The Action Function of Adiabatic Systems

Hans Detlef Hüttenbach

Abstract. The action function of a relativistic macroscopic adiabatic (or closed) system of particles, described as a continuously differentable function of energy-momentum in space-time, is shown to exist. It is shown to be a plane wave, wheras its 2^{nd} integral satisfies the covariant Maxwell's equations. It is shown then, how to restate these results in terms of Functional Analysis of Hilbert spaces.

With it, we show a.o. that $\mathcal{PCT} = -\mathcal{CPT} = \pm 1$ holds for this system, which is a strong form of the PCT-theorem. It is shown that - in order to capture the concept of mass - the standard model gauge group could be augmented by a factor group $U(2)$, such that the complete gauge group would become $U(4)$.

1. Introduction

1.1. Synopsis of Action in Classical Mechanics

In classical mechanics, a dynamical system is described w.r.t. one time coordinate t and n location coordinates $q_1, \dots, q_n$ by a Lagrangian function $L(t, q_1, \dots, q_n, \dot{q}_1, \dots, \dot{q}_n)$, which for fixed, real $t_0 < t$ defines a (linear) functional on the (vector) space of all (piecewise/continuously) differentiable paths $\omega : [t_0, t] \ni s \mapsto (q_1(s), \dots, q_n(s)) \in \mathbb{R}^n$ by

$$S(\omega) := \int_{t_0}^{t} L(s, q_1(s), \dots, q_n(s), \dot{q}(s)_1, \dots, \dot{q}_n(s))ds.$$

This is called the action functional, and it is demanded to be extremal on the physically possible paths. If it can be solved globally, keeping the start point, $t_0, q_1(t_0) \dots, q_n(t_0)$, fixed, it results in S being expressed as action function $S(t, q_1, \dots, q_n)$, which often is termed as "Hamilton's principal function".

When the energy E is conserved, then $S = \sum_{1 \le i \le n} \int p_i dq_i - Edt$, where the p_i are the momentum coordinates for the location coordinates q_i. Inverting time, one gets $\tilde{S} = \int Edt + \sum_{1 \le i \le n} \int p_i dq_i$. In other words: if the dynamical system is conserving energy and can be solved completely, then the vector field $(E, p_1(t, q_1, \ldots, q_n), \ldots p_n(t, q_1, \ldots, q_n))$ is integrable, and its integral is $\tilde{S}$, which is what in the following will be called "**action function**".

1.2. Definition of the Adiabatic Dynamical System

The above mechanical model is limited to systems containing only a very few particles, whereas in nearly all situations circumstances, billions of particles are involved, resulting into equations with billions of variables. In these cases, the system is to be modelled as a quadrupel of energy and momentum densities $j = (j_0, \ldots, j_3)$, where $j_\mu : \mathbb{R}^4 \ni (t, \mathbf{x}) \mapsto j_\mu(t, \mathbf{x}) \in \mathbb{R}$ is the energy density for $\mu = 0$ and momentum density component for $\mu = 1, 2, 3$:

Definition 1.1. Let $j_\mu j^\mu := j_0^2 - \cdots - j_3^2$, where the speed of light $c \equiv 1$ is understood throughout. Then an **adiabatic system** of (massive) particles is a 4-vector $j = (j_0, \ldots, j_3)$ of continuously differentiable functions

$$j_\mu : \mathbb{R}^4 \ni x := (t, \mathbf{x}) \mapsto j_\mu(t, \mathbf{x}) \in \mathbb{R}$$

of energy j^0 and momentum (j_1, j_3, j_3), such that the following conditions are met:

1. (*Massiveness*) The image $Im(j) := \{j(x) \mid x \in \mathbb{R}^4\}$ of j is disjoint with the light cone $\mathcal{C} := \{p \in \mathbb{R}^4 \mid p_0^2 - \cdots - p_3^2 = 0\}$.
2. (*Adiabaticity*) $\sum_{0 \le \mu \le 3} \partial_\mu j_\mu \equiv 0$, where $\partial_\mu := \partial / \partial x^\mu$.

Remark 1.2. There is no sense in demanding $j_0 \ge 0$, because time inversion transforms a positive energy into a negative one, anyhow.

Remark 1.3. The first condition states that all particles in the system have a mass unequal zero, so that no particle will move at the speed of light (massiveness). The second condition states the isolatedness or closedness of the system: there is no energy energy created or lost by the system (adiabaticness).

Remark 1.4. The energy momentum $j(t, \mathbf{x})$ is the (experimentally detectable) energy momentum at the space time point $(t, \mathbf{x}) \in \mathbb{R}^4$. There is no qualifying statement as to how this value is composed of.

2. Integability of Adiabatic Systems

Theorem 2.1. 1. *Let j be an adiabatic system, and let γ_μ be the Dirac matrices (see e.g.* [4], *Sec. 19.5.1 or - preferrably - wikipedia.org). Then $\not{j}(\not{x}) := \sum_\mu j_\mu(x_0\gamma_0, \ldots, x_3\gamma_3)\gamma_\mu$ is integrable w.r.t. the differential form $d\omega := \gamma_0 dx_0 + \gamma_1 dx_1 + \gamma_2 dx_2 + \gamma_3 dx_3$, if and only if the partial derivatives of $j = (j_0, \ldots, j_3)$ anti-commute for all x, i.e: $\partial_\nu j_\mu(x) = -\partial_\mu j_\nu(x)$ for $0 \le \mu, \nu \le 3$ and $x \in \mathbb{R}^4$.*

2. *The action function $\Phi := \int \not{j} d\omega$ of the 4-vector field j then is a plane wave, i.e.: $\Box\Phi = 0$, where $\Box := \partial_0^2 - \cdots - \partial_3^2$ is the wave operator.*
3. *Φ can be integrated again w.r.t. $d\omega$ along the time and space coordinates $x_0, \ldots, x_3$, yielding a 4-vector (spinor) field $\not{A} := (A_0\gamma_0, \ldots, A_3\gamma_3)$, for which $\Box\not{A} = \not{j} := (j_0\gamma_0, \ldots, j_3\gamma_3)$ holds.*

Proof. The proof is via the following lemma:

Lemma 2.2. *In order to be a Lorentz covariant system, the (Euclidean) derivative* $Dj := (j)_{\mu\nu} = (\partial_\mu j_\nu)_{0\leq\mu,\nu\leq 3}$ *of an adiabatic system* $j = (j_0, \ldots, j_3)$ *is anti-commuting for all its off-diagonal elements, i.e.:* $(Dj)_{\mu\nu} = -(Dj)_{\nu\mu}$ *for* $0 \leq \mu \neq \nu \leq 3$.

Proof. Since j is continuously differentiable, its derivative, $Dj = (\partial_\mu j_\nu)_{\mu\nu}$ exists and can be split into the sum of a symmetric matrix $(f)_{\mu\nu}$, i.e.: $f_{\mu\nu} := \frac{1}{2}(\partial_\mu j_\nu + \partial_\nu j_\mu)$ for $0 \leq \mu, \nu \leq 3$ and an antisymmetric matrix $(g)_{\mu\nu} := (j)_{\mu\nu} - (f)_{\mu\nu}$.

$(f)_{\mu\nu}$ defines a 2-form $\omega = \sum_{\mu,\nu} f_{\mu\nu} dx_\mu \wedge dx_\nu$, which rewrites into $\omega = \sum_{0\leq\mu<\nu\leq 3}(f_{\mu\nu} - f_{\nu\mu})dx_\mu \wedge dx_\nu \equiv 0$ because of the symmetry of $(f)_{\mu\nu}$. So, its external derivative $d\omega$ likewise vanishes, and ω therefore is closed (see: [1]). And because the domain $\mathbb{R}^4$, on which $(f)_{\mu\nu}$ is defined, is locally convex, so star-shaped, ω itself is exact, i.e.: integrable into a 1-form $I\omega = f_0 dx_0 + \cdots + f_3 dx_3$ (again, see [1, Sec. 2.12-2.13]). In other words, the symmetric matrix $(f)_{\mu\nu}$ is (path) integrable to a vector function $(f_0, \ldots, f_3)$. And again, since $\omega \equiv 0$ is the external derivative of $f_0 dx_0 + \cdots + f_3 dx_3$, $f_0 dx_0 + \cdots + f_3 dx_3$ is an exact differential form, so $(f_0, \ldots, f_3)$ is path integrable to a function F, say. Because $\sum_\mu f_{\mu\mu} = 0$, we have:

$$\Delta F := (\partial_0^2 + \cdots + \partial_3^2)F \equiv 0.$$

So, $F \in ker(\Delta)$, where $ker(\Delta)$ is the kernel of Δ, which is the vector space of all harmonic functions on $\mathbb{R}^4$. And as such, these are real parts of analytic functions (see e.g [**?**]), and therefore maximum and minimum within a (4-dimensional) ball of Euclidean radius r are always positioned on the r-sphere itself. Now with F, also the partial derivatives are harmonic; and these are energy and momentum (contributions), since F is an action function. In particular, as a minimum requirement, we expect momentum and energy to be constant in space-like separated regions from the source as large spatial distances. But then, due to harmoniticity of energy and momentum density contributions, these functions must be constant throughout! So, apart from ΔF not being Lorentz covariant, that means that unless zero or constant throughout, the forces that result in a change of energy and momentum in space and time, will never converge to zero for this action function at large distances.
In all, not just for covariance, but for the physical principle of locality of the sources, we have to demand $(f)_{\mu\nu} \equiv 0$. □

An immediate consequence of the anti-symmetry is:

Corollary 2.3. $\nabla j_0 + \partial_0 \mathbf{j} = 0$, *i.e.:* $\partial_k j_0 = -\partial_0 j_k$ *for* $k = 1, 2, 3$.

Remark 2.4. This is the law of inertia, and, for charges that is the law of inductivity (as will become clear below). And, because ∇j_0 is (relativistically) a force, caused by the 4^{th} component j_0, which is internal to the system, we have to include it into the system. With it, (even under relativistic conditions) the sum of all forces in an adiabatic system adds up to zero. (In the non-relativistic limit ∇j_0 goes to zero, and the above converges to Newton's 3^{rd} law: actio equals reactio. (As a cross-check, note that if the matrix Dj was symmetric, we would get $\nabla j_0 - \partial_0 \mathbf{j} = 0$, which clearly gets the sign wrong.)

We can now proceed with the proof of the theorem:
Because $g_{\mu\nu} = -g_{\nu\mu}$ for $0 \leq \mu \neq \nu \leq 3$, $(g_{\mu\nu}\gamma_\mu\gamma_\nu)_{0\leq\mu,\nu\leq 3}$ is a symmetric matrix. So, substituting $x = (x_0, \ldots, x_3) \to \not{y} = (y_0\gamma_0, \ldots, y_3\gamma_3)$,

$$\not{j}(y) := (j_0(\gamma_0 y_0, \ldots, \gamma_3 y_3)\gamma_0, \ldots, j_3(\gamma_0 y_0, \ldots, \gamma_3 y_3)\gamma_3)$$

has a symmetric derivative matrix, where the derivative is taken w.r.t. y, hence again Poincaré's lemma applies, so there is a function $\Phi(\not{y})$, such that $\nabla\Phi := (\partial/\partial y_0, \cdots, \partial/\partial y_3)\Phi(\not{y}) = \not{j}(\not{y})$. In other words: $\not{j}$ is integrable to Φ w.r.t. the differential form $d\omega := \gamma_0 dy_0 + \cdots + \gamma_3 dy_3$.

This proves the theorem's first statement. And, inserting this equation into the adiabaticity condition, we get $\Box\Phi(\not{x}) = 0$, which proves the second statement.

To prove the third statement, we choose a fixed $a = (a_0, \cdots, a_3) \in \mathbb{R}^4$ and define

$$\not{A}(\not{x}) := \int_{a_0\gamma_0}^{x_0\gamma_0} \Phi(y_0\gamma_0, x_1\gamma_1, \ldots, x_3\gamma_3) dy_0\gamma_0 + \cdots + \int_{a_3\gamma_3}^{x_3\gamma_3} \Phi(x_0\gamma_0, \cdots, x_2\gamma_2, y_3\gamma_3) dy_3\gamma_3.$$

Then $\not{A} = (A_0\gamma_0, \ldots, A_3\gamma_3)$ is a spinor-valued 4-vector, and we get a (spinor-valued) 4-vector field $A = \gamma_0 A_0 + \cdots + \gamma_3 A_3$, for which

$$(\gamma_0\partial_0 + \cdots + \gamma_3\partial_3)^2 (A_0, \ldots, A_3) = (j_0, \ldots, j_3)$$

holds. $\Box$

Remark 2.5. The above proof's strategy is straightforward: By replacement of $dx = \sum_\mu dx_\mu$ with $d\omega := \gamma_0 dx_0 + \gamma_1 dx_1 + \gamma_2 dx_2 + \gamma_3 dx_3$, the external derivative of a scalar function f becomes the 1-form $d\omega f = \sum_\mu \partial_\mu f \gamma_\mu dx_\mu$, a 1-form then is generally defined by $\omega f := \sum_\mu f_\mu \gamma_\mu dx_\mu$, where the f_μ are (continuously differentiable) scalar functions, and its external derivative then becomes the 2-form

$$d\omega f := \sum_{\mu,\nu} \partial_\mu f_\nu \gamma_\mu\gamma_\nu dx_\mu \wedge dx_\nu = \sum_{\mu<\nu} (\partial_\mu f_\nu + \partial_\nu f_\mu)\gamma_\mu\gamma_\nu dx_\mu \wedge dx_\nu,$$

which is zero, if and only if $\partial_\mu f_\nu = -\partial_\nu f_\mu$ for all $\mu \neq \nu$. With this, a differential k-form is said to be closed, if and only if its external derivative is zero, it is defined to be exact, if and only if it is the external derivative of a (k-1)-form, and Poincaré's lemma applies again.

Remark 2.6. The essence of the above proof is that, instead of bothering with curls in 4-dimensional space-time and non-integrable Euclidean vector fields, to bypass that by mapping j to the spinor-field $\not{j} = (j_0\gamma_0, \ldots, j_3\gamma_3)$, do the integration there, and after integration inversely map $\not{A} = (A_0\gamma_0, \ldots, A_3\gamma_3)$ into $A = (A_0, \ldots, A_3)$ (see below for details).

Remark 2.7. It should be noted that $\not{A}$ is unique up to the addition of the external derivative $dG = \sum_\mu \partial_\mu G(\not{x}) dx_\mu \gamma_\mu$ of a 0-form $G : \not{x} \mapsto G(\not{x}) \in \mathbb{R}$. And with this, the preferred choice of G is such that the resulting 4-vector field $A'(\not{x})$ becomes itself adiabatic, i.e. $\sum_\mu \frac{\partial A'(\not{x})}{\partial(x_\mu \gamma_\mu)} = \sum_\mu \partial_\mu A'(x) \equiv 0$.

3. Formulation in Terms of Functional Analyis of Hilbert Spaces

3.1. Preliminaries

For the following, some basic notions on Hibert spaces are needed which are assumed to be complex throughout (see [5], Ch.VI-VII, p. 182 ff.): An (unbounded linear) operator "on" a Hilbert space $\mathcal{H}$ is a linear mapping T of a subspace $D(T) \subset \mathcal{H}$ into $\mathcal{H}$. $D(T)$ is called domain of definition of T, T is said to be densely defined, if $D(T)$ is dense in $\mathcal{H}$, it is said to be bounded, if $D(T) = \mathcal{H}$, and it is called closed, if its graph, $\{(x, Tx)) \mid x \in D(T)\}$, is a closed subset of $\mathcal{H} \times \mathcal{H}$. A projection of $\mathcal{H}$ is defined as a bounded linear operator π on $\mathcal{H}$, such that $\pi = \pi^2$. Let $\Pi(\mathcal{H})$ denote the set of all projections of $\mathcal{H}$. Let $\mathcal{B}(\mathbb{R})$ be the Borel algebra of $\mathbb{R}$, which by itself is partially ordered. A spectral measure of $\mathcal{H}$ is a mapping $dE : \mathcal{B}(\mathbb{R}) \ni X \mapsto \int_X dE_\lambda := E(X) \in \Pi(\mathcal{H})$, such that $E(\mathbb{R}) = id_{\mathcal{H}}$ is the identity of $\mathcal{H}$ and such that for all Borel sets $X, Y \subset \mathbb{R}$: $E(X \cap Y) = E(X)E(Y)$ holds. With this, a selfadjoint operator on $\mathcal{H}$ can be defined as a densely defined and closed operator $T : D(T) \to \mathcal{H}$ for which a spectral measure dE_λ exists, such that $Tx = \int_{-\infty}^{\infty} \lambda dE_\lambda x$ for $x \in D(T)$. A densely defined operator that is uniquely extendable to a selfadjoint operator is called essentially selfadoint. Two selfadjoint operators are said to be commuting, if their spectral measures commute, and a complex combination of two commuting self-adjoint operators is said to be a normal operator.

Definition 3.1. A densely defined and closed operator $T : D(T) \to \mathcal{H}$ will be called **quasi-selfadjoint**, if there exists a finite dimensional subspace $X \subset \mathcal{H}$, a spectral measure dE_λ that commutes with the canonical projection $\pi : \mathcal{H} \to X$, and n inversions on X, $I_1, \ldots, I_n$, such that

$$T = \int_{-\infty}^{\infty} (\lambda_1 I_1 + \cdots + \lambda_n I_n) dE_{\lambda_1 + \cdots + \lambda_n} = \int_{\mathbb{R}^n} (\lambda_1 I_1 + \cdots + \lambda_n I_n) dE_{\lambda_1} \cdots dE_{\lambda_n}.$$

(An inversion on X is an automorphism for which its square is the identity id_X.) If the I_k are even allowed to be such that $I_k^2 = \pm id_X$, then T will be called **quasi-normal**.

Remark 3.2. A selfadjoint operator is quasi-selfadoint. Conversely, for $n = 1$, i.e. if only one inversion I is involved, a quasi-selfadjoint operator is self-adjoint. Moreover, a quasi-selfadjoint operator T, for which the n inversions all commute with eachother, is the sum of n commuting selfadjoint operators, hence selfadjoint, too.

3.2. The Pullback Topology

We exactly have that situation with relativistic operators Q, which are 4-vectors $(Q_0, \dots, Q_3)$, such that $Q_0^2 - \cdots - Q_3^2$ is preserved. Here, X is the 4-dimensional vector space $\mathbb{C}^4$, equipped with the Minkowski metrics $d : \mathbb{C}^4 \ni x \mapsto \bar{x}_0 x_0 - \cdots - \bar{x}_3 x_3 \in \mathbb{R}$, and $Q = \int_{\mathbb{R}^4} (x_0\gamma_0 + \cdots + x_3\gamma_3) dE_{x_0} \cdots dE_{x_3}$ then is a quasi-normal operator (supposed it is closed and densely defined).

But now we can do more: Because the γ_μ anti-commute, they are linearly independent, so $\Theta : \mathbb{R}^4 \ni x \mapsto \sum_\mu x_\mu \gamma_\mu \in \mathcal{M}$ is a vector space isomorphism of $\mathbb{R}^4$ onto $\mathcal{M}$.

Remark 3.3. To be precise, $\mathcal{M}$ is not a vector space over the field $\mathbb{R}$, but over the field $\mathbb{R} \cdot 1_4$, where 1_4 stands for the 4×4 unit matrix, that is: the field are the real multiples of 1_4, and an inner product on $\mathcal{M}$ will then map into that field.

We can now pull back from the Euclidean geometry by basing the Minkowski space on $x_0\gamma_0, \dots x_3\gamma_3$:

Θ extends naturally as an isomorphism $\Theta : \mathbb{C}^4 \ni x + iy \mapsto \Theta x + i\Theta y \in \mathcal{M}_\mathcal{C} := \mathcal{M} + i\mathcal{M}$. Let $L^2(\mathcal{M})$ be the space of all functions $f : \mathcal{M} \to \mathcal{M}_\mathbb{C}$ with $\Theta^{-1} f \Theta \in L^2(\mathbb{R}^4, \mathbb{C}^4)$. This defines an isomorphism ι from $L^2(\mathcal{M})$ onto $L^2(\mathbb{R}^4, \mathbb{C}^4)$, so that $\|f\|^2_{L^2(\mathcal{M})} := \|\iota f\|^2_{L^2(\mathbb{R}^4,\mathbb{C}^4)}$ makes $L^2(\mathcal{M})$ become a Hilbert space. Written in terms of $f = \sum_\mu f_\mu \gamma_\mu \in L^2(\mathcal{M})$:

$$\begin{aligned} \|f\|^2 &= \int (\sum_\mu \overline{f_\mu(x_0\gamma_0, \dots, x_3\gamma_3)} f_\mu(x_0\gamma_0, \dots, x_3\gamma_3)) 1_4 \gamma_0 \cdots \gamma_3 d^4x \\ &= \int \big(f(x_0\gamma_0, \dots, x_3\gamma_3)\big)^* f(x_0\gamma_0, \dots, x_3\gamma_3) \gamma_0 \cdots \gamma_3 d^4x. \end{aligned} \tag{3.1}$$

The isomorphism ι has the property to map matrices that are anti-symmetric in their off-diagonal elements into symmetric matrices and vice versa. Dj with its anti-symmetric off-diagonal elements might not be integrable within the Euclidean metric, but under ι^{-1} it is.

Also, the derived relation $\Box A = j$ becomes in the pulled-back Euclidean metrics $\Delta A = j$, which now just trivially states that j is the source of the vector field A.

The Dirac equation follows from this:

The operator $\not{\partial} := i\partial_0\gamma_0 - \cdots - i\partial_3\gamma_3$ with the Schwartz space of rapidly decreasing smooth functions on $\mathbb{R}^4$ chosen as domain of definition $D(\not{\partial})$ then makes it a densely defined, symmetric operator on $L^2(\mathcal{M})$, the Fourier transform, which is an isometric automorphism on $L^2(\mathcal{M})$, transforms it to its spectral resolution as a multiplication operator, the graph of which can be closed in $L^2(\mathcal{M})$, so $\not{\partial}$ is essentially self-adjoint. Let $\mathcal{D}$ be the Fourier inverse of all $f \in D(\not{\partial})$, such that $supp(f) \cap \{0\} = \emptyset$, i.e. those functions that vanish in an ϵ-environment of the origin. Then $\not{\partial}$ is invertible on $\mathcal{D}$, which itself is a dense subspace of $L^2(\mathcal{M})$. So, $\not{\partial}^{-1}$ is a densely defined symmetric operator. Then, trivially, $\not{\partial}\Phi = j$ for $\Phi = \not{\partial}^{-1} j$ with $j \in \mathcal{D}$, which can be rewritten into the eigenvalue equation $\not{\partial}\Phi = m\Phi$, which is Dirac's equation. (It means that, basically, up to phase symmetry, the quantum mechanical waves can be identified with classical action functions.)

4. Masses and Charges

The reason for not calling the adiabaticity condition by its common name "law of mass conservation" is that this condition is not only about mass, but of charge either: By integrating the action another time along each of the 4 components to a vector field $(A_0, \ldots, A_3)$, we saw that the A_μ obey Maxwell's covariant equations, $\Box A_\mu = j_\mu$. Now, one might suspect that these equations might not be a "real Maxwell electrodynamics" at all.
Just to prove that these relations really make a Maxwell theory, take the anti-symmetric part $(g_{\mu\nu})_{\mu\nu}$ of the Euclidean derivative Dj as in the proof of the theorem, and integrate each term $g_{\mu\nu}$) with the Green's function $G(x, y)$ (which inverts the wave operator $\Box$) as in [3, Vol. II, Ch. 21-3]. The result is an anti-symmetric matrix again, which is just the electrodynamical field tensor, made of electric and magnetic field components. So, there is no difference to Maxwell's theory.

There is more to say:
$\mathcal{M}$ is not just a vector space, but a vector space of mappings on another vector space, $\mathbb{C}^4$, which has been disregarded sofar. So, $\mathbb{C}^4$ is a degeneracy (or "defect") for $\mathcal{M}$, from which one can deliberately pick any vector $(\chi_1, \ldots, \chi_4) \in \mathbb{C}^4$. Now let $p := E\gamma_0 + \cdots + p_3\gamma_3$ be a non-zero energy-momentum from $\mathcal{M}$. Then $\gamma_5 := i\gamma_0 \cdots \gamma_3$ transforms p into $-p$, so that γ_5 is (equivalent to) the space-time reflection. But γ_5 has two (2-fold degenerate) eigenspaces $\Xi_\pm$ for the two eigenvalues ± 1. Therefore, according to whether $\chi \in \Xi_\pm$, either $\gamma_5 p = \mp p$.

So, if we identify mass with energy (which explains the name mass conservation), then there are two types of masses: one which retains its (positive) value under space-time inversion, and one which is positive and negative and is inverted under space-time inversions. Obviously, the first one is what one expects to be "the mass". Since masses are neutral composites of charged particles, this suggests the second type of mass to be the electric charge. So, γ_5 will be the charge inversion $\mathcal{C}$, and the adiabatic system is a neutral theory for $\chi \in \Xi_+$ and a charged one with $\chi \in \Xi_-$.

5. CPT

Because γ_0 is symmetric and anti-commutes with $\gamma_1, \ldots, \gamma_3$, it represents space-inversion, i.e. parity $\mathcal{P}$. Likewise, $\mathcal{T} := i\gamma_1\gamma_2\gamma_3$ represents the time-inversion. So, $\mathcal{C} = i\gamma_0 \cdots \gamma_3 = \mathcal{PT}$, the inversions $\mathcal{P}, \mathcal{C}, \mathcal{T}$ anti-commute, and, up to a factor ± 1 each of the three inversions is the product of the other two.

Let $\Pi_\pm$ be the eigenspaces of $\mathcal{P}$ for the eigenvalues ± 1. Then with $\chi \in \Pi_+$ the adiabatic system is called bosonic, and for $\chi \in \Pi_-$ it is called fermionic.

6. Forces: Interaction of Adiabatic Systems

The rationale behind the above $\mathcal{PCT}$-relation is that any pair of these discrete inversions resolves the 2-fold degeneracy of the eigenvalues ± 1, which each of the inversions has: Let's pick $\mathcal{C}$ and $\mathcal{P}$. The 2-dimensional eigenspaces $\Xi_\pm$ for $\mathcal{C}$ each split in 1-dimensional subspaces, which either preserve or invert parity $\mathcal{P}$; these are usually termed as spin-up/down states. So, the adiabatic system splits into combinations of charged/uncharged and spin-up/spin-down theories, which are conserved with time. And, assuming that the systems are parity-invariant, the four possible scaling parameters reduce to two: one for mass (the mechanical one), and one for charges (the electromagnetic one). Using the fine structure constant $e^2/(\hbar c)$, we can scale both, neutral and charged adiabatic systems in units of $\hbar$.

Till here, the A_μ obey the covariant Maxwell equations, i.e. $\Box A_\mu = j_\mu$. But it is known, that the non-covariant Maxwell equations, in terms of electric and magnetic field strengths **E** and **B** are invariant as to the transformation $A_\mu \mapsto A_\mu + \partial_\mu F$, where F is an arbitrary scalar and smooth function in spacetime $\mathbb{R}^4$. This is called the gauge invariance of the Maxwell equations.

Clearly, by substituting $G(\not{x}) := F(x)$, dG becomes an arbitrary additive integration constant for $A(\not{x})$, for which $\Box G(\not{x}) \equiv 0$ holds, and we can choose G such that $\sum_\mu \partial_\mu(\partial_\mu(F(x) + A_\mu(x)) \equiv 0$. So, the particular choice of G decides over whether A itself is adiabatic, i.e. obeys $\sum_\mu \partial_\mu A_\mu \equiv 0$, or not. In electrodynamics, this condition is known by the name "Lorentz gauge".

It is nothing but superficial, to accept this gauge symmetry as an abstract, mathematical truth. But it will be substantial to ask, what this symmetry is physically about:
When we think of j as a 4-vector of charge densities, then it is tempting to think of the addition dG to $\sum_\mu A_\mu dx_\mu \gamma_\mu$ of neutral masses being added. But then the addition of neutral mass to the charged system would break the adiabiticity of the A_μ, and the gravitational field will not be adiabatic itself. So we exclude that hypothesis.
What we know is that neither $A(x)$ nor $A(\not{x})$ are path integrable to a scalar field (unless $j \equiv 0$), however G or F are scalar fields: the latter define phase transformations $\psi(\not{x}) \mapsto e^{iG(\not{x})}\psi(\not{x})$, so capture a local phase symmetry $U(1)$, and according to the phase chosen at each $\not{x}$, the fields become adiabatic or not.

Hence, unless contrary to all physical experience and evidence, both gravitational and electrodynamic fields are non-adiabatic (in case of which closed subsystems would not exist at all), neither gravitational nor electrodynamical fields are scalar gauge fields, based on $U(1)$.
Then, where are the neutral masses and the gravitational fields?
The masses must be contained in the neutral part of the above described

adiabatic system, i.e. that subspace of $\mathbb{C}^4$-valued functions $f : x \mapsto f(x)$, for which $f(x) = \mathcal{C}f(x) = \gamma_5 f(x)$ for all $x = (x_0, \ldots, x_3)$. The complementary subspace is the space of all f, for which $\mathcal{C}f = -f$, the subspace of charged particles. As a result, charged and neutral adiabatic systems are complementary direct sums, independent from eachother.

And then the question is: Given any adiabatic system of charged particles, electrons, say, which part of its rest energy stems from charges, and which part comes from electrodynamically inactive, neutral masses?
Of course, by measuring the weight of electrons and protons, etc., it is well-known that this ratio is specific and constant for each type of particles, so that it is just a spontaneously broken symmetry.

In all above, we derived that the group $U(1)$ is not capable to deliver neither a gauge theory for electromagnetism nor gravitation: if it was, the vector field $A = (A)_\mu$ could be expressed as $A_\mu = \partial_\mu F$ for some complex-valued function F, which we know is wrong.
This in particular, falsifies the standard model's claim of electrodynamics to be a $U(1)$-gauge theory.

To be non-destructive, let's now ask, what that $U(1)$ gauge symmetry as phase symmetry is physically good for:
To do so, let's consider the minimal requirements for electrodynamics to become a gauge theory: What we know is that a unitary gauge theory for electrodynamics minimally demands the the anti-symmetry of the matrix Dj. (Note that the anti-symmetric 4×4-matrices span a 4-dimensional vector space, whereas the complementary vector space of symmetric matrices is of dimension 12.) So, for a gauge group to capture electrodynamics, we minimally need 4 dimensions, and that group is to be $U(2) \equiv U(1) \times SU(2)$! Now, note that in that minimal $U(2)$ theory, the $SU(2)$ subgroup captures the notion of opposite charges, and there is just one non-trivial inversion defined in $U(2)$, which is either charge or time/energy/mass inversion. Classical physics would be considered to be incomplete without time and charge inversion (symmetries) to be both in place. So, the minimalistic gauge theory will have to be built upon $U(1) \times U(2)$, and then it will capture gravitation and electrodynamics altogether: The subspace of $\Xi_- \subset \mathbb{C}^2$ of eigenvectors of the charge inversion $\mathcal{C}$ for the eigenvalue -1 gives the electrodynamics, and the restriction on the complementary subspace Ξ_+ of eigenvectors of $\mathcal{C}$ for the eigenvalue $+1$ defines an independent gravitational gauge theory.

7. Adding mass to the Standard Model

The Standard Model states $SU(3) \times SU(2) \times U(1)$ as the fundamental symmetry group. In it, $SU(3)$ captures the symmetry of the theory of strong force and $SU(2) \times U(1)$ the symmetry of the electro-weak theory, a.k.a. Salam-Weinberg theory, in which $U(1)$ captures the electromagnetic charge symmetry. As we saw above, it is not containing electrodynamics, unless the standard model group is extended to $SU(3) \times SU(2) \times U(1) \times SU(2) \times U(1)$, which happens to be of dimension 16, and isomorphic to $U(4)$!

Now, let me come back to the Dirac spinors of section 4:
In there it was shown that we have quadruples $(\chi_1, \dots, \chi_4) \in \mathbb{C}^4$ at our free disposal, on which the Dirac matrices $\gamma_0, \dots, \gamma_3$ operate, and that these quadruples allow to determine whether the quadruple is invariant as to charge inversion and parity. So, these quadruples are states that track charge and parity. And because the norm of these vectors already goes into the scalar functions j_μ or A_μ, we can make them unit vectors, that is: members of a 4-dimensional complex unit ball. Next, we expect an adiabatic system to be globally symmetric as to space, time, and charge inversion. Then it follows that all unit vectors $\chi_1, \dots, \chi_4$ from the 4-dimensional unit ball are in symmetry, which makes the symmetry group of these unit vectors become $U(4)$. And it is not by accident that this group coincides with our extension of the symmetry group of the Standard Model: Even though gluons and some leptons have positive masses which confine the reach of their forces, their composites must show up as even bigger masses to the outside, then taking their share in the macroscopic world of gravitation.

8. Outlook

The above exclusively dealt with adiabatic systems. These are closed sytems, free of exterior forces. Therefore, all (internal) forces add up to zero (see the above corollary 2.3 and its subsequent remark). This is what allowed the calculation of the action function Φ. But, perhaps surprisingly, it turned out to be a plane wave ($\Box\Phi = 0$). So, in the absence of external forces, it spreads freely at the speed of light, and because it is sourceless, it cannot interact with the source itself, and may only interact with a target that it hits. A non-trivial interaction will cause a change of the target sources, which means that the target's action function will spread that change sourcelessly over to the original source, and likewise causes a change of action there. And as a result of iterations, one would get a superposition of action functions, which is just Feynman's ingenious idea of path integration. One can then identify Φ with a field of virtual photons that travel along the path of extremal action (as Feynman did). But then, the photons will not have any impact on their sources, which does conflict with Einstein's conception of photons (see [2]). Einstein's conception of photons as real particles interacting with its source as they leave it, raises essential problems: The adiabatic system above will leak energy, because the photons carry away energy. It needs an infinite bare mass/energy distinct from the observed charge/mass to stabilize the observed masses, which otherwise would unstably resolve, leading to small-scale divergencies to be overcome, etc... Many of these problems have been solved during the last century through renormalization.

However, whatever the final successful calculation will be, the result must yield an adiabatic system of particles of observed charge/mass with the very same stable energy momentum as in a theory with zero interaction of field with its source. At best, a theory built on the assumption of non-zero interaction between field and source will therefore result in a complicated calculation of zero with additional parameters and constants to be determined.

A century ago, the vast majority of phyicists would keep with the simplicity. Current physics holds (for good reasons) that simplicity might not lead to truth.

So, the ultimate question is: Is there a way to truely determine whether the interaction of an electromagnetic field with its source is zero or non-zero? And there is:

To its answer, I propose a simple experiment:
It needs a large container filled with cool gas of some well-known total rest energy m and to inject into it (slowly) cool electrons and positrons of equal rest energy m_1 from opposite sides. Annihilation processes will set in, and what is to detect is whether the system's total rest mass after annihilation has dropped to m or lower, or whether it is approximately $m+2m_1$ as it was before annihilation. This experiment has never been carried out.

Appendix

A. Generalization (Straightforward)

In the form given so far, the state vectors $\chi = (\chi_1, \dots, \chi_4) \in \mathbb{C}^4$ do not depend on spacetime. The equations can however be written more generally by replacing the scalar components j_μ with 4-tuples $j_\mu(x) := (j_{\mu,1}, \dots, j_{\mu,4})$, on which the Dirac matrices γ_ν operate from the left. Integration w.r.t. $\sum \gamma_\mu dx_\mu$ must then be done from the left, in accordance with the left sided differentiation. There won't be any change or twist otherwise, as the change results into four adiabatic scalar systems, which superimpose, each one added within its own component.

Note that by this inclusion of the state vectors $(\chi_1, \dots, \chi_4)$ into the $\gamma_\mu j_\mu$, the resulting quadruple components $j_{\mu,k}$, $k = 1, \dots, 4$, become complex, phase symmetric vectors, and their action integral Φ, then turn into a complex, phase symmetric vector of four components of action functions. And if it was irrelevant from where the components came from and one could disregard the norm on $\mathbb{C}^4$, then one could pick any one of the complex, phase symmetric function components.
Keep in mind: Complexity and phase symmetry are exclusively caused by the phase symmetric states, represented by the unit vectors $(\chi_1, \dots, \chi_4)$.

B. A Glimpse Beyond Adiabaticity

Because of $\mathcal{C} = \mathcal{PT}$, there is hope that electromagnetic and gravitational interaction might be can be written within a single equation:
If the interaction $V(j', j)$ of two spatially separated adiabatic systems j' and j is to make sense, then only if the action of j' on j is the reaction of j

on j' and vice versa. Now, if that is to be $\mathcal{T}$-symmetric under relativistic conditions, then $V(j',j)$ must be rewritable into $V(j',j) \to U(\mathcal{T}j',j)$, where U is symmetric: why? - because an action of j on j' needs time to cause a reaction on j', and upon time-inversion action and reaction switch meaning, as the former reaction becomes the source of action on the former action. So $U(j',\mathcal{C}j) = -U(j',j)$ does not enforce $U(\mathcal{T}j',j)$ to vanish for neutral j or j', as long as $U(j',\mathcal{P}j) = U(j',j)$ is valid. With j and j' being either quadruples of or scalar phase-symmetric complex functions, one may transform $\gamma_\mu \mapsto i\gamma_\mu$. Then complex conjugation does invert the $i\gamma_0$ but keeps the $i\gamma_k$ invariant, and w.r.t. that choice, therefore the time inversion $\mathcal{T}$ becomes the complex conjugation.

That said,

$$V(j',j)(x) := (Const) \int \frac{\overline{\hat{j'}_\mu(x')}\hat{j}^\mu(x)}{(x_0 - x_0')^2 - \cdots - (x_3 - x_3')^2} d^4x'$$

with $Const$ to be chosen according to whether j and j' are both either charged or neutral (i.e. map to either Ξ_- or $\Xi+$), where $\hat{}: j \mapsto \hat{j}$ denotes the Fourier transform, becomes interesting.

References

[1] H. Cartan, *Differential Forms* Herman Kershaw, 1971.

[2] A. Einstein, *Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?*, Annalen der Physik. **18**:639, 1905.

[3] R. P. Feynman, *Lectures on Physics*, Vol. I-III, Addison Wesley, 1977.

[4] H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics...*, World Scientific Publishing Co., 2009.

[5] M. Reed, B. Simon,*Methods of Modern Mathematical Physics*, Vol. I, Academic Press, 1980.

[6] Wikipedia, *Harmonic function*, `https://en.m.wikipedia.org/wiki/Harmonic_function`, 2016.

Hans Detlef Hüttenbach

e-mail: `detlef.huettenbach@computacenter.com`