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 exists. 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, which is a strong form of the PCT-theorem; we show that - in order to capture the concept of mass - the standard model gauge group has to be augmented by a factor group $U(2)$, such that the complete gauge group becomes $U(4)$.

It is shown that the sourceless action field in itself suffices to describe the long ranged interaction of matter, both, electromagnetic and gravitational.

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*) $\sum_{0 \leq \mu \leq 3} \partial_\mu j_\mu \equiv 0$, where $\partial_\mu := \partial / \partial x^\mu$.

Remark 1.2. There is no sense in demanding $j_0 \geq 0$, 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 $\not{j}(x_0, \ldots, x_3) := \sum_\mu j_\mu(x) \gamma_\mu$ is integrable w.r.t. the differential form $d\omega := \gamma_0 dx_0 + \gamma_1 dx_1 + \gamma_2 dx_2 + \gamma_3 dx_3$.*
2. *The action function $\Phi := \int \not{j} d\omega$ of the 4-vector field j 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. *The (Euclidean) derivative* $Dj := (j)_{\mu\nu} = (\partial_\mu j_\nu)_{0\le\mu,\nu\le 3}$ *of an adiabatic system* $j = (j_0, \dots, j_3)$ *is anti-commuting for all its off-diagonal elements, i.e.:* $(Dj)_{\mu\nu} = -(Dj)_{\nu\mu}$ *for* $0 \le \mu \ne \nu \le 3$.

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 $(f)_{\mu\nu}$ with zero diagonal elements, i.e.: $f_{\mu\nu} := \frac{1}{2}(\partial_\mu j_\nu + \partial_\nu j_\mu)$ for $\mu \ne \nu$ and $f_{00} = \cdots f_{33} = 0$, and a matrix $(g)_{\mu\nu} := (j)_{\mu\nu} - (f)_{\mu\nu}$, which is anti-symmetric in its off-diagonal elements.

It remains to prove that $(f)_{\mu\nu} = 0$ for all $0 \le \mu, \nu \le 3$:
$(f)_{\mu\nu}$ defines a 2-form $\omega = \sum_{\mu,\nu} f_{\mu\nu} dx_\mu \wedge dx_\nu$, which rewrites into $\omega = \sum_{0\le\mu\ne\nu\le 3}(f_{\mu\nu} - f_{\nu\mu}) dx_\mu \wedge dx_\nu \equiv 0$ 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 $(f_0, \dots, f_3)$. 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 $(f_0, \dots, f_3)$ is path integrable to a function F, say. Because $f_{00} = \cdots = f_{33} = 0$, 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 consists in the vector space of all linear mappings on $\mathbb{R}^4$, so $f = \nabla F$ is a quadrupel of constant functions, and therefore its derivative vanishes, i.e.: $(f)_{\mu\nu} \equiv 0$. □

An immediate consequence 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).

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

$$\not{f}(y) := (f_0(\gamma_0 y_0, \dots, \gamma_3 y_3)\gamma_0, \dots, f_3(\gamma_0 y_0, \dots, \gamma_3 y_3)\gamma_3)$$

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 $\nabla\Phi := (\partial/\partial y_0, \cdots, \partial/\partial y_3)\Phi(\not{y}) = \not{f}(\not{y})$. In other words: $\not{f}$ 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

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

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. □

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

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

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. Note that Φ might generally not be a scalar function, but rather the sum $\Phi = \Phi_0 + \sum_{0 \leq \mu < \nu \leq 3} \Phi_{\mu\nu} \gamma_\mu \gamma_\nu$, where Φ_0 and the $\Phi_{\mu\nu}$ are scalar functions: That depends on the choice of the space-time coordinates:
A function $\phi : x \mapsto \phi(x)$ can be deliberately rewritten as a function $\phi : \not{x} \mapsto \phi(\not{x})$ on Dirac spinors $\not{x} := \sum_\mu \gamma_\mu x_\mu$. Frankly, then the spinor differential $d\omega$ would have to be replaced with the Euclidean differential dx again, and care must be taken as to the the order of right or left multiplication with the non-commuting factors γ_μ.

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, $I_1, \dots, I_n$, 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 $f = \sum_\mu f_\mu \gamma_\mu \in L^2(\mathcal{M})$:

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

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 $\partial\!\!\!/ := i\partial_0\gamma_0 - \cdots - i\partial_3\gamma_3$ with the Schwartz space of rapidly decreasing smooth functions on $\mathbb{R}^4$ chosen as domain of definition $D(\partial\!\!\!/)$ 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 $\partial\!\!\!/$ is essentially self-adjoint. Let $\mathcal{D}$ be the Fourier inverse of all $f \in D(\partial\!\!\!/)$, such that $supp(f) \cap \{0\} = \emptyset$, i.e. those functions that vanish in an ϵ-environment of the origin. Then $\partial\!\!\!/$ is invertible on $\mathcal{D}$, which itself is a dense subspace of $L^2(\mathcal{M})$. So, $\partial\!\!\!/^{-1}$ is a densely defined symmetric operator. Then, trivially, $\partial\!\!\!/\Phi = j$ for $\Phi = \partial\!\!\!/^{-1} j$ with $j \in \mathcal{D}$, which can be rewritten into the eigenvalue equation $\partial\!\!\!/\Phi = m\Phi$, which is Dirac's equation. (It means that 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 the following:

$\mathcal{M}$ 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 $(\chi_1, \dots, \chi_4) \in \mathbb{C}^4$. 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 eigenspeces $\Xi_\pm$ 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 $\hbar$.
The problem now is: How do two adiabatic systems themselves interact (to first order)?

Definition 6.1. As pointed out in Remark 2.7, the action Φ splits into the sum of a scalar action Φ_0 and a spinor-valued action $\sum_{0\le\mu<\nu\le 3} \Phi_{\mu\nu}\gamma_\mu\gamma_\nu$. Φ_0 is called **mechanical action**, and the spinor-valued part is the electromagnetic action. Similarly, the 4-vector potential A is the sum of $(F_0\gamma_0, \ldots, F_3\gamma_3)$, called **mechanical energy-momentum**, and $(iG_0\gamma_5\gamma_0, \ldots, iG_3\gamma_5\gamma_3)$, which will be called **electromagnetic energy-momentum**. Let $j = (j_0\gamma_0, \ldots, j_3\gamma_3)$ and

$j' = (j'_0\gamma_0, \ldots, j'_3\gamma_3)$ be two adiabatical systems, and let

$$A := (iG_0\gamma_5\gamma_0, \ldots, iG_3\gamma_5\gamma_3)$$

be the electromagnetic energy-potential of j. By defining

$$(\epsilon_0 j'_0 A_0, (1/\mu_0) j'_1 A_1, \ldots, (1/\mu_0) j'_3 A_3)$$

to be the interacting **electromagnetic 4-potential** of j and j', where ϵ_0 is the electric permitivity and μ_0 is the magnetic susceptibility, Maxwell's equations follow: this is, because the electromagnetic action $\sum_{0\leq\mu<\nu\leq 3}\Phi_{\mu\nu}\gamma_\mu\gamma_\nu = \sum_{0\leq\mu<\nu\leq 3}((1/2)\Phi_{\mu\nu}\gamma_\mu\gamma_\nu - (1/2)\Phi_{\mu\nu}\gamma_\nu\gamma_\mu) = \sum_{0\leq\mu,\nu\leq 3}(1/2)\Phi_{\mu\nu}\gamma_\mu\gamma_\nu$ with an anti-symmetric matrix $((1/2)\Phi_{\mu\nu})_{0\leq\mu,\nu\leq 3}$, which is the electromagnetic field tensor (up to the two electric and magnetic constants ϵ_0 and μ_0). Preliminating the independence of the interaction of two neutral, mechanical adiabatic systems from the particles' velocity, the simplest guess for that interaction then is that for a constant $g_0 \in \mathbb{R}$ the **interacting energy for two mechanical/neutral adiabatic systems** j and j' will be given by $g_0 j'_0 A_0$.

Remark 6.2. Because both, electromagnetic and mechanical interaction, are proportional to the charges/masses of the two interacting systems, the interaction of an electromagnetic system with a mechanical system is zero. This is exactly what is needed in order that the Lorentz gauge works in electromagnetics:
Let there be a positron at rest at some space-time instance $x \in \mathbb{R}^4$. Then its charge q is proportional to its (rest) energy E_0. Seen from another system with constant speed v w.r.t. the one at rest, that particle now has the energy $E' = E_0(1-\beta^2)^{-1/2}$, although its observed charge still is q. The covariance of the Maxwell equations hence demands that the increase of energy $E' - E_0$ must be an electromagnetic invariant. It is energy though, due to the laws of Lorentz covariance.

In particular, j decomposes into the sum of $j_n := \Box F$ and the complementary $j_e := j - j_n = i\Box(G_0\gamma_5, \ldots, G_3\gamma_3)$, where j_n is the neutral, electromagnetically invariant part, and j_e are sources of an electromagnetic field. Because of its optical inactivity, j_n qualifies as **dark matter**.
By the above remark, both j_n and j_e are no 4-vectors, unless the kinetical energy of the charges is permitted to be added to j_e, instead of to j_n. However, j itself is a 4-vector, so $j_0^2 - \cdots - j_3^2$ must be the relativistically invariant square of the total rest energy (or mass). And now the question is, given any particle system, which part of its rest energy is electromagnetically inactive or dark, and which part is electromagnetically active? Because in neither classical electromagnetism nor gravitational theory there is any restriction as to this, this is a local symmetry. In fact, this has to be a (local) $U(2)$-symmetry:

7. Gauge Symmetry

It may be rewarding to view the choice of the vectors $\chi \in \mathbb{C}^4$ as a gauge and turn the above into a gauge theory. That way, assuming the same scale

of units in the four components, one gets the group $U(4)$ of unitary 4×4 matrices as a gauge symmetry:
For each point $x \in \mathbb{R}^4$ in space-time and each unit vector $\chi \in \mathbb{C}^4$, $(\gamma_0 j_0(x) + \cdots + \gamma_3 j_3(x))\chi$ is just one root of $j_0^2(x) - \cdots - j_3^2(x)$, but from this we can get all the other roots $(\gamma_0 j_0(x) + \cdots + \gamma_3 j_3(x))U(x)\chi$ with $U(x) \in U(4)$ (for each x)!
$U(4)$ is a connected and reducible group, and can be factored into the product $SU(3) \times U(2) \times U(2)$, where $U(2)$ itself is isomorphic to $SU(2) \times U(1)$. So, that gauge symmetry is a supergroup of the standard model group $SU(3) \times SU(2) \times U(1)$: $U(4) = (SU(3) \times SU(2) \times U(1)) \times (SU(2) \times U(1))$.
We can now identify the extra $U(2)$ group with the above (local) symmetry group of electromagnetically active/inactive rest energy:
Let's consider a fixed point $x \in \mathbb{R}$ in space-time. Then $j(x)$ is purely charged at x, i.e. contains no electromagnetically inactive rest mass $m_0(x)$ (in the particular reference frame), if and only if its derivative $Dj(x)$ is anti-symmetric. Now let $j'(x)$ be purely neutral, i.e.: $Dj'(x)$ is a symmetric matrix at x. Let $j_0^2(x) - \cdots j_3^2(x) = j'(x)_0^2 - \cdots - j_3^2(x) \neq 0$. Then $\Theta(\lambda_1, \lambda_2) : j(x) \mapsto \lambda_1 j(x) + \lambda_2 j'(x)$ defines this $U(2)$-symmetry for $\lambda_1, \lambda_2 \in \mathbb{C}$ with $|\lambda_1^2| + |\lambda_2^2| = 1$.

Again, this group $U(2)$ is isomorphic to the factor group $SU(2) \times U(1)$, in which $U(1)$ takes on the role of symmetry of mass/energy inversion. We have this symmetry, because all dynamical laws are invariant w.r.t the to the multiplication of energy with a global phase factor $e^{i\lambda}$.
Differently put: it is this symmetry which guarantees that we can think of masses to be all positive.

Remark 7.1. Note that with $U(4)$, a super group of standard model, we even more need some method of spontaneous symmetry breaking, but this time not only to explain, why particles have a mass at all, but now, also to explain, why electrons always have the same, particular charge/mass ratio, the so called fine structure constant $\frac{e^2}{\hbar c}$, and how that relates to the mass ratio of positrons and protons.

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. This is what enforces the action function Φ to be a plane wave ($\Box\Phi = 0$). We can identify Φ with a field of virtual photons. 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 of 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 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.

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.

Hans Detlef Hüttenbach
e-mail: detlef.huettenbach@computacenter.com