\nu 0} J^{\mu\nu} m \sim (0, W_k), \quad \textit{rest frame}, \tag{3.6}$$

$$W_k = \frac{1}{2}\, m\, \epsilon_{ijk} J^{ij} = m J_k = m s_k \tag{3.7}$$

Where $J_k = \frac{1}{2}\,\epsilon_{ijk} J^{ij}$ is the *angular momentum* operator, which is the Hilbert space representation of rotation generator: $j_{ij} = x_i \partial_j - x_j \partial_i = \frac{\partial}{\partial \theta^k}$, here θ^k parameterizes rotation angle around the k–th axis. Physically, *spin* is exactly the intrinsic angular momentum of a particle: $J_k = s_k$, therefore the spin 4–vector shall be rigorously defined as the eigenvalue of Pauli–Lubanski pseudovector.

To sum up, assuming that there is no extra symmetry beyond Poincaré $\mathfrak{iso}(3,1)$, particles can be completely reduced and classified by representations of $\mathfrak{iso}(3,1)$ on $\mathcal{H}$. We have:

$$P^\mu \left|p^\mu, s^\mu\right\rangle = p^\mu \left|p^\mu, s^\mu\right\rangle, \quad J^\mu \left|p^\mu, s^\mu\right\rangle \approx s^\mu \left|p^\mu, s^\mu\right\rangle, \quad J^\mu = \frac{W^\mu}{m} \tag{3.8}$$

Here $\left|p^\mu, s^\mu\right\rangle \in \mathcal{H}$ denotes a particle state in the Hilbert space.

---

[14] Conventionally, $H = P^0$ serves as the energy operator, i.e. *Hamiltonian*, while E denotes its eigenvalue.

[15] For Poincaré algebra in higher dimensions, i.e. $\mathfrak{iso}(d,1)$ with $d \geq 4$, there are more Casimirs, see e.g. [15]. Unfortunately, we (i.e. authors of this review) haven't found any reference about the physical interpretation of these higher order Casimirs.

[16] By one of the fundamental postulates of relativity, there is no faster-than-light travel, which implies $m^2 = -p_\mu p^\mu \geq 0$. This is represented in the Hilbert space as $M^2 = -P_\mu P^\mu \geq 0$.

However, the above notation is *not* quite exact (as is indicated by the "$\approx$" sign). Note that W_μ's do *not* commute, hence it is impossible to find simultaneous eigenstates for all four W_μ's. The best we can manage is to pick a special W_μ, say $\mu = z$ in the rest frame, and then index the eigenstates with eigenvalues of $J_z = \frac{W_z}{m}$. (3.8) is then modified to be:

$$P^\mu \left|p^\mu, s_z\right\rangle = p^\mu \left|p^\mu, s_z\right\rangle, \quad J_z \left|p^\mu, s_z\right\rangle = s_z \left|p^\mu, s_z\right\rangle \tag{3.9}$$

The action of W^μ on a resting particle is the representation of stabilizer subgroup[17] $\mathrm{SO}(3) \subset \mathrm{ISO}^+(3,1)$, which keeps $p^\mu \sim (m, 0)$ fixed. Since $\mathrm{SO}(3)$ is a compact Lie group, according to Peter–Weyl theorem, its irreducible representations (*irrep*s) are finite dimensional. It is natural to define *elementary* particles as irreps, since they cannot be decomposed further (physically or mathematically); hence s^μ for an elementary particle can only take on discrete values.

Therefore, the rigorous label of a particle state should be:

$$\begin{gathered} \mathcal{H} \supset \mathcal{H}_{m,j} \ni \left|m, p^\mu; j, s_z\right\rangle \equiv \left|p^\mu, s_z\right\rangle, \\ p_\mu p^\mu = -m^2, \quad s_z = -j, -j+1, \cdots, +j, \quad j \in \mathbb{Z}_+ \end{gathered} \tag{3.10}$$

Here m, j labels the representation $\mathcal{H}_{m,j}$, which stays invariant under Poincaré transformations, while p, s_z labels the state in such irrep. The allowed values of s_z come from the representation of $\mathrm{SO}(3)$.

Now we are finally prepared to embrace another quirk from quantum mechanics — *projective representation*, which leads directly to the existence of spinors.

3.1 Projective Representation and Spinor

As mentioned before, in quantum mechanics the state and dynamics of a system is described with a $\mathbb{C}$–valued *wave function* $\Psi(\xi)$, or equivalently an abstract *state* $|\Psi\rangle$ in Hilbert space $\mathcal{H}$. The variable ξ labels simultaneous eigenstate $|\xi\rangle$ of some Hermitian *observables*, e.g. momentum P^μ and spin-z component W^z. The set of all such eigenstates $\{|\xi\rangle\}$ forms an orthonormal[18] basis of $\mathcal{H}$.

Wave function $\Psi(\xi)$ can be regarded as the projection of state Ψ onto the orthonormal basis $|\xi\rangle$, i.e. $\Psi(\xi) = \langle\xi|\Psi\rangle$, where $\langle\cdot|\cdot\rangle$ is the Hermitian inner product. Here we adopt the physics convention that $\langle\cdot|\cdot\rangle$ is $\mathbb{C}$ linear in its *second* entry, while conjugate-linear in its first entry.

[17] In physics literature this is the so-called Wigner's *little group*.

[18] Although mathematically imprecise, it is common in physics to treat *distributions* such as Fourier basis $\{e^{ikx}\}_{k\in\mathbb{R}}$ as "orthonormal", with the modified orthonormal relations $\frac{1}{2\pi}\int \mathrm{d}x\, e^{-ik'x} e^{ikx} = \delta(k-k')$, where δ is the Dirac delta distribution. This can be made mathematically rigorous in functional analysis, hence there is no need to worry.

Projective representation arises from Born's *statistical interpretation* of the wave function, which relates the probability of observed outcome with its wave function, namely,

$$|\langle\xi|\Psi\rangle|^2 \sim \Big(\text{ Probability of observing state } |\Psi\rangle \text{ in eigenstate } |\xi\rangle\Big) \tag{3.11}$$

This means that states differ by an overall phase $e^{i\alpha}$ represents the same physical state, i.e. $e^{i\alpha}|\Psi\rangle \sim |\Psi\rangle$. Mathematically speaking, the physical Hilbert space is actually the projective space $\mathrm{P}\mathcal{H}$.

The projective nature of physical Hilbert space has two direct consequences. First, representations of symmetries in Hilbert space should be *unitary* to preserve probability $|\langle\cdot|\cdot\rangle|^2$, i.e.,

$$\mathcal{U}\colon G \to \mathrm{U}(\mathcal{H}) \subset \mathrm{GL}(\mathcal{H}) \tag{3.12}$$

Here $\mathcal{U}$ and $U(\cdot)$ stands for unitary representation and unitary group, respectively. Second, since we do not require invariance of $\langle\cdot|\cdot\rangle$ but only its norm $|\langle\cdot|\cdot\rangle|$, $\mathcal{U}$ need not be an "exact" representation, but can differ by a phase; this is a so-called *projective representation*:

$$\mathcal{U}(\Lambda')\,\mathcal{U}(\Lambda) = e^{i\alpha}\,\mathcal{U}(\Lambda'\Lambda), \quad \Lambda \in G \tag{3.13}$$

At first, it seems that the extra phase adds great complexity to possible representations. Fortunately, projective representations have been well understood in mathematics (even before its quantum mechanical origin, e.g. by Schur [16] in 1911 for finite groups). In fact, a projective representation can always be *lifted* to an ordinary representation, see e.g. [14, 17]. Generally, we have:

Theorem (Lifting Projective Representations) A unitary representation of group G on projective Hilbert space $\mathcal{U}\colon G \to \mathrm{U}(\mathrm{P}\mathcal{H})$ can always be lifted to a unitary representation of its *central extension* $\tilde{G}$ on linear Hilbert space $\tilde{\mathcal{U}}\colon \tilde{G} \to \mathrm{U}(\mathcal{H})$, i.e. the following diagram commutes:

$$\begin{array}{ccc} \tilde{G} & \xrightarrow{\ \pi\ } & G \\ {\scriptstyle\tilde{u}}\downarrow & & \downarrow{\scriptstyle u} \\ \mathrm{U}(\mathcal{H}) & \longrightarrow & \mathrm{U}(\mathrm{P}\mathcal{H}) \end{array}$$

Moreover, central extension $\tilde{G}$ of Lie group G can be constructed as follows:

1. Extension by discrete $\mathbb{Z}_n$, in this case $\tilde{G}$ is the covering group of G; the maximal extension of this kind is naturally the *universal covering* of G, with $\mathbb{Z}_n = \pi_1(G)$ its fundamental group. This kind of extension is *topological* and does not change the corresponding Lie algebra, i.e. $\mathfrak{g} = \mathrm{Lie}\,G = \mathrm{Lie}\,\tilde{G}$.
2. Extension by continuous $\mathrm{U}(1)$, which descends into a non-trivial central extension of its Lie algebra:

$$\mathrm{d}\pi:\ \tilde{\mathfrak{g}} = \mathrm{Lie}\,\tilde{G} \longrightarrow \mathfrak{g} = \mathrm{Lie}\,G \tag{3.14}$$

$\tilde{\mathfrak{g}}, \tilde{G}$ therefore has an extra dimension generated by such U(1) subgroup.

For Poincaré symmetry, there is no need to consider central extension of the Lie algebra $\mathfrak{iso}(3,1)$; in fact, representation of such Lie algebra central extension is always equivalent of some representation of the original $\mathfrak{iso}(3,1)$, by a redefinition of basis[19]. However, $\mathrm{ISO}^+(3,1)$ has non-trivial π_1 due to its rotation subgroup SO(3); topologically,

$$\begin{gathered}\mathrm{ISO}^+(3,1) \cong \mathbb{R}^4 \rtimes \mathrm{SO}^+(3,1) \approx \mathbb{R}^4 \times \mathbb{R}^3 \times \mathrm{SO}(3) \approx \mathbb{R}^4 \times \mathbb{R}^3 \times \big(S^3/\mathbb{Z}_2\big), \\ \cong\text{: isomorphic as groups}, \quad \approx\text{: diffeo as manifolds}\end{gathered} \tag{3.15}$$

Where $\mathbb{R}^4$, $\mathbb{R}^3$ and SO(3) corresponds to translations P^μ, Lorentz boosts $J^{0i} = -J^{i0}$ and spatial rotations $J_k = \frac{1}{2}\epsilon_{ijk}J^{ij}$. Therefore, $\mathrm{ISO}^+(3,1)$ can be centrally extended by $\mathbb{Z}_2 = \pi_1\big(\mathrm{ISO}^+(3)\big)$.

Generally, We define *spin group* $\mathrm{Spin}(p,q)$ as the $\mathbb{Z}_2$ extension (double cover) of $\mathrm{SO}(p,q)$[20]. Note that $\mathrm{Spin}(p,q)$ is *not* always connected, and its identity component $\mathrm{Spin}^+(p,q)$ is *not* always simply connected[21]. However, for $p > 2 > q \geq 0$, $\mathrm{Spin}^+(p,q)$ is indeed simply connected, therefore it's the universal cover of $\mathrm{SO}^+(p,q)$. For $(3+1)$ dimensional spacetime $\mathbb{R}^{3,1}$, we have the following diagram:

$$\begin{array}{ccccc}
\mathrm{Spin}(3,1) & \longrightarrow & \mathrm{Spin}^+(3,1) \cong \mathrm{SL}(2,\mathbb{C}) & \longrightarrow & \mathrm{Spin}(3) \cong \mathrm{SU}(2) \\
\Big\downarrow{\scriptstyle \tilde{\pi}}\Big|{\scriptstyle /\mathbb{Z}_2} & & \Big\downarrow{\scriptstyle /\mathbb{Z}_2} & & \Big\downarrow{\scriptstyle \pi}\Big|{\scriptstyle /\mathbb{Z}_2} \\
\mathrm{SO}(3,1) & \longrightarrow & \mathrm{SO}^+(3,1) & \longrightarrow & \mathrm{SO}(3)
\end{array}$$

All arrows in this diagram are homomorphic projections: the vertical downwards arrows are $\mathbb{Z}_2$ covering maps, while the horizontal right arrows are projections onto subgroups.

[19] See [14, 17]. A general criterion is given by Bargmann's Theorem [18]; it states that *every projective representation of a connected, simply connected, finite-dimensional Lie group $\tilde{G}$ with Lie algebra second cohomology $H^2(\mathrm{Lie}\,\tilde{G}, \mathbb{R}) = 0$ can be simply lifted as a unitary representation*, i.e. such $\tilde{G}$ is the "maximal" extension.

[20] More concretely, it can be explicitly constructed via Clifford algebra $\mathrm{C}\ell(p,q)$.

[21] For example, $\pi_1\big(\mathrm{Spin}(2,1)\big) = \mathbb{Z}$ infinite cyclic, this gives rise to fractional spin $s \in \mathbb{Q}_+$ (*anyons*) in $(2+1)$ dimensions [19]. If we consider central extension of the Lie algebra, then the particle spin can even take arbitrary $s \in \mathbb{R}_+$ values.

For $(3+1)$ dimensions, we have accidental isomorphisms:

$$\mathrm{Spin}^+(3,1) \cong \mathrm{SL}(2,\mathbb{C}) \approx \mathbb{R}^3 \times S^3 \tag{3.16}$$

$$\mathrm{Spin}(3) \cong \mathrm{SU}(2) \approx S^3 \tag{3.17}$$

Where $\mathrm{SL}(2,\mathbb{C})$, $\mathrm{SU}(2)$ are classical $\mathbb{R}$–Lie groups.

Back to physics, an elementary particle in $(3+1)$ dimensional spacetime is therefore a unitary irrep of $\mathrm{Spin}^+(3,1)$, which is, furthermore[22], an induced representation from the irrep of $\mathrm{Spin}(3) \cong \mathrm{SU}(2)$.

3.2 Vector and Spinor Representations

As is evidenced before, we shall only consider irreps of $\mathrm{Spin}(3) \cong \mathrm{SU}(2)$; general representations of spin group is then closely related (induced by or composed of) $\mathrm{SU}(2)$ irreps. Here we list some key facts about irreps of $\mathrm{Spin}(3) \cong \mathrm{SU}(2)$.

- Recall $\mathrm{SO}(3)$ irreps mentioned in (3.10); they are basically spin-z eigenstates:

  $$|s_z\rangle\,, \quad s_z = -j, -j+1, \cdots, +j, \quad j \in \mathbb{Z}_{\geq 0} \tag{3.18}$$

  Here p^μ label is suppressed for sake of simplicity. All $\mathrm{SO}(3)$ irreps can always be lifted as a $\mathrm{Spin}(3) \cong \mathrm{SU}(2)$ irrep, by allowing its $\mathbb{Z}_2$ center to act trivially. In this way, we obtain the *tensor representations* of $\mathrm{SU}(2)$.

- Since $\mathrm{SU}(2)$ is simply connected, to obtain *all* of its $\mathbb{R}$-irreps, we need only consider irreps of Lie algebra $\mathfrak{spin}(3) \cong \mathfrak{su}(2)$. It turns out they are structurally identical to $\mathrm{SO}(3)$ irreps (naturally, since $\mathfrak{so}(3) \cong \mathfrak{su}(2)$ are identical Lie algebras), but with allowed j values extended to include half integers: $j \in \mathbb{Z}/2$. Irreps of $\mathrm{SU}(2)$ that are *not* irreps of $\mathrm{SO}(3)$ are called *spinor representations*.

- Therefore, all of $\mathrm{Spin}(3) \cong \mathrm{SU}(2)$ irreps, and equivalently $\mathfrak{so}(3) \cong \mathfrak{su}(2)$ irreps, are of the following form:

  $$|s_z\rangle\,, \quad s_z = -j, -j+1, \cdots, +j, \tag{3.19}$$

  $$j \in \left\{0, \tfrac{1}{2}, 1, \tfrac{3}{2}, 2, \cdots\right\} = \begin{cases} \mathbb{Z}_{\geq 0}, & \textit{tensor repr.} \\ \mathbb{Z}_{\geq 0} + \frac{1}{2}, & \textit{spinor repr.} \end{cases} \tag{3.20}$$

  The set of all tensor irreps of $\mathrm{SU}(2)$ descends to the full irreps of $\mathrm{SO}(3)$, while spinor irreps are only irreps of $\mathrm{SU}(2)$, not irreps of $\mathrm{SO}(3)$. However, they are indeed projective representations of $\mathrm{SO}(3)$.

[22] For more discussion about the irreducibility of such induced representation, see [20].

- Note that the dimension of irrep is given by:

$$\dim_{\mathbb{C}} \operatorname{span}_{\mathbb{C}}\{|s_z\rangle\} = 2j+1 \tag{3.21}$$

Weyl spinor is then defined to be the fundamental representation V with $j = \frac{1}{2}$, which is simply the defining module SU(2):

$$V = \operatorname{span}_{\mathbb{C}}\left\{\left|s_z = +\tfrac{1}{2}\right\rangle, \left|s_z = -\tfrac{1}{2}\right\rangle\right\} \equiv \operatorname{span}_{\mathbb{C}}\left\{|\uparrow\rangle_z, |\downarrow\rangle_z\right\} \tag{3.22}$$

- Rotation generators $J_k = \frac{1}{2}\,\epsilon_{ijk} J^{ij}$ is represented on V by Pauli matrices:

$$J_k = \frac{1}{2}\,\sigma_k, \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{3.23}$$

Notice how 2π rotation is not $+\mathbb{1}$ but $-\mathbb{1}$, which again confirms that V is an irrep for SU(2), but only a projective representation for SO(3):

$$e^{-iJ_z\theta} = e^{-i\sigma_z\frac{\theta}{2}} \overset{\theta=2\pi}{=\!=} -\mathbb{1} \tag{3.24}$$

In physics we would like J_k to be Hermitian or self-adjoint, hence the $-i$ coefficient in the exponential map $e^{-iJ_z\theta}$. This is one key feature of spinor: 2π rotation action is not identity $\mathbb{1}$, but will produce a minus sign on the spinor! In some sense we can think of spinor as the $\mathbb{C}$omplex "square root" of a vector [21].

- In general, transformation of operators (e.g. Pauli–Lubanski W_μ) in Hilbert space is related to its transformation as a tensor in spacetime. If we neglect projective representation for now, we have:

$$\mathcal{U}^\dagger W_\mu \mathcal{U} = \Lambda^\nu{}_\mu W_\nu, \quad \mathcal{U}\colon \Lambda \in \mathrm{SO}(3,1) \to \mathrm{U}(\mathcal{H}), \quad (\textit{without projective repr.}) \tag{3.25}$$

This is a direct consequence of $\mathcal{U}$ being a unitary representation, i.e. preserving group multiplication and having $\mathcal{U}^\dagger = \mathcal{U}^{-1}$.

- However, due to the projective nature of the Hilbert space, $\mathcal{U}(\Lambda)$ in the LHS should be lifted to $\tilde{\mathcal{U}}(\tilde{\Lambda})$, while the Λ in the RHS remains unchanged, i.e.,

$$\tilde{\mathcal{U}}^\dagger W_\mu \tilde{\mathcal{U}} = \Lambda^\nu{}_\mu W_\nu, \quad \tilde{\mathcal{U}} = \tilde{\mathcal{U}}(\tilde{\Lambda}), \quad \tilde{\Lambda} \in \mathrm{SL}(2,\mathbb{C}) \tag{3.26}$$

- The natural lift from Λ to $\tilde{\Lambda}$ while preserving Lie group structure is by considering homotopic paths from $\mathbb{1}$ to Λ. As is discussed before, we need only consider the doubly connected rotation subgroup $\mathrm{SO}(3) \subset \mathrm{SO}(3,1)$. Consider axis–angle parametrization of SO(3), we have:

$$\Lambda = e^{iJ_k\theta^k} \longmapsto (\theta^k) = (\theta^x, \theta^y, \theta^z), \quad \theta^k \in [0, 2\pi) \tag{3.27}$$

Here $iJ_k \in \mathfrak{so}(3)$ belongs to the defining module, then parametrization of SU(2) can be easily achieved by extending the range of θ^k, i.e., we have $\tilde{J}_k \in \mathfrak{su}(2)$ defining module and:

$$\tilde{\Lambda} = e^{i\tilde{J}_k\theta^k} \longmapsto (\theta^k) = (\theta^x, \theta^y, \theta^z), \quad \theta^k \in [0, 4\pi) \tag{3.28}$$

- $\tilde{\Lambda}$ is related to unit quartenions $S^3 \subset \mathbb{H} = \mathrm{span}_{\mathbb{R}}\{\mathbf{1}, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$ by the following identity, which is a direct analog of Euler's identity $e^{i\theta} = \cos\theta + i\sin\theta$:

$$\begin{gathered} \tilde{\Lambda} = e^{i\tilde{J}_k\theta^k} \sim \cos\frac{|\theta|}{2} + \hat{\theta}\sin\frac{|\theta|}{2}, \\ |\theta| = \sqrt{\theta_k\theta^k} = \sqrt{\theta_x^2+\theta_y^2+\theta_z^2}, \qquad \hat{\theta} = \big(\mathbf{i}\,\theta_x + \mathbf{j}\,\theta_y + \mathbf{k}\,\theta_z\big)/|\theta| \end{gathered} \tag{3.29}$$

  θ is the rotation angle, while $\hat{\theta}$ describes the rotation axis, hence axis–angle parametrization. Replace $\tilde{\Lambda}$ in (3.26) with (θ^k), and we have:

$$\tilde{\mathcal{U}}^\dagger W_\mu \tilde{\mathcal{U}} = \Lambda^\nu{}_\mu W_\nu, \quad \tilde{\mathcal{U}} = \tilde{\mathcal{U}}(\theta), \quad \theta = (\theta^k), \quad \theta_k \in [0, 4\pi) \tag{3.30}$$

4 Spin Structure and Spinor Field

From our previous discussions, we have established that in quantum mechanics, an elementary particle $|p^\mu, s_z\rangle \in \mathcal{H}$ is something that lives in a unitary irrep of extended Poincaré group $\mathbb{R}^4 \rtimes \mathrm{Spin}^+(3,1)$. By definition, particles become eigenstates in the Hilbert space $\mathcal{H}$ of wave functions, which is often de-localized in spacetime.

On the other hand, in field theory, a particle should be an excitation of some local field $\phi(x)$, like a ripple in the pond. Is it possible to relate these two pictures?

The answer is yes, and the idea behind this relation is in fact quite basic. Roughly speaking, $|p^\mu, s_z\rangle$ serves as *Fourier modes* of field $\phi(x)$. Details of this correspondence in quantum field theory are give in *Weinberg*'s chapter: *Quantum Fields and Antiparticles* [14], but similar holds for ordinary quantum mechanics. The end result is that similar to an elementary particle $|p^\mu, s_z\rangle$, elementary fields $\phi(x)$ also becomes a representation of the Poincaré symmetry, where not only the symmetry acts on coordinates x in the regular way:

$$\phi(x) \quad \longmapsto \quad \phi(x - \delta x) \tag{4.1}$$

but also the Lorentz symmetry $\mathrm{Spin}^+(3,1) \cong \mathrm{SL}(2,\mathbb{C})$ gets represented pointwise on the *fiber* $\phi(x_0)$. This naturally leads us to consider *fiber bundles* based on $\mathcal{M}$.

4.1 Introduction to Principal Bundle

To introduce spin structure in a natural way, we follow the discussions of [21–24]. For a mathematically rigorous understanding of spin geometry, see [5, 25, 26]. Our notation will mostly conform to *Lee* [27]. We shall use the language of fiber bundles, introduced in [23, 27].

To simplify the situation, we will again work on hypersurface $\Sigma \subset \mathcal{M}$. It is illuminating to first consider the classical situation, then we have a natural poinwise fundamental representation of SO(3) — the tangent space $T_x\Sigma$. Union of tangent spaces at each point gives the tangent bundle:

$$T\Sigma = \coprod_{p\in\Sigma} T_p\Sigma \tag{4.2}$$

SO(3) action is reflected in the transformation $T_p\Sigma$ under isometric coordinate transformations, or equivalently, isometric *transition function* between charts $\{(U_\alpha, \Phi_\alpha)\}_\alpha$:

$$\Phi_\alpha \circ \Phi_\beta^{-1}\big|_p = \tau_{\alpha\beta}(p), \quad \tau_{\alpha\beta} \in \mathrm{SO}(3) \subset \mathrm{GL}(3,\mathbb{R}) \tag{4.3}$$

Where Φ_α is the local trivialization on U_α. In other words, SO(3) is hence the *structure group* of tangent bundle, when the transitions between charts is restricted to isometries defined by the metric.

However, there is one subtlety that needs to be addressed: the refinement of $\mathrm{GL}(3,\mathbb{R})$ transition to SO(3) is locally feasible, but might fail if we consider *global* transition around a manifold that is not orientable. To see this, we first observe that transition functions should always satisfy such *cocycle condition* (*Lee* [27], Problem 10-5):

$$\tau_{UV} \circ \tau_{VW} \circ \tau_{WU} = \mathbb{1} \tag{4.4}$$

For a Möbius strip however, this is impossible if we restrict $\tau \in \mathrm{SO}(3)$. It is not hard to see that $\tau_{UV} \circ \tau_{VW} \circ \tau_{WU} = -\mathbb{1}$, if we only allow $\tau \in \mathrm{SO}(3)$. In fact, given a metric, we can only guarantee the refinement of $\mathrm{GL}(3,\mathbb{R}) \supset \mathrm{O}(3)$.

From now on we shall assume that our manifold $\Sigma \in \mathcal{M}$ is orientable, therefore the tangent bundle becomes a natural realization of the local SO(3) isometry w.r.t. the metric. In fact, any representation of SO(3) can be expressed as tensor products and direct sums of the fundamental representations, therefore the *tensor fields* realize all possible SO(3) representations:

$$\phi^{\mu_1\mu_2\cdots\mu_k}{}_{\nu_1\nu_2\cdots\nu_l} \in \Gamma(\Sigma, T^{(k,l)}\Sigma) : \ \Sigma \to T^{(k,l)}\Sigma \tag{4.5}$$

If we require only local SO(3) symmetry, then the job is done; everything is naturally provided by the tensor bundle. However, after quantization the symmetry must be extended to Spin(3), therefore our job is to somehow *lift* the tensor bundle to a *spinor bundle*, in a natural manner that is compatible to the $\mathbb{Z}_2$ covering:

$$\pi\colon \mathrm{Spin}(3) \cong \mathrm{SU}(2) \to \mathrm{SO}(3) \tag{4.6}$$

In order to obtain a compatible spinor bundle, it is helpful to re-examine the common structure of all tensor bundles — the SO(3) transition between charts. In general, it is possible to remove all unnecessary data of a fiber bundle, leaving only the information of how it transitions between charts. This gives the so-called *principal bundle*.

Definition (Principal G–Bundle) A principal G–bundle $P_G \to \Sigma$ is a fiber bundle whose model fiber is a manifold diffeo to the structure group G.

For tensor bundles transforming under SO(3) isometry, they correspond to the same principal SO(3) bundle $P_{\mathrm{SO}(3)}$. In fact, it can be explicitly constructed by considering the orthonormal frames fields on the tangent bundle, i.e. the *frame bundle*.

Since its fiber is diffeo to the structure group itself, principal bundle has a natural *left and right* group action, defined by the usual group multiplication on G. This two-sided group action corresponds to the action on the contra-variant and covariant component of a tensor.

In the physics community, we often say that *a tensor is something that transform like a tensor*. Now we can see that this saying is not without its merits, since indeed the essence of a tensor is its transformation property, captured by the principal bundle.

Even better, given a principal G–bundle and a G–module F as the model fiber, it is possible to reconstruct the *associated fiber bundle* up to equivalence. This is precisely the VECTOR BUNDLE CONSTRUCTION THEOREM given in *Lee* [27], Problem 10-6, and we will not restate it here. We need only these two conditions:

- Transition function is a G representation on F:

$$\tau_{UV}(p) \in G \subset \operatorname{End} F \tag{4.7}$$

- Cocycle condition:

$$\tau_{UV} \circ \tau_{VW} \circ \tau_{WU} = \mathbb{1} \tag{4.8}$$

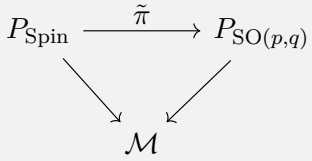
With the idea of principal bundle and construction theorem of associated bundle, our task of defining a spinor field can now be decomposed into the following two steps:

- Find a natural and compatible lift of $P_{\mathrm{SO}(3)}$ to some principal Spin(3)–bundle $P_{\mathrm{Spin}(3)}$; this is called the *spin structure* on Σ;
- Construct a spinor bundle associated with $P_{\mathrm{Spin}(3)}$, then its section is precisely the spinor field we want to obtain.

4.2 Spin Structure

What kind of lift of $P_{\mathrm{SO}(3)}$ to $P_{\mathrm{Spin}(3)}$ is considered natural? This is characterized by the definition of spin structure. We follow closely the discussions in [22, 23].

Definition (Spin Structure) A spin structure is a principal $\mathrm{Spin}(p,q)$–bundle P_{Spin} together with a bundle morphism:



Which restricts fiber-wise to the $\mathbb{Z}_2$ covering homomorphism $\pi\colon \mathrm{Spin}(p,q) \to \mathrm{SO}(p,q)$.

Spin structures need not exist and even if they do they need not be unique. We shall see the obstruction to such lifting by explicit construction.

Again, we choose to work on the hypersurface Σ. Starting from the SO(3) structure, transition function $\tau \in \mathrm{SO}(3)$ satisfies the cocycle condition:

$$\tau_{UV} \circ \tau_{VW} \circ \tau_{WU} = \mathbb{1}, \quad \tau \in \mathrm{SO}(3) \tag{4.9}$$

The lifting of $\tau \in \mathrm{SO}(3)$ to $\tilde{\tau} \in \mathrm{Spin}(3)$ always exists locally. First, we simply choose one of the pre-images of the $\mathbb{Z}_2$ covering at some point $p \in \Sigma$, i.e.

$$\pi\big(\tilde{\tau}(p)\big) = \tau(p), \quad \pi\colon \mathrm{Spin}(3) \to \mathrm{SO}(3), \quad p \in \Sigma \tag{4.10}$$

Then we need to extend such choice of $\tilde{\tau}$ smoothly around the manifold. There will be no obstruction of such extension around a small neighborhood of $p \in U \subset \Sigma$, but it may fail globally if two different "paths" of extension yields incompatible results. In the end, we can only guarantee that:

$$\pi\big(\tilde{\tau}_{UV} \circ \tilde{\tau}_{VW} \circ \tilde{\tau}_{WU}\big) = \mathbb{1}, \tag{4.11}$$

$$\tilde{\tau}_{UV} \circ \tilde{\tau}_{VW} \circ \tilde{\tau}_{WU} = f_{UVW} = \pm\mathbb{1} \in \ker \pi = \mathbb{Z}_2 \tag{4.12}$$

But for a spin structure $P_{\mathrm{Spin}} \to P_{\mathrm{SO}(3)}$, $\tau \in \mathrm{SO}(3)$ and $\tilde{\tau} \in \mathrm{Spin}(3)$ should both satisfy the cocycle condition. This is possible iff. $f_{UVW} = 1$. It turns out that f_{UVW} as a function of charts is in fact independent of the choice of local frames; it is completely fixed by the manifold Σ, i.e. $f_{UVW} = f(\Sigma)$ is a *characteristic class* of Σ.

Notice that our discussions about spin structure greatly resemble those in the last section, about the existence of SO(3) transition. Note that the $\mathrm{GL}(3,\mathbb{R}) \to \mathrm{O}(3) \to \mathrm{SO}(3)$ refinement is also locally trivial, but obstruction occurs in the form of (*non-*) orientability when we try to extend it globally on the whole manifold. The situation with spin structure is quite similar, with the obstruction given by $f(\Sigma)$.

In fact, these two things can be described in a unifying way by *Stiefel–Whitney class* w_n and *Čech cohomology* $H^n(\Sigma, \mathbb{Z}_2)$. $f_{UVW} = f(\Sigma)$ defined above is precisely the 2nd Stiefel–Whitney class $w_2(T\Sigma)$.

We will not dive too deep into this subject; for further discussions, see [23, 25]. Rather, we simply summarize some relevant results below:

- Given any bundle $E \to \Sigma$, $w_n(E) \in H^n(\Sigma, \mathbb{Z}_2)$;
- Σ orientable iff. $w_1(T\Sigma)$ is trivial;
- Σ has spin structures iff. $w_2(T\Sigma)$ is trivial; furthermore, if this is the case, then the spin structures on Σ are in one-to-one correspondence with elements of $H^1(\Sigma, \mathbb{Z}_2)$.
- $H^1(\Sigma, \mathbb{Z}_2) \cong \mathrm{Hom}(\pi_1(\Sigma), \mathbb{Z}_2)$, i.e. ways of assigning $(\pm)$ signs to non-contractible loops (S^1) in Σ; here $\pi_1(\Sigma)$ is the fundamental group.

In general, the refinement and lifting of structure group happens in the following steps:

- We start with a C^∞ manifold Σ with $\tau \in \mathrm{GL}(n, \mathbb{R})$;
- With Riemannian metric g_{ij}, by restricting to isometries, $\tau \in \mathrm{O}(n) \subset \mathrm{GL}(n, \mathbb{R})$;
- With orientability we can further restrict $\tau \in \mathrm{SO}(n) \subset \mathrm{O}(n)$;
- With spin structure, we can lift $\tau \in \mathrm{SO}(n)$ to $\tilde{\tau} \in \mathrm{Spin}(n)$.

To better understand spin structure, we shall look at the most basic example $\Sigma = S^1$ with unit radius the usual flat metric. Note that S^1 is an oriented 1–fold, $w_1(\mathrm{T}S^1) = \mathbb{1}$, and its tangent bundle is just the trivial line bundle:

$$\mathrm{T}S^1 \cong S^1 \times \mathbb{R} \tag{4.13}$$

whose isometric transition function is trivial $\tau \equiv \mathbb{1} = \mathrm{SO}(1)$, so the frame bundle is just $S^1 \times \{\mathbb{1}\} \cong S^1$. The spin group, however, is non-trivial: $\mathrm{Spin}(1) \cong \mathbb{Z}_2$, hence there might be non-trivial liftings to P_{Spin}.

Indeed, $w_2(\mathrm{T}S^1) = \mathbb{1}$, but $H^1(\Sigma, \mathbb{Z}_2) \cong \mathrm{Hom}(\pi_1(\Sigma), \mathbb{Z}_2) = \pm\mathbb{1}$, and we have two inequivalent spin structures on S^1. It is more intuitive to look at their associated line bundles. One of them is simply $S^1 \times \mathbb{R}$, but the other one consists of a non-trivial *twist* on the fiber:

$$\begin{gathered} E = S^1 \times_{\mathbb{Z}_2} \mathbb{R}, \quad \psi \in \Gamma(\Sigma, E) : \Sigma \to E, \\ \psi(\theta + 2\pi) = -\psi(\theta) \end{gathered} \tag{4.14}$$

This is precisely the *Möbius bundle* studied extensively in *Lee* [27]. In fact, the lifting of $\tau \equiv \mathbb{1}$ to $\tilde{\tau} \in \mathbb{Z}_2$ is explicitly given in *Lee* [27], Problem 10-13. By introducing the $\mathbb{C}$omplex label $z = e^{i\theta} \in S^1$ and two charts $U_\pm = S^1 \setminus \{\pm 1\}$, we have:

$$\tilde{\tau}_{+,-}(z) = \begin{cases} +\mathbb{1}, & \mathrm{Im}\, z > 0, \\ -\mathbb{1}, & \mathrm{Im}\, z < 0, \end{cases} \quad z = e^{i\theta} \in S^1 \tag{4.15}$$

In summary, S^1 has two distinct spin structures, and their associated $\mathbb{R}$ line bundles are respectively, $S^1 \times \mathbb{R}$ and $S^1 \times_{\mathbb{Z}_2} \mathbb{R}$, which also correspond to periodic and anti-periodic boundary conditions on $\theta \in [0, 2\pi]$.

In some literature (incl. most physics texts), these two spin structures are affectionally named *Neveu–Schwarz* (NS) and *Ramond* (R). In general, a genus g Riemann surface has $2g$ distinct periods, which give rise to 2^{2g} inequivalent spin structures.

4.3 Dirac Operator and Spin Connection

Back to Witten's proof of the positive energy theorem, we have an asymptotic flat 3–fold $\Sigma \subset \mathcal{M}$, and indeed $w_{1,2}(T\mathcal{M}) = w_{1,2}(T\Sigma) = \mathbb{1}$ are all trivial, so the existence of spin structure is guaranteed [4].

With a spin structure, we can construct associated *spinor bundles* just like the S^1 case. However, the model fiber F of a spinor bundle is generally some $\mathbb{C}$ vector space, since the related Hilbert space on Σ is a $\mathbb{C}$ linear space.

In general, the model fiber F can be any spin representation. In Witten's original proof, the (complexified) Clifford $\mathrm{C}\ell(3,1)^{\mathbb{C}}$ module is chosen, with $F \cong \mathbb{C}^4$ as a vector space. This is in fact the defining module of $\mathrm{Spin}(3,1)$, and corresponds to the famous *Dirac spinor*. Dirac spinor bundle is then pulled back to the hypersurface $\Sigma \subset \mathcal{M}$ and treated as a $\mathrm{Spin}(3)$ spinor bundle.

Dirac spinor first arose in the effort of finding a wave equation compatible with relativity. Dirac noted that relativistic wave equation for an electron should be 1^{st} order in time dependence, so that it behaves nicely as an initial value problem [28]. This leads to the need of taking a "square root" of the d'Alembert operator $\partial^a\partial_a$, i.e. finding some scalar operator $i\not{\partial}$ so that:

$$(i\not{\partial})^2 = \partial^a\partial_a \tag{4.16}$$

Here i is introduced to ensure that $i\not{\partial}$ is self-adjoint, much like $\partial^a\partial_a$ itself. Dirac discovered that such $i\not{\partial}$ could be realized using $\gamma^a \in \mathrm{C}\ell(3,1)$, namely,

$$\not{\partial} = \gamma^a\partial_a, \quad \not{\partial}^2 = -\mathbb{1}\,\partial^a\partial_a, \tag{4.17}$$

$$\mathbb{1}, \gamma^a \in \mathrm{C}\ell(3,1), \tag{4.18}$$

$$\tfrac{1}{2}\{\gamma^a, \gamma^b\} = \tfrac{1}{2}(\gamma^a\gamma^b + \gamma^a\gamma^b) = -\eta^{ab} \tag{4.19}$$

$\not{\partial}$ is the so-called *Dirac operator*. $\mathrm{C}\ell(3,1)$ is faithfully represented by 4×4 $\mathbb{C}$omplex matrices, and the field ψ it acts on shall be a column vector $\psi \in F \cong \mathbb{C}^4$. This is the Dirac spinor.

Dirac operator and Clifford γ^a can then be generalized in curved spacetime; we have:

$$\not{\nabla} = \gamma^a\nabla_a \equiv \theta^a.\nabla_a \tag{4.20}$$

Where θ^a is the orthonormal coframe defined in (1.2), and "." denotes some Clifford multiplication. Without referencing γ^a, the latter expression can then be generalized to arbitrary fiber F, as long as F is some spin representation. *Dirac's equation* in curved spacetime can then be expressed in the following form[23]:

$$\left(i\not{\nabla} - m\right)\psi(x) = 0 \tag{4.21}$$

In Witten's proof, since we are only working on 3–fold $\Sigma \in \mathcal{M}$, this is further truncated to the *hypersurface Dirac operator*:

$$\mathcal{D} = \theta^i.\nabla_i = \sum_{i=1}^{3} \theta^i.\nabla_i \tag{4.22}$$

[23] There are numerous conventions for $\mathrm{C}\ell(3,1)$ and $\not{\nabla}$. We have chosen a convention that is compatible with Witten's proof [1, 4]. Note that this differs from Weinberg's convention [14].

In fact, on such 3–fold Σ, Witten's proof can actually be simplified using basic Spin(3) representations, i.e. *Weyl spinors* introduced in Section 3.2. Then, the θ^i multiplications can be expressed with Pauli matrices σ^i, defined in (3.23), and the fiber is thus reduced to $F = V \cong \mathbb{C}^2$. See [2] for this simplified version of Witten's proof.

Note that we haven't addressed the nature of covariant derivative ∇. Since we are no longer working on a tensor bundle, this could not be the usual Levi-Civita connection. In fact, the lifting of $P_{\mathrm{SO}(3)}$ to P_{Spin} naturally induces a lifting of the usual Levi-Civita connection; this is the so-called *spin connection*. See [22, 24] for further discussions.

Generally, for $\Sigma \subset \mathcal{M}$, ∇: spin connection on $\mathcal{M}$, $\mathcal{D}$: hypersurface Dirac operator, (4.16) is modified to be the *Lichnerowicz–Weitzenböck formula*:

$$\mathcal{D}^\dagger \mathcal{D} = D^2 = \nabla^\dagger \nabla + \mathcal{R} \tag{4.23}$$

Where "†" denotes $\mathbb{C}$–adjoint w.r.t. the standard Hermitian form on $F \cong \mathbb{C}^n$, and $\mathcal{R}$ is some curvature corrections [4].

5 Sketch of the Proof

With previous preparations, we are finally ready to sketch the main ideas of Witten's proof, following [1, 2, 4].

- First, assume there is matter described by spinor field ψ on the manifold $\mathcal{M}$.

  The motivation and validity of this assumption is the main content of this review. It is indeed the essence of Witten's proof; following steps, though technically difficult, are conceptually simple compared to this first step.

- The dynamics of spinor field respects spacetime symmetries, and $\psi(x)$ transforms as a *spinor representation* on its fiber, specified in section 3.2. Just as is discussed in section 2.1, we can construct a conserved Noether current of ψ:

  $$j^\mu = \psi^\dagger . \theta^0 . \theta^\mu . \psi \tag{5.1}$$

  This is a null (lightlike) vector near the asymptotic boundary, hence the contraction of j^μ with the ADM energy (2.6) yields:

  $$j^\mu P_\mu \propto E - |P| \tag{5.2}$$

- On the other hand, by definition, $j^\mu P_\mu$ can be expanded with curvature $R_{\mu\nu}$ and field ψ. Using Einstein's equations, $R_{\mu\nu}$ can be further expressed in local energy–momentum $T_{\mu\nu}$ in the following form:

  $$j^\mu P_\mu = \int_\Sigma \left(-T_{0\nu} j^\nu + \cdots \right) \mathrm{d}\, Vol \tag{5.3}$$

  Where the $(\cdots)$ part contains only $\nabla_\mu \psi$ terms. This can be seen as the integral form of the Lichnerowicz–Weitzenböck formula (4.23), where $\mathcal{R}$ is replaced by $T_{\mu\nu}$.

  Calculations from now on relies heavily on tricks from spinor analysis, including extensive use of (4.23) and *Bochner formula*, which is well-reviewed in math texts such as [5, 26]. Witten, on the other hand, refers to supergravity for inspirations [1].

- $T_{\mu\nu}$ satisfies the dominant energy condition (DEC) (2.8), so the first term of (5.3) is positive. There are also spinor bilinear $\|\nabla\psi\|^2$ in (5.3), which are also automatically positive. The only remaining part is proportional to $\not{D}\psi$. If we can choose some non-zero ψ so that:

$$\not{D}\psi = 0 \tag{5.4}$$

  Then the proof is done.

- (5.4) is called the *Dirac–Witten* equation. Proof of the theorem now comes down to analyzing this PDE. In fact, on asymptotically flat $\Sigma \subset \mathcal{M}$, if ψ vanishes at infinity *fast enough* ($\sim \frac{1}{r^{1-\epsilon}}$), then it has to be zero $\psi = 0$. This gives $E = 0$ and $\mathcal{M}$ flat.

  Furthermore, if ψ is asymptotically constant, then the right-hand side of (5.3) is positive, since it contains spinor bilinear $\|\nabla\psi\|^2$.

  These results are the so-called *Witten's vanishing theorem* [4]. It can be proven rigorously using Green's function of operator $\not{D}$, as is proved and well-reviewed in [1, 4]. This concludes Witten's proof of the positive energy theorem.

□

6 Acknowledgements

We thank Dr. Luis Apolo for his extensive proof-reading and valuable suggestions for this review. We thank Prof. Mauricio Romo, fellow student Yuan Zhong and Kai Xu for many inspiring discussions. We thank Prof. Hui Ma for her great teachings in the *Smooth Manifolds* class.

References

[1] Edward Witten. *A Simple Proof of the Positive Energy Theorem*. *Commun. Math. Phys.* 80:381, 1981.

[2] Norbert Straumann. *General Relativity*. Graduate Texts in Physics. Springer, Dordrecht, 2013.

[3] R. Schon & Shing-Tung Yau. *On the Proof of the positive mass conjecture in general relativity*. *Commun. Math. Phys.* 65:45–76, 1979.

[4] Thomas Parker & Clifford Henry Taubes. *On Witten's Proof of the Positive Energy Theorem*. *Commun. Math. Phys.* 84:223, 1982.

[5] J. Jost. *Riemannian Geometry and Geometric Analysis*. Universitext. Springer Berlin Heidelberg, 2011. ISBN: 9783642212987.
URL: https://link.springer.com/book/10.1007/978-3-642-21298-7.

[6] Robert M. Wald. *General Relativity*. Chicago Univ. Pr., Chicago, USA, 1984.

[7] Xiao Zhang. *Recent progress on the positive energy theorem*. *Int. J. Mod. Phys.* A30(28n29):1545018, 2015. arXiv: 1508.04230 [math.DG].

[8] Máximo Bañados & Ignacio A. Reyes. *A short review on Noether' s theorems, gauge symmetries and boundary terms*. *Int. J. Mod. Phys.* D25(10):1630021, 2016. arXiv: 1601.03616 [hep-th].

[9] Cedomir Crnkovic & Edward Witten. *Covariant description of canonical formalism in geometrical theories. Three hundred years of gravitation*.676–684, 1987.

[10] Daniel Harlow & Jie-Qiang Wu. *Covariant phase space with boundaries*, 2019. arXiv: 1906.08616 [hep-th].

[11] Vivek Iyer & Robert M. Wald. *Some properties of Noether charge and a proposal for dynamical black hole entropy*. *Phys. Rev.* D50:846–864, 1994. arXiv: gr-qc/9403028 [gr-qc].

[12] Richard L. Arnowitt, Stanley Deser & Charles W. Misner. *The Dynamics of general relativity*. *Gen. Rel. Grav.* 40:1997–2027, 2008. arXiv: gr-qc/0405109 [gr-qc].

[13] Erik Curiel. *A Primer on Energy Conditions*. *Einstein Stud.* 13:43–104, 2017. arXiv: 1405.0403 [physics.hist-ph].
URL: http://strangebeautiful.com/papers/curiel-primer-energy-conds.pdf.

[14] Steven Weinberg. *The Quantum theory of fields. Vol. 1: Foundations*. Cambridge University Press, 2005. ISBN: 9780521670531, 9780511252044.

[15] Xavier Bekaert & Nicolas Boulanger. *The Unitary representations of the Poincare group in any spacetime dimension*. In *2nd Modave Summer School in Theoretical Physics Modave, Belgium, August 6-12, 2006*, 2006. arXiv: hep-th/0611263 [hep-th].

[16] J Schur. *Über die Darstellung der symmetrischen und der alternierenden Gruppe durch gebrochene lineare Substitutionen. Journal für die reine und angewandte Mathematik*. 139:155–250, 1911.

[17] Martin Schottenloher. *A mathematical introduction to conformal field theory*. *Lect. Notes Phys.* 759:1–237, 2008.

[18] V. Bargmann. *On Unitary ray representations of continuous groups*. *Annals Math.* 59:1–46, 1954.

[19] Frank Wilczek. *Quantum Mechanics of Fractional Spin Particles*. *Phys. Rev. Lett.* 49:957–959, 1982.

[20] Wu-Ki Tung. *Group theory in physics: an introduction to symmetry principles, group representations, and special functions in classical and quantum physics*. World Scientific Publishing Company, 1985.

[21] Luis Alvarez-Gaume, Gregory W. Moore & Cumrun Vafa. *Theta Functions, Modular Invariance and Strings*. *Commun. Math. Phys.* 106:1–40, 1986.

[22] José Figueroa-O' Farrill. *Spin geometry*.
URL: https://empg.maths.ed.ac.uk/Activities/Spin/.

[23] M. Nakahara. *Geometry, topology and physics*. 2003.

[24] Konstantin Wernli. *Lecture Notes on Spin Geometry*, 2019. arXiv: 1911.09766 [math.DG].

[25] H Blaine Lawson & Marie-Louise Michelsohn. *Spin geometry*, volume 38. Princeton university press, 2016.

[26] Jean-Pierre Bourguignon, Oussama Hijazi, Jean-Louis Milhorat, Andrei Moroianu & Sergiu Moroianu. *A Spinorial Approach to Riemannian and Conformal Geometry*. 2015.

[27] J. Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer New York, 2012. ISBN: 9781441999825.
URL: https://link.springer.com/book/10.1007/978-1-4419-9982-5.

[28] Paul Adrien Maurice Dirac. *The principles of quantum mechanics*, number 27. Oxford university press, 1981.