Modave Schools and Workshops

Volume 2

Second Modave Summer School in Mathematical Physics

August 6th-12th, 2006

PROCEEDINGS

International Solvay Institutes

Modave Schools and Workshops

Volume 1 - First Modave Summer School
in Mathematical Physics,
June 19th-25th, 2005

Volume 2 - Second Modave Summer School
in Mathematical Physics,
August 6th-12th, 2006

Foreword

After the success of the first Modave Summer School in Mathematical Physics, a second edition was organized by a group of young physicists from the Service de Physique Théorique et Mathématique of the ULB, the Theoretical Particle Physics Group of the VUB and the Service de Mécanique et Gravitation of the UMH. The school took place in the center "les Cent Fontaines" in the small Belgian town of Modave as for the first edition. The event benefited from the support of the International Solvay Institutes.

In modern theoretical physics, a very strong mathematical background is required for the understanding of many profound physical issues. Unfortunately, in usual schools and workshops addressed to young physicists these mathematical tools are not covered in detail. The Modave philosophy is to provide the beginners in the field with these tools necessary to get closer to the recent developments in physics.

Another aim of the school is to promote the exchange of knowledge between the participants. In this respect, the Modave lectures are given by the young researchers themselves, in the domain where they are becoming experts. The geographical situation of Modave, far from everything, and the countryside atmosphere highly favour informal discussions and interactions among participants.

The summer school consisted of about 5 hours of lectures a day, during the morning and the late afternoon. This schedule encouraged the participants to interact during the afternoon in informal discussions. Various social events were proposed during the evening.

Some of the lecturers made their contributions available for the international community of physicists by publishing an extended electronic version of their talk on the Internet. We hope that the Modave lectures will be helpful for young researchers in theoretical physics.

We would like to thank all lecturers and participants for their contribution to the success of the school. Thanks also to the International Solvay Institutes to have made copies of the proceedings available for the lecturers in printed format.

The organizers,

N. Bouatta, S. Cnockaert, G. Compère,
S. de Buyl, S. Detournay,
V. Mathieu, A. Wijns.

Local Organizing Committee

N. Bouatta, S. Cnockaert, G. Compère,
S. de Buyl, S. Detournay,
V. Mathieu, A. Wijns.

List of Participants

Joke Adam (KUL),
Paolo Aschieri (München, Germany)
Xavier Bekaert (IHES),
Nazim Bouatta (ULB),
Nicolas Boulanger (UMH),
Laurent Claessens (UCL),
Sandrine Cnockaert (ULB),
Andres Collinucci (KUL),
Geoffrey Compère (ULB),
Thomas Creutzig (DESY, Hamburg, Germany),
Sophie de Buyl (ULB),
Stéphane Detournay (UMH),
Jarah Evslin (ULB),
Wen Jiang (Oxford),
Euihun Joung (Paris7, France),
Stanislav Kuperstein (ULB),
Serge Leclercq (UMH),
Mauricio Leston (ULB),
Carlo Maccaferri (ULB),
Vincent Mathieu (UMH),
Jakob Palmkvist (Albert Einstein Institute, Golm, Germany)
Daniel Persson (ULB),
Christoffer Petersson (Chalmers U., Gothenburg),
Jan Rosseel (KUL),
Paul Smyth (KUL),
Nassiba Tabti (ULB),
Dieter Van den Bleeken (KUL),
Vincent Wens (ULB),
Alexander Wijns (VUB).

Contents

The unitary representations of the Poincaré group in any spacetime dimension

Xavier Bekaert[a,1] and Nicolas Boulanger[b,2]

[a] Laboratoire de Mathématiques et Physique Théorique,
Unité Mixte de Recherche 6083 du CNRS, Fédération Denis Poisson,
Université François Rabelais,
Parc de Grandmount, 37200 Tours (France)

[b] Service de Mécanique et Gravitation,
Université de Mons-Hainaut, Académie Wallonie-Bruxelles,
6 avenue du Champ de Mars, 7000 Mons (Belgium)

Abstract. An extensive group-theoretical treatment of linear relativistic wave equations on Minkowski spacetime of arbitrary dimension $D \geqslant 3$ is presented in these lecture notes. To start with, the one-to-one correspondence between linear relativistic wave equations and unitary representations of the isometry group is reviewed. In turn, the method of induced representations reduces the problem of classifying the representations of the Poincaré group $ISO(D-1,1)^{\uparrow}$ to the classication of the representations of the stability subgroups only. Therefore, an exhaustive treatment of the two most important classes of unitary irreducible representations, corresponding to massive and massless particles (the latter class decomposing in turn into the "helicity" and the "infinite-spin" representations) may be performed via the well-known representation theory of the orthogonal groups $O(n)$ (with $D-3 \leqslant n \leqslant D-1$). Finally, covariant wave equations are given for each unitary irreducible representation of the Poincaré group with non-negative mass-squared. Tachyonic representations are also examined. All these steps are covered in many details and with examples. The present notes also include a self-contained review of the unitary representation theory of the general linear and (in)homogeneous orthogonal groups in terms of Young diagrams.

[1]E-mail address: `xavier.bekaert@lmpt.univ-tours.fr`.
[2]E-mail address: `nicolas.boulanger@umh.ac.be`; Chargé de Recherches FNRS (Belgium).

1 Group-theoretical preliminaries

Elementary knowledge of the theory of Lie groups and their representations is assumed (see *e.g.* the textbooks [1, 2] or the lecture notes [3]). The basic definitions of the Lorentz and Poincaré groups together with some general facts on the theory of unitary representations are reviewed in order to fix the notation and settle down the prerequisites.

1.1 Universal covering of the Lorentz group

The group of linear homogeneous transformations $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ ($\mu,\nu = 0,1,\dots,D-1$) preserving the Minkowski metric $\eta_{\mu\nu}$ of "mostly plus" signature $(-,+,\dots,+)$,

$$\Lambda^T \eta \Lambda = \eta\,,$$

where Λ^T denotes the matrix transpose of Λ, is called the *Lorentz group* $O(D-1,1)$.

A massless particle propagates on the light-cone $x^2 = 0$. Without loss of generality, one may consider that its momentum points along the $(D-1)$th spatial direction. Then it turns out to be convenient to make use of the *light-cone coordinates*

$$x^\pm = \frac{1}{\sqrt{2}}\,(\,x^{D-1} \pm x^0\,) \quad \text{and} \quad x^m \quad (m = 1,\dots,D-2)\,,$$

where the Minkowski metric reads $\eta_{++} = 0 = \eta_{--}$, $\eta_{+-} = 1 = \eta_{-+}$ and $\eta_{mn} = \delta_{mn}$ ($m,n = 1,\dots,D-2$).

On physical grounds, one will mainly be interested in the matrices Λ's with determinant $+1$ and such that $\Lambda^0{}_0 \geqslant 0$. It can be shown that such matrices Λ's also form a group that one calls the *proper orthochronous Lorentz group* denoted by $SO(D-1,1)^\uparrow$. It is connected to the identity, but not *simply connected*, that is to say, there exist loops in the group manifold $SO(D-1,1)^\uparrow$ which are not continuously contractible to a point. In order to study the representations (reps) of $SO(D-1,1)^\uparrow$, one may first determine its universal covering group, *i.e.* the Lie group which is simply connected and whose Lie algebra is isomorphic to $\mathfrak{so}(D-1,1)$, the Lie algebra of $SO(D-1,1)^\uparrow$. For $D \geqslant 4$, the universal covering group, denoted $Spin(D-1,1)$, is the double cover of $SO(D-1,1)^\uparrow$. The spin groups $Spin(D-1,1)$ have no intrinsically projective representations. Therefore, a single (or double) valued "representation" of $SO(D-1,1)^\uparrow$ is meant to be a genuine representation of $Spin(D-1,1)$.

Warning: The double cover of $SO(2,1)^\uparrow$ is the group $SU(1,1)$, in close analogy to the fact that the double cover of $SO(3)$ is $SU(2)$. The group $SU(2)$ is also the universal covering group of $SO(3)$, but beware that the universal cover of $SO(2,1)^\uparrow$ is actually $\mathbb{R}^3$, covering $SO(2,1)^\uparrow$ infinitely often. Thus one may not speak of the spin group for the "degenerate case" of the proper orthochronous Lorentz group in spacetime dimension three. The analogue is that the universal cover of $SO(2) \cong U(1)$ is $\mathbb{R}$ covering it infinitely often, so that one may not speak of the spin group for the "degenerate case" of the rotation group in two spatial dimensions.

1.2 The Poincaré group and algebra

The transformations

$$x'^\mu = \Lambda^\mu{}_\nu x^\mu + a^\mu$$

where a is a spacetime translation vector, form the group of all inhomogeneous Lorentz transformations. If one denotes a general transformation by (Λ, a), the multiplication law in the Poincaré group is given by

$$(\Lambda_2, a_2)\cdot(\Lambda_1, a_1) = (\Lambda_2\Lambda_1, a_2 + \Lambda_2 a_1)\,,$$

so that the *inhomogeneous Lorentz group* is the semi-direct product denoted by

$$IO(D-1,1) = \mathbb{R}^D \rtimes O(D-1,1)\,.$$

The subgroup $ISO(D-1,1)^\uparrow$ of inhomogeneous proper orthochronous Lorentz transformations is called the *Poincaré group*. The Lie algebra $\mathfrak{iso}(D-1,1)$ of the Poincaré group is presented by the generators $\{ P_\mu \,,\, M_{\nu\rho} \}$ and by the commutation relations

$$\begin{aligned} i\,[M_{\mu\nu}, M_{\rho\sigma}] &= \eta_{\nu\rho} M_{\mu\sigma} - \eta_{\mu\rho} M_{\nu\sigma} - \eta_{\sigma\mu} M_{\rho\nu} + \eta_{\sigma\nu} M_{\rho\mu} && (1.1)\\ i\,[P_\mu, M_{\rho\sigma}] &= \eta_{\mu\rho} P_\sigma - \eta_{\mu\sigma} P_\rho\,, && (1.2)\\ i\,[P_\mu, P_\rho] &= 0\,. && (1.3) \end{aligned}$$

Two subalgebras must be distinguished: the Lie algebra $\mathfrak{so}(D-1,1)$ of the Lorentz group presented by the generators $\{ M_{\nu\rho} \}$ and by the commutation relations (1.1), and the Lie algebra $\mathbb{R}^D$ of the Abelian translation group presented by the generators $\{ P_\mu \}$ and by the commutation relations (1.3). The latter algebra $\mathbb{R}^D$ is an ideal of the Poincaré algebra, as can be seen from (1.2). Altogether, this implies that the Lie algebra of the Poincaré group is the semi-direct sum $\mathfrak{iso}(D-1,1) = \mathbb{R}^D \niplus \mathfrak{so}(D-1,1)$.

The Casimir elements of a Lie algebra $\mathfrak{g}$ are homogeneous polynomials in the generators of $\mathfrak{g}$ providing a distinguished basis of the center $\mathcal{Z}\Big(\mathcal{U}(\mathfrak{g})\Big)$ of the universal enveloping algebra $\mathcal{U}(\mathfrak{g})$ (see *e.g.* the part V of the lecture notes [3]). The quadratic Casimir operator of the Lorentz algebra $\mathfrak{so}(D-1,1)$ is the square of the generators $M_{\mu\nu}$:

$$\mathcal{C}_2\Big(\mathfrak{so}(D-1,1)\Big) \;=\; \frac{1}{2}\, M^{\mu\nu} M_{\mu\nu}\,. \qquad (1.4)$$

The quadratic Casimir operator of the Poincaré algebra $\mathfrak{iso}(D-1,1)$ is the square of the momentum

$$\mathcal{C}_2\Big(\mathfrak{iso}(D-1,1)\Big) \;=\; -P^\mu P_\mu\,, \qquad (1.5)$$

while the quartic Casimir operator is

$$\mathcal{C}_4\Big(\mathfrak{iso}(D-1,1)\Big) \;=\; -\frac{1}{2} P^2 M_{\mu\nu} M^{\mu\nu} + M_{\mu\rho} P^\rho M^{\mu\sigma} P_\sigma\,,$$

which, for $D = 4$, is the square of the Pauli-Lubanski vector W^μ,

$$W^\mu := \frac{1}{2}\, \varepsilon^{\mu\nu\rho\sigma} M_{\nu\rho} P_\sigma\,.$$

1.3 ABC of unitary representations

The mathematical property that all non-trivial unitary reps of a non-compact simple Lie group must be infinite-dimensional has some physical significance, as will be reviewed later.

Finite-dimensional unitary reps of non-compact simple Lie groups: *Let* $\mathrm{U} : G \to U(n)$ *be a unitary representation of a Lie group* G *acting on a (real or complex) Hilbert space* $\mathcal{H}$ *of finite dimension* $n \in \mathbb{N}$*. Then* U *is completely reducible. Moreover, if* U *is irreducible and if* G *is a connected simple non-compact Lie group, then* U *is the trivial representation.*

Proof: For the property that U is completely reducible, we refer *e.g.* to the proof of the proposition 5.15 in [1]. The image $\mathrm{U}(G)$ for any unitary representation U defines a subgroup of $U(n)$. Moreover, the kernel of U is a normal subgroup of the simple group G. Therefore, either the representation is trivial and $\ker \mathrm{U} = G$, or it is faithfull and $\ker \mathrm{U} = \{e\}$. In the latter case, U is invertible and its image is isomorphic to its domain, $\mathrm{U}(G) \cong G$. Actually, the image $\mathrm{U}(G)$ is a non-compact subgroup of $U(n)$ because if $\mathrm{U}(G)$ was compact, then $\mathrm{U}^{-1}\big(\mathrm{U}(G)\big) = G$ would be compact since U^{-1} is a continuous map. But the group $U(n)$ is compact, thus it cannot contain a non-compact subgroup. Therefore the representation cannot be faithful, so that it is trivial. (A different proof of the second part of the theorem may be found in the section 8.1.B of [2].) □

Another mathematical result which is of physical significance is the following theorem on unitary irreducible representations (UIRs) of compact Lie groups.

Unitary reps of compact Lie groups: *Let* U *be a UIR of a compact Lie group* G*, acting on a (real or complex) Hilbert space* $\mathcal{H}$*. Then* $\mathcal{H}$ *is finite-dimensional. Moreover, every unitary representation of* G *is a direct sum of UIRs (the sum may be infinite).*

Proof: The proofs are somewhat lengthy and technical so we refer to the section 7.1 of [2] for complete details. □

Examples of (not so) simple groups:
• On the one hand, all (pseudo)-orthogonal groups $O(p,q)$ are either Abelian ($p+q=2$) or simple ($p+q>2$). Moreover, orthogonal groups ($p\,q=0$) are compact, while pseudo-orthogonal groups ($p\,q\neq 0$) are non-compact.
• On the other hand, the inhomogeneous Lorentz group $IO(D-1,1)$ is non-compact (both $\mathbb{R}^D$ and $O(D-1,1)$ are non-compact) but it is *not* semi-simple (because its normal subgroup $\mathbb{R}^D$ is Abelian).

2 Elementary particles as unitary irreducible representations of the isometry group

Except for the final remarks, this section is based almost *ad verbatim* on the introduction of the illuminating work of Bargmann and Wigner [4], modulo some changes of notation and terminology in order to follow the modern conventions.

The wave functions $|\,\psi\,\rangle$ describing the possible states of a quantum-mechanical system form a linear vector space $\mathcal{H}$ which, in general, is infinite-dimensional and on which a positive-definite inner product $\langle\,\phi\,|\,\psi\,\rangle$ is defined for any two wave functions $|\,\phi\,\rangle$ and $|\,\psi\,\rangle$ (*i.e.* they form a Hilbert space). The inner product usually involves an integration over the whole configuration or momentum space and, for particles of non-vanishing spin, a summation over the spin indices.

If the wave functions in question refer to a free particle and satisfy relativistic wave equations, there exists a correspondence between the wave functions describing the same state in different Lorentz frames. The transformations considered here form the group of all *inhomogeneous* Lorentz transformations (including translations of the origin in space and time). Let $|\,\psi\,\rangle$ and $|\,\psi\,\rangle'$ be the wave functions of the same state in two Lorentz frames L and L', respectively. Then $|\,\psi\,\rangle'=U(\Lambda,a)\,|\,\psi\,\rangle$, where $U(\Lambda,a)$ is a linear unitary operator which depends on the transformation (Λ,a) leading from L to L'. By a proper normalization, U is determined by Λ up to a factor ± 1. Moreover, the operators U form a single- or double-valued representation of the inhomogeneous Lorentz group, *i.e.*, for a succession of two transformations (Λ_1,a_1) and (Λ_2,a_2), we have

$$U(\Lambda_2,a_2)U(\Lambda_1,a_1)=\pm U(\Lambda_2\Lambda_1,a_2+\Lambda_2 a_1)\,. \tag{2.1}$$

Since all Lorentz frames are equivalent for the description of our system, it follows that, together with $|\,\psi\,\rangle$, $U(\Lambda,a)\,|\,\psi\,\rangle$ is also a possible state viewed from the original Lorentz frame L. Thus, the vector space $\mathcal{H}$ contains, with every $|\,\psi\,\rangle$, all transforms $U(\Lambda,a)\,|\,\psi\,\rangle$, where (Λ,a) is any inhomogenous Lorentz transformation.

The operators U may also replace the wave equation of the system. In our discussion, we use the wave functions in the "Heisenberg" representation, so that a given $|\,\psi\,\rangle$ represents the system for all times, and may be chosen as the "Schrödinger" wave function at time $t=0$ in a given Lorentz frame L. To find $|\,\psi\,\rangle_{t_0}$, the Schrödinger function at time t_0, one must therefore transform to a frame L' for which $t'=t-t_0$, while all other coordinates remain unchanged. Then $|\,\psi\,\rangle_{t_0}=U(\Lambda,a)\,|\,\psi\,\rangle$, where (Λ,a) is the transformation leading from L to L'.

A classification of all unitary representations of the inhomogeneous Lorentz group, *i.e.* of all solution of (2.1), amounts, therefore, to a classification of all possible relativistic wave equations. Two reps U and $\widetilde{U}=VUV^{-1}$, where V is a fixed unitary operator, are equivalent. If the system is described by wave functions $|\,\psi\,\rangle$, the description by

$$\widetilde{|\,\psi\,\rangle}=V\,|\,\psi\,\rangle \tag{2.2}$$

is isomorphic with respect to linear superposition, with respect to forming the inner product of two wave functions, and also with respect to the transition from one Lorentz frame to another. In fact, if $|\,\psi\,\rangle'=U(\Lambda,a)\,|\,\psi\,\rangle$, then

$$\widetilde{|\,\psi\,\rangle}'=V\,|\,\psi\,\rangle'=VU(\Lambda,a)V^{-1}\widetilde{|\,\psi\,\rangle}=\widetilde{U}(\Lambda,a)\widetilde{|\,\psi\,\rangle}\,.$$

Thus, one obtains classes of equivalent wave equations. Finally, it is sufficient to determine the irreducible representations (irreps) since any other may be built from them.

Two descriptions which are equivalent according to (2.2) may be quite different in appearance. The best known example is the description of the electromagnetic field by the fieldstrength and the vector potential, respectively. It cannot be claimed either that equivalence in the sense of (2.2) implies equivalence in every physical aspect. It should be emphasized that any selection of one among the equivalent systems involves an explicit or implicit assumption as to possible interactions, *etc.* Our analysis is necessarily restricted to free particles and does not lead to any assertion about possible interactions.

The present discussion is not based on any hypothesis about the structure of the wave equations provided that they be covariant. In particular, it is not necessary to assume differential equations in configuration space. But it is a result of the group-theoretical analysis that every irreducible wave equation is equivalent, in the sense of (2.2), to a system of differential equations for fields on Minkowski spacetime.

Remarks:

• An important theorem proved by Wigner is that any symmetry transformation that is continuously related to the identity must be represented by a linear unitary operator (see *e.g.* the appendix A of [5]). Strictly speaking, physical states are represented by *rays* in a Hilbert space. Therefore the unitary representations of the symmetry group are actually understood to be *projective* representations. In spacetime dimensions $D \geqslant 4$, this subtlety reduces to the standard distinction between single and double valued representations of the Poincaré group, as was taken for granted in the text.

• Notice that the previous discussion remains entirely valid if the Minkowski spacetime $\mathbb{R}^{D-1,1}$ is replaced everywhere by any other maximally symmetric spacetime (*i.e.* de Sitter spacetime dS_D, or anti de Sitter spacetime AdS_D) under the condition that the inhomogeneous Lorentz group $IO(D-1,1)$ be also replaced everywhere by the corresponding group of isometries (respectively, $O(D,\,1)$ or, $O(D-1,\,2)$).

• In first-quantization, particles are described by fields on the spacetime and isometries are represented by unitary operators. A particle is said to be "elementary" if the representation is irreducible, and "composite" if the representation is made of a product of irreps. In second-quantization, a unitary representation of the isometry group describes the one-particle Hilbert space of states.

Another point of view: On the one hand, the rules of quantum mechanics imply that quantum symmetries correspond to unitary representations of the symmetry group carried by the Hilbert space of physical states. Furthermore, if time translations are a one-parameter subgroup of the symmetry group, then the Schrödinger equation is essentially a unitary representation of this subgroup. On the other hand, the principle of relativity dictates that the isometries of spacetime be symmetries of the physical system. All together, this implies that linear relativistic wave equations may be identified with unitary reps of the isometry group.

3 Classification of the unitary representations

3.1 Induced representations

The method of induced reps was introduced by Wigner in his seminal paper [6] on the unitary representations of the inhomogeneous Lorentz group $IO(3,1)$ in four spacetime dimensions, which admits a straightforward generalization to any spacetime dimension D, as reviewed now. The content of this subsection finds its origin in the section 2.5 of the comprehensive textbook [5].

From (1.3) one sees that all the translation generators commute with each other, so it is natural to express physical states $|\,\psi\,\rangle$ in terms of eigenvectors of the translation generators P^μ. Introducing a label σ to denote all other degrees of freedom, one thus considers states $\Psi_{q,\sigma}$ with $P_\mu\Psi_{q,\sigma} = q_\mu\Psi_{q,\sigma}$. From the infinitesimal translation $U = \mathbb{1} - iP^\mu\epsilon_\mu$ and repeated applications of it, one finds that finite translations are represented on $\mathcal{H}$ by $U(\,\mathbb{1},a) = \exp(-i\,P^\mu a_\mu)$, so one has

$$U(\,\mathbb{1},a)\,\Psi_{q,\sigma} \;=\; e^{-i\,q\cdot a}\,\Psi_{q,\sigma}\,.$$

Using (1.2), one sees that the effect of operating on $\Psi_{p,\sigma}$ with a quantum homogeneous transformation $U(\Lambda,0)\equiv U(\Lambda)$ is to produce an eigenvector of the translation generators with eigenvalue

Λp :

$$\begin{aligned} P^\mu U(\Lambda)\Psi_{p,\sigma} &= U(\Lambda)[U^{-1}(\Lambda)P^\mu U(\Lambda)]\Psi_{p,\sigma} = U(\Lambda)({\Lambda^{-1}}^\mu{}_\rho P^\rho)\Psi_{p,\sigma} \\ &= \Lambda^\mu{}_\rho p^\rho U(\Lambda)\Psi_{p,\sigma}\,, \end{aligned}$$

since $(\Lambda^{-1})^\rho{}_\nu = \Lambda_\nu{}^\rho$. Hence $U(\Lambda)\Psi_{p,\sigma}$ must be a linear combination of the states $\Psi_{\Lambda p,\sigma}$:

$$U(\Lambda)\Psi_{p,\sigma} = \sum_{\sigma'} C_{\sigma'\sigma}(\Lambda,p)\Psi_{\Lambda p,\sigma'}\,. \tag{3.1}$$

In general, it is possible by using suitable linear combinations of the $\Psi_{p,\sigma}$ to choose the σ labels in such a way that the matrix $C_{\sigma'\sigma}(\Lambda,p)$ is block-diagonal; in other words, so that the $\Psi_{p,\sigma}$ with σ within any one block *by themselves* furnish a representation of the Poincaré group. It is natural to identify the states of a specific particle type with the components of a representation of the Poincaré group which is irreducible, in the sense that it cannot be further decomposed in this way. It is clear from (3.1) that all states $\Psi_{p,\sigma}$ in an irrep of the Poincaré group have momenta p belonging to the orbit of a single fixed momentum, say q^μ.

One has to work out the structure of the coefficients $C_{\sigma'\sigma}(\Lambda,p)$ in irreducible representations of the Poincaré group. In order to do that, note that the only functions of p^μ that are left invariant by all transformations $\Lambda^\mu{}_\nu \in SO(D-1,1)^\uparrow$ are, of course, $p^2 = \eta_{\mu\nu}p^\mu p^\nu$ and, for $p^2 \leqslant 0$, also the sign of p^0 . Hence, for each value of p^2 , and (for $p^2 \leqslant 0$) each sign of p^0 , one can choose a standard four-momentum, say q^μ , and express any p^μ of this class as

$$p^\mu = L^\mu{}_\nu(p)q^\nu\,,$$

where $L^\mu{}_\nu$ is some standard proper orthochronous Lorentz transformation that depends on p^μ , and also implicitly on our choice of q^μ . One can define the states $\Psi_{p,\sigma}$ of momentum p^μ by

$$\Psi_{p,\sigma} \equiv N(p)\,U\Big(L(p)\Big)\,\Psi_{q,\sigma}\,, \tag{3.2}$$

where $N(p)$ is a numerical normalization factor. Operating on (3.2) with an arbitrary homogeneous Lorentz transformation $U(\Lambda)$, one now finds

$$\begin{aligned} U(\Lambda)\Psi_{p,\sigma} &= N(p)\,U\Big(\Lambda L(p)\Big)\,\Psi_{q,\sigma} \\ &= N(p)\,U\Big(L(\Lambda p)\Big)\,U\Big(L^{-1}(\Lambda p)\Lambda L(p)\Big)\Psi_{q,\sigma}\,. \end{aligned} \tag{3.3}$$

The point of this last step is that the Lorentz transformation $L^{-1}(\Lambda p)\Lambda L(p)$ takes q to $L(p)q = p$, then to Λp , and finally back to q , so it belongs to the subgroup of the Lorentz group consisting of Lorentz transformations $W^\mu_{\ \nu}$ that leave q^μ invariant : $W^\mu{}_\nu q^\nu = q^\mu$. This stability subgroup is called the *little group* corresponding to q . For any $W, \bar{W}$ in the little group, one has

$$U(W)\Psi_{q,\sigma} = \sum_{\sigma'} D^q_{\sigma'\sigma}(W)\Psi_{q,\sigma'} \tag{3.4}$$

and

$$D^q_{\sigma'\sigma}(\bar{W}W) = \sum_{\sigma''} D^q_{\sigma'\sigma''}(\bar{W})D^q_{\sigma''\sigma}(W)\,,$$

that is to say, the coefficients $D^q(W)$ furnish a representation of the little group. In particular, for $W(\Lambda,p) \equiv L^{-1}(\Lambda p)\Lambda L(p)$, (3.3) becomes

$$U(\Lambda)\Psi_{p,\sigma} = N(p)\sum_{\sigma'} D_{\sigma'\sigma}(W(\Lambda,p))U\Big(L(\Lambda p)\Big)\Psi_{q,\sigma'}$$

or, recalling the definition (3.2),

$$U(\Lambda)\Psi_{p,\sigma} = \frac{N(p)}{N(\Lambda p)}\sum_{\sigma'} D_{\sigma'\sigma}\big(W(\Lambda,p)\big)\Psi_{\Lambda p,\sigma'}\,. \tag{3.5}$$

Apart from the question of normalization, the problem of determining the coefficients $C_{\sigma'\sigma}$ in the transformation rule (3.1) has been reduced to the problem of determining the coefficients

$D_{\sigma'\sigma}$. In other words, the problem of determining all possible irreps of the Poincaré group has been reduced to the problem of finding all possible irreps of the little group, depending on the class of momentum to which q^μ belongs. This approach, of deriving representations of a semi-direct product like the inhomogeneous Lorentz group from the representations of the stability subgroup, is called the *method of induced representations*.

The wave function $\Psi_{p,\sigma}$ depends on the momentum, therefore its Fourier transform $\Psi_{x,\sigma}$ depends on the spacetime coordinate, so that the wave function is called a (complex) *field* on Minkowski spacetime $\mathbb{R}^{D-1,1}$ and the quantities $\Psi_{x,\sigma}$ at fixed x and for varying σ are referred to as its *physical components* at x.

3.2 Orbits and stability subgroups

The orbit of a given non-vanishing vector q^μ of Minkowski spacetime $\mathbb{R}^{D-1,1}$ under Lorentz transformations is, by definition, the hypersurface of constant momentum square p^2. Geometrically speaking, it is a quadric of curvature radius $m>0$. More precisely, the hypersurface

- $p^2=-m^2$ is a two-sheeted hyperboloid, each sheet of which is called a *mass-shell*. The corresponding UIR is said to be *massive*.
- $p^2=0$ is a cone, each half of which is called a *light-cone*. The corresponding UIR is said to be *massless* ($m=0$).
- $p^2=+m^2$ is a one-sheeted hyperboloid. The corresponding UIR is said to be *tachyonic*.

Orthochronous Lorentz transformations preserve the sign of the time component of vectors of non-positive momentum-squared, thus the orbit of a time-like (light-like) vector is the mass-shell (respectively, light-cone) to which the corresponding vector belongs.

Remarks:
• Notice that the Hilbert space carrying the irrep is indeed an eigenspace of the quadratic Casimir operator (1.5), the eigenvalue of which is $\mathcal{C}_2=\pm\, m^2$ (the eigenvalue is real because the representation is unitary), as it should according to Schur's lemma. Moreover, the quadratic Casimir classifies the UIRs but does not entirely characterize them.
• As quoted in Section 2, it is not necessary to assume differential equations in position space, because the group-theoretical analysis leads directly to a wave function which is a function of the momenta on the orbit, the Fourier transform of which is a function in position space obeying the Klein–Gordon equation $\Box\,\Psi_{x,\sigma}\,=\,\pm\, m^2\,\Psi_{x,\sigma}$.

By going to a rest-frame, it is easy to show that the stabilizer of a time-like vector $q^\mu=(m,\,\overrightarrow{0}\,)\neq 0$ is the rotation subgroup $SO(D-1)\subset SO(D-1,1)^\uparrow$. For a space-like vector, one may choose a frame such that the non-vanishing momentum is along the $(D-1)$th spatial axis: $q^\mu=(0,0,\dots,0,m)\neq 0$. Thus its stabilizer is the subgroup $SO(D-2,1)^\uparrow\subset SO(D-1,1)^\uparrow$. In the case of a light-like vector, the little group "*is not quite so obvious*" to determine, as was stressed by Wigner himself [7]. It clearly contains the rotation group $SO(D-2)$ in the space-like hyperplane $\mathbb{R}^{D-2}$ transverse to the light-ray along the momentum. Now, we will provide an algebraic proof that the stabilizer of a light-like vector is the Euclidean group $ISO(D-2)$. According to Wigner, reviewing his $D=4$ analysis, "*no simple argument is known (...) to show directly that the group of Lorentz transformations which leave a null vector invariant is isomorphic to the two-dimensional Euclidean group, desirable as it would be to have such an argument. Clearly, there is no plane in the four-space of momenta in which these transformations could be interpreted directly as displacements (...) because all transformations considered here are homogeneous*" [7]. Even though there is no simple geometric way to understand this fact, the algebraic proof reviewed here is rather straightforward.

<u>Proof:</u> By going in a light-cone frame (see Section 1.1), it is possible to express the components of a momentum p^μ obeying $p^2=0$ as $p_\mu=(p_-,0,0,\dots,0)$. In words, one can set the component p_+ to zero, as well as all the transverse components p_m $(m=1,\dots,D-2)$. The condition that the component p_- be unaffected by a Lorentz transformation translates as $0\stackrel{!}{=}i[p_-,M_{\nu\rho}]=\eta_{-\nu}\,p_\rho-\eta_{-\rho}\,p_\nu$ due to (1.2). Obviously, the transformation generated by M_{+-} does modify p_-, hence it cannot be part of the little group for p. The other Lorentz generators preserve p_-, but they should also satisfy the equations $[p_m,M_{\mu\nu}]=0=[p_+,M_{\mu\nu}]$. It is readily seen that

$i[p_m, M_{n-}] = \delta_{mn} p_- \neq 0$ (for $m = n$), therefore M_{n-} does not belong to the little group of p_μ either. We are left with the generators $\{M_{mn}, M_{+n}\}$ which preserve the (vanishing) value of p_+. It turns out to be more convenient for later purpose to work with the generators $\pi_n := p_- M_{+n} = p^\mu M_{\mu\, n}$ instead. This redefinition does not modify the algebra since p_- commutes with all the generators of the little group. From the Poincaré algebra (1.1)–(1.3) one finds, in the light-cone frame,

$$i\,[M_{mn}, M_{pq}] \;=\; \delta_{np} M_{mq} - \delta_{mp} M_{nq} - \delta_{qm} M_{pn} + \delta_{qn} M_{pm}\,, \tag{3.6}$$

$$i\,[\pi_m, M_{np}] \;=\; \delta_{mn} \pi_p - \delta_{mp} \pi_n\,, \tag{3.7}$$

$$i\,[\pi_m, \pi_n] \;=\; 0\,. \tag{3.8}$$

As can be seen, the generators $\{M_{mn}, \pi_m\}$ span the Lie algebra of the inhomogeneous orthogonal group $ISO(D-2)$. □

For later purpose, notice that the quadratic Casimir operator of the Euclidean algebra $\mathfrak{iso}(D-2)$ presented by the generators $\{M_{mn}, \pi_m\}$ and the relations (3.6)-(3.8) is the square of the "translation" generators

$$\mathcal{C}_2\Big(\mathfrak{iso}(D-2)\Big) \;=\; \pi^m \pi_m\,. \tag{3.9}$$

To end up this discussion, one should consider the case of a vanishing momentum. Of course, the orbit of a vanishing vector under linear transformations is itself while its stabilizer is the whole linear subgroup. Therefore, the subgroup of $SO(D-1,1)^\uparrow$ leaving invariant the zero-momentum vector $p^\mu = 0$ is the whole group itself. This ends up the determination of the orbit and stabilizer of any possible vector $\in \mathbb{R}^{D-1,1}$.

Remark: The zero-momentum ($q^\mu = 0$) representations are essentially UIRs of the little group $SO(D-1,1)^\uparrow$ because the translation group acts trivially. The proper orthochronous Lorentz group may be identified with the isometry group of the de Sitter spacetime dS_{D-1}. In other words, the wave function of the zero-momentum representation may be interpreted as a wave function on a lower-dimensional de Sitter spacetime, and conversely. Even though their physical meaning may differ, both UIRs may be mathematically identified.

3.3 Classification

To summarize the previous subsection, the UIRs of the Poincaré group $ISO(D-1,1)^\uparrow$ have been divided into four classes according to the four possible orbits of the momentum, summarized in the following table (where $m^2 > 0$):

Gender	Orbit	Stability subgroup	UIR
$p^2 = -m^2$	Mass-shell	$SO(D-1)$	Massive
$p^2 = 0$	Light-cone	$ISO(D-2)$	Massless
$p^2 = +m^2$	Hyperboloid	$SO(D-2,1)^\uparrow$	Tachyonic
$p_\mu = 0$	Origin	$SO(D-1,1)^\uparrow$	Zero-momentum

The problem of classifying the UIRs of the Poincaré group $ISO(D-1,1)^\uparrow$ has been reduced to the classication of the UIRs of the stability subgroup of the momentum, which are either a unimodular orthogonal group, an Euclidean group or a proper orthochronous Lorentz group.

Actually, the method of induced representation may also be applied to the classification of the UIRs of the Euclidean group $ISO(D-2)$, the little group of a massless particle. The important thing to understand is that the light-like momentum p^μ is fixed and that what should be examined is the action of the little group on the physical components characterized by σ. From (3.8) one sees that the $D-2$ "translation" generators π^i commute with each other, so it is natural to express physical states $\Psi_{p,\sigma}$ in terms of eigenvectors ξ^m of these generators π^m. Introducing a label ς to denote all remaining physical components, one thus considers states $\Psi_{p,\,\xi,\,\varsigma}$ with $\pi_m \Psi_{p,\,\xi,\,\varsigma} = \xi_m \Psi_{p,\,\xi,\,\varsigma}$. The discussion presented in Subsection 3.1 may be repeated almost identically, up to the replacement of the momentum p by the eigenvector ξ, the label σ by ς, the Poincaré group $ISO(D-1,1)^\uparrow$ by the Euclidean group $ISO(D-2)$ and the proper orthochronous Lorentz group $SO(D-1,1)^\uparrow$ by the unimodular orthogonal group $SO(D-2)$. The conclusion is therefore similar: the problem of determining all possible irreps of the massless little group $ISO(D-2)$ has been reduced to the problem of finding all possible irreps of the stability subgroup of the $(D-2)$-vector ξ, called the *short little group* in the literature [8].

The massless representations induced by a non-trivial representation of the little group may therefore be divided into distinct categories, depending on the class of momentum to which ξ^m belongs. The situation is simpler here because there exist only two possible classes of orbits for a vector in the Euclidean space $\mathbb{R}^{D-2}$: either the origin $\xi^m = 0$, or a $(D-3)$-sphere of radius $\mu > 0$. In the first case, the action of the elusive "translation" operators π^m is trivial and, effectively, the little group is identified with the short little group $SO(D-2)$. These representations are most often referred to as *helicity* representations by analogy with the $D = 4$ case. In the second case, the corresponding representations are most often referred to as *continuous spin* representations [8], even though Wigner also used the name *infinite spin* [7]. The former name originates from the fact that the transverse vector ξ^m has a continuous range of values. Nevertheless, the latter name is more adequate in some respect, as will be argued later on. Roughly speaking the point is that, on the orbit $\xi^2 = \mu^2$, the components spanned by the internal vector ξ^m take values on the sphere $S^{D-3} \subset \mathbb{R}^{D-2}$ of radius $\mu = |\xi|$. The "radius" μ of this internal sphere has actually the dimension of a mass parameter (the reason is that the sphere S^{D-3} is somehow in internal "momentum" space). Indeed, for massless representations, the parameter μ classifying the various irreps should be understood as the analogue of the mass for massive irreps, while the angular coordinates on the sphere S^{D-3} are the genuine "spin" degrees of freedom, the Fourier conjugates of which are discrete variables as is more usual for spin degrees of freedom. This point of view motivates the name "infinite spin."[3]

To summarize, the UIRs of the Euclidean group $ISO(D-2)$ are divided into two classes according to the two possible orbits of the $(D-2)$-vector ξ_m, summarized in the following table:

Gender	Orbit	Stability subgroup	Massless UIR
$\xi^2 = \mu^2$	Sphere	$SO(D-3)$	Infinite spin
$\xi_m = 0$	Origin	$SO(D-2)$	Helicity

As a consequence of the method of induced representations, the physical components of a first-quantized elementary particle span a UIR of the little group. The number of local degrees of freedom (or of physical components) of the field theory is thus given by the dimension of the Hilbert space carrying the UIR of the little group. In the light of the standard results of representation theory (reviewed in Subsection 1.3) and using the method of induced representation, the UIRs of the Poincaré group may alternatively be divided into two distinct classes: the *finite-component* ones (the massive and the helicity reps) for which the (short) little group is compact, and the *infinite-component* ones (the infinite-spin, the tachyonic and the zero-momentum reps) for which the little group is non-compact.

Remarks:

- More precisely, the lower-dimensional cases $D = 2, 3$ are degenerate in the following sense (when $p^\mu \neq 0$). In $D = 2$, all little groups are trivial, thus all physical fields are scalars. In $D = 3$, all little groups are Abelian (massive: $SO(2)$, massless: $\mathbb{R}$, tachyonic: $SO(1,1)^\uparrow \cong \mathbb{R}$) hence all their UIRs have (complex) dimension one: generically, fields have one physical degrees of freedom. Notice that the helicity reps may be assigned a "*conformal* spin" if they are extended to irreps of the group $SO(D,2) \supset SO(D-1,1)^\uparrow$. Notice also that the "spin" of a massive representation is not discretized in $D = 3$ but is an arbitrary real number[4] [9] because the universal cover of $SO(2,1)^\uparrow$ covers it infinitely often.
- For massive and helicity representations, the number of local physical degrees of freedom may be determined from the well known formulas for the dimension of any UIR of the orthogonal groups (reviewed in Subsection 4.3 for the tensorial irreps).
- This group-theoretical analysis does not probe topological theories (such as Chern-Simons theory) because such theories correspond to identically vanishing representations of the little group since they have no *local* physical degrees of freedom.

The following corollary provides a group-theoretical explanation of the fact that combining the principle of relativity with the rules of quantum mechanics necessarily leads to *field* theory.

Corollary: *Every non-trivial unitary irreducible representation of the isometry group of any maximally-symmetric spacetime is infinite-dimensional.*

[3] Actually, in Subsection 5.3 an explicit derivation of the continuous spin representation from a proper "infinite spin" limit of a massive representation is reviewed. All the former comments find their natural interpretation in this point of view.

[4] This peculiarity is related to the existence of anyons in three spacetime dimensions.

Proof: The Hilbert space carrying a non-trivial unitary representation of the Poincaré group is infinite-dimensional because (i) in the generic case, $q_\mu \neq 0$, the orbit is either a hyperboloid ($p^2 \neq 0$) or a cone ($p^2 = 0$) and the space of wave functions on the orbit is infinite-dimensional, (ii) the zero-momentum representations of the Poincaré group are unitary representations of the de Sitter isometry group. Thus, the proof is ended by noticing that all non-trivial unitary representations of the isometry group of (anti) de Sitter spacetimes $(A)dS_D$ also are infinite-dimensional, because their isometry groups are *pseudo*-orthogonal Lie groups. $\square$

4 Tensorial representations and Young diagrams

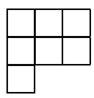
Most of the material reviewed here may be found in textbooks such as [10]. Nevertheless, large parts of this section are either copied or adapted from the reference [11] because altogether it provides an excellent summary, both for its pedagogical and comprehensive values. The material collected in the present section goes slightly beyond what is strictly necessary for these lectures, but the reader may find it useful in specific applications.

4.1 Symmetric group

An (unlabeled) *Young diagram*, consisting of n boxes arranged in r (left justified) rows, represents a *partition* of the integer n into r *parts*:

$$n = \sum_{a=1}^{r} \lambda_a \qquad (\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_r)\,.$$

That is, λ_a is the number of boxes in the ath row. Successive row lengths are non-increasing from top to bottom. A simpler notation for the partition is the list of its parts: $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_r\}$. For instance,

$\{3,3,1\} =$  .

Examples: There are five partitions of 4:

$$\{4\}, \{3,1\}, \{2,2\}, \{2,1,1\}, \{1,1,1,1\}\,. \tag{4.1}$$

Partitions play a key role in the study of the symmetric group $\mathfrak{S}_n$. This is the group of all permutations of n objects. It has $n!$ elements and *its inequivalent irreducible representations may be labeled by the partitions of* n. [In the following, Greek letters λ, μ and ν will be used to specify not only partitions and Young diagrams but also irreducible representations of the symmetric group and other groups.]

The connection between the symmetric group and tensors was initially developed by H. Weyl. This connection can be approached in (at least) two equivalent ways.

A. Let $T_{\mu_1 \ldots \mu_n}$ be a 'generic' n-index tensor, without any special symmetry property. [For the moment, 'tensor' just means a function of n indices, not necessarily with any geometrical realization. It must be meaningful, however, to *add* — and form linear combinations of — tensors of the same rank.] A *Young tableau*, or labeled Young diagram, is an assignment of the numbers $1, 2, \ldots, n$ to the n boxes of a Young diagram λ. The tableau is *standard* if the numbers are increasing both along rows from left to right and down columns from top to bottom. The entries $1, \ldots, n$ in the tableau indicate the n successive indices of $T_{\mu_1 \ldots \mu_n}$. The tableau defines a certain symmetrization operation on these indices: *symmetrize* on the set of indices indicated by the entries in each row, then *antisymmetrize* the result on the set of indices indicated by the entries in each column. The resulting object is a tensor, $\widetilde{T}$, with certain index symmetries. Now let each permutation of $\mathfrak{S}_n$ act (separately) upon $\widetilde{T}$. The $n!$ results are not linearly independent; they span a vector space $V_\lambda^{\mathfrak{S}_n}$ which supports an irreducible representation of $\mathfrak{S}_n$. Different tableaux corresponding to the same diagram λ yield equivalent (by not identical) representations.

Example: The partition $\{2, 2\}$ of 4 has two standard tableaux:

$$\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & 4 \\ \hline\end{array} \quad \text{and} \quad \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & 4 \\ \hline\end{array} \quad . \tag{4.2}$$

Let us construct the symmetrized tensor $\widetilde{T}_{abcd} := R_{ab|cd}$ corresponding to the second of these:

$$\begin{array}{|c|c|}\hline \mathrm{a} & \mathrm{c} \\ \hline \mathrm{b} & \mathrm{d} \\ \hline\end{array} \quad . \tag{4.3}$$

First symmetrize over the first and third indices (a and c), and over the second and fourth (b and d):

$$\frac{1}{4}\left(T_{abcd} + T_{cbad} + T_{adcb} + T_{cdab}\right).$$

Now antisymmetrize the result over the first and second indices (a and b), and over the third and fourth (c and d);[5] dropping the combinatorial factor $\frac{1}{16}$, we get

$$\begin{aligned} R_{ab|cd} &= T_{abcd} + T_{cbad} + T_{adcb} + T_{cdab} - T_{bacd} - T_{cabd} - T_{bdca} - T_{cdba} \\ &- T_{abdc} - T_{dbac} - T_{acdb} - T_{dcab} + T_{badc} + T_{dabc} + T_{bcda} + T_{dcba}\,. \end{aligned}$$

It is easy to check that R possesses the symmetries of the Riemann tensor. There are two independent orders of its indices (e.g. $R_{ab|cd}$ and $R_{ac|bd}$), and applying any permutation to the indices produces some linear combination of those two basic objects. On the other hand, performing on T the operations prescribed by the first tableau in (4.2) produces a different expression $P_{ac|bd}$, which, however, generates a two-dimensional representation of $\mathfrak{S}_4$ with the same abstract index structure as that generated by R. A non-standard tableau would also yield such a representation, but the tensors within it would be linear combinations of those already found. One finds

$$\begin{aligned} P_{ac|bd} &= T_{abcd} + T_{bacd} + T_{abdc} + T_{badc} - T_{cbad} - T_{bcad} - T_{cbda} - T_{bcda} \\ &- T_{adcb} - T_{dacb} - T_{adbc} - T_{dabc} + T_{cdab} + T_{dcab} + T_{cdba} + T_{dcba}\,. \end{aligned}$$

As the reader may check, no linear combinations of P can reproduce R. The objects $P_{ab|cd}$, $P_{ac|bd}$, $R_{ab|cd}$ and $R_{ac|bd}$ are linearly independent. Although R and P are characterized by the same Young *diagram*, they are associated with different standard Young *tableaux* and therefore span two *different* irreducible representations of $\mathfrak{S}_n$.

Example: Define a *symmetrized Riemann tensor* (the *Jacobi tensor*) by $J_{ad;bc} := \frac{1}{2}\,(R_{ab|cd} + R_{ac|bd})$. It obeys $J_{ab;cd} = J_{ba;cd} = J_{ab;dc}$. Then it is easy to show that $R_{ab|cd} = \frac{2}{3}\,(J_{ad;bc} - J_{bd;ac})$. Thus the tensor J has no fewer independent components and contains no less information than the tensor R, despite the extra symmetrization; R is recovered from J by an antisymmetrization. The tensors R and J are really the same tensor expressed with respect to different bases.

B. The *regular representation* of $\mathfrak{S}_n$ is the $n!$-dimensional representation obtained by letting $\mathfrak{S}_n$ act by left multiplication on the formal linear combinations of elements of $\mathfrak{S}_n$. [That is, one labels the basis vectors of $\mathbb{R}^{n!}$ by elements of $\mathfrak{S}_n$, defines that action of each permutation on the basis vectors in the natural way, and extends this definition to the whole space by linearity.] Equivalently, the vector space of the regular representation is the space of real-valued functions defined on $\mathfrak{S}_n$. [In general the regular representation is defined with complex scalars, but for $\mathfrak{S}_n$ it is sufficient to work with real coefficients.]

Regular representation: *The regular representation contains every irreducible representation with a multiplicity equal to its dimension. Each Young diagram λ corresponds to an irreducible representation of $\mathfrak{S}_n$. Its dimension and multiplicity are equal to the number of standard tableaux of diagram λ.*

[5] Here we adopt the convention that the second round of permutations interchanges indices with the same *names*, rather than indices in the same *positions* in the various terms. The opposite convention is tantamount to antisymmetrizing *first*, which leads to a different, but mathematically isomorphic, development of the representation theory. The issue here is analogous to the distinction between space-fixed and body-fixed axes in the study of the rotation group (active or passive transformations).

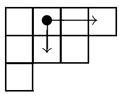
The symmetrization procedure described under **A.** can be transcribed to the more abstract context **B.** to construct a projection operator onto the subspace of $\mathbb{R}^{n!}$ supporting each representation. [The numerical coefficient needed to normalize the tableau operation as a projection — an operator whose square is itself — is not usually the same as that accumulated from the individual symmetrization operations. For example, to make R_{abcd} into a projection of T_{abcd}, one needs to divide by 12, not 16.]

Example: In (4.1), the partition $\{4\}$ corresponds to the totally symmetric four-index tensors, a one-dimensional space $V^{\mathfrak{S}_4}_{\{4\}}$. Similarly, $\{1,1,1,1\}$ yields the totally antisymmetric tensors. A generic rank-four tensor, T_{abcd}, can be decomposed into the sum of its symmetric and antisymmetric parts, plus a remainder. The theory we are expounding here tells how to decompose the remainder further. The partition $\{2,2\}$ yields two independent two-dimensional subrepresentations of the regular representation; in more concrete terms, there are two independent pieces of T_{abcd} ($\frac{1}{12}\,R_{ab|cd}$ and $\frac{1}{12}\,P_{ac|bd}$) constructed as described in connection with (4.2). One of these ($R_{ab|cd}$) has exactly the symmetries of the Riemann tensor; the other ($P_{ac|bd}$, coming from the first tableau of (4.2)) has the same abstract symmetry as the Riemann tensor, but with the indices ordered differently. Finally, each of the remaining partitions in (4.1) can be made into a standard tableau in three different ways. Therefore, each of these two representations has three separate pieces of T corresponding to it, and each piece is three-dimensional (has three independent index orders after its symmetries are taken into account). Thus the total number of independent tensors which can be formed from the irreducible parts of T_{abcd} by index permutations is

$$1^2+1^2+2^2+3^2+3^2=24=4!$$

which is simply the total number of permutations of the indices of T itself, as it must be.

To state a formula for the dimension of an irreducible representation $V^{\mathfrak{S}_n}_{\lambda}$ of $\mathfrak{S}_n$, we need the concept of the hook length of a given box in a Young diagram λ. The *hook length* of a box in a Young diagram is the number of squares directly below or directly to the right of the box, including the box once:



Example: In the following diagram, each box is labeled by its hook length:

$$\begin{array}{|c|c|c|c|}\hline 6&4&3&1\\ \hline 4&2&1\\ \cline{1-3} 1\\ \cline{1-1}\end{array}\ .$$

One then has the following *hook length formula* for the dimension of the representation $V^{\mathfrak{S}_n}_{\lambda}$ of $\mathfrak{S}_n$ corresponding to the Young diagram λ:

$$\dim V^{\mathfrak{S}_n}_{\lambda}=\frac{n!}{\prod(\text{ hook lengths})}\,. \tag{4.4}$$

Remark: Note carefully that the "dimension" we have been discussing up to now is the number of independent *index orders* of a tensor, not the number of independent *components* when the tensor is realized geometrically with respect to a particular underlying vector space or manifold. The latter number depends on the dimension (say D) of that underlying space, while the former is independent of D (so long as D is sufficiently large, as we tacitly assume in generic discussions). For example, the number of components of an antisymmetric two-index tensor is $\frac{D(D-1)}{2}$, but the number of its index orders is always 1, except in dimension $D=1$ where no non-zero antisymmetric tensors exist at all.

4.2 General linear group

We now turn to the representation theory of the general linear and orthogonal groups, where the (spacetime) dimension D plays a key role. The theory of partitions and of the representations of the permutation groups is the foundation on which this topic is built.

Let $\{v_a\}$ represent a generic element of $\mathbb{R}^D$ (or of the cotangent space at a point of a D-dimensional manifold). The action of non-singular linear operators on this space gives a D-dimensional irreducible representation V of the general linear group $GL(D)$; indeed, this representation defines the group itself. The rank-two tensors, $\{T_{ab}\}$, carry a larger representation of $GL(D)$ ($V \otimes V$, of dimension D^2), where the group elements act on the two indices simultaneously. The latter representation is reducible: it decomposes into the subspace of symmetric and antisymmetric rank-two tensors $V \otimes V \cong (V \odot V) \oplus (V \wedge V)$, of respective dimensions $\frac{D(D+1)}{2}$ and $\frac{D(D-1)}{2}$. Similarly, the tensor representation of rank n, $V^{\otimes n}$, decomposes into irreducible representations of $GL(D)$ which are associated with the irreducible representations of $\mathfrak{S}_n$ acting on the indices, which in turn are labeled by the partitions of n, hence by Young diagrams. Young diagrams with more than D rows do not contribute [if λ is a partition of n into more than D parts, then the associated index symmetrization of a D-dimensional rank-n tensor yields an expression that vanishes identically; in particular, there are no non-zero totally antisymmetric rank-n tensors if $n > D$].

More precisely, let λ be a Young *tableau*. The *Schur module* $V_\lambda^{GL(D)}$ is the vector space of all rank-n tensors $\tilde{T}$ in $V^{\otimes n}$ such that:

(i) the tensor $\tilde{T}$ is completely antisymmetric in the entries of each column of λ,

(ii) complete antisymmetrization of $\tilde{T}$ in the entries of a column of λ and another entry of λ that is on the right-hand side of the column vanishes.

This construction is equivalent to the construction **A.**

Example: Associated with the Young tableau (4.3), the tensor $R_{ab|cd}$ introduced in the subsection 4.1 obeys to the conditions (i) and (ii): $R_{ab|cd} = -R_{ba|cd} = -R_{ab|dc}$ and $R_{ab|cd} + R_{bc|ad} + R_{ca|bd} = 0$.

As explained in the footnote 5, if one interchanges everywhere in the previous constructions the words "symmetric" and "antisymmetric," then the (reducible) representation spaces characterized by the same Young *diagram* [but not by the same Young *tableau*] are isomorphic and the conditions (i)-(ii) must be replaced with:

(a) the tensor is completely symmetric in the entries of each column of λ,

(b) complete symmetrization of the tensor in the entries of a row of λ and another entry of λ that sits in a lower row vanishes.

Example: Taking the standard Young tableau (4.3) and constructing, following the "manifestly symmetric convention", the irreducible tensor associated with it, one obtains a tensor $\mathcal{R}$ with the same abstract index symmetries as J [*i.e.* obeying the constraints (a) and (b)] but which is however linearly independent from J, thence linearly independent from R alone. The tensor $\mathcal{R}$ can be expressed as a linear combination of *both* R *and* P. Similarly, taking the first standard Young tableau in (4.2) and following the manifestly symmetric convention, one obtains a tensor $\mathcal{P}$ obeying (a) and (b). This tensor is linearly independent from P alone as it is a linear combination of *both* P and R. Summarizing, associated with the Young *diagram* $\{2,2\}$ we have the (reducible) representation space spanned by either $\{R, P\}$ in the manifestly antisymmetric convention or by $\{\mathcal{R}, \mathcal{P}\}$ in the manifestly symmetric convention.

Remarks:

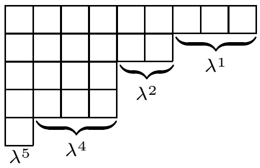
• An important point to note is that, by the previous construction featuring irreducible tensors with definite symmetry properties, we have got *all* the finite-dimensional irreducible representations of $GL(D,\mathbb{R})$.

• In order to make contact with an alternative road to the representation theory of $GL(D)$, one says that the irreducible representation $\Gamma_{\lambda^1 \ldots \lambda^{D-1}}$ of $\mathfrak{sl}(D,\mathbb{C}) \equiv A_{D-1}$ with highest weight $\Lambda = \lambda^1\Lambda_{(1)} + \lambda^2\Lambda_{(2)} + \ldots + \lambda^{D-1}\Lambda_{(D-1)}$ [see *e.g.* the Part II of the lecture notes [3] for definitions and notations] is obtained by applying the Schur functor $\mathbb{S}_\lambda$ [*i.e.* the construction presented above] to the standard representation V, where the Young diagram is

$$\lambda = \{\lambda^1 + \ldots + \lambda^{D-1}\,,\, \lambda^2 + \ldots + \lambda^{D-1}\,,\, \ldots\,,\, \lambda^{D-1}\,,\, 0\}\,.$$

In terms of the Young diagram for λ, the Dynkin labels λ^a ($1 \leqslant a \leqslant D-1$) are the differences of lengths of rows: $\lambda^a = \lambda_a - \lambda_{a+1}$.

Example: If $D=6$, then



is the Young diagram corresponding to the irrep $\Gamma_{3,2,0,3,1}$ of $A_5 \equiv \mathfrak{sl}(6,\mathbb{C})$.

The dimension of the representation $V_\lambda^{GL(D)}$ of $GL(D)$ corresponding to the Young diagram λ is:

$$\dim V_\lambda^{GL(D)} = \prod \frac{D - \text{row} + \text{column}}{\text{hook length}}\,, \tag{4.5}$$

where the product is over the n boxes while "row" and "column" respectively give the place of the corresponding box. As was underlined before, the formula (4.5) is distinct from the hook length formula (4.4).

Examples:
• In the following diagram

$$\begin{array}{|c|c|c|c|}\hline 5 & 6 & 7 & 8 \\ \hline 4 & 5 & 6 \\ \cline{1-3} 3 \\ \cline{1-1}\end{array}\quad ,$$

each box is labeled by its value in the numerator of (4.5) for $D=5$. Observe that, for the corresponding diagram λ, $\dim V_\lambda^{GL(5)} = 1050 \neq 70 = \dim V_\lambda^{\mathfrak{S}_8}$.
• The space of (anti)symmetric tensors of V of rank n are denoted by $\odot^n(V)$ (respectively, $\wedge^n(V)$). It carries an irreducible representation of $GL(D)$ labeled by a Young diagram made of one row (respectively, column) of length n. The dimensions

$$\dim \odot^n(V) = \left(\begin{array}{c} D+n-1 \\ n \end{array}\right), \qquad \dim \wedge^n(V) = \left(\begin{array}{c} D \\ n \end{array}\right), \tag{4.6}$$

are easily computed from the formula (4.5) and reproduce the standard results obtained from combinatorial arguments.

If T_1 and T_2 are tensors of ranks n_1 and n_2, then their product is a tensor of rank n_1+n_2. Each factor T_j transforms under index permutation according to some representation of $\mathfrak{S}_{n_j}$, and under linear transformation by the corresponding representation of $GL(D)$. It follows immediately that the product tensor $T_1 \otimes T_2$ transforms as some representation of $\mathfrak{S}_{n_1} \times \mathfrak{S}_{n_2}$. This *induces* a representation of the full permutation group $\mathfrak{S}_{n_1+n_2}$ which is associated with a corresponding representation of $GL(D)$, called *Kronecker product*. It is possible to reduce these last two representations into a sum of irreducible ones. We may assume that the factor representations are irreducible, since the original tensors T_j could have been broken into irreducible parts at the outset.

Littlewood–Richardson rule: The decomposition of an "outer product" $\mu \cdot \nu$ of irreducible representations μ and ν of $\mathfrak{S}_{n_1}$ and $\mathfrak{S}_{n_2}$, respectively, into irreducible representations of $\mathfrak{S}_{n_1+n_2}$ can be determined by means of the following algorithm involving Young diagrams. The product is commutative, so it does not matter which factor is regarded as the "right-hand" one. [In practice, on should choose the simpler Young diagram for that role.]

(I) Label each box in the top row of the right-hand diagram, ν, by "a", each box in the second row by "b", etc.

(II) Add the labeled boxes of ν to the left-hand diagram μ, one at a time, first the as, then the bs, ..., subject to these constraints:

(A) No two boxes in the same column are labeled with the same letter;

(B) At all stages the result is a legitimate Young diagram;

(C) At each stage, if the letters are read right-to-left along the rows, from top to bottom, one never encounters more bs than as, more cs than bs, etc.

(III) Each of the distinct diagrams constructed in this way specifies an irreducible subrepresentation λ, appearing in the decomposition of the outer product. The same labeled Young diagram may arise in more than one way; the multiplicity of that representation must be counted accordingly.

Remarks:
• This rule enables products of *distinct* tensors to be decomposed. When the factors are the same tensor, the list is further restricted by the requirement of symmetry under interchange of the factors. This is the problem of *plethysm*, whose solution requires more complicated techniques than the Littlewood–Richardson rule.
• Representations with too many parts (columns of length greater than D) must be deleted from the list of subrepresentations of the $GL(D)$. [If irreducible representations of the special linear group $SL(D)$ are considered instead, every column of length D must be removed from the corresponding Young diagram.]

4.3 Orthogonal group

It remains to consider index contractions. Up to now we considered only covariant tensors, because in the intended application there is a metric tensor which serves to relate contravariant and covariant tensors. Contractions are mediated by this metric. Implicitly, therefore, one is restricting the symmetry group of the problem from the general linear group to the subgroup that leaves the metric tensor invariant, the orthogonal group $O(D)$. [If the metric has indefinite signature, the true symmetry group is a non-compact analogue of the orthogonal group, such as the Lorentz group. This does not affect the relevant aspects of the *finite-dimensional* representation theory.] Each irreducible $GL(D)$ representation $V_\lambda^{GL(D)}$ decomposes into irreducible $O(D)$ representations $V_\nu^{O(D)}$, labeled by Young diagrams ν obtained by removing an even number of boxes from λ. The *branching rule* for this process involves a sort of inverse of the Littlewood–Richardson rule:

Restriction from $GL(D)$ to $O(D)$: *The irreps of $GL(D)$ may be reduced to direct sums of irreps of $O(D)$ by extracting all possible trace terms formed by contraction with products of the metric tensor and its inverse.*

The reduction is given by the branching rule for $GL(D) \downarrow O(D)$:

$$V_\lambda^{GL(D)} = V_{\lambda/\Delta}^{O(D)} \equiv V_\lambda^{O(D)} \oplus V_{\lambda/\{2\}}^{O(D)} \oplus V_{\lambda/\{4\}}^{O(D)} \oplus V_{\lambda/\{2,2\}}^{O(D)} \oplus \dots \tag{4.7}$$

where Δ is the formal infinite sum [12]

$$\Delta = 1 + \begin{array}{|c|c|}\hline & \\ \hline \end{array} + \begin{array}{|c|c|c|c|}\hline & & & \\ \hline \end{array} + \begin{array}{|c|c|}\hline & \\ \hline & \\ \hline \end{array} + \dots$$

corresponding to the sum of all possible plethysms of the metric tensor, and where λ/μ means the sum of the Young diagrams ν such that $\nu \cdot \mu$ contains λ according to the Littlewood–Richardson rule (with the corresponding multiplicity).

Examples:
• The $GL(D)$ irreducible representation labeled by the Young diagram $\{2,2\}$ decomposes with respect to $O(D)$ according to the direct sum $\{2,2\}/\Delta = \{2,2\}+\{2,0\}+\{0,0\}$ which corresponds to the decomposition of the Riemann tensor into the Weyl tensor, the traceless part of the Ricci tensor and the scalar curvature, respectively.
• The $GL(D)$ irreducible representation labeled by the Young diagram $\{n\}$ decomposes with respect to $O(D)$ according to the direct sum $\{n\}/\Delta = \{n\}+\{n-2\}+\{n-4\}+\dots$, corresponding to the decomposition of a completely symmetric tensor or rank n into its traceless part, the traceless part of its trace, *etc.* This provides an alternative proof of the obvious fact that the number of independent components of a traceless symmetric tensor of rank n is equal to the number of independent components of a symmetric tensor of rank n minus the number

of independent components of a symmetric tensor of rank $n-2$ (its trace): $\dim V^{O(D-2)}_{\{n\}} = \dim V^{GL(D)}_{\{n\}} - \dim V^{GL(D)}_{\{n-2\}}$. Using the formula (4.6) allows to show that

$$\dim V^{O(D)}_{\{n\}} = \frac{(D+2n-2)(D+n-3)!}{n!(D-2)!}\,. \tag{4.8}$$

The very useful formula (4.8) contains as a particular case the well-known fact that all the traceless symmetric tensorial representations of $O(2)$ are two-dimensional (indeed, any UIR of an Abelian group is of complex dimension one). Moreover, the traceless symmetric tensorial representations of rank n of the rotation group $O(3)$ are the well-known integer spin representations of dimension equal to $2n+1$.

The following theorem is very important (see e.g. the first reference of [10]):

Vanishing irreps for (pseudo-)orthogonal groups: *Whenever the sum of the lengths of the first two columns of a Young diagram λ is greater than D, then the irreducible representation of $O(D)$ labeled by λ is identically zero.*

Young diagrams such that the sum of the lengths of the first two columns does not exceed D are said to be *allowed*.

Finite-dimensional irreps of (pseudo-)orthogonal groups: *Each non-zero finite-dimensional irreducible representation of $O(p,q)$ is isomorphic to a completely traceless tensorial representation, the symmetry properties of which are labeled by an allowed Young diagram λ .*

The dimension of the tensorial irrep is determined by the following rule due to King [13]:

(α) The numbers $D-1$, $D-3$, $D-5$, …, $D-2r+1$are place in the end boxes of the 1st, 2nd, 3rd, …, rth rows of the diagram λ . A labeled Young diagram of n numbers is then constructed by inserting in the remaining boxes of the diagram, numbers which increase by one in passing from one box to its left-hand neighbor.

(β) This labeled Young diagram is extended to the limit of the triangular Young diagram τ of r rows. This produces a Young diagram $\widetilde{\lambda}$ the ath row of which has length equal the maximum between the two integers $\tau_a = r-a+1$ and λ_a.

(γ) The series of numbers in any row of the Young diagram $\widetilde{\lambda}$ is then extended by inserting in the remaining boxes of the diagram, numbers which decrease by one in passing from one box to its right-hand neighbor. The resulting numbers will be called the "King length."

(δ) The row lengths λ_1, λ_2, …, λ_r are then added to all of the numbers of the Young diagram $\widetilde{\lambda}$ which lie on lines of unit slope passing through the first box of the 1st, 2nd, …, rth rows, respectively, of the Young diagram λ .

The dimension is equal to the product of the integers in the resulting labeled Young diagram $\widetilde{\lambda}$ divided by the product of

- the hook length of each box of λ, and of

- the King length of each box of $\widetilde{\lambda}$ outside λ .

Examples:

• In the following diagram, allowed for $D=5$,

$$\begin{array}{|c|c|c|c|}\hline 7&6&5&4\\ \hline 4&3&2\\ \cline{1-3} 0\\ \cline{1-1}\end{array}\quad ,$$

each box is labeled by its King length, while in the diagram

$$\begin{array}{|c|c|c|c|}\hline 11&9&6&4\\ \hline 7&4&2\\ \cline{1-3} 1\\ \cline{1-1}\end{array}\quad ,$$

each box is labeled by the number obtained at the very end of King's rule. Observe that, for the corresponding diagram λ, it was not necessary to perform the steps (β)-(γ) and that, $\dim V_\lambda^{O(5)} = 231 < 1050 = \dim V_\lambda^{GL(5)}$.

• In the following Young diagram $\lambda = \{2,2,1\}$, allowed for $D = 5$,

$$\begin{array}{|c|c|}\hline 5 & 4 \\ \hline 3 & 2 \\ \hline \multicolumn{1}{|c|}{0} & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array}\qquad ,$$

each box is labeled by the number obtained after step (α). The step (β) is now necessary and gives the Young diagram $\widetilde{\lambda} = \{3,2,1\}$. At the end of steps (γ) and (δ), respectively, the result is

$$\xrightarrow{(\gamma)}\quad \begin{array}{|c|c|c|}\hline 5 & 4 & 3 \\ \hline 3 & 2 & \multicolumn{1}{c}{} \\ \cline{1-2} \multicolumn{1}{|c|}{0} & \multicolumn{2}{c}{} \\ \cline{1-1}\end{array}\quad \xrightarrow{(\delta)}\quad \begin{array}{|c|c|c|}\hline 7 & 6 & 4 \\ \hline 5 & 3 & \multicolumn{1}{c}{} \\ \cline{1-2} \multicolumn{1}{|c|}{1} & \multicolumn{2}{c}{} \\ \cline{1-1}\end{array}\qquad ,$$

so that $\dim V_\lambda^{O(5)} = \frac{7\cdot6\cdot5\cdot4\cdot3}{(4\cdot3\cdot2)\cdot(3)} = 35 < 75 = \dim V_\lambda^{GL(5)}$.

• The space of traceless symmetric tensors of V of rank n carries an irreducible representation of $O(D)$ labeled by a Young diagram made of one row of length n for which the dimension (4.8) is easily reproduced from the King rule, since the rules (β)-(γ) may be omitted

• Computing the number of components of the Weyl tensor and of a symmetric, traceless, rank-two tensor in $D = 4$ dimensions, enables one to give the decomposition $\{2,2\}/\Delta = \{2,2\} + \{2,0\} + \{0,0\}$ of the Riemann tensor into the Weyl tensor, the traceless part of the Ricci tensor and the scalar curvature, respectively, in terms of the corresponding dimensions. This gives the well-known result $20 = 10 + 9 + 1$.

Unitary irreps of orthogonal groups: *Each non-zero inequivalent UIR of $O(D)$ corresponds to an allowed Young diagram λ, and conversely.*

Proof: The orthogonal group is compact, thence any UIR is finite-dimensional (see Subsection 1.3). Furthermore, any finite-dimensional irrep of the orthogonal group is labeled by an allowed Young diagram. Moreover, an important result is that any finite-dimensional representation may be endowed with a sesquilinear form which makes it unitary. □

The quadratic Casimir operator of the orthogonal algebra $\mathfrak{so}(D)$ presented by its generators and its commutation relations

$$i\,[M_{\mu\nu}, M_{\rho\sigma}] = \delta_{\nu\rho}M_{\mu\sigma} - \delta_{\mu\rho}M_{\nu\sigma} - \delta_{\sigma\mu}M_{\rho\nu} + \delta_{\sigma\nu}M_{\rho\mu} \tag{4.9}$$

is the sum of square of the generators (similarly to the definition (1.4) for $\mathfrak{so}(D-1,1)$ since these two *complex* algebras are isomorphic). Its eigenvalue on a finite-dimensional irrep labeled by an allowed Young diagram $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_r\}$ is given in the subsection 9.4.C of [2]:

$$\Big[\,\mathcal{C}_2\Big(\mathfrak{so}(D)\Big) - \sum_{a=1}^{r}\lambda_a(\lambda_a + D - 2a)\Big]\,V_\lambda^{O(D)} = 0\,. \tag{4.10}$$

Examples:

• The UIRs of the Abelian $O(2) \cong U(1)$ are labeled by one integer only, which is the eigenvalue of the single generator on the irrep, say $h \in \mathbb{Z}$. The only allowed Young diagrams are made of a single row of length equal to the non-negative integer $s = |h|$. The traceless symmetric tensorial representations of $O(2)$ are two-dimensional, the sum of the two irreps labeled by $h = \pm s$. The formula (4.10) with $D = 2$, $r = 1$ and $\lambda_1 = s$ gives the obvious eigenvalue s^2, since the quadratic Casimir operator of the rotation group $O(2)$ is equal to the square of the single generator.

• The quadratic Casimir operator of the rotation group $O(3)$ is the square of the angular momentum. The irrep of $O(3)$ with spin $s \in \mathbb{N}$ is labeled by the allowed Young diagram made of a single row of length equal to the integer s. The formula (4.10) with $D = 3$, $r = 1$ and $\lambda_1 = s$ gives the celebrated eigenvalue $s(s+1)$.

• The irrep of $O(D)$ carried by the space of traceless symmetric tensors of rank n is labeled by the allowed Young diagram $\{n\}$ made of a single row of length equal to an integer n. The formula (4.10) with $r = 1$ and $\lambda_1 = n$ gives the eigenvalue $n(n+D-2)$ for the quadratic Casimir operator.

The following branching rule is extremely useful in the process of dimensional reduction.

Restriction from $GL(D)$ to $GL(D-1)$: *The restriction to the subgroup $GL(D-1) \subset GL(D)$ of a finite-dimensional irrep of $GL(D)$ determined by the Young diagram λ contains each irrep of $GL(D-1)$ labeled by Young diagrams μ such that*

$$\lambda_1 \geqslant \mu_1 \geqslant \lambda_2 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_{r-1} \geqslant \lambda_r \geqslant \mu_r \geqslant 0\,,$$

with multiplicity one. The same theorem holds for the restriction $O(D) \downarrow O(D-1)$ where λ is an allowed Young diagram.

These rules are discussed in the section 8.8.A of [2]. They may be summarized in the following branching rule for $GL(D) \downarrow GL(D-1)$,

$$V_\lambda^{GL(D)} = V_{\lambda/\Sigma}^{GL(D-1)} \equiv V_\lambda^{GL(D-1)} \oplus V_{\lambda/\{1\}}^{GL(D-1)} \oplus V_{\lambda/\{2\}}^{GL(D-1)} \oplus V_{\lambda/\{3\}}^{GL(D-1)} \oplus \ldots \tag{4.11}$$

where Σ is the formal infinite sum of all Young diagrams made of a single row.

Example: The branching rule applied to symmetric irrep labeled by a Young diagram $\{n\}$ made of one row of length n gives as a result:

$$\{n\}/\Sigma \,=\, \{\,n\,\} + \{n-1\} + \{n-2\} + \ldots + \{1\} + \{0\}\,.$$

This implies the obvious fact that a completely symmetric tensor of rank n whose indices run over D values may be decomposed as a sum of completely symmetric tensors of rank n, $n-1$, $\ldots$, 1, 0 whose indices run over $D-1$ values. A non-trivial instance of the branching rule for $O(D) \downarrow O(D-1)$ is that the same result is true for *traceless* symmetric tensors as well.

4.4 Auxiliary variables

Let λ be a Young diagram with s columns and r rows.

The Schur module $V_\lambda^{GL(D)}$ in the "manifestly antisymmetric convention" can be built *via* a convenient construction in terms of polynomials in $s \times D$ graded variables satisfying appropriate conditions. More precisely, the vector space $V_\lambda^{GL(D)}$ is isomorphic to a subspace of the algebra

$$\odot^s\,\big(\wedge(V)\big) \equiv \underbrace{\wedge(V)\odot\ldots\odot\wedge(V)}_{s\text{ factors}}\,. \tag{4.12}$$

of s symmetric tensor products of antisymmetric forms $\in \wedge(V)$. The elements of $\odot^s\big(\wedge(V)\big)$ are usually called *multiforms*. The D generators of the Ith factor $\wedge(V)$ are written $d_I x^\mu$ ($I = 1, 2, \ldots, s$). By definition, the multiform algebra is presented by the graded commutation relations

$$d_I x^\mu\, d_J x^\nu \;=\; (-)^{\delta_{IJ}}\, d_J x^\nu\, d_I x^\mu\,, \tag{4.13}$$

where the wedge and symmetric products are not written explicitly. The condition (i) of Subsection 4.2 is automatically verified for any element $\Phi \in \odot^s\big(\wedge(V)\big)$ due to the fact that the variables are anticommuting in a fixed column ($I = J$). The $GL(D)$-irreducibility condition (ii) of Subsection 4.2 is implemented by the conditions

$$\Big(d_I x\cdot\frac{\partial^L}{\partial(d_J x)} \,-\, \delta_{IJ}\,\ell_I\Big)\Phi \,=\, 0\,,\quad (I \leqslant J) \tag{4.14}$$

where the dot stands for the contraction of the indices, ℓ_I for the length of the Ith column in the Young diagram λ and ∂^L stands for "left" derivative. By the Weyl construction, an element $\Phi \in \odot^s\big(\wedge(V)\big)$ satisfying (4.14) belongs to the Schur module $V_\lambda^{GL(D)}$. Following the discussion of Subsection 4.3, if λ denotes an allowed Young diagram, such an element $\Phi \in V_\lambda^{GL(D)}$ is irreducible under the (pseudo)-orthogonal group $O(p,q)$ ($p+q=D$) if it is traceless, that is

$$\Big(\frac{\partial^L}{\partial(d_I x)}\cdot\frac{\partial^L}{\partial(d_J x)}\Big)\Phi \,=\, 0\,,\quad (\forall\, I, J) \tag{4.15}$$

where the dot stands now for the contraction of indices via the use of the metric preserved by $O(p,q)$. An element $\Phi\in\odot^s\big(\wedge(V)\big)$ such that (4.14)-(4.15) are fulfilled belongs to the Schur module $V_\lambda^{O(p,q)}$ labeled by the Young diagram λ.

The Schur module $V_\lambda^{GL(D)}$ admits another convenient realization in terms of polynomials in $r\times D$ commuting variables. In other words, the vector space $V_\lambda^{GL(D)}$ is isomorphic to a subspace of the polynomial algebra in the variables u_a^μ $(a=1,2,\dots,r)$ where the index a corresponds to each row. The condition (a) of Subsection 4.2 is automatically verified for any such polynomial due to the fact that the variables are commuting in a fixed row. The $GL(D)$-irreducibility condition (b) of Subsection 4.2 is implemented by the conditions

$$\Big(u_a\cdot\frac{\partial}{\partial u_b}-\delta_{ab}\,\lambda_a\Big)\Phi\,=\,0\,,\quad(a\leqslant b)\tag{4.16}$$

where the dot still stands for the contraction of the indices. The degree of homogeneity of the polynomial Φ in the variables u_a^μ (for fixed a) is λ_a. The corresponding coefficients are tensors irreducible under the general linear group. By the Weyl construction, a polynomial $\Phi(u_a)$ satisfying (4.16) belongs to the Schur module $V_\lambda^{GL(D)}$. Again, such an element $\Phi\in V_\lambda^{GL(D)}$ is irreducible under the (pseudo)-orthogonal group $O(p,q)$ $(p+q=D)$ iff it is traceless, that is

$$\Big(\frac{\partial}{\partial u_a}\cdot\frac{\partial}{\partial u_b}\Big)\Phi\,=\,0\,,\quad(\forall\,a,b)\tag{4.17}$$

where the dot stands for the contraction of indices via the use of the metric preserved by $O(p,q)$. A polynomial $\Phi(u_a)$ such that (4.16)-(4.17) are fulfilled belongs to the Schur module $V_\lambda^{O(p,q)}$ labeled by an allowed Young diagram λ.

Example: Consider an irreducible representation of the orthogonal group $O(D)$ labeled by the Young diagram $\{n\}$ made of a single row of length equal to an integer n. The polynomial $\Phi(u)\in V_{\{n\}}^{O(D)}$ obeys to the irreducibility conditions

$$\Big(u\cdot\frac{\partial}{\partial u}-n\Big)\Phi\,=\,0\,,\qquad\Big(\frac{\partial}{\partial u}\cdot\frac{\partial}{\partial u}\Big)\Phi\,=\,0\,.\tag{4.18}$$

They mean that the polynomial is homogeneous (of degree equal to n) and harmonic, so that its components correspond to a symmetric traceless tensor of rank n:

$$\Phi(u)\;=\;\frac{1}{n!}\Phi_{\mu_1\dots\mu_n}\,u^{\mu_1}\dots u^{\mu_n}\,,\qquad\delta^{\mu_1\mu_2}\Phi_{\mu_1\mu_2\mu_3\dots\mu_n}=0\,.$$

Of course the integral of the square of such a polynomial over $\mathbb{R}^D$ is, in general, infinite. But the restriction of an harmonic polynomial on the unit sphere $\overrightarrow{u}^2=1$ is square integrable on S^{D-1}. This restriction is called a *spherical harmonic* of degree n. Therefore the space of spherical harmonics of degree n provides an equivalent realization of the Schur module $V_{\{n\}}^{O(D)}$. For $D=3$, the space $V_{\{n\}}^{O(3)}$ is spanned by the usual spherical harmonics $Y_n^m(\theta,\phi)$ on the two-sphere with $|m|\leqslant n$.

Remarks:

• The infinitesimal generators of the pseudo-orthogonal group $O(p,q)$ are represented by the operators

$$M_{\mu\nu}=i\sum_{a=1}^r u_a^\rho\Big(g_{\rho\mu}\frac{\partial}{\partial u_a^\nu}-g_{\rho\nu}\frac{\partial}{\partial u_a^\mu}\Big)\,.$$

Reordering the factors and making use of (4.16)-(4.17) allows to reproduce the formula (4.10) for the eigenvalues of the quadratic Casimir operator.

• Instead of polynomial functions in the commuting variables, one may equivalently consider *distributions* obeying to the same conditions. The space of solutions would carry an equivalent irrep, as follows from the highest-weight construction of the representation. However, it does not make sense any more of talking about the "coefficients" of the homogeneous distribution so that the link with the equivalent tensorial representation is more intricate.

The example of the spherical harmonics suggests that it might be convenient to realize any unitary module of the orthogonal group $O(D)$ as a space of functions on the unit hypersphere

S^{D-1} satisfying some linear differential equations. Better, the symmetry under the orthogonal group would be made manifest by working with homogeneous harmonic functions on the *ambient space* $\mathbb{R}^D$, evaluated on any hypersphere $S^{D-1}\subset\mathbb{R}^D$.

Spherical harmonics: *To any UIR of the isometry group $O(D)$ of a hypersphere S^{D-1}, one may associate manifestly covariant differential equations for functions on S^{D-1} embedded in $\mathbb{R}^D$ whose space of solutions carry the corresponding UIR.*

<u>Proof:</u> Any UIR of the isometry group $O(D)$ corresponds to a Schur module $V_\lambda^{O(D)}$ which may be realized as the space of polynomials $\Phi(\overrightarrow{u}_a)$ such that (4.16)-(4.17) are obeyed. Let us introduce the notation: $\overrightarrow{x}:=\overrightarrow{u}_1$ and $\overrightarrow{t}_{a-1}:=\overrightarrow{u}_a$ for $a=2,\dots,r$. One interprets the polynomial $\Phi(\overrightarrow{x},\overrightarrow{t}_{\underline{a}})$ (where the index $\underline{a}$ runs from 1 to $r-1$) as a tensor field on the Euclidean space $\mathbb{R}^D$ parametrized by the Cartesian coordinates $\overrightarrow{x}$, with some auxiliary variables $\overrightarrow{t}_{\underline{a}}$ implementing the tensor components. The conditions (4.16)-(4.17) for a and b strictly greater than 1 imply that

$$\Big(t_{\underline{a}}\cdot\frac{\partial}{\partial t_{\underline{b}}}-\delta_{\underline{ab}}\,\lambda_{\underline{a}}\Big)\Phi\,=\,0\,,\quad(\underline{a}\leqslant\underline{b})\qquad\Big(\frac{\partial}{\partial t_{\underline{a}}}\cdot\frac{\partial}{\partial t_{\underline{b}}}\Big)\Phi\,=\,0\,,\tag{4.19}$$

where $\underline{\lambda}=\{\lambda_2,\dots,\lambda_r\}$ is the Young diagram obtained from λ by removing its first row. Thus the components of the "tensor field" $\Phi(\overrightarrow{x},\overrightarrow{t}_{\underline{a}})$ carry an irreducible representation of $O(D)$ labeled by $\underline{\lambda}$. The conditions (4.16) for $a=b=1$ imply that

$$(x\cdot\frac{\partial}{\partial x}-\lambda_1)\Phi=0\,,$$

so the polynomial $\Phi(\overrightarrow{x},\overrightarrow{t}_{\underline{a}})$ is homogeneous of degree λ_1 in the radial coordinate $|\overrightarrow{x}|$. The condition (4.17) for $a=b=1$ is interpreted as the Laplace equation

$$\Big(\frac{\partial}{\partial x}\cdot\frac{\partial}{\partial x}\Big)\Phi\,=\,0\tag{4.20}$$

on the ambient space $\mathbb{R}^D$, it imples that the tensor field Φ is harmonic in ambient space. The condition (4.16) for $b>a=1$ states that the radial components vanish,

$$\Big(x\cdot\frac{\partial}{\partial t_{\underline{a}}}\Big)\Phi\,=\,0\,,\tag{4.21}$$

so the tensor components are longitudinal to the hyperspheres S^{D-1} . Therefore the evaluation of the non-vanishing components of $\Phi(\overrightarrow{x},\overrightarrow{t}_{\underline{a}})$ on the unit hypersphere $|\overrightarrow{x}|=1$ is an *intrinsic* tensor field living on the hypersphere S^{D-1} and whose tensor components carry an irrep of the stability subgroup $O(D-1)$ labeled by $\underline{\lambda}$. These tensor fields generalize the spherical harmonics to the generic case $r\geqslant 1$. Finally, the condition (4.17) for $b>a=1$ states that the tensor field is divergenceless in ambient space,

$$\Big(\frac{\partial}{\partial x}\cdot\frac{\partial}{\partial t_{\underline{a}}}\Big)\Phi\,=\,0\,.\tag{4.22}$$

The differential equations (4.20) and (4.22) are written in ambient space but they may be reformulated in intrinsic terms on the hypersphere, at the price of losing the manifest covariance under the full isometry group $O(D)$. □

4.5 Euclidean group

The method of induced representations was introduced in Subsection 3.1 for the Poincaré group $ISO(D-1,1)^\uparrow$ and applied to the Euclidean group $ISO(D-2)$ in Subsection 3.3. Focusing on the genuine (*i.e.* with a non-trivial action of the translation generators) irreps of the *inhomogeneous* orthogonal group, all of them are induced from an UIR of the stability subgroup. Using the results of the previous section 4.3, one may summarize the final result into the following classification.

Unitary irreps of the inhomogeneous orthogonal groups: *Each inequivalent UIR of the group $IO(D)$ with a non-trivial action of its Abelian normal subgroup is associated with a positive real number μ and an allowed Young diagram of the subgroup $O(D-1)$, and conversely.*

The orbits of the linear action of the orthogonal group $O(D)$ on the Euclidean space $\mathbb{R}^D$ are the hyperspheres S^{D-1} of radius R. The isometry group of any such hypersphere S^{D-1} is precisely $O(D)$. Considering a region of fixed size on these hyperspheres, in the limit $R\to\infty$ the sphere becomes a hyperplane $\mathbb{R}^{D-1}$. Therefore the homogeneous and inhomogeneous orthogonal groups are related by some infinite radius limit: $O(D)\to IO(D-1)$. Such a process is frequently referred to as an *Inönü-Wigner contraction* in the physics literature [14]. This is better seen at the level of the Lie algebra. Specializing the Dth directions, the commutation relations (4.9) take the form

$$
\begin{aligned}
i\,[M_{mn},M_{pq}] &= \delta_{np}M_{mq}-\delta_{mp}M_{nq}-\delta_{qm}M_{pn}+\delta_{qn}M_{pm}\,, &\quad (4.23)\\
i\,[M_{mD},M_{pq}] &= \delta_{mn}M_{pD}-\delta_{mp}M_{nD}\,, &\quad (4.24)\\
i\,[M_{mD},M_{pD}] &= M_{pm}\,. &\quad (4.25)
\end{aligned}
$$

where the latin letters take $D-1$ values. Defining $M_{mD}\,=\,R\,P_m$ and taking the limit $R\to\infty$ (with P_m fixed) in the relations (4.23)-(4.25) lead to

$$
\begin{aligned}
i\,[M_{mn},M_{pq}] &= \delta_{np}M_{mq}-\delta_{mp}M_{nq}-\delta_{qm}M_{pn}+\delta_{qn}M_{pm}\,, &\quad (4.26)\\
i\,[P_m,M_{pq}] &= \delta_{mn}P_p-\delta_{mp}P_n\,, &\quad (4.27)\\
i\,[P_m,P_p] &= 0\,. &\quad (4.28)
\end{aligned}
$$

As can be seen, the generators $\{M_{mn},P_m\}$ span the Lie algebra of the inhomogeneous orthogonal group $IO(D-1)$. The former argument proves the contraction $\mathfrak{so}(D)\to\mathfrak{iso}(D-1)$.

The limit of a sequence of irreps of the homogeneous orthogonal group $O(D)$, in which one performs an Inönü-Wigner contraction, is automatically a representation of the inhomogeneous orthogonal group $IO(D-1)$ (if the limit is not singular). An interesting issue is the inverse problem: which irreps of $IO(D-1)$ may be obtained as the limit of such a sequence of irreps of $O(D)$? The problem is non-trivial because, generically, the limit of a sequence of irreps is a *reducible* representation.

Contraction of UIRs of the homogeneous orthogonal groups: *Each inequivalent UIR of the group $IO(D-1)$ with a non-trivial action of its Abelian normal subgroup may be obtained as the contraction of a sequence of UIRs of the group $O(D)$.*

More precisely, the Inönü-Wigner contraction $R\to\infty$ of a sequence of UIRs of $O(D)$, labeled by allowed Young diagrams $\nu=\{s,\lambda_1,\ldots,\lambda_r\}$ such that the limit of the quotient s/R is a fixed positive real number μ, is the UIR of $IO(D-1)$ labeled by the parameter μ and the Young diagram $\lambda=\{\lambda_1,\ldots,\lambda_r\}$.

Proof: The use of the spherical harmonics construction discussed at the end of Subsection 4.4 is very convenient here. The main idea is to solve the homogeneity condition in a neighborhood of $x^D\neq 0$ as follows:

$$\Phi(x^m,x^D,t_{\underline{a}})\;=\;z^s\,\phi\left(\frac{x^m}{z},t_{\underline{a}}\right)\,, \qquad (4.29)$$

where $\overrightarrow{x}\,=\,(x^m,x^D)$ and $\phi(y^m,t_{\underline{a}}):=\Phi(y^m,\frac{s}{\mu},t_{\underline{a}})$. In other words, one may perform a convenient change of coordinates from the *homogenous coordinates* (x^m,x^D) to the set (y^m,z) where

$$y^m=\frac{x^m}{z}$$

are the *inhomogenous coordinates* (on the projective space $\mathbb{PR}^{D-1}$ minus the point at infinity $z=0$) and

$$z=\frac{\mu\,x^D}{s}$$

is a scale variable.. The magic is that the equations for the generalized spherical harmonics have a well-behaved limit $x^D\to\infty$ in terms of $\phi(y^m,t_{\underline{a}})$ when x^D/s is fixed to be equal to the ratio z/μ, where z and μ are finite [15]. To see that, one should use the relations

$$
\begin{aligned}
\frac{\partial}{\partial x^m} &= \frac{1}{z}\,\frac{\partial}{\partial y^m}\,,\\
\frac{\partial}{\partial x^D} &= \frac{\mu}{s}\left(\frac{\partial}{\partial z}-\frac{1}{z}\,y^m\frac{\partial}{\partial y^m}\right)\,. \qquad (4.30)
\end{aligned}
$$

Moreover, the equation in tis limit may be identified with equations for the proper UIR of the inhomogeneous orthogonal group $IO(D-1)$ realized homogeneously in terms of the inhomogenous coordinates. □

Example: The simplest instance is when $\lambda = \{0\}$ because one considers the sequence of harmonic functions $\Phi(x^m, x^D)$ of homogeneity degree s. The Laplace operator acting on $\Phi(x^m, x^D)$ reads in terms of $\phi(y^m)$ as follows

$$\Delta_{\mathbb{R}^D}\Phi = z^{s-2}\left[\frac{\partial}{\partial y}\cdot\frac{\partial}{\partial y} + \frac{\mu^2}{s^2}\left(s(s-1) - (2s-1)\left(y\cdot\frac{\partial}{\partial y}\right) + \left(y\cdot\frac{\partial}{\partial y}\right)^2\right)\right]\phi\,,$$

due to the homogeneity condition (4.29) and the relations (4.30). The Laplace equation $\Delta_{\mathbb{R}^D}\Phi = 0$ is thus equivalent to the equation

$$\left[\frac{\partial}{\partial y}\cdot\frac{\partial}{\partial y} + \frac{\mu^2}{s^2}\left(s(s-1) - (2s-1)\left(y\cdot\frac{\partial}{\partial y}\right) + \left(y\cdot\frac{\partial}{\partial y}\right)^2\right)\right]\phi = 0\,,$$

whose limit for $s\to\infty$ is the Helmholtz equation $[\,\Delta_{\mathbb{R}^{D-1}} + \mu^2\,]\,\phi = 0$, where $\Delta_{\mathbb{R}^{D-1}} = \frac{\partial}{\partial y}\cdot\frac{\partial}{\partial y}$. The space of solutions of the Helmholtz equation carries an UIR of $IO(D-1)$ induced from a trivial representation of the stability subgroup $O(D-2)$.

5 Relativistic wave equations

The *Bargmann-Wigner programme* amounts to associating, with any given UIR of the Poincaré group, a manifestly covariant differential equation whose positive-energy solutions transform according to the corresponding UIR. Physically, it might be natural to restrict this programme to the two most important classes of UIRs: the massive and massless representations. Mathematically, this restriction is convenient because the group-theoretical analysis is simpler since any of these UIRs is induced from an UIR of a unimodular orthogonal group $SO(n)$ (with $D-3 \leqslant n \leqslant D-1$), as can be checked easily on the tables of Subsection 3.3.

In 1948, this restricted programme was completed by Bargmann and Wigner in four dimensions when, for each such UIR of $ISO(3,1)^\uparrow$, a relativistic wave equation was written whose positive-energy solutions transform according to the corresponding UIR [4]. But this case ($D=4$) will not be reviewed here in details because it may cast shadow on the generic case. Indeed, it is rather peculiar in many respects:

- The quadratic and quartic Casimir operators essentially classify the UIRs, but this is no more true in higher dimensions where more Casimir operators are necessary and the classification quickly becomes technically cumbersome in this way. Moreover, one should stress that the eigenvalues of the Casimir operators do not characterize uniquely an irreducible representation (for instance, the quadratic and quartic Casimir operators vanish for all helicity representations).
- The (complex) Lorentz algebra $\mathfrak{so}(3,1)$ is isomorphic to the direct sum of two (complex) rotation algebras $\mathfrak{so}(3)\cong\mathfrak{sp}(2)$. These isomorphisms allow the use of the convenient "dotted-undotted" formalism for the finite-dimensional (non-unitary) irreps of the spin group $Spin(3,1)$.
- The symmetric tensor-spinor fields are sufficient to cover all inequivalent cases.
- The helicity short little group $SO(2)$ is Abelian, therefore its irreps are one-dimensional, for fixed helicity. Notice that the helicity is discretized because the representation of the "little group" $SO(2)$ is a restriction of the representation of the group $Spin(3)\cong SU(2)$ which has no intrinsically projective representations.
- The infinite-spin short little group $SO(1)$ is trivial, thus there are only two inequivalent infinite-spin representations (single- or double-valued) [6].
- *etc.*

Moreover, there exists an extensive literature on the subject of UIRs of $ISO(3,1)^\uparrow$ and we refer to the numerous pedagogical reviews available for more details on the four-dimensional case (see *e.g.* the inspiring presentations of [5] and [16]).

It is standard to require time reversal and parity symmetry of the field theory. More precisely, the wave equations we will consider are covariant under the two previous transformations. As a consequence of the time reversal symmetry, the representation is *irreducible* under the group $ISO(D-1,1)$ but *reducible* under the Poincaré group $ISO(D-1,1)^\uparrow$: the Hilbert space of solutions contain both positive and negative energy solutions. Furthermore, the parity symmetry implies that the representation is *irreducible* under the inhomogeneous Lorentz group $IO(D-1,1)$ but *reducible* under the group $ISO(D-1,1)$ (for instance, both chiralities are present in the massless case for D even). To conclude, the Bargmann-Wigner programme is actually understood as associating, with any given UIR of the inhomogeneous Lorentz group, a manifestly covariant differential equation whose solutions transform according to the corresponding UIR.

5.1 General procedure

The lesson on induced representations that we learned from Wigner implies the following strategy:

1. Pick a unitary representation of the (short) little group.

2. Introduce a wave function on $\mathbb{R}^{D-1,1}$ taking values in some (possibly non-unitary) representation of the Lorentz group $O(D-1,1)$ the restriction of which to the (short) little group contains the representation of step 1.

3. Write a system of linear covariant equations, differential in position space x^μ thus algebraic in momentum space p_ν , for the wave function of step 2. These equations may not be independent.

4. Fix the momentum and check in convenient coordinates that the wave equations of step 3 put to zero all "unphysical" components of the wave function. More precisely, verify that its non-vanishing components carry the unitary representation of step 1.

Proof: The fact that the set of linear differential equations is taken to be manifestly covariant ensures that the Hilbert space of their solutions carries a (infinite-dimensional) representation of $IO(D-1,1)$. The fourth step determines the representation of the little group by which it is induced. □

In the physics literature, the fourth step is referred to as "looking at the physical degrees of freedom." If the (possibly reducible) representation is proven to be unitary, then this property is summarized in a "no-ghost theorem."

The Klein-Gordon equation $(p^2 \pm m^2)\Psi = 0$ is always, either present in the system of covariant equations or a consequence thereof. Consequently, the Klein-Gordon equation will be assumed implicitly from now on in the step 3. Therefore, the step 4 will be immediately performed in a proper Lorentz frame. (We refer the reader to the Subsection 3.2 for more details.)

The completion [17] of the Bargmann-Wigner programme for finite-component representations is reviewed in the subsections 5.2 and 5.3 for single-valued UIRs of the Poincaré group.[6] The tachyonic case[7] is more briefly discussed in Subsection 5.4. The zero-momentum representation is not considered here since it essentially is a unitary representation of the de Sitter spacetime dS_{D-1}.

5.2 Massive representations

The Bargmann-Wigner programme is easy to complete for massive UIRs because the massive stability subgroup is the orthogonal group $O(D-1) \subset O(D-1,1)$. By going to a rest-frame,

[6] Spinorial irreps may be adressed analogously by supplementing the system of differential equations with Dirac-like equations and gamma-trace constraints (see *e.g.* [15, 18] for more details). The Bargmann-Wigner programme has been completed for anyonic representations in three-dimensional Minkowski spacetime [19].

[7] The discussion presented in the section 5.4 was not published before, it directly derives from private conversations between X.B. and J. Mourad.

the time-like momentum vector takes the form $p^\mu = (m, \overrightarrow{0}) \neq 0$. The physical components of the field are thus carrying a tensorial irrep of the group $O(D-1)$ of orthogonal transformations in the spatial hyperplane $\mathbb{R}^{D-1}$. In other words, the linear wave equations should remove all components including time-like directions. These unphysical components are responsible for the fact that the Fock space is not endowed with a positive-definite norm.

Step 1. From the sections 1.3 and 4, one knows that any unitary representation of the orthogonal group $O(D-1)$ is a sum of UIRs which are finite-dimensional and thus, equivalent to a tensorial representation. Let us consider the UIR of $O(D-1)$ labeled by the allowed Young diagram $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_r\}$ (*i.e.* the sum of the lengths of its first two columns does not exceed $D-1$).

Step 2. The simplest way to perform the Bargmann-Wigner programme in the massive case is to choose a covariant wave function whose components carry the (finite-dimensional and non-unitary) tensorial irrep of the Lorentz group $O(D-1,1)$ labeled by the Young diagram λ. As explained in the subsection 4.4, a convenient way of realizing this is in terms of a wave function $\Phi(p, u_a)$ polynomial in the auxiliary commuting variables u^μ_a satisfying the irreducibility conditions (4.16)-(4.17).

Step 3. The massive Klein-Gordon equation has to be supplemented with the transversality conditions

$$\Big(p \cdot \frac{\partial}{\partial u_a}\Big)\Phi \,=\, 0\,, \tag{5.1}$$

of the wave function.

Step 4. Looking at a fixed-momentum mode in its corresponding rest-frame $p^\mu = (m, \overrightarrow{0})$ leads to the fact that the components of the wave function along the timelike momentum are set to zero by (5.1): $\Phi = \Phi(p, \overrightarrow{u}_a)$. In words, Φ does not depend on the time components u^0_a, $\forall\, a$. In this case, the conditions (4.16)-(4.17) read as irreducibility conditions under the orthogonal group $O(D-1)$. □

Example: Massive symmetric representations with "spin" equal to s correspond to Young diagrams $\lambda = \{s\}$ made of one row of length equal to the integer s. In four spacetime dimensions, this representation is precisely what is usually called a "massive spin-s field."[8] The covariant wave function $\Phi(p, u)$ obeys to the irreducibility conditions (4.16)-(4.17) of the components

$$\Big(u \cdot \frac{\partial}{\partial u} - s\Big)\Phi \,=\, 0\,, \qquad \Big(\frac{\partial}{\partial u} \cdot \frac{\partial}{\partial u}\Big)\Phi \,=\, 0\,. \tag{5.2}$$

The wave function Φ is homogeneous of degree s and harmonic in the auxiliary variable u. If the wave function $\Phi(p, u)$ is polynomial in the auxiliary variable u, then its components correspond to a symmetric traceless tensor of rank s

$$\Phi(p, u) \,=\, \frac{1}{s!}\Phi_{\mu_1\dots\mu_s}(p)\, u^{\mu_1}\dots u^{\mu_s}\,, \qquad \eta^{\mu_1\mu_2}\Phi_{\mu_1\mu_2\mu_3\dots\mu_s}(p) = 0\,.$$

The covariant wave equations are the massive Klein-Gordon equation together with the transversality condition

$$\Big(p \cdot \frac{\partial}{\partial u}\Big)\Phi \,=\, 0\,, \tag{5.3}$$

which reads in components as

$$p^{\mu_1}\Phi_{\mu_1\mu_2\dots\mu_s}(p) = 0\,. \tag{5.4}$$

The non-vanishing components of a solution of (5.4) must be along the spatial directions, *i.e.* only $\Phi_{i_1\dots i_s}(p)$ may be $\neq 0$. This symmetric tensor field is traceless with respect to the spatial metric: $\delta^{i_1i_2}\Phi_{i_1i_2i_3\dots i_s}(p) = 0$, thus the physical components carry a symmetric irrep of the orthogonal group $O(D-1)$, the dimension of which can be computed by making use of the formula (4.8). The polynomial wave function $\Phi(p, u)$ evaluated on the internal unit hypersphere $u^iu_i = 1$ corresponds to a decomposition of the physical components in terms of the spherical harmonics on the internal hypersphere S^{D-2}, which is an equivalent, though rather unusual, way of representing the physical components (usually, the use of spherical harmonics is reserved to the "orbital" part of the wave function).

[8]To our knowledge, the Bargmann-Wigner programme for the massive integer-spin representations in four-dimensional Minkowski spacetime was adressed along the lines reviewed here for the first time by Fierz in [20].

The quartic Casimir operator of the Poincaré algebra is easily evaluated in components in the rest frame

$$\begin{aligned}-\tfrac{1}{2}\quad & P^2 M_{\mu\nu}M^{\mu\nu} \ + \ M_{\mu\rho}P^\rho\, M^{\mu\sigma}P_\sigma \\ &= \frac{1}{2}\, m^2(M_{ij}M^{ij} + 2M_{i0}M^{i0}) \ - \ m^2 M_{i0}M^{i0} \ = \ m^2\,\frac{1}{2}\, M_{ij}M^{ij}\,,\end{aligned}$$

giving as a final result for a massive representation associated with a Young diagram λ

$$\begin{aligned}\mathcal{C}_4\Big(\mathfrak{iso}(D-1,1)\Big) \quad &= \quad \mathcal{C}_2\Big(\mathfrak{iso}(D-1,1)\Big)\,\mathcal{C}_2\Big(\mathfrak{so}(D-1)\Big)\,, \\ &= \quad m^2\sum_{a=1}^{r}\lambda_a(\lambda_a + D - 2a - 1)\,, \end{aligned} \tag{5.5}$$

where the eigenvalues of the quadratic Casimir operator of the rotation algebra are given by the formula (4.10).

Example: In any dimension D, the eigenvalue of the quartic Casimir operator for a massive symmetric representation of rank s is equal to $m^2\, s(s + D - 3)$. In four spacetime dimensions, the square of the Pauli-Lubanski vector acting on a massive field of spin-s is indeed equal to $m^2\, s(s + 1)$.

Each massive representation in $D \geqslant 4$ dimensions may actually be obtained as the first Kaluza–Klein mode in a dimensional reduction from $D + 1$ down to D dimensions. There is no loss of generality because the massive little group $SO(D - 1)$ in D dimension is identified with the $(D + 1)$-dimensional helicity (short) little group. Such a Kaluza–Klein mechanism leads to a Stückelberg formulation of the massive field.

The massless limit $m \to 0$ of a massive irrep with λ fixed is, in general, reducible because the irrep of the massive little group $SO(D - 1)$ is restricted to the helicity (short) little group $SO(D-2) \subset SO(D-1)$. This argument combined with the known branching rule for $O(D-1) \downarrow O(D - 2)$ (reviewed in Subsection 4.3) allows to prove that the massless limit of a massive irrep of the homogeneous Lorentz group labeled by a fixed Young diagram λ contains each helicity irrep labeled by Young diagrams μ such that

$$\lambda_1 \geqslant \mu_1 \geqslant \lambda_2 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_{r-1} \geqslant \lambda_r \geqslant \mu_r \geqslant 0\,,$$

with multiplicity one. The zero modes of a dimensional reduction from $D + 1$ down to D dimensions are determined by the same rule.

Example: The zero modes of the dimensional reduction of a massive symmetric representations with "spin" equal to s are all helicity symmetric representations with integer "spins" not greater than the integer s, each with multiplicity one. For the dimensional reduction of a gravitational theory (*i.e.* a spin-two particle), one recovers the usual result that the massless spectrum is made of one "graviton" (spin-2), one "photon" (spin-1) and one "dilaton" (spin-0).

5.3 Massless representations

The quartic Casimir operator of the Poincaré algebra is evaluated easily in components in the light-cone coordinates (see Subsection 3.2 for notations),

$$-\,\frac{1}{2}\,P^2\, M_{\mu\nu}M^{\mu\nu} \ + \ M_{\mu\rho}P^\rho\, M^{\mu\sigma}P_\sigma \ = \ 0 \ + \ M_{m+}P^+\, M^{m-}P_- \ = \ \pi_m\pi^m\,,$$

giving as a final result for a massless representation

$$\mathcal{C}_4\Big(\mathfrak{iso}(D-1,1)\Big) \ = \ \mathcal{C}_2\Big(\mathfrak{iso}(D-2)\Big) \ = \ \mu^2 \tag{5.6}$$

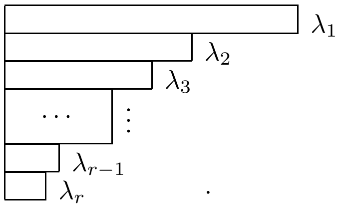
where the quadratic Casimir operator of the massless little group is written in (3.9).

Helicity representations

Helicity representations correspond to the case $\mu = 0$, so that $\pi^m = 0$ and in practice the representation is induced from a representation of the orthogonal group $O(D-2)$.

Step 1. Again, any unitary representation of the orthogonal group $O(D-2)$ is a sum of finite-dimensional UIRs. Let us consider the UIR of the helicity short little group $O(D-2)$ labeled by the allowed Young diagram $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_r\}$ (that is, the sum of the lengths of its first two columns does not exceed $D-2$):

$\lambda =$



. (5.7)

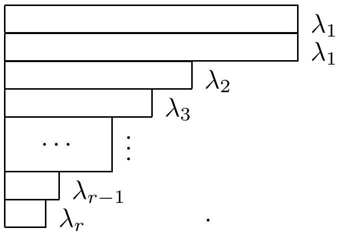
The step 2 is more subtle to perform than for massive representations because the field equations must set to zero all components along the light-cone of the covariant wave function, because they are unphysical. In other words, the covariant wave equations should remove *two* directions, and not only one like in the massive case. This fact implies that the transversality is not a sufficient condition any more, it must be supplemented either by other equations or by gauge symmetries asserting that one may quotient the solution space by pure gauge fields. In these lecture notes, one focuses on two *gauge-invariant* formulations which may be respectively referred to as "Bargmann-Wigner formulation" in terms of the fieldstrength and "gauge-fixed formulation" in terms of the potential.

Bargmann-Wigner equations

The so-called "Bargmann-Wigner equations" were actually first written by Dirac [21] in four-dimensional Minkowski spacetime in spinorial form. Their name originates from their decisive use in the completion of the Bargmann-Wigner programme [4]. The generalization of the Bargmann-Wigner equations to any dimension was presented in [17] for tensorial irreps (reviewed here) and in [18] for spinorial irreps.

Step 2. Let $\overline{\lambda} = \{\lambda_1, \lambda_1, \lambda_2, \ldots, \lambda_r\}$ be the Young diagram depicted as

$\overline{\lambda} =$



. (5.8)

It is obtained from the Young diagram λ represented in (5.7) by adding a row of equal length on top of the first row of λ. The Young diagram $\overline{\lambda}$ has at least two rows of equal lengths and the sum of the lengths of its first two columns does not exceed D. The covariant wave function is chosen to take values in the Schur module $V_{\overline{\lambda}}^{O(D-1,1)}$ realized in the manifestly antisymmetric convention. Following Subsection 4.4, the wave function $\mathcal{K}(p, d_{_I}x)$ is taken to be a polynomial in the graded variables $d_{_I}x^\mu$ ($I = 1, 2, \ldots, \lambda_1$) obeying the commutation relations (4.13). Moreover, the irreducibility conditions of the components under the Lorentz group $O(D-1,1)$ are

$$\Big(d_{_I}x^\mu \frac{\partial^L}{\partial(d_{_J}x^\mu)} - \delta_{_{IJ}}\,\overline{\ell}_{_I} \Big)\mathcal{K} = 0\,, \quad (I \leqslant J) \tag{5.9}$$

where $\overline{\ell}_I$ stands for the length of the Ith column in the Young diagram $\overline{\lambda}$, and

$$\Big(\eta^{\mu\nu} \frac{\partial^L}{\partial(d_{_I}x^\mu)} \frac{\partial^L}{\partial(d_{_J}x^\nu)} \Big)\mathcal{K} = 0\,. \tag{5.10}$$

Step 3. The covariant wave equations may be summarized in the assertion that the wave function is a "harmonic" multiform in the sense that, $\forall\, I$, it is "closed"

$$\Big(p_\mu\, d_{_I}x^\mu \Big)\mathcal{K} = 0\,, \tag{5.11}$$

and "coclosed" (*i.e.* transverse)

$$\Big(p^\mu \frac{\partial^L}{\partial(d_{_I}x^\mu)} \Big)\mathcal{K} = 0\,. \tag{5.12}$$

The operators $p \cdot d_I x$ act as "exterior differentials" (or "curls"), they are nilpotent and obey graded commutation relations. As one can easily see, the wave equations (5.11) and (5.12), considered together, imply the massless Klein-Gordon equation. Actually, the equations (5.11) may even be imposed off-shell, whereas the equations (5.12) only hold on-shell [17].

Step 4. In the light-cone frame (see Section 1.1), the components of the momentum may be taken to be $p_\mu = (p_-, 0, 0, \ldots, 0)$ with $p_- \neq 0$. On the one hand, the transversality condition (5.12) implies that the wave function does not depend on the variables $d_I x^+$. On the other hand, the closure condition (5.11) reads $(p_- d_I x^-)\, \mathcal{K} = 0$, the general solution of which is $\mathcal{K} = (\prod_I p_- d_I x^-)\, \phi$, where ϕ depends neither on $d_I x^-$ nor on $d_I x^+$ (due to the transversality condition). In other words, the directions along the light-cone have been removed, since $\phi = \phi(p, d_I x^m)$ $(m = 1, 2, \ldots, D-2)$. Focusing on this field, one may show that the irreducibility conditions (5.9) become, in terms of the function ϕ,

$$\Big(d_I x^m \frac{\partial^L}{\partial(d_J x^m)} - \delta_{IJ}\, \ell_I \Big) \phi = 0\,, \quad (I \leqslant J) \tag{5.13}$$

where $\ell_I = \overline{\ell}_I - 1$, and the trace conditions (5.10) implies

$$\Big(\delta^{mn} \frac{\partial^L}{\partial(d_I x^m)} \frac{\partial^L}{\partial(d_J x^n)} \Big) \phi = 0\,. \tag{5.14}$$

Since ℓ_I is the length of the Ith column of the Young diagram λ, the system of equations (5.13)-(5.14) states that the components of the function ϕ carry a tensorial irrep of the orthogonal group $O(D-2)$. Therefore, the same is true for the physical components of the wave function $\mathcal{K}$. □

This may be reformulated covariantly by saying that the closure (5.11) of the wave function implies that

$$\mathcal{K} = \Big(\prod_{I=1}^{\lambda_1} p_\mu d_I x^\mu \Big)\, \phi\,. \tag{5.15}$$

In components, this means that the tensor $\mathcal{K}$ is equal to λ_1 curls of the tensor ϕ. This motivates the name "fieldstrength" for the wave function $\mathcal{K}(p, d_I x)$, the components of which are irreducible under the Lorentz group (when evaluated on zero-mass shell) and labeled by $\overline{\lambda}$, and the name "potential" or "gauge field" for the wave function $\phi(p, d_I x)$, the components of which may be taken to be irreducible under the general linear group, with symmetries labeled by the Young diagram λ.

Examples:

• The helicity vectorial representation corresponds to a Young diagram $\lambda = \{1\}$ made of a single box. In four spacetime dimensions, this representation is precisely what is usually called a "vector gauge field". The Young diagram $\overline{\lambda} = \{1, 1\}$ is a single column made of two boxes. The wave function in momentum space is given by

$$\mathcal{K} = \frac{1}{2}\, \mathcal{K}_{\mu\nu}(p)\, dx^\mu dx^\nu$$

which carries an irrep of $GL(D, \mathbb{R})$: the antisymmetric rank-two representation. As one can see, the wave function actually is a differential two-form, the components of which transforming as an antisymmetric tensor of rank two. The wave equations (5.11) and (5.12), respectively, read in components

$$p_\mu \mathcal{K}_{\nu\rho} + p_\nu \mathcal{K}_{\rho\mu} + p_\rho \mathcal{K}_{\mu\nu} = 0 \quad \text{(Bianchi identities)}$$

and

$$p^\mu \mathcal{K}_{\mu\nu} = 0 \qquad \text{(transversality conditions)}\,.$$

The differential two-form $\mathcal{K}$ is indeed harmonic (closed and coclosed). In physical terms, one says that the fieldstrength $\mathcal{K}_{\mu\nu}$ obeys to the Maxwell equations. As usual, the Bianchi identities imply that the fieldstrength derives from a potential: $\mathcal{K}_{\mu\nu} = p_\mu \phi_\nu - p_\nu \phi_\mu$. In the light-cone coordinates, the transversality implies that the components $\mathcal{K}_{+\nu}$ vanish, thus the only non-vanishing components are $\mathcal{K}_{-n} = p_- \phi_n$. Therefore the only physical components correspond to a $(D-2)$-vector in the hyperplane transverse to the light-cone.

• Helicity symmetric representations with "helicity" (or "spin") equal to s correspond to Young diagrams $\lambda = \{s\}$ made of one row of length equal to the integer s. In four spacetime dimensions, this representation is precisely what is usually called a "massless spin-s field." The Young diagram $\overline{\lambda} = \{s, s\}$ is a rectangle made of two row of length equal to the integer s. The wave function is thus a polynomial in the auxiliary variables

$$\mathcal{K} \,=\, \frac{1}{2^s}\, \mathcal{K}_{\mu_1\nu_1\,|\ldots|\,\mu_s\nu_s}\, d_1x^{\mu_1} d_1x^{\nu_1}\, \ldots\, d_sx^{\mu_s} d_sx^{\nu_s}$$

satisfying the irreducibility equations (5.9)-(5.10) with $\ell_I = 2$ $(I = 1, \ldots, s)$. The fieldstrength tensor $\mathcal{K}_{\mu_1\nu_1\,|\ldots|\,\mu_s\nu_s}$ was first introduced by Weinberg in four spacetime dimensions [22]. The tensor $\mathcal{K}$ is, by construction, antisymmetric in each of the s sets of two indices

$$\mathcal{K}_{\mu_1\nu_1\,|\ldots|\,\mu_s\nu_s} \,=\, -\,\mathcal{K}_{\nu_1\mu_1\,|\ldots|\,\mu_s\nu_s} \,=\, \ldots \,=\, -\,\mathcal{K}_{\mu_1\nu_1\,|\ldots|\,\nu_s\mu_s}\,. \tag{5.16}$$

Moreover, the complete antisymmetrization over any set of three indices gives zero and all its traces are zero, so that the tensor indeed belongs to the space irreducible under the Lorentz group $O(D-1,1)$ characterized by a two-row rectangular Young diagram of length s. In four-dimensional Minkowski spacetime, the irrep of the Lorentz group $O(3,1)$ carried by the (on-shell) Weinberg tensor is usually denoted as $(s,0)\oplus(0,s)$. More precisely, the symmetry properties of the tensor $\mathcal{K}_{\mu_1\nu_1\,|\ldots|\,\mu_s\nu_s}$ are labeled by the Young tableau

$$\begin{array}{|c|c|c|c|}\hline \mu_1 & \mu_2 & \cdots & \mu_s \\ \hline \nu_1 & \nu_2 & \cdots & \nu_s \\ \hline \end{array}\;.$$

The equation (5.15) means that the components of the tensor $\mathcal{K}_{\mu_1\nu_1\,|\ldots|\,\mu_s\nu_s}$ are essentially the projection of $p_{\mu_1}\ldots p_{\mu_s}\phi_{\nu_1\ldots\nu_s}$ on the tensor field irreducible under $GL(D,\mathbb{R})$ with symmetries labeled by the above Young tableau. The physical components $\phi_{n_1\ldots n_s}$ of the symmetric tensor gauge field $\phi_{\nu_1\ldots\nu_s}$ are along the $D-2$ directions transverse to the light-cone. The number of physical degrees of freedom of a helicity symmetric field of rank s can be computed by making use of the formula (4.8).

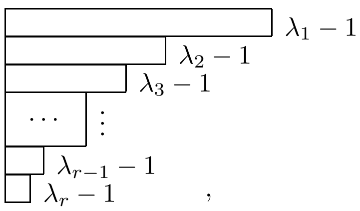
• The helicity symmetric representation with "spin" equal to 2 corresponds to the graviton. The fieldstrength has the symmetry properties of the Riemann tensor. Its on-shell tracelessness indicates that it corresponds to the Weyl tensor. The equations (5.11) are the Bianchi identities for the linearized Riemann tensor on flat spacetime, whereas the equations (5.12) hold as a consequence of the sourceless Einstein equations linearized around flat spacetime.

Gauge-fixed equations

The following equations are somewhat unusual, but they proved to be determinant in the completion of the Bargmann-Wigner programme for the infinite spin representations [15].

Step 2. Let $\widehat{\lambda} = \{\lambda_1 - 1, \lambda_2 - 1, \ldots, \lambda_r - 1\}$ be the Young diagram depicted as

$\widehat{\lambda} =$



, (5.17)

obtained from the Young diagram λ represented in (5.7) by removing the first column of λ. Therefore the sum of the length of the first two columns of the Young diagram $\widehat{\lambda}$ does not exceed $D-2$. The covariant wave function is chosen to take values in the Schur module $V^{O(D-1,1)}_{\widehat{\lambda}}$ realized in the manifestly symmetric convention. Actually, as anticipated in Subsection 4.4, it turns out to be crucial to regard the wave function $\Phi(p, u_a)$ as a *distribution* in the commuting auxiliary variables u^μ_a, obeying to

$$\left[\left(u_a \cdot \frac{\partial}{\partial u_b}\right) - \widehat{\lambda}_a\, \delta_{ab}\right]\Phi \;=\; 0\,, \quad (a \leqslant b)\,. \tag{5.18}$$

$$\left(\frac{\partial}{\partial u_a} \cdot \frac{\partial}{\partial u_b}\right)\Phi \;=\; 0\,, \tag{5.19}$$

Step 3. Proper wave equations are the transversality condition (5.1) combined with the equation

$$(p \cdot u_a)\,\Phi = 0\,. \tag{5.20}$$

The equations (5.20) and (5.1) are the respective analogues of the closure and coclosure conditions (5.11)-(5.12). A drastic difference is that the operators $p \cdot u_a$ are not nilpotent (thus there is no underlying cohomology). Actually, the equation (5.20) has no solution if Φ is assumed to be a polynomial in all the variables.

Step 4. Equation (5.20) can be solved as

$$\Phi = \delta(u_a \cdot p)\,\phi\,, \tag{5.21}$$

where the distribution $\phi(p, u_a)$ may actually be assumed to be a function depending polynomially on the auxiliary variables u_a for the present purpose. The Dirac delta is a distribution of homogeneity degree equal to minus one, hence the irreducibility conditions (5.18)-(5.19) imply that

$$\left[\left(u_a \cdot \frac{\partial}{\partial u_b}\right) - \lambda_a\, \delta_{ab}\right]\phi \;=\; 0 \qquad (a \leqslant b)\,, \tag{5.22}$$

$$\left(\frac{\partial}{\partial u_a} \cdot \frac{\partial}{\partial u_b}\right)\phi \;=\; 0\,. \tag{5.23}$$

The function ϕ is defined from (5.21) modulo the equivalence relation

$$\phi \sim \phi + \sum_{a=1}^{r} (u_a \cdot p)\,\epsilon_a \tag{5.24}$$

where ϵ_a are arbitrary functions. This means that (5.21) is equivalent to the alternative road towards the Bargmann-Wigner programme: the gauge symmetry principle with the irreducible components of $(u_a \cdot p)\,\epsilon_a$ being pure gauge fields. As mentioned before, this path will not be addressed here (see *e.g.* [17] and refs therein for more discussions on the gauge-invariance issue). Therefore, one may say that the equation (5.20) is the "remnant" of the gauge symmetries (5.24). In the light-cone coordinates, the gauge symmetries (5.24) imply that one may choose a representative ϕ which does not depend on the variables u_a^- (the gauge is "fixed"). The transversality condition (5.1) implies that ϕ is also transverse, implying no dependence on u_a^+ ("gauge shoots twice"). Thus ϕ depends only on the transverse auxiliary variables u_a^m, so one concludes by observing that the physical components of ϕ carry a tensorial irrep of $O(D-2)$ labeled by λ. □

Infinite spin representations

Infinite spin representations correspond to the case $\mu \neq 0$ and, in practice, the representation of the massless little group $IO(D-2)$ is induced from a representation of the orthogonal group $O(D-3)$. The parameter μ is a real parameter with the dimension of a mass. Wigner proposed a set of manifestly covariant equations to describe fields carrying these UIR in four spacetime dimensions [23]. Recently, they were generalized to arbitrary infinite-spin representations in any dimension [15].

Step 1. Again, any unitary representation of the orthogonal group $O(D-3)$ is a sum of finite-dimensional UIRs. Let us consider the UIR of the helicity short little group $O(D-3)$ labeled by the allowed Young diagram $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_r\}$ (that is, the sum of the lengths of its first two columns does not exceed $D-3$).

Step 2. In order to have manifest covariance, it is necessary to lift the eigenvalues ξ^m of the generators π^m in the massless little group to a D-vector ξ^μ. In practice, the covariant wave function is taken to be a distribution $\Phi(p, \xi, u_a)$ satisfying the conditions (4.16)-(4.17). The tensorial components associated with the commuting variables u_a belong to the Schur module of the Lorentz group $O(D-1,1)$ labeled by an allowed Young diagram λ.

Step 3. Relativistic equations describing a first-quantized particle with infinite spin are

$$(p \cdot \xi)\,\Phi \;=\; 0\,, \tag{5.25}$$

$$\left(p \cdot \frac{\partial}{\partial \xi} - i\right)\Phi \;=\; 0\,, \tag{5.26}$$

$$(\xi^2 - \mu^2)\,\Phi \;=\; 0\,, \tag{5.27}$$

together with the transversality conditions

$$
\begin{aligned}
(p \cdot u_a)\,\Phi &= 0\,, && (5.28)\\
\left(p \cdot \frac{\partial}{\partial u_a}\right)\Phi &= 0\,, && (5.29)\\
\left(\xi \cdot \frac{\partial}{\partial u_a}\right)\Phi &= 0\,. && (5.30)
\end{aligned}
$$

This system of equations is far from being independent. For instance, compatibility condition of the systems (5.25)-(5.26) or (5.28)-(5.29) is the massless Klein-Gordon equation.

Step 4. The equation (5.26) reflects the fact that the couples $(p\,,\xi)$ and $(p\,,\xi+\alpha p)$ are physically equivalent for arbitrary $\alpha \in \mathbb{R}$. Indeed, one gets

$$\Phi(p\,,\xi+\alpha p) \,=\, e^{i\alpha}\,\Phi(p\,,\xi) \tag{5.31}$$

from Equation (5.26). The equation (5.27) states that the internal vector ξ is a space-like vector while the mass-shell condition states that the momentum is light-like. From the equation (5.25), one obtains that the internal vector is transverse to the momentum. All together, one finds that ξ may be taken to live on the hypersphere S^{D-3} of radius μ embedded in the transverse hyperplane $\mathbb{R}^{D-2}$. In brief, the "continuous spin" degrees of freedom essentially correspond to $D-3$ angular variables, whose Fourier conjugates are discrete variables analogous to the usual spin degrees of freedom. Finally, proceeding analogously to the "gauge-fixed" wave equations of the helicity representations, one may show [15] that the conditions (5.28)-(5.30) concretely remove three unphysical directions in the components, so that the final result is a tensorial irrep of the short little group $O(D-3)$ fixing both the momentum p and the internal vector ξ.

From the group theoretical point of view, the UIR of the homogeneous and inhomogeneous orthogonal groups are related by an Inönü-Wigner contraction $O(D-1) \to IO(D-2)$ (see Subsection 4.5). It follows that one can obtain the continuous spin representations from the massive ones in a suitable massless limit $m \to 0$ since their little group UIRs are related by a contraction. The quartic Casimir operator of the Poincaré group for the massive representation is related to its Young diagram ν labeling the UIR of the little group $O(D-1)$ via the formula (5.5):

$$\mathcal{C}_4\Big(\mathfrak{iso}(D-1,1)\Big) = m^2 \sum_{a=1}^{r} \nu_a(\nu_a + D - 2a - 1)\,, \tag{5.32}$$

In order to keep $\mathcal{C}_4$ non-vanishing, the massless limit must be such that the product of the "spin" $\nu_1 = s$ and the mass m remains finite. More precisely, one needs $sm \to \mu$ in order to reproduce (5.6), so that the spin goes to infinity while the row lengths ν_a for $a \neq 1$ are kept equal to λ_{a-1} [24, 15]. The Fourier transform (in the internal space spanned by ξ) of the wave equations (5.25)-(5.30) may be obtained in this way from the wave equations of a massive representation in "gauge-fixed" form (see [15] for more details). This limit is very similar to the contraction of Subsection 4.5.

5.4 Tachyonic representations

The tachyonic representations have some similarities with the massive representations. The simpler one is the analogue of the Klein-Gordon equation, up to a change of sign for the mass term. The other similarity is that the linear equations should remove the components along the momentum. Of course, the major difference is that the momentum is space-like. The quartic Casimir operator of the Poincaré algebra is also evaluated easily in components, giving as a final result for a tachyonic representation,

$$\mathcal{C}_4\Big(\mathfrak{iso}(D-1,1)\Big) \;=\; \mathcal{C}_2\Big(\mathfrak{iso}(D-1,1)\Big)\,\mathcal{C}_2\Big(\mathfrak{so}(D-2,1)\Big)\,, \tag{5.33}$$

where the eigenvalues of the quadratic Casimir operator of the rotation algebra are given by the formula (4.10).

Step 1. The first step is more involved for the tachyonic case since it requires the exhaustive knowledge of the UIR theory for the groups $SO(D-2,1)^{\uparrow}$. Fortunately, complete results are available [25]. However, the steps 2-3 further require the completion of the Bargmann-Wigner programme for the isometry group $SO(D-2,1)^{\uparrow}$ of the de Sitter spacetime dS_{D-2} , which is still an open problem in full generality.

Let us assume that this programme has been performed through an ambient space formulation, analogous to the one of the spherical harmonics, as discussed in the subsection 4.4. More explicitly, let us consider that the physical components of the wave function have been realized via a function on the hyperboloid dS_{D-2} of radius $\mu>0$ embedded in $\mathbb{R}^{D-2,1}$ with some set of auxiliary commuting vectors of $\mathbb{R}^{D-2,1}$ (for the spin degrees of freedom) and the corresponding $O(D-2,1)$-covariant wave equations of the UIR are known explicitly. The step 1 is therefore assumed to be performed.

Step 2. In order to have manifest Lorentz invariance, all auxiliary variables are lifted to D-vectors: the coordinates of the internal de Sitter spacetime are denoted by ξ^μ and the auxiliary variables by u^μ_A . The wave function is taken to be $\Phi(p,\xi,u_A)$, where the internal vector ξ plays a role similar to the one in the infinite-spin representations. An important distinction is that in the ambient space formulation, one would evaluate the wave function on the hypersurface $\xi^2=\mu^2$ instead of imposing this relation on the wave function, as in (5.27). The $O(D-2,1)$-covariant wave equations for the UIR of the little group $O(D-2,1)$ must be $O(D-1,1)$-covariantized accordingly. Concretely, this implies that the components of the covariant wave function carry an (infinite-dimensional) irrep of the Lorentz group.

Step 3. These covariantized wave equations and the tachyonic Klein-Gordon equation $(p^2-m^2)\psi=0$ must be supplemented by two equations: say the orthogonality condition (5.25), similarly to the infinite spin representation, and the transversality condition (5.1), similarly to the massive representation. The orthogonality condition (5.25) may be replaced by another transversality equation for the vector ξ .

Step 4. Now, the equation (5.25) implies that the internal vector belongs to the hyperplane $\mathbb{R}^{D-2,1}$ orthogonal to the momentum p . Its intersection with the hypersurface $\xi^2=\mu^2$ restricts ξ to the internal de Sitter space $dS_{D-2}\subset\mathbb{R}^{D-2,1}$. Moreover, the condition (5.1) sets to zero all components of the wave function along the momentum. Therefore, the remaining components are physical and carry an UIR of the little group $O(D-2,1)$ by construction (see step 2). $\square$

Example: The simplest non-trivial example corresponds to a tachyonic representation of the inhomogeneous Lorentz group $IO(D-1,1)$ induced by a representation of the little group $O(D-2,1)$ corresponding to "massive scalar field" on the "internal de Sitter spacetime" dS_{D-2} with $D\geqslant 4$. This UIR belongs to the *principal continuous series* of UIR of the group $O(D-2,1)$ and it may be realized as the space of harmonic functions on $\mathbb{R}^{D-2,1}$ of (complex) homogeneity degree s equal to $\frac{3-D}{2}+i\,\sigma$ (with σ a positive real parameter) evaluated on the unit one-sheeted hyperboloid $dS_{D-2}\subset\mathbb{R}^{D-2,1}$. They can be regarded as a generalization of the spherical harmonics in the Lorentzian case, where the degree is a complex number. The eigenvalue of the quadratic Casimir operator (1.4) of the little group $O(D-2,1)$ on this representation is equal to

$$\mathcal{C}_2\Big(\mathfrak{so}(D-2,1)\Big) \;=\; \left(\frac{D-3}{2}\right)^2+\sigma^2\,. \tag{5.34}$$

The d'Alembertian on the unit hyperboloid evaluated on such functions is precisely equal to the former eigenvalue (as is true for the Laplacian on the unit sphere evaluated on spherical harmonics) so the corresponding fields on the internal spacetime dS_{D-2} are indeed "massive". Inserting the above result in (5.33), one sees that the quartic Casimir operator is negative for the corresponding tachyonic representation. In four-dimensional Minkowski spacetime, this implies that the Pauli-Lubanski vector is time-like. The Lorentz-covariant wave function is taken to be $\Phi(p,\xi)$ evaluated on $\xi^2=1$ and the corresponding relativistic equations for the induced tachyonic representation may be chosen as

$$\left(p^2-m^2\right)\Phi \;=\; 0\,, \tag{5.35}$$

$$\left(p\cdot\frac{\partial}{\partial\xi}\right)\Phi \;=\; 0\,, \tag{5.36}$$

$$\left(\frac{\partial}{\partial\xi}\cdot\frac{\partial}{\partial\xi}\right)\Phi \;=\; 0\,, \tag{5.37}$$

$$\left(\xi\cdot\frac{\partial}{\partial\xi}-s\right)\Phi \;=\; 0\,, \tag{5.38}$$

where one should remember that $s = \frac{3-D}{2} + i\,\sigma$. Notice the formal analogy with the equations (5.2) for the massive scalar field.

Remark: There might be sometimes confusion in the folklore surrounding the tachyons. We would like to insist on the fact that the tachyonic representations are indeed *unitary* (by definition). Still, their physical interpretation is problematic because they are *not causal* in the sense that one may show that the support of their propagator requires superluminal propagation. Roughly speaking, the acausality is obvious because the momentum is space-like, $p^2 = +m^2$. The confusing point is that one may try to circumvent this problem in the following way: solving $p^2 - m^2 = 0$ by $p^\mu = (im, \overrightarrow{0})$ enforces causality, but the price to pay is the loss of unitarity. Indeed, the energy is pure imaginary, hence a naive plane-wave $e^{\pm i\, p_0\, x^0}$ is actually a non-integrable exponential $e^{\pm m x^0}$. These remarks are summarized in the following table:

$E = p_0$	$\lvert\overrightarrow{p}\rvert$	Unitarity	Causality
0	m	OK	KO
$\pm im$	0	KO	OK

Nevertheless, the tachyonic representations should not be discarded too quickly on such physical grounds. Actually, if tachyonic representations appear in the spectrum of a theory, then it merely signals a local instability of the field theory in the sense that the perturbation theory is performed around an unstable vacuum, and the tachyon might roll to a stable vacuum (if any). For instance, the Higgs particle is described by nothing but a tachyonic scalar field (induced by the trivial representation of the little group). By analogy, one may wonder if some infinite-component tachyonic field (induced by a non-trivial representation of the little group) could not play a role in some huge Brout–Englert–Higgs mechanism providing mass to an infinite tower of gauge fields in various massless irreps.

Acknowledgments

X. B. is especially grateful to his collaborator J. Mourad for numerous fruitful discussions on group theory and relativistic wave equations. He also thanks K. Noui for stimulating exchanges on the unitary representations in the case of three-dimensional spacetimes.

The authors acknowledge the Institut des Hautes Études Scientifiques de Bures-sur-Yvette and the Université de Mons-Hainaut for hospitality during the first stage of this work.

References

[1] B. C. Hall, "An elementary introduction to groups and representations" `math-ph/0005032`. Expanded version published as *Graduate Texts in Mathematics* **222** (Springer-Verlag, 2003).

[2] A. O. Barut and R. Raczka, *Theory of group representations and applications* (Polish Scientific Publishers, 1977).

[3] X. Bekaert, N. Boulanger, S. de Buyl, F. Dolan and J.-H. Park, "Lie algebras: representation and applications," to be published in the proceedings of the *First Modave Summer School in Mathematical Physics* (Modave, June 2005).

[4] V. Bargmann and E. P. Wigner, "Group theoretical discussion of relativistic wave equations," Proc. Nat. Acad. Sci. **34** (1948) 211.

[5] S. Weinberg, *The quantum theory of fields, Vol. 1: Foundations* (Cambridge University Press, 1995) Chapter 2.

[6] E. P. Wigner, "On unitary representations of the inhomogeneous Lorentz group," Annals Math. **40** (1939) 149.

[7] E. P. Wigner, "Invariant quantum mechanical equations of motion," in *International Atomic Energy Agency* (Vienna, 1963) 59.

[8] L. Brink, A. M. Khan, P. Ramond and X. Z. Xiong, "Continuous spin representations of the Poincaré and super-Poincaré groups," J. Math. Phys. **43** (2002) 6279 [`hep-th/0205145`].

[9] B. Binegar, "Relativistic field theories in three dimensions," J. Math. Phys. **23** (1982) 1511.

[10] M. Hamermesh, *Group theory and its application to physical problems* (Addison-Wesley, 1962) Chapter 10;
H. Bacry, *Leçons sur la théorie des groupes et les symétries des particules élémentaires* (Dunod, 1967) Chapter 4;
W. Fulton and J. Harris, *Representation theory: a first course* (Springer, 2004) Chapters 4, 6, 15 and 19.

[11] S. A. Fulling, R. C. King, B. G. Wybourne and C. J. Cummins, "Normal forms for tensor polynomials. 1: The Riemann tensor," Class. Quantum Grav. **9** (1992) 1151.

[12] R.C. King, "Branching rules for classical Lie groups using tensor and spinor methods," J. Phys. A: Math. Gen. **8** (1975) 429.

[13] R. C. King, "The dimensions of irreducible tensor representations of the orthogonal and symplectic groups," Canad. Jour. Math. **23** (1971) 176.

[14] E. Inönü and E. P. Wigner, "On the contraction of groups and their representations," Proc. Nat. Acad. Sci. **39** (1953) 510.

[15] X. Bekaert and J. Mourad, "The continuous spin limit of higher spin field equations," JHEP **0601** (2006) 115 [hep-th/0509092].

[16] I. L. Buchbinder and S. M. Kuzenko, *Ideas and methods of supersymmetry and supergravity: Or a walk through superspace* (Institute of Physics Publishing, 1998) Chapters 1 and 6.

[17] X. Bekaert and N. Boulanger, "Tensor gauge fields in arbitrary representations of $GL(D,\mathbb{R})$," Commun. Math. Phys. **245** (2004) 27 [hep-th/0208058]; [hep-th/0606198].

[18] I. Bandos, X. Bekaert, J. A. de Azcarraga, D. Sorokin and M. Tsulaia, "Dynamics of higher spin fields and tensorial space," JHEP **0505** (2005) 031 [hep-th/0501113].

[19] R. Jackiw and V. P. Nair, "Relativistic wave equations for anyons," Phys. Rev. D **43** (1991) 1933;
J. L. Cortes and M. S. Plyushchay, "Anyons as spinning particles," Int. J. Mod. Phys. A **11** (1996) 3331 [hep-th/9505117].

[20] M. Fierz, "Über die relativistische theorie krafterfreier teilchen mit beliebigem spin," Helv. Phys. Acta **XII** (1939) 3.

[21] P. A. M. Dirac, "Relativistic wave equations," Proc. Roy. Soc. Lond. **155A** (1936) 447.

[22] S. Weinberg, "Photons and gravitons in perturbation theory: derivation of Maxwell's and Einstein's equations," Phys. Rev. **138** (1965) B988.

[23] E.P. Wigner, "Relativistische Wellengleichungen" Z. Physik **124** (1947) 665.

[24] A. M. Khan and P. Ramond, "Continuous spin representations from group contraction," J. Math. Phys. **46** (2005) 053515 [Erratum-ibid. **46** (2005) 079901] [hep-th/0410107].

[25] V. Bargmann, "Irreducible unitary representations of the Lorentz group," Annals Math. **48** (1947) 568;
E. A. Thieleker, "The unitary representations of the generalized Lorentz groups," Trans. Amer. Math. Soc. **199** (1974) 327.

The BRST antifield formalism
Part I: Theory

Serge Leclercq

Service de Mécanique et Gravitation,
Université de Mons-Hainaut, Académie Wallonie-Bruxelles,
6 avenue du Champ de Mars, 7000 Mons (Belgium)

Abstract. In this lecture, we recall the notions of gauge theories, differential and (co)homologies. Then, we present the antifield formalism for gauge theories, illustrated by the example of the electromagnetic theory.

In this lecture, we present the antifield formalism, which is a cohomological formalism for Lagrangian gauge theories developed in quantum field theory by Batalin and Vilkovisky [1]. The first cohomological formalism, the BRST formalism, described Hamiltonian gauge theories and was developed by Becchi, Rouet and Stora [2] and independently by Tyutin [3]. Both formalisms have a classical version, which has been reviewed in a book by Henneaux and Teitelboim [4], on which this lecture is based. More recent results and notations can be found, for exemple, in [5, 6, 7, 8].

The lecture is organized as follows. In section 1, we recall the main ingredients of local Lagrangian gauge theories. We then introduce the algebraic concepts that we need in order to build a cohomological formalism in section 2. We proceed by introducing the building blocks of the formalism : the longitudinal derivative and the ghosts, the Koszul-Tate differential and the antifields. In section 4, finally, we set the formalism and present the main theorem justifying the construction, as well as some considerations about the local (co)homologies. We conclude by illustrating the construction on Maxwell's electromagnetic theory.

1 Local Lagrangian gauge theories

History space versus jet space

Let us consider a local action, i.e. a functional that is the integral over a spacetime domain of a Lagrangian density, where the latter depends on some fields ϕ^i and their derivatives up to finite order k :

$$S[\phi] = \int_{\mathcal{D}\subset\mathbb{R}^n} \mathcal{L}(x^\mu, \phi^i, \partial_\mu\phi^i, ..., \partial_{\mu_1\ldots\mu_k}\phi^i) d^n x \ .$$

Of course, since this is a functional, the integration makes sense only if a particular "history" of the fields $\phi^i(x^\mu)$ is given. Indeed, S is defined on the history space I, the functional space of the values of the fields.

At a given point x^μ, the information about a history is contained in the values of the fields and all their derivatives. But, if we are interested in functions that depend only on a finite number of these derivatives and if we assume that the rest of the information can be recovered when making the integration, then we can trade the dependence on the functions $\phi^i(x^\mu)$ for a dependence on the value of these functions and a finite number of their derivatives at each point, i.e. for the finite set $\{x^\mu, \phi^i, \partial_\mu\phi^i, ..., \partial_{\mu_1\ldots\mu_k}\phi^i\}$ where each variable must now be considered as independent. Such sets $\{x^\mu, \phi^i, \partial_\mu\phi^i, ..., \partial_{\mu_1\ldots\mu_k}\phi^i\}$ are the coordinates of a finite dimensional vectorial space called J_k, the jet space at order k.

So, since we will consider local Lagrangians, which depend only on a finite number of derivatives of the fields, we will see them as functions over jet spaces. Many computations in the history space would be difficult, and it is even a bit hasardous to define the antifield formalism in I. On the other hand, the structure of J_k is very convenient, but boundary terms that arise from the integration of divergences will have to be taken into account.
Finally, let us define the partial derivative in J_k :

$$\partial_\mu = \frac{\partial}{\partial x^\mu} + \partial_\mu\phi^i \frac{\partial}{\partial\phi^i} + ... + \partial_{\nu_1\ldots\nu_{k-1}\mu}\phi^i \frac{\partial}{\partial(\partial_{\nu_1\ldots\nu_{k-1}}\phi^i)} \ .$$

Of course this looks obvious, but it is necessary because x^μ, the fields and their derivatives are independent coordinates in J_k.

Equations of motion, stationary surface

The dynamical equations, which select the histories that extremize S, are related to functional derivatives of the action. In the case of a local theory, the relevant derivatives are the Euler-Lagrange variational derivatives :

$$\frac{\delta S}{\delta\phi^i} = \frac{\delta\mathcal{L}}{\delta\phi^i} = \sum_{j=0}^{k} (-1)^j \partial^j_{\mu_1\ldots\mu_j} \frac{\partial\mathcal{L}}{\partial(\partial^j_{\mu_1\ldots\mu_j}\phi^i)} \ .$$

The vanishing of these derivatives gives the dynamical equations. These equations determine a surface in I called the stationary surface :

$$\Sigma \equiv \{\phi^i(x^\mu) \in I | \frac{\delta S}{\delta \phi^i} = 0\} \ .$$

Let us note that, in J_k, the stationary surface is rather determined by the dynamical equations and their derivatives up to a finite order l :

$$\Sigma_J \equiv \{a \in J_k | \forall j < l < +\infty : \partial_{\mu_1 \ldots \mu_j} \frac{\delta \mathcal{L}}{\delta \phi^i} = 0\} \ .$$

Indeed, since the variables $\phi^i, \partial_\mu \phi^i, ...$ are independent, there is no reason why an equation $f(x^\mu, \phi^i, \partial_\mu \phi^i, ..) = 0$ should imply that its derivative $\partial_\nu f(x^\mu, \phi^i, \partial_\mu \phi^i, ..) = 0$ be true as well.

As we will be working in the complete (history or jet) space and not just on the stationary surface, it is convenient to introduce the weak equality :

$$f \approx 0 \Leftrightarrow f = f^{\mu_1 \ldots \mu_j | i} \partial_{\mu_1 \ldots \mu_j} \frac{\delta \mathcal{L}}{\delta \phi^i} \ .$$

Gauge transformations and Noether identities

Gauge transformations are transformations of the fields that depend on arbitrary functions and leave the action invariant. They have the form :

$$\delta_\varepsilon \phi^i = R^i_\alpha \varepsilon^\alpha \ ,$$

where $R^i_\alpha = \mathcal{R}^i_\alpha + \mathcal{R}^{i\mu}_\alpha \partial_\mu + ... + \mathcal{R}^{i\mu_1 \ldots \mu_j}_\alpha \partial_{\mu_1 \ldots \mu_j}$ is a differential operator called generator of the gauge transformations, whose coefficients can depend on the coordinates, the fields and their derivatives up to order k (as usual, j is a finite order). The ε^α's are arbitrary functions defined on the spacetime domain $\mathcal{D}$. The above transformations are gauge transformations if they satisfy

$$\delta_\varepsilon S = 0 \ .$$

A set of generators that allows to generate all the gauge symmetries of an action is called a generating set.

Let us detail what the variation of the action looks like:

$$\begin{aligned} \delta_\varepsilon S &= \int_{\mathcal{D}} \delta_\varepsilon \mathcal{L} \ d^n x = \int_{\mathcal{D}} \frac{\delta \mathcal{L}}{\delta \phi^i} R^i_\alpha \varepsilon^\alpha \ d^n x \\ &= \int_{\mathcal{D}} \partial_\mu j^\mu \ d^n x + \int_{\mathcal{D}} (\bar{R}^i_\alpha \frac{\delta \mathcal{L}}{\delta \phi^i}) \varepsilon^\alpha \ d^n x = 0 \ . \end{aligned}$$

The last step is obtained by integration by parts, with $\bar{R}^i_\alpha$ being the adjoint operator of R^i_α. The first term is the integral of a divergence and thus a boundary term, which depends only on the values of the fields, the ε^α's and their derivatives up to a finite order on the boundary of $\mathcal{D}$. In order for the gauge transformations to have a meaning, the boundary terms must be zero. Thus, it is necessary to assume that the arbitrary functions and their derivatives up to order j vanish on the border, which is not too restrictive.

Computing the variation of the action for appropriate choices of the arbitrary functions, it is obvious that the quantities $\bar{R}^i_\alpha \frac{\delta \mathcal{L}}{\delta \phi^i}$ have to vanish identically everywhere in $\mathcal{D}$. These relations among the equations of motion are called the Noether identities :

$$\bar{R}^i_\alpha \frac{\delta \mathcal{L}}{\delta \phi^i} \equiv 0 \ .$$

There is one and only one Noether identity for each gauge transformation. Conversely, studying the algebraic structure of the equations of motion is the best way to find every gauge transformation of a Lagrangian theory.

Let us note that every theory possesses some trivial gauge transformations:

$$\forall S : \delta_\mu \phi^i = \mu^{[ij]} \frac{\delta S}{\delta \phi^j} \Rightarrow \delta_\mu S = \int_{\mathcal{D}} \frac{\delta S}{\delta \phi^i} \frac{\delta S}{\delta \phi^j} \mu^{[ij]} d^n x = 0 \ ,$$

where $\mu^{[ij]}$ is some antisymmetric parameter.[1] Such transformations have no physical meaning and can be discarded.

Reducibility, gauge algebra

The generating set can sometimes be overcomplete (for example, to ensure Lorentz invariance or, trivially, because it includes generators of trivial transformations). In these cases, there exist classes of gauge parameters ε^α that depend on arbitrary functions and that lead to trivial gauge transformations. In other words, one can construct differential operators Z^α_A such that the gauge transformations with gauge parameter $\varepsilon^\alpha = Z^\alpha_A \eta^A$, where η^A is an arbitrary function, are trivial: $\exists\ \mu^{[ij]} \mid \delta_\eta \phi^i = R^i_\alpha Z^\alpha_A \eta^A = \mu^{[ij]} \frac{\delta S}{\delta \phi^i}$. Such theories are called reducible of order (at least) 1. It is possible to have further reducibilities if the Z^α_A's form themselves an overcomplete set of reducibility conditions.

Similarly to the Noether identities associated with gauge transformations, there are relations among the gauge generators associated with reducibilities. Indeed, by integrating by parts in the variation of the field, the identities $\bar{Z}^\alpha_A \bar{R}^i_\alpha = C^{[ij]}_A \frac{\delta S}{\delta \phi^j}$ are obtained for some coefficients $C^{[ij]}_A(\phi^i)$.

Let us now define the gauge algebra. It is very interesting to study the commutator of two gauge transformations. The requirement that this commutator be still a gauge transformation gives rise to relations among the generating set :

$$R^j_\alpha \frac{\delta R^i_\beta}{\delta \phi^j} - R^j_\beta \frac{\delta R^i_\alpha}{\delta \phi^j} = R^i_\gamma C^\gamma_{\alpha\beta}(\phi^i) + M^{[ij]}_{\alpha\beta}(\phi^i) \frac{\delta S}{\delta \phi^j} \ .$$

This operator has to be seen as acting on the product of arbitrary functions $\varepsilon^\alpha \eta^\beta$, and the generators are acting only on the function with which they are contracted. The operators $C^\gamma_{\alpha\beta}(\phi^i)$ and $M^{[ij]}_{\alpha\beta}(\phi^i)$ are called structure functions and characterize the gauge algebra. The Jacobi identity gives rise to on-shell conditions for these coefficients : e.g. $C^\alpha_{\beta[\mu} C^\beta_{\nu\rho]} \approx 0$. Many more relations among the gauge generators and the new operators like $C^\alpha_{\beta[\mu}$ arise when more and more commutators are taken. They constitute conditions for the consistency of the theory.

2 Gradings, differentials and homologies

Grassmann parity, bosonic and fermionic fields

Even when the fields are commuting, some anticommuting fields will have to be introduced, so let us introduce some general tools to handle commuting and anticommuting fields.

Consider a set of fields $\{\phi^i\} = \{q^a\ ,\ \theta^\alpha\}$ such that the q^a's are bosonic and the θ^α's fermionic: $q^a q^b = q^b q^a$, $q^a \theta^\alpha = \theta^\alpha q^a$ and $\theta^\alpha \theta^\beta = -\theta^\beta \theta^\alpha$. It is natural to introduce the Grassmann parity : $\epsilon_a \equiv \epsilon(q^a) = 0$ and $\epsilon_\alpha \equiv \epsilon(\theta^\alpha) = 1$. The q^a's are then said to be even, bosonic or commuting. The θ^α's are odd, fermionic or anticommuting. With this notation, the commutation rule of two fields ϕ^i is simply

$$\phi^i \phi^j = (-1)^{\epsilon_i \epsilon_j} \phi^j \phi^i \ .$$

Let us now consider the set of all polynomials in the ϕ^i's (with real or complex coefficients). Given the product and the commutation rule of the basic fields, it is clearly an associative algebra A. The Grassmann parity is also called a $\mathbb{Z}_2$-grading (because the value of the parity is 0 or 1), the algebra is said to be $\mathbb{Z}_2$-graded-commutative, or supercommutative. It is natural to consider that polynomials of power zero in the fields are in the algebra, so the number 1 is the unit for the product of the algebra. When the polynomials are composed only of terms of odd/even powers of the θ^α's, they have a definite Grassmann parity. It is easy to check that

$$\forall x, y \in A \mid \exists \epsilon_x, \epsilon_y \ :\ \epsilon_{xy} = \epsilon_x + \epsilon_y \ .$$

Even elements define a subset $A_0 \subset A$, odd ones define $A_1 \subset A$, and of course $A = A_0 \oplus A_1$. This remains consistent if the algebra is extended to any function of the bosonic fields (Note that because of the finite number, say k, of anticommuting θ^α's, there are no polynomials of degree higher than k in the fermionic fields).

[1] The fields have been considered here as commuting, which is the reason why the integrand of the last integral is identically zero. Of course, anticommuting fields can appear (for example Dirac spinors), but this requires some considerations that will be made in the next section.

Operators, differentials and gradings

We will consider the set $End(A)$ of endomorphisms from A to A (these are morphisms for the sum in A). Given the composition of operators, $End(A)$ is an associative algebra with unit. Some operators can have a definite Grassmann parity : $\tau \in End(A)$ is said to be even if $\forall x \in A \mid \exists \epsilon_x \;:\; \epsilon_{\tau(x)} = \epsilon_x$, it is said to be odd if $\forall x \in A \mid \exists \epsilon_x \;:\; \epsilon_{\tau(x)} = 1 - \epsilon_x$. $End(A)$ is obviously the direct sum of its subsets of definite parity. It is found that

$$\forall \tau, \sigma \in End(A) \mid \exists \epsilon_\tau, \epsilon_\sigma \;:\; \epsilon_{\tau\sigma} = \epsilon_\tau + \epsilon_\sigma \ .$$

In general, $End(A)$ is not supercommutative, so it is useful to introduce the graded commutator of two operators τ, σ with definite parity: $[\tau, \sigma] = \tau\sigma - (-1)^{\epsilon_\tau \epsilon_\sigma} \sigma\tau$.
This commutator satisfies a graded Jacobi identity :

$$\begin{array}{l} \forall \rho, \sigma, \tau \in End(A) : \\ [[\rho, \sigma], \tau] + (-1)^{\varepsilon_\rho(\varepsilon_\sigma + \varepsilon_\tau)}[[\sigma, \tau], \rho] + (-1)^{\varepsilon_\tau(\varepsilon_\rho + \varepsilon_\sigma)}[[\tau, \rho], \sigma] = 0 \ . \end{array}$$

$\mathbb{Z}_2$-gradings are not the only ones that can be defined. Indeed, it is often possible to introduce $\mathbb{N}$-gradings or $\mathbb{Z}$-gradings. Generally, A is then the direct sum of a set of classes labelled by a natural or integer number :

$$A = \bigoplus_{n \in \mathbb{N}\, or\, \mathbb{Z}} A_n \qquad \text{such that} \qquad \forall\ n, m : A_n A_m \subset A_{n+m} \ .$$

The label is called a grading or a degree, and is noted : $x \in A_n \Leftrightarrow deg\ x = n$.
A common exemple is the polynomial degree in some of the generators ϕ^i of A (in particular, the form degree related to the exterior differential is such a grading). More generally, any sum of entire multiples of such polynomial degrees is a grading. Let us note that odd operators have to raise or lower the number of anticommuting generators by an odd number. Finally, the grading can be extended to operators : $deg\ \tau = n \Leftrightarrow \forall m, \forall x \in A_m : \tau(x) \in A_{n+m}$. $End(A)$ is obviously the direct sum of its subsets of definite degree.
Remark : 0 has no definite degree, as it belongs to every class A_n (and of course, the same holds for the 0 operator).

Derivatives are operators with a definite parity, that satisfy a $\mathbb{Z}_2$-graded Leibniz rule. In this lecture, only derivatives acting from the left will be considered :

$$\begin{array}{l} d \in End(A) \text{ is a left derivative} \\ \Leftrightarrow \forall x, y \in A \mid \exists \epsilon_x, \epsilon_y \;:\; d(xy) = (dx)y + (-1)^{\epsilon_x \epsilon_d} x dy \ . \end{array}$$

Derivatives form a subalgebra of $End(A)$, that will be noted $Der(A)$, and which also decomposes into classes of definite degree. Let us note that the commutator is internal to $Der(A)$ (i.e. the commutator of two derivatives is a derivative).

A differential D is a nilpotent odd derivative :

$$D \in Der(A) \text{ is a differential} \Leftrightarrow D^2 = 0 \text{ and } \epsilon_D = 1 \ .$$

In practice, we will assume that there is a grading such that $deg\ D = 1$ or -1 (in the case of a $\mathbb{Z}$-grading, there is a liberty on the sign of the label and it will always be chosen such that $deg\ D = 1$).

The kernel of D is the set $Ker\ D = \{x \in A \mid Dx = 0\}$ and the image of D is the set $Im\ D = \{x \in A \mid \exists y \in A \mid x = Dy\}$. Elements of $Ker\ D$ are called D-closed objects ; elements of $Im\ D$ are called D-exact objects.

Because of the nilpotency of D, it is obvious that $Im\ D \subset Ker\ D$. Furthermore, $Im\ D$ is an ideal of $Ker\ D$: $\forall x \in Ker\ D, y \in Im\ D \;:\; \exists z \mid yx = (Dz)x = D(zx) \in ImD$. We can thus define the coset space of $Ker\ D$ modulo $Im\ D$:

$$H(D) = \frac{Ker\ D}{Im\ D} = \{[a] \mid Da = 0\ ,\ a' \in [a] \Leftrightarrow \exists b | a' = a + Db\} \ .$$

If $degD = -1$, this space is called the homology of D and is noted $H_*(D)$; if $degD = 1$, it is called the cohomology of D and is noted $H^*(D)$. (Co)homologies are the direct sum of their classes of definite degree, which will often be considered invidually

$$H_*(D) = \bigoplus H_n(D) \quad \text{or} \quad H^*(D) = \bigoplus H^n(D) \ .$$

It is also possible to define the (co)homology of a differential D in $Der(A)$. When a derivative d (anti)commutes with D : $[d, D] = 0$, d is said to be D-closed. When there is a derivative Δ such that $d = [\Delta, D]$, d is said to be D-exact. The set of D-closed derivatives is a subalgebra of $Der(A)$ (for the commutator).

We find that

$$\begin{array}{l} \forall\, d \ : \ [D,[D,d]] = \pm\frac{1}{2}[d,[D,D]] = 0 \\ \forall\, d \,|\, [d,D] = 0 \ , \ \forall\, \Delta \ : \ [d,[D,\Delta]] = \pm[D,[\Delta,d]] \ , \end{array}$$

thanks to the Jacobi identity. This clearly means that the D-exact derivatives form an ideal of the D-closed ones. The (co)homology of D in $Der(A)$ is the obvious coset space and is noted $\mathcal{H}_*(D)$ or $\mathcal{H}^*(D)$.

Some derivatives are diffentials only in a subspace $B \subset A$. A derivative γ is called a differential in B if the projection of its square on B vanishes. For example, we will consider in the sequel the algebra of functions of the fields $C^\infty(I)$ and the subset of functions on the stationary surface $C^\infty(\Sigma)$, both in tensor product with a polynomial space in some fermionic fields C^α. Any derivative γ such that $\gamma^2 \approx 0$ will then be called a differential in $B = C^\infty(\Sigma) \otimes \mathbb{R}[C^\alpha]$ (remember that $\approx$ means "equal on the stationary surface"). It will also be possible to define the cohomology of γ in B, which will be noted $H^*(\gamma, B)$.

Finally, let us define a differential modulo a differential. Let D be a differential in A, of negative grading. If γ is an odd derivative such that $[D, \gamma] = 0$ and $\exists\Delta \ : \ \gamma^2 = -[D, \Delta]$, then γ is called a differential modulo D. It is clear that this implies that γ is a differential in a space of representatives of $H_*(D)$: If $Da = 0$ then $\gamma^2 a = D(\Delta a) = 0$ in $H_*(D)$. In the sense of $\mathcal{H}_*(D)$, γ is D-closed and γ^2 is D-exact. The cohomology of γ modulo D is noted $H(\gamma|D)$ or $H(\gamma, H_*(D))$.

$$H(\gamma|D) = \{[a] \mid Da = 0 \ , \exists b|\gamma a + Db = 0 \ , \ a' \in [a] \Leftrightarrow \exists c, e|a' = a + \gamma c + De\} \ .$$

Remark : in the case where γ is a true differential (i.e. $\Delta = 0$), the cohomology of γ is of course defined in A, the condition $Da = 0$ is removed and the cohomology of γ modulo D is then

$$H(\gamma|D) = \{[a] \mid \exists b|\gamma a + Db = 0 \ , \ a' \in [a] \Leftrightarrow \exists c, e|a' = a + \gamma c + De\} \ .$$

3 Ghosts and antifields

We are now ready to introduce the ingredients of the antifield formalism. First the ghosts related to the so-called "longitudinal derivative" γ, which is a differential on-shell (but not necessarily off-shell). Then the antifields and the Koszul-Tate differential δ, introduced to work off-shell. We will show that the longitudinal derivative can be considered as a differential modulo the Koszul-Tate differential. The third step, detailed in section 4, will consist in combining γ and δ into a global differential s containing all the information about the theory.

3.1 Longitudinal derivative γ

Given a history $\phi^i(x^\mu)$, any history $\phi^i(x^\mu) + R^i_\alpha \varepsilon^\alpha$ gives the same value of the action, by definition of the gauge tranformations. This is true in particular for histories extremising the action: if $\phi^i(x^\mu) \in \Sigma$, then $\forall \varepsilon^\alpha(x^\mu) \ : \ \phi^i(x^\mu) + R^i_\alpha \varepsilon^\alpha \in \Sigma$. The gauge transformations thus generate submanifolds on the stationary surface, which will be called "gauge orbits", with a "dimension" equal to the number M of gauge tranformations. This dimension has to be seen as the number of generators ε_α, which is not very satisfying, because of the functional nature of the history space. But, as usually, the trouble can be avoided by working in J_k, where arbitrary values of the ε_α's and their derivatives up to the appropriate finite order are considered (in this case, of course, the dimension is higher that the number of gauge transformations).

Since the gauge orbits are functional manifolds, we can formally consider maps on those manifolds, with M "coordinates" X_α. This is quite similar to a usual map $\{e_\alpha\}$ on a geometrical manifold. In that case, the concept of arbitrary move on the manifold is replaced by the introduction of the tangent space and the dual coordinates $\underline{\theta}^\alpha$. The latter, which are just the usual dx^α in the natural basis, constitute the basic fermionic one-forms. The exterior differential can then be introduced, which is fermionic and raises by one the form degree, i.e. the number of dx^α's. The action of d on functions reproduces the first-order effect of an arbitrary move, with

the advantage that there are no further orders because $d^2 = 0$, and that the basic one-forms are well-defined independent coordinates.
Let us now try to apply the same pattern to the gauge orbits. We introduce a set of fermionic fields C^α, dual to the coordinates X_α, and we call them ghosts. Then, we introduce a derivative γ, that acts along the gauge orbits, and is thus called the longitudinal derivative. It is meant to reproduce an "arbitrary move" on the orbits, i.e. a gauge transformation, when acting on the fields. Its action on the latter thus takes the form

$$\gamma\phi^i = R^i_\alpha C^\alpha \ .$$

As the coordinates X_α are related to the gauge transformations, it is quite natural that the Lie bracket of the coordinates is the same as the commutator of the transformations : $[X_\alpha, X_\beta] \approx C^\gamma_{\alpha\beta} X_\gamma$, where the $C^\gamma_{\alpha\beta}$'s are the structure functions of the gauge algebra. The dual version of this is the action of γ on the ghosts :

$$\gamma C^\alpha = \frac{1}{2} C^\alpha_{\beta\gamma} C^\gamma C^\beta \ .$$

Let us now study γ^2. On ϕ^i, the action of γ^2 is related to the structure functions of the gauge algebra, but, due to cancellations, only a term proportional to the equations of motion remains. It can also be seen that

$$\gamma^2 C^\alpha = C^\alpha_{\beta\gamma} \gamma C^\gamma C^\beta = \frac{1}{2} C^\alpha_{\beta\gamma} C^\gamma_{\mu\nu} C^\beta C^\mu C^\nu \approx 0 \ .$$

So, γ is not a differential in the whole space but only on Σ.

The grading related to γ is the polynomial degree in the ghosts, which will be called the pure ghost number :

$$pgh\ C^\alpha = 1 \ , \ pgh\ \phi^i = 0 \ , \ pgh\ \gamma = 1 \ .$$

One can compute the cohomology of γ in Σ, which we note $H^*(\gamma, C^\infty(\Sigma) \otimes \mathbb{R}[C^\alpha])$ (we have simply considered an algebra of functions on Σ and polynomial in the ghosts). The cohomology in *pgh* 0 is the set of functions closed under γ, which are none other that the gauge invariant functions. There are obviously no γ-exact objects in *pgh* 0.

This construction holds for J_k, where values of the fields C^α and their derivatives up to a finite order are considered. These variables can be added to build an extended jet space, where γ is an algebraic operator. We will choose γ to commute with the partial derivatives (or that $\gamma d + d\gamma = 0$, where d is the exterior differential of the spacetime manifold).

3.2 Koszul-tate differential δ

Up to this point, we have managed to define a symmetry on the stationary surface that replaces the arbitrary gauge transformation. The second step is to relate the fact of being on-shell to a rigid transformation that is a differential off-shell. To do so we will build what is called a homological resolution of $C^\infty(\Sigma) \otimes \mathbb{R}[C^\alpha]$, the algebra where γ is a differential.

A homological resolution of an algebra A is realized when there is a differential δ acting in an algebra $A' \supset A$, related to a grading labelled by a natural number k with $deg\ \delta = -1$, such that $\forall k > 0 : H_k(\delta) = 0$ and $H_0(\delta) = A$. The latter equality must be understood as an isomorphism, the elements of A being representatives of the homology classes.

Our groups A and A' will be $C^\infty(\Sigma) \otimes \mathbb{R}[C^\alpha]$ and $C^\infty(I) \otimes \mathbb{R}[C^\alpha]$. The grading of δ will be called the antighost number, and noted *antigh*. By definition, if we want $C^\infty(\Sigma) \otimes \mathbb{R}[C^\alpha]$ to be the *antigh* 0 class, we have to require that

$$antigh\ \phi^i = 0 \text{ and } antigh\ C^\alpha = 0 \ .$$

As *antigh* $\delta = -1$ and since there is no field of negative degree (by definition of an $\mathbb{N}$-grading), one must have $\delta\phi^i = 0$ and $\delta C^\alpha = 0$. If there were no further fields, there would not be any δ-exact combinations of ϕ^i's, and the homological class in degree zero would be $C^\infty(I) \otimes \mathbb{R}[C^\alpha]$. If we want to restrict to $C^\infty(\Sigma)$, we must somehow make the equations of motion δ-exact. The solution is to extend the space so as to include a new set of fields with *antigh* 1, called the antifields and noted ϕ^*_i, in one-to-one correspondance with the fields ϕ^i. The nature of those fields does not matter, we will just consider them through the differential δ. We define the action of δ on the antifields by

$$\delta\phi_i^* = \frac{\delta S}{\delta\phi^i}\ ,\ antigh\phi_i^* = 1\ .$$

Notice that this implies that the parity ot the antifields is the opposite of that of the corresponding fields : $\varepsilon(\phi_i^*) = 1 - \varepsilon(\phi^i)$. Furthermore, we see that $\delta^2\phi_i^* = 0$ automatically. Then, we have to make sure that the other homology classes are zero. For the moment, this is not the case: the Noether identities imply that $\delta\bar{R}^i_\alpha\phi_i^* = 0$. The combinations $\bar{R}^i_\alpha\phi_i^*$ are δ-closed but not δ-exact, so they would appear in the $H_1(\delta) = 0$ if nothing is done. The only way out is to introduce some new fields C^*_α such that

$$\delta C^*_\alpha = \bar{R}^i_\alpha\phi_i^*\ ,\ antigh\ C^*_\alpha = 2\ .$$

By construction, $\delta^2 C^*_\alpha = 0$. If the theory is irreducible, there are no relations among the (adjoints of the) gauge generators, so no combinations of C^*_α's can be δ-closed and the constuction stops.

If the theory is reducible of order one, one needs to introduce a further set of antifields of *antigh* 3, C^*_A, such that $\delta C^*_A = -\bar{Z}^a_A C^*_a - \frac{1}{2}C^{[ij]}_A\phi_i^*\phi_j^*$ in order to compensate for the redundancy of the gauge generators. Unfortunately, when the theory is reducible, the longitudinal derivative cannot be easily extended to the whole space. It has to be replaced by another derivative, $\tilde{\gamma}$, called a "model" for γ, the homology of which is isomorphic to $H(\gamma)$ but the action of which on the different families of fields is a bit more complicated. Furthermore, new families of ghosts, corresponding to the different generations of antifields, have to be introduced. For exemple, at first order of reducibility, a family of bosonic fields C^A of *pgh* 2 are introduced and one has $\tilde{\gamma}C^\alpha = \frac{1}{2}C^\alpha_{\beta\gamma}C^\beta\ C^\gamma + Z^\alpha_A\ C^A + ...$ We will not give more details about that in this lecture, and we will only consider irreducible gauge theories from now on.

As usual, this was a bit formal but everything is very well defined in J_k. The jet space is just extended to the different families of antifields and their derivatives up to a finite order, then δ is a simple derivation commuting with ∂_μ (or anticommuting with the exterior derivative d) :

$$\delta = \sum_{j=0}^{j_{max}} \partial_{\mu_1\ldots\mu_j}(\bar{R}^i_\alpha\phi_i^*)\frac{\partial}{\partial(\partial_{\mu_1\ldots\mu_j}C^*_\alpha)} + \sum_{m=0}^{m_{max}} \partial_{\nu_1\ldots\nu_m}\frac{\delta\mathcal{L}}{\delta\phi^i}\frac{\partial}{\partial(\partial_{\nu_1\ldots\nu_m}\phi_i^*)}\ .$$

An important fact is that our construction not only ensures that δ provides a resolution of $C^\infty(\Sigma)\otimes\mathbb{R}[C^\alpha]$ in $\mathbb{R}[C^*_\alpha,\phi^*_i]\otimes C^\infty(I)\otimes\mathbb{R}[C^\alpha]$, but it also provides a resolution of $Der(C^\infty(\Sigma)\otimes\mathbb{R}[C^\alpha])$ in $Der(\mathbb{R}[C^*_\alpha,\phi^*_i]\otimes C^\infty(I)\otimes\mathbb{R}[C^\alpha])$. This means that $\forall k > 0\ :\ \mathcal{H}_k(\delta) = 0$ and $\mathcal{H}_0(\delta) = Der(C^\infty(\Sigma)\otimes\mathbb{R}[C^\alpha])$. The proof is very general for every differential δ such that each variable of strictly positive *antigh* is either not δ-exact or δ-closed, but we will not show it here.

4 Construction of the formalism

4.1 The differential s

First, we have to define the action of γ on the antifields. In fact, this is quite arbitrary, but the most logical choice is to fix $\gamma\phi_i^*$ and γC^*_α such that γ and δ anticommute. It is obvious that this is true for the fields and the ghosts, because $[\gamma,\delta]$ is an *antigh* -1 operator. By taking $\gamma\phi_i^* = \phi_j^*\frac{\delta[R^j_\alpha C^\alpha]}{\delta\phi^i}$, we find $(\delta\gamma+\gamma\delta)\phi_i^* = 0$ thanks to the Noether identities and to the fact that the variational derivatives of a divergence are identically zero. The value of γC^*_α is chosen similarly (we will not write it here explicitely but it will be easy to compute given the differential s a bit later).

Since γ anticommutes with δ, it is δ-closed. In addition to that, γ^2 is δ-exact, i.e. there exists an auxiliary derivative Δ of *antigh* 1 such that $\gamma^2 = -\delta\Delta - \Delta\delta$. It is easy (at least in principle) to build the latter using the property that $\gamma^2 a \approx 0$ if $\delta a = 0$. Let us construct the action of Δ on objects with increasing *antigh*. Given an *antigh* 0 object a_0, which trivially satisfies $\delta a_0 = 0$: $\exists\ b\ |\ \gamma^2 a_0 = \delta b$. We can simply define $\Delta a_0 = -b$ and the relation $\gamma^2 = -\delta\Delta - \Delta\delta$ holds (when acting on *antigh* 0 quantities). Then, if a_1 is an *antigh* 1 object, δa_1 is of *antigh* 0, so it satisfies $\gamma^2\delta a_1 = -\delta\Delta\delta a_1$. This can be rewritten as $\delta[\gamma^2 a_1 + \Delta\delta a_1] = 0$. The vanishing of $H_1(\delta)$ now implies that $\exists c|\gamma^2 a_1 = -\Delta\delta a_1 - \delta c$, and the action of Δ in *antigh* 1 is defined as $\Delta a_1 = c$. The same kind of argument gives the value of Δ for higher *antigh*. Putting the two properties of γ with respect to δ together, we have shown that γ is a differential modulo δ.

Now, the conditions $\delta^2 = 0$, $\gamma\delta+\delta\gamma = 0$ and $\gamma^2+\delta\Delta+\Delta\delta = 0$ could be seen as the first three terms of the *antigh* expansion of the equation s^2=0 with $s = \delta + \gamma + \Delta +$ higher *antigh* terms. We will prove that, given that δ provides a resolution in the derivative space, such a differential s actually exists. This differential, defined on the whole functional space and which encodes every characteristic of the gauge theory, is the central object of the formalism.

Homological perturbations

Let us show that if :

- δ is a differential related to the $\mathbb{N}$-grading *antigh* noted k, with *antigh* $\delta = -1$, such that $\forall k > 0\ :\ H_k(\delta) = 0$ and $\mathcal{H}_k(\delta) = 0$;
- γ is a differential modulo δ (of *antigh* 0), i.e. $\gamma\delta + \delta\gamma = 0$ and $\exists \overset{(1)}{s}\ \mid \gamma^2 = -[\delta, \overset{(1)}{s}]$;
  γ is associated with a $\mathbb{N}$-grading *pgh*, with *pgh* $\gamma = 1$ and *pgh* $\delta = 0$;

Then there exists a differential s associated with the $\mathbb{Z}$-grading $gh = pgh - antigh$ with $gh\ s = 1$, and such that $s = \delta + \gamma + \sum_{k\geq 1} \overset{(k)}{s}$ with *antigh* $\overset{(k)}{s} = k$.

Furthermore, the cohomology classes of s in positive gh are isomorphic to those of γ in *antigh* 0 :

$$\forall i > 0\ :\ H^i(s) \cong H^i(\gamma, H_0(\delta))$$

where i is the gh number for s and is the pgh number for γ. In particular, gauge-invariant functions correspond to cosets of the group $H^0(s) \cong H^0(\gamma, H_0(\delta))$.

<u>Proof</u>

1. Let us consider $s_n = \delta + \gamma + \overset{(1)}{s} + ... + \overset{(n)}{s}$ and let us assume that its square has no terms of *antigh* $< n$: $s_n^2 = \overset{(n)}{\rho} + \overset{(n+1)}{\rho} + \overset{(n+2)}{\rho} + ...$
   The hypothesis tells us that it is true for $n = 1$. It is sufficient to show that for any n, there exists $\overset{(n+1)}{s}$ such that s_{n+1}^2 begins at *antigh* $(n + 1)$.
   It is trivial that $[s_n^2, s_n] = s_n^3 - s_n^3 = 0$. The term of lowest *antigh* of this expression is $[\overset{(n)}{\rho}, \delta] = 0$, so $\overset{(n)}{\rho}$ is δ-closed.
   But we know that $\mathcal{H}_n(\delta) = 0$, thus $\exists \overset{(n+1)}{s}\ \mid\ \overset{(n)}{\rho} = -[\overset{(n+1)}{s}, \delta]$.
   If $s_{n+1} = s_n + \overset{(n+1)}{s}$,
   then $s_{n+1}^2 = \overset{(n)}{\rho} + \delta \overset{(n+1)}{s} + \overset{(n+1)}{s} \delta + \overset{(n+1)}{\rho'} + ... = \overset{(n+1)}{\rho'} + ...$
2. Let us consider any element x of the algebra, it can be expanded according to the *antigh* number : $x = \sum_{k\geq 0} \overset{(k)}{x}$. The isomorphism needed to prove the second statement is simply given by the map π that applies x on $\overset{(0)}{x}$. The reason is that $sx = \gamma \overset{(0)}{x} + \delta \overset{(1)}{x} + ...$ so $\pi s x = \gamma \pi x$ in $H_0(\delta)$. It is clearly a morphism : $\forall x, y\ :\ \pi(xy) = \overset{(0)}{x}\overset{(0)}{y} = \pi(x)\pi(y)$.
   (a) <u>π is surjective :</u> We need to prove that any *antigh* 0 object x_0 that is γ-closed in $H_0(\delta)$ can be deformed into an s-closed x. By assumption, $\exists \overset{(1)}{x}\ \mid \gamma \overset{(0)}{x} + \delta \overset{(1)}{x} = 0$.
   This tells us that $s(\overset{(0)}{x} + \overset{(1)}{x}) = \overset{(1)}{t} + \overset{(2)}{t} + ...$
   Now, it is sufficient to prove that if $x_n = \overset{(0)}{x} + \overset{(1)}{x} + ... + \overset{(n)}{x}$ is such that $s x_n = \overset{(n)}{t} + \overset{(n+1)}{t} + ...$ then there exists a $\overset{(n+1)}{x}$ such that $s x_{n+1}$ begins at *antigh* $n + 1$.

As $s^2 x_n = 0$, its lowest *antigh* term vanishes too : $\delta \overset{(n)}{t} = 0$, and as $H_n(\delta) = 0$: $\exists \overset{(n+1)}{x} \mid \overset{(n)}{t} = -\delta \overset{(n+1)}{x}$.
If $x_{n+1} = x_n + \overset{(n+1)}{x}$,
then $sx_{n+1} = \overset{(n)}{t} + \delta \overset{(n+1)}{x} + \overset{(n+1)}{t'} + ... = \overset{(n+1)}{t'} + ...$

(b) $\underline{\pi \text{ is injective :}}$ We have to prove that: $sx = 0$, $\pi x \in [0] \subset H^i(\gamma, H_0(\delta)) \Rightarrow x \in [0] \subset H^i(s)$. More explicitely, the left-hand side means that $\exists \overset{(0)}{z}, \overset{(1)}{z} \mid \overset{(0)}{x} = \gamma \overset{(0)}{z} + \delta \overset{(1)}{z}$. If $x' = x - s(\overset{(0)}{z} + \overset{(1)}{z})$, then $sx' = 0$ and x' begins at *antigh* 1, so $\delta \overset{(1)}{x'} = 0$. But $H_1(\delta) = 0$ thus $\exists \overset{(2)}{z} \mid \overset{(1)}{x'} = \delta \overset{(2)}{z}$. Then $x'' = x - s(\overset{(0)}{z} + \overset{(1)}{z} + \overset{(2)}{z})$ is such that $sx'' = 0$ and x'' begins at *antigh* 2. Going on like this recursively, the different *antigh* components of x are removed one by one. It is finally found that x is s-exact, which proves that π is injective.

4.2 The antibracket and the generator W

Even more interesting is the fact that the differential s admits a generating functional. Before exhibiting it, we have to introduce the antibracket. The latter is very similar to a Poisson bracket, but with pairs of variables of opposite parity. For an irreducible theory, its action on two functionals is :

$$(A, B) = \frac{\delta^R A}{\delta \phi^i} \frac{\delta^L B}{\delta \phi_i^*} - \frac{\delta^R A}{\delta \phi_i^*} \frac{\delta^L B}{\delta \phi^i} + \frac{\delta^R A}{\delta C^\alpha} \frac{\delta^L B}{\delta C_\alpha^*} - \frac{\delta^R A}{\delta C_\alpha^*} \frac{\delta^L B}{\delta C^\alpha} \ .$$

The indices L and R just indicate whether the derivatives act from the left or from the right, which is not equivalent for fermionic fields. The antibracket is also well-defined on local functions, for which the functional derivatives are replaced by the variational derivatives. The antibracket clearly raises the *gh* number by one (the first two terms lower the *antigh* by one, the others lower the *pgh* by one and the *antigh* by two), and it is clearly fermionic. It satisfies the following rules :

Symmetry : $(A, B) = -(-1)^{(\varepsilon_A+1)(\varepsilon_B+1)}(B, A)$
Jacobi identity : $(-1)^{(\varepsilon_A+1)(\varepsilon_C+1)}(A, (B, C)) +$ cyclic permutations $= 0$
"Leibniz rule" : $(A, BC) = (A, B)C + (-1)^{\varepsilon_B(\varepsilon_A+1)} B(A, C)$

It can be shown that there exists a definite local functional W such that for any functional or local function A:

$$sA = (W, A) \ .$$

W is a *gh* 0 bosonic functional. It is often considered as an extended action for the antifield formalism, because its *antigh* 0 component is none other than the action S :

$$W = S + \overset{(1)}{W} + \overset{(2)}{W} + \dots \ .$$

Indeed, this choice reproduces the right *antigh* -1 part of the action of s (i.e. the differential δ) on the antifields ϕ_i^* : $\delta \phi_i^* = \frac{\delta S}{\delta \phi^i}$. The second term of W, in *antigh* 1, can also be easily written: it is $\overset{(1)}{W} = \int_{\mathcal{D}} \phi_i^* R_\alpha^i C^\alpha d^n x$, which generates the relations $\gamma \phi^i = R_\alpha^i C^\alpha$ and $\delta C_\alpha^* = \bar{R}_\alpha^i \phi_i^*$.

An important feature of W is related to the nilpotency of the differential s. Using the Jacobi identity, the latter means that, for any A, $0 = s^2 A = (W, (W, A)) = \pm\frac{1}{2}(A, (W, W))$. This is true iff

$$(W, W) = 0 \ .$$

This very important condition is called the master equation.

The master equation is the key constraint that the generating functional W has to satisfy. We will prove recursively that if the *antigh* expansion of W starts with the first two terms written above, then one can always construct the further orders one by one in such a way that the master equation is satisfied. To do so, it is enough to show that, if $\overset{(n-1)}{R} = \overset{(0)}{W} + \overset{(1)}{W} + \dots + \overset{(n-1)}{W}$ satisfies

the master equation up to its *antigh* component $n-2$, then one can build $\overset{(n)}{R}$ that satisfies it up to *antigh* $n-1$. Indeed, $\overset{(1)}{R}= S+ \overset{(1)}{W}$ satisfies the *antigh* 0 component of the master equation, which is equivalent to the Noether identities.
Let us define $\overset{(n-1)}{D}$ as the component of *antigh* $n-1$ of $(\overset{(n-1)}{R}, \overset{(n-1)}{R})$, which is its first non zero component. The *antigh* $n-1$ term of $(W, W) = 0$ is $2\delta \overset{(n)}{W} + \overset{(n-1)}{D} = 0$. On the other hand, the Jacobi identity implies that $(\overset{(n-1)}{R}, (\overset{(n-1)}{R}, \overset{(n-1)}{R})) = 0$, the lower *antigh* term of which is $\delta \overset{(n-1)}{D} = 0$. Thanks to the vanishing of $H_{n-1}(\delta)$, it is now obvious that $\overset{(n)}{W}$ exists.

For an irreducible theory, the component of *antigh* 2 obtained is $\overset{(2)}{W}= \int_{\mathcal{D}} (\frac{1}{2} C^*_\alpha C^\alpha_{\gamma\beta} C^\beta C^\gamma - \frac{1}{4}\phi^*_i \phi^*_j M^{ij}_{\alpha\beta} C^\alpha C^\beta) d^n x$. When the theory is reducible, the reducibility constants appear in terms like $C^*_\alpha Z^\alpha_A C^A$ or $\phi^*_i \phi^*_j C^{ij}_A C^A$, and so forth. We see that the first terms of the generator involve all the various coefficients characterizing the theory : the action, the generators of the gauge transformations, the structure functions, etc. These coefficients are very easily singled out, as the terms are all linearly independent. In fact, W is a single functional which contains all the infomation about the theory in a very simple way. The consistency of the whole is secured by a single constraint: the master equation. This is a fundamental advantage, which makes it possible for exemple to study deformations of an action with very few hypotheses (see Part II). But we have to worry a bit more about the local nature of the objects. s is very well defined in J_k but the homologies of the different derivatives are not quite equivalent to those in I.

4.3 Locality, homologies modulo d

Exterior differential d

The usual exterior differential $d = dx^\mu \partial_\mu$ can be introduced on the space-time manifold. It is easily extended to the jet space thanks to the definition of the partial derivative on J_k. The form degree is the $\mathbb{N}$-grading of d and is given by the number of anticommuting 1-forms dx^μ. This grading is bounded from above by n because there are only n independent dx^μ's. The Poincaré lemma states that, on a domain with no holes or periodicity, the only d-closed object that are not d-exact are the constant zero forms (i.e. the numbers). In other words :

$$\forall\ i > 0\ :\ H^i(d) = 0 \text{ and } H^0(d) = \mathbb{R}\ .$$

The same result holds in a jet space, except in form degree n, for bosonic and/or fermionic fields and is called the algebraic Poincaré lemma. We will use the dual notation
$d^n x = \frac{1}{n!}\varepsilon_{\mu_1 \dots \mu_n} dx^{\mu_1} \wedge \dots . \wedge dx^{\mu_n}$, $d^{n-1} x_\mu = \frac{1}{(n-1)!}\varepsilon_{\mu\mu_1 \dots \mu_{n-1}} dx^{\mu_1} \wedge \dots \wedge dx^{\mu_{n-1}}$.
Any n-form is obviously closed. A $(n-1)$-form can be written $v = V^\mu d^{n-1} x_\mu$ and its exterior derivative is the d-exact n-form : $dv = \partial_\mu V^\mu d^n x$. It is well-known that the Euler-Lagrange derivatives of a divergence are identically zero, which is equivalent to saying that divergences are trivial terms in a Lagrangian. The homology of d in form degree n in J_k is thus isomorphic to the set of functions which have the same variational derivatives:

$$\begin{array}{c} H^n(d) = \left\{ [a]\ |deg\ a = n, a' \in [a] \Leftrightarrow \forall \phi^i\ :\ \dfrac{\delta a'}{\delta \phi^i} = \dfrac{\delta a}{\delta \phi^i} \right\} \\ \forall 0 < i < n\ :\ H^i(d) = 0 \\ H^0(d) = \mathbb{R} \end{array} .$$

The local Lagrangian and any function defined on J_k can be seen as the coefficient of a n-form. Let us note that if the metric is not flat (as for the Einstein-Hilbert theory of course), coefficients of n-forms must be tensorial densities (for exemple, $\sqrt{|g|}$, which appears in the Einstein-Hilbert action). So, in that case, the Lagrangian, the equations of motion and the antifields all have that behaviour. Trivial terms in the Lagrangian are just d-exact n-forms the integrals of which are of course boundary terms in the action, which are assumed to vanish.

Homologies modulo d

The fact that we consider only vanishing boundary terms means that two functionals are equal if they are the integral of n-forms equal modulo d : if $A = \int a$ and $B = \int b$, then $A = B \Leftrightarrow \exists v \mid a = b + dv$. This implies that a s-closed functional is related to an n-form s-closed modulo d :

$$sA = 0 = s \int a \Rightarrow \exists b \; | sa + db = 0 \; .$$

The same holds for s-exact functionals :

$$A = sB \Rightarrow \exists c | a = sb + dc$$

δ and s are differentials, constructed to commute with d in J_k, so δ and s are differentials modulo d and it is very natural to define $H^i_*(\delta|d)$ and $H^{*,i}(s|d)$, where i is the form degree. The equivalent in J_k of $H^*(s)$ in I is in fact $H^{*,n}(s|d)$. It is also possible to define the homology of γ modulo d in $H_0(\delta)$ and it can be shown that :

$$\forall k \geq 0 \;\; : \;\; H^{k,n}(s|d) \cong H^{k,n}(\gamma|d, H_0(\delta)) \; .$$

It is also found that

$$\forall k < 0 \;\; : H^{k,n}(s|d) \cong H^n_{-k}(\delta|d) \; .$$

Finally, let us say a few words about $H^n_k(\delta|d)$. It is not vanishing as $H_k(\delta)$ does. But there are never many independent classes, and the non vanishing part is always at *pgh* 0. The Poincaré lemma and $H_k(\delta) = 0$ provide isomorphisms between classes of $H^n_k(\delta|d)$:

$$\forall n \geq 1, k > 1 \;\; : H^n_k(\delta|d) \cong H^{n-1}_{k-1}(\delta|d) \; .$$

If $k > n$, it is found that $H^n_k(\delta|d) \cong H^0_{k-n}(\delta|d) = H_{k-n}(\delta) = 0$. It is also found that $\forall k \geq 1, n \geq 1$ (but $(k,n) \neq (1,1)$) : $H^k_n(\delta|d) \cong H^{k-1}_{n-1}(d|\delta)$ and $H^1_1(\delta|d) \cong H^0_0(d|\delta)/\mathbb{R}$. The homology of d modulo δ in *antigh* 0 is called the characteristic cohomology, it contains the non trivial on-shell d-closed forms. Those are generalized conserved currents, which have to be known for exemple in some general relativity problems. The isomorphisms that we cited are very useful because $H^n_k(\delta|d)$ is rather easy to compute for $k \geq 2$. These cohomology classes are also needed in order to compute deformations of Lagrangians.

5 A little example

A little example might be useful at this stage, so let us consider the theory of electromagnetism.

The basic settings are as follows: the fields are just the A_μ's, the action is $S = -\frac{1}{4}\int F_{\mu\nu}F^{\mu\nu}\, d^n x$, where $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$. It is invariant under the gauge transformations $\delta_\epsilon A_\mu = \partial_\mu \epsilon$, which correspond to the Noether identity $\partial^\mu \partial^\nu F_{\mu\nu} = 0$. The theory has no reducibility. The equations of motion are $\partial_\nu F^{\nu\mu} = 0$.

Let us now apply the rules to build the BRST formalism.

First: a ghost should be associated with each gauge parameter. So let us introduce the ghost C, corresponding to the gauge parameter ϵ .

Secondly: define the longitudinal derivative γ. It acts on the fields like a gauge transformation where the gauge parameter is replaced by its ghost, so we take $\gamma A_\mu = \partial_\mu C$. Since the gauge algebra is abelian, the action of γ on the ghost vanishes: $\gamma C = 0$. In this case, we see that γ is a true differential: $\gamma^2 = 0$ even off-shell.

Thirdly: introduce the first generation of antifields. An antifield has to be introduce for each field, so we introduce the antifields $A^{*\mu}$.

Fourthly: build up the Koszul-Tate differential δ. The action of δ on the fields and the ghosts vanishes: $\delta A_\mu = 0$ and $\delta C = 0$. The action of δ on the antifields yields the left-hand side of the equations of motion: $\delta A^{*\mu} = \partial_\nu F^{\nu\mu}$.

Fifthly: check whether a next generation of antifields is needed. A new generation of antifields has to be introduced if the homology of δ in nonvanishing *antigh* is nontrivial. This reduces to ask whether there are δ-closed combinations of antifields. At this stage, it is equivalent to look for relations among the equations of motion, i.e. Noether identities. In our case, there is a Noether identity: $\partial_\mu \partial_\nu F^{\mu\nu} = 0$, which implies that $\delta \partial_\mu A^{*\mu} = 0$. A further ghost must be introduced in order for $\partial_\mu A^{*\mu}$ to be nontrivial in $H_1(\delta)$.

Sixthly: iterate the last three steps until there is no need for further ghosts. In general, the number of generations of antifields to be introduced corresponds to the reducibility level of the theory plus two. We introduce a new antifield for the Noether identity and note it C^*. The action of δ on this antifield is $\delta C^* = -\partial_\mu A^{*\mu}$, so that $\partial_\mu A^{*\mu}$ becomes trivial in the homology. Since the theory is irreducible, the iteration stops here.

Seventhly: extend the action of the longitudinal derivative to the ghosts, in such a way that it anticommutes with the Koszul-Tate differential. In our case, it is enough to choose that the action of γ on the antifields vanishes.

To summarize, the set of fields and antifields is composed of

- the fields A_μ, of even parity, $pgh = 0$, $antigh = 0$ and $gh = 0$;
- the ghost C, of odd parity, $pgh = 1$, $antigh = 0$ and $gh = 1$;
- the antifields $A^{*\mu}$, of odd parity, $pgh = 0$, $antigh = 1$ and $gh = -1$;
- the antifield C^*, of even parity, $pgh = 0$, $antigh = 2$ and $gh = -2$.

The pgh, $antigh$ and gh numbers have been attributed using the following rules: $pgh = 0$ for the fields and the antifields; $antigh = 0$ for the fields and the ghosts; $gh = pgh - antigh$; increase the pgh, resp. the $antigh$, by one for each new generation of ghosts, resp. antifields. The Grassmann parity is alternating for each new generation. The symmetry between the fields+ghosts and the antifields is no accident, it is a general rule.

The action of γ and δ is zero on any (anti)field and ghost, except in the following cases:

$$\gamma A^\mu = \partial^\mu C \ , \qquad \delta A^*_\mu = \partial^\nu F_{\nu\mu} \ , \qquad \delta C^* = -\partial^\mu A^*_\mu \ .$$

One easily checks that $\gamma^2 = 0$, $\delta^2 = 0$ and $\gamma\delta + \delta\gamma = 0$.

The differential s is simply $s = \gamma + \delta$. No further terms are needed since $s^2 = 0$ is already satisfied. This differential is generated by the generating functional

$$W = \int \Big(-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + A^*_\mu \partial^\mu C \Big) \, d^n x \ ,$$

where the first term is easily identified as the action. The second term is built as the antifield corresponding to the field multiplied by a gauge transformation, where the gauge parameter has been replaced by its ghost. One can check that the master equation $(W, W) = 0$ is satisfied, and that W generates s correctly, so there is no need for further terms.

One now has the tools to compute the various cohomology groups or apply the formalism to whatever problem of electromagnetism one has in mind. The slightly more complicated case of Yang-Mills is developped in the Part II of the BRST antifield formalism lectures.

Acknowledgements

I would like to thank the organisers of the second Modave Summer School for giving me the opportunity of presenting this lecture. In particular, I thank Sandrine Cnockaert for giving much useful advice to me for writing the present notes. I would also like to thank Philippe Spindel and Nicolas Boulanger for useful discussions about the antifield formalism.

References

[1] I.A. Batalin and G.A. Vilkovisky, *Phys. Lett.* **B102** (1981) 27 ; *Phys. Rev.* **D28** (1983) 2567 ; *Phys. Rev.* **D30** (1984) 508.

[2] C. Becchi, A. Rouet and R. Stora, *Commun. Math. Phys.* **42** (1975) 127 ; *Ann. Phys.* (NY) **98** (1976) 287.

[3] I.V. Tyutin, *Gauge invariance in field theory and statistical mechanics*, Lebedev preprint FIAN, n^{0}39 (1975).

[4] M. Henneaux and C. Teitelboim, "Quantization of gauge systems," Princeton, USA: Univ. Pr. (1992) 520 p.

[5] G. Barnich and M. Henneaux, *Phys. Lett.* **B311** (1993) 123, hep-th/9304057.
M. Henneaux, *Contemp. Math.* **219** (1998) 93, hep-th/9712226.

[6] G. Barnich, F. Brandt and M. Henneaux, *Commun. Math. Phys.* **174** (1995) 57, hep-th/9405109.

[7] G. Barnich, F. Brandt and M. Henneaux, *Commun. Math. Phys.* **174** (1995) 93, hep-th/9405194.

[8] G. Barnich, F. Brandt and M. Henneaux, *Phys. Rept.* **338** (2000) 439, hep-th/0002245.

The BRST antifield formalism
Part II: Applications

Sandrine Cnockaert[1]

Physique Théorique et Mathématique, Université Libre de Bruxelles
& International Solvay Institutes
ULB, Campus Plaine C.P. 231, B–1050 Bruxelles, Belgium.

Abstract. In these lectures, we present an introduction to two applications of the Lagrangian BRST formalism (also called field-antifield or Belavin-Vilkovisky formalism): the quantization of systems with gauge freedom and the construction of consistent interactions.

[1]Aspirant du F.N.R.S.

The BRST formalism is mainly known because it offers a systematic way to quantize theories with a gauge freedom. For these theories, the most naive implementation of the path integral does not work. Indeed, the integration over gauge orbits makes the path integral infinite, as e.g. for electromagnetism:

$$\int \mathcal{D}A_\mu \, e^{\frac{i}{\hbar}S[A_\mu]} = \infty \; .$$

One solution is to remove all the redundant degrees of freedom to work with the physical fields only: one chooses a gauge before making the integration and integrates only over the physical degrees of freedom. However, this procedure is not liked because it spoils the gauge invariance, makes locality non manifest, is not covariant, etc. A way out was found by Faddeev and Popov. They turned the infinities into new terms in the action, which involve a new kind of fields called "ghosts". The Faddeev-Popov action, which is a functionnal of the gauge fields and of the ghost fields, has no gauge invariance (that is, no symmetry transformations depending on an arbitrary function), but it exhibits a new global symmetry, called BRST symmetry, which mixes gauge fields and ghosts. In spite of this success, the quantization problem was not completely solved because the procedure is limited to "simple" gauge theories. For example, it does not work for general reducible theories or when the gauge algebra is open. Furthermore, the mathematical scheme behind the BRST symmetry and the whole construction was at first poorly understood. Their meaning became only clear with the developpement of the BRST approach and of the antifield formalism.

In these lectures, we will first show how the antifield formalism is used to build actions without gauge invariance that can be used in the path integral quantization procedure. This construction is systematic and can be done for any gauge theory. It can be related to the Faddeev-Popov results when the latter exist.

The second topic covered in these lectures is the application of the antifield formalism to the "consistent local interaction problem". The problem one wants to solve is the following: given a "free" theory of gauge fields, what are all the consistent couplings that one can introduce to make these fields interact locally? By "consistent", one means that the number of degrees of freedom is unchanged. The antifield formalism offers a very systematic way to answer this question, through a reformulation in terms of cohomology groups for various differentials related to the BRST-differential and presented in the part I of these lectures (s, d, γ and δ). We will briefly present this topic and illustrate it on an example.

Summary of Part I :

The main achievement of Part I is that the description of a theory by some fields and an action S is replaced by a new description : the set of fields is enlarged to include new fields called ghosts and antifields, while new terms are added to the action to build the "extended action" W. The latter generates the BRST[2] differential s through the antibracket : $sA = (W, A)$, where the antibracket is given by

$$(A, B) = \frac{\delta^R A}{\delta \Phi^A}\frac{\delta^L B}{\delta \Phi^*_A} - \frac{\delta^R A}{\delta \Phi^*_A}\frac{\delta^L B}{\delta \Phi^A} \; .$$

Here we have used a new notation: the fields and ghosts are collectively denoted by Φ^A, while the antifields are denoted by Φ^*_A. The extended action W no longer admits the original gauge transformations as symmetries, but it is invariant under the rigid symmetry s:

$$sW = (W, W) = 0 \; . \tag{0.1}$$

This crucial equation is called the master equation. Another important feature is that, by construction, the gauge invariant observables are in one-to-one correspondance with the equivalence classes of $H^0(s)$ in ghost number zero.

1 Quantization

The purpose of this section is to bridge the gap between the antifield formalism presented in the Part I of the antifield lectures and its use for the path integral quantization. On our way, we will meet Faddeev and Popov.

[2]Actually this is not exactly the BRST differential, the latter being what this differential becomes when the gauge is completely fixed. But we will nevertheless use the same terminology.

The outline of this section is as follows. In the path integral approach, one needs an action that has no gauge invariance. The extended action could be a better candidate than the original action, but it also possesses some gauge invariance (Section 1.1). One can however fix this invariance in a clever way, by chosing a "gauge-fixing fermion", which removes the antifields. To build the latter fermion, it is necessary to introduce new fields, called *antighosts* (Section 1.2). In the end, one is left with an action that has no gauge invariance, but admits a global symmetry mixing the remaining fields (including ghosts and antighosts), called the BRST symmetry (Section 1.3). This action can be used in the path integral, modulo one last check: the correlation functions computed must be independent of the gauge-fixing procedure (Section 1.5). This requirement imposes conditions on the theory through the "quantum master equation". We illustrate the gauge-fixing procedure on an example in Section 1.4.

1.1 Gauge invariance and gauge-fixing fermion

The extended action built in Part I still possesses gauge invariances, though different from those of the original theory. Indeed, differentiating the master equation

$$(W,W) = 2\frac{\delta^R W}{\delta\Phi^A}\frac{\delta^L W}{\delta\Phi^*_A} = 0$$

with respect to the fields (or to the antifields) yields the following relations among the left-hand sides of the equations of motion:

$$0 = \frac{1}{2}\frac{\delta^R(W,W)}{\delta z^c} = \frac{\delta^R W}{\delta\Phi^A}\,\frac{\delta^L\delta^R W}{\delta\Phi^*_A\delta z^c} - \frac{\delta^R W}{\delta\Phi^*_A}\,\frac{\delta^L\delta^R W}{\delta\Phi^A\delta z^c}\ , \tag{1.1}$$

where z^c can be either Φ^C or Φ^*_C. Each of these Noether identities indicates the presence of a gauge symmetry. (Exercice: write them!)

One can show (see [2]) that the number of independent gauge transformations matches the number of fields (or equivalently the number of antifields), which we call N. In order to remove the gauge freedom, one thus has to fix the gauge by N gauge conditions. One usually uses these conditions to gauge-fix the antifields, that is, to express them in terms of the fields. This is implemented by choosing a function $\psi(\Phi^A)$ depending only on the fields and imposing the gauge conditions

$$\Phi^*_A = \frac{\delta\psi}{\delta\Phi^A}\ . \tag{1.2}$$

From this equation, one reads that ψ must be a fermion, and have the ghost number -1:

$$\epsilon(\psi) = 1\ , \quad gh(\psi) = -1\ .$$

(Indeed, Φ^A and Φ^*_A have opposite parity, and $gh(\Phi^*_A) = -gh(\Phi^A) - 1$.) Accordingly, the function ψ is called the "gauge-fixing fermion". Note that in Eq.(1.2) it is equivalent to take left or right derivatives (cf the rules of Appendix A).

We will now consider two issues. First, we will try to build a gauge-fixing fermion ψ and see that it is necessary to introduce new fields, the antighosts. We will illustrate this construction on an example. Secondly, we will check under which conditions the path integral is independent of the choice of ψ.

1.2 Antighost fields

We have said above that the gauge-fixing fermion should depend only on the fields and have ghost number -1. However, all the fields of the minimal set[3] have positive ghost number (only the antifields have negative ghost number). It is thus necessary to add new fields (and their corresponding antifields) in order to build an appropriate gauge-fixing fermion.

It is allowed to add new fields by pairs $\{\alpha,\beta\}$ (so-called "trivial pairs") that satisfy the following relations:

$$\begin{aligned} s\alpha &= 0\,, & s\beta &= \alpha\,, \\ \epsilon(\alpha) &= \epsilon(\beta)+1\,, & gh(\alpha) &= gh(\beta)+1\,. \end{aligned} \tag{1.3}$$

[3]As can be guessed, the minimal set is the smallest set of fields and antifields that verifies the properties listed in Part I. One can arbitrarily enlarge it, e.g. by adding redundant gauge transformation generators, and in fact the minimal set has no deep meaning.

(one must also include their antifield counterparts $\{\alpha^*, \beta^*\}$). Such a couple will add no new contribution to the cohomology of s: all s-cocycles involving these fields are s-exact and thus trivial. This means that, while adding these fields, one keeps the nice property that the set of observables of the theory is isomorphic to the cohomology of s in ghost number zero, $H^0(s)$. Enlarging the minimal set of fields to include this trivial pair thus contains the same information as the minimal set itself. So let us do it!

One easily generates the BRST variations (1.3) for α and β by adding a BRST-exact term to the extended action:

$$W^{minimal} \to W = W^{minimal} + W^{trivial} ,$$

where $W^{trivial} = -\alpha\beta^*$. One can then read off the BRST variation of α^* and β^*, which is

$$s\alpha^* = (-)^{\epsilon(\beta)}\beta^* \quad and \quad s\beta^* = 0 .$$

We will call the new fields "antighosts" because they correspond to the antighosts of Faddeev and Popov. In specific cases they have other nicknames like B-fields, auxiliary fields, Nakanishi-Lautrup auxiliary field, "extra" ghosts, Stuckelberg fields, etc.

The freedom to add trivial pairs is used in the gauge-fixing procedure to introduce fields with negative ghost number. The number of pairs added depends on the reducibility level of the theory. So, e.g., in the irreducible case one adds one trivial pair $\{\bar{C}_{0\alpha_0}, \bar{\pi}_{0\alpha_0}\}$ of fields that have the same index structure as the ghost $C_0^{\alpha_0}$. Their ghost numbers and parities are

$$gh[\bar{C}_{0\alpha_0}] = -1 , \quad gh[\bar{\pi}_{0\alpha_0}] = 0 , \quad \epsilon[\bar{C}_{0\alpha_0}] = \epsilon[C_0^{\alpha_0}] , \quad \epsilon[\bar{\pi}_{0\alpha_0}] = \epsilon[C_0^{\alpha_0}] + 1 .$$

For a first-stage reducible theory, one adds two more pairs $\{\bar{C}_{1\alpha_1}, \bar{\pi}_{1\alpha_1}\}$ and $\{C^1_{1\alpha_1}, \pi^1_{1\alpha_1}\}$. For an L-th stage reducible theory, for each integer s ranging from 0 to L, one adds $s+1$ trivial pairs having the same indices as the s-th stage ghost.

We will not enter into the particulars about why exactly these antighost fields should be added. Let us just state that, if the gauge-fixing fermion is properly chosen, they allow to remove all gauge and non-propagating degrees of freedom. More details about this can be found e.g. in [1].

As we just implied by using the words "properly chosen", not every gauge-fixing fermion is allowed, one must make sure that it kills all the unwanted degrees of freedom. Nevertheless, in the end some freedom is left in the choice of the gauge-fixing fermion. One then chooses (or tries to choose !) the fermion that most simplifies the calculation one is to perform.

1.3 Gauge-fixed action and BRST symmetry

Having chosen the gauge-fixing fermion and used the gauge-fixing conditions, one is left with the gauge-fixed action

$$W_\psi = W(\Phi^A, \Phi^*_A = \frac{\delta\psi}{\delta\Phi^A}) = W(\Phi^A, \Phi^*_A)|_{\Sigma_\psi} ,$$

where we have denoted by Σ_ψ the surface on which $\Phi^*_A = \frac{\delta\psi}{\delta\Phi^A}$ holds.

Quite naturally, one defines the gauge-fixed BRST symmetry by

$$s_\psi X(\Phi^A) \equiv (W, X)|_{\Sigma_\psi} .$$

The gauge-fixed action is invariant under the gauge-fixed BRST symmetry. Indeed, using

$$\frac{\delta^L W_\psi}{\delta\Phi^A} = \frac{\delta^L W}{\delta\Phi^A}\Big|_{\Sigma_\psi} + \frac{\delta^L\delta^L\psi}{\delta\Phi^A\delta\Phi^B}\frac{\delta^L W_\psi}{\delta\Phi^*_B}\Big|_{\Sigma_\psi} ,$$

one has

$$\begin{aligned} s_\psi W_\psi &= (W, W_\psi)|_{\Sigma_\psi} = -\frac{\delta^R W}{\delta\Phi^*_A}\Big|_{\Sigma_\psi}\frac{\delta^L W_\psi}{\delta\Phi^A} \\ &= -\Big[\frac{\delta^R W}{\delta\Phi^*_A}\frac{\delta^L W}{\delta\Phi^A}\Big]\Big|_{\Sigma_\psi} - \Big[\frac{\delta^R W}{\delta\Phi^*_A}\frac{\delta^L\delta^L\psi}{\delta\Phi^A\delta\Phi^B}\frac{\delta^L W}{\delta\Phi^*_B}\Big]\Big|_{\Sigma_\psi} \\ &= 0 . \end{aligned}$$

The first term vanishes because of the master equation. The second term vanishes because of the statistics: using the formulae of the appendix A and reordering the factors, one gets $\frac{\delta^R W}{\delta\Phi^*_A}\frac{\delta^L\delta^L\psi}{\delta\Phi^A\delta\Phi^B}\frac{\delta^L W}{\delta\Phi^*_B} = -\frac{\delta^R W}{\delta\Phi^*_B}\frac{\delta^L\delta^L\psi}{\delta\Phi^B\delta\Phi^A}\frac{\delta^L W}{\delta\Phi^*_A}$, which must thus vanish.

The gauge-fixed BRST transformation is not always a differential. One has

$$\begin{aligned} s^2_\psi(X(\Phi^A)) &= (W, s_\psi X)|_{\Sigma_\psi} = -\frac{\delta^R W}{\delta\Phi^*_B}\Big|_{\Sigma_\psi}\frac{\delta^L s_\psi X}{\delta\Phi^B} \\ &= \frac{\delta^R W}{\delta\Phi^*_B}\Big|_{\Sigma_\psi}\Big[\frac{\delta^L\delta^R W}{\delta\Phi^B\delta\Phi^*_A}\frac{\delta^L X}{\delta\Phi^A} + \frac{\delta^L\delta^L\psi}{\delta\Phi^B\delta\Phi^C}\frac{\delta^L\delta^R W}{\delta\Phi^*_C\delta\Phi^*_A}\frac{\delta^L X}{\delta\Phi^A} \\ &\qquad + (-)^{\epsilon_B(\epsilon_A+1)}\frac{\delta^R W}{\delta\Phi^*_A}\frac{\delta^L\delta^L X}{\delta\Phi^B\delta\Phi^A}\Big]\Big|_{\Sigma_\psi}. \end{aligned} \tag{1.4}$$

Notice that by ϵ_A, we mean the Grassmann parity of Φ^A, which is opposite to the parity of Φ^*_A. Using equation (1.1), the first term can be rewritten as $\frac{\delta^R W}{\delta\Phi^B}\frac{\delta^L\delta^R W}{\delta\Phi^*_B\delta\Phi^*_A}\frac{\delta^L X}{\delta\Phi^A}$. It can then be combined with the second term to yield

$$\Big(\frac{\delta^R W}{\delta\Phi^C} + \frac{\delta^R W}{\delta\Phi^*_B}\frac{\delta^R\delta^L\psi}{\delta\Phi^C\delta\Phi^B}\Big)\frac{\delta^L\delta^R W}{\delta\Phi^*_C\delta\Phi^*_A}\frac{\delta^L X}{\delta\Phi^A} = \frac{\delta^R W_\psi}{\delta\Phi^C}\frac{\delta^L\delta^R W}{\delta\Phi^*_C\delta\Phi^*_A}\frac{\delta^L X}{\delta\Phi^A}.$$

The last term of (1.4) identically vanishes because of symmetry properties. The final answer is thus

$$s^2_\psi(X(\Phi^A)) = \frac{\delta^R W_\psi}{\delta\Phi^C}\frac{\delta^L\delta^R W}{\delta\Phi^*_C\delta\Phi^*_A}\frac{\delta^L X}{\delta\Phi^A}, \tag{1.5}$$

which means that generically the gauge-fixed action is nilpotent only on-shell, i.e. when the gauge-fixed equations of motion

$$\frac{\delta^R W_\psi}{\delta\Phi^C} = 0$$

are satisfied.

1.4 Example of a gauge-fixing fermion

Let us now illustrate this construction on an example. We will consider the Yang-Mills theory.

The set of fields is given by $\{A^a_\mu\}$. Their equations of motion are obtained from the action

$$S(A^a_\mu) = -\frac{1}{4}\int d^n x F^a_{\mu\nu}F^{\mu\nu}_a ,$$

where

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - f_{bc}{}^a A^b_\mu A^c_\nu ,$$

and the structure functions $f_{bc}{}^a$ satisfy the Jacobi identity

$$f_{ab}{}^e f_{ec}{}^d + f_{bc}{}^e f_{ea}{}^d + f_{ca}{}^e f_{eb}{}^d = 0 .$$

The equations of motion read

$$D^\mu{}_a{}^b F_{b\mu\nu} = 0 ,$$

where

$$D^\mu{}_a{}^b \equiv \delta_a{}^b\partial^\mu + f_{ca}{}^b A^{c\mu} .$$

The equations of motion and the action are invariant under the gauge transformation

$$A^a_\mu \to A^a_\mu + D^a_{\mu b}\Lambda^b ,$$

where Λ^b is an arbitrary function and

$$D_\mu{}^a{}_b \equiv \delta_b{}^a\partial_\mu - f_{cb}{}^a A^c_\mu .$$

There are no reducibility relations.

The gauge invariance of the action implies the Noether identity

$$D^\nu{}_c{}^a D^\mu{}_a{}^b F_{b\mu\nu} = 0 .$$

Following the antifield construction, one introduces a ghost C^a for each gauge parameter Λ^a. For each field and ghost, one further adds an antifield. The set of fields (including ghosts) and antifields is thus $\{A^a_\mu,\, C^a\}$ and $\{A^{*\mu}_a,\, C^*_a\}$. Their parity and ghost numbers are

$$\begin{array}{llll} \epsilon(A^a_\mu)=0\,, & \epsilon(C^a)=1\,, & \epsilon(A^{*\mu}_a)=1\,, & \epsilon(C^*_a)=0\,, \\ gh(A^a_\mu)=0\,, & gh(C^a)=1\,, & gh(A^{*\mu}_a)=-1\,, & gh(C^*_a)=-2\,. \end{array}$$

The extended action reads

$$W=\int d^nx\Big(-\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a+A^*_{a\mu}D^{\mu a}{}_bC^b+\frac{1}{2}C^*_cf_{ab}{}^cC^bC^a\Big)\,. \tag{1.6}$$

One can check that W satisfies the master equation

$$(W,W)=0\,.$$

The action of the BRST differential on the various fields is computed from its generator (1.6) and yields

$$\begin{aligned} sA^{a\mu} &= D^{\mu a}{}_bC^b\,, \\ sC^a &= -\frac{1}{2}f_{bc}{}^aC^cC^b\,, \\ sA^*_{a\mu} &= D^\nu{}_a{}^bF_{b\nu\mu}-A^*_{c\mu}f_{ab}{}^cC^b\,, \\ sC^*_a &= D^\mu{}_a{}^bA^*_{b\mu}-C^*_cf_{ab}{}^cC^b\,. \end{aligned}$$

Let us now fix the gauge invariance of the action (1.6). Since the theory is irreducible, one just introduces one pair of antighosts: $\{\bar{C}_a,\bar{\pi}_a\}$ (note that $\bar{\pi}_a$ is often denoted by B_a and called B-field), with

$$\epsilon(\bar{C}_a)=1\,,\quad gh(\bar{C}_a)=-1\,,\quad \epsilon(\bar{\pi}_a)=0\,,\quad gh(\bar{\pi}_a)=0\,,$$

as well as the corresponding antifields, $\{\bar{C}^{*a},\bar{\pi}^{*a}\}$, with

$$\epsilon(\bar{C}^{*a})=0\,,\quad gh(\bar{C}^{*a})=0\,,\quad \epsilon(\bar{\pi}^{*a})=1\,,\quad gh(\bar{\pi}^{*a})=-1\,.$$

The BRST transformations are now generated by the extended action

$$W=\int d^nx\,[-\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a+A^*_{a\mu}D^{\mu a}{}_bC^b+\frac{1}{2}C^*_cf_{ab}{}^cC^bC^a-\bar{\pi}_a\bar{C}^{*a}]\,,$$

which is (1.6) where we have added the trivial term $-\bar{\pi}_a\bar{C}^{*a}$. The BRST transformations of the antighosts and their antifields is

$$s\,\bar{C}_a=\bar{\pi}_a\,,\quad s\,\bar{\pi}_a=0\,,\quad s\,\bar{C}^{*a}=0\,,\quad s\,\bar{\pi}^{*a}=\bar{C}^{*a}\,.$$

The gauge-fixing fermion can for example be chosen to be

$$\psi=\int d^nx\,\bar{C}_a(\bar{\pi}^a/2\xi+\partial^\mu A^a_\mu)\,,$$

where ξ is a parameter. The gauge-fixing conditions (1.2) then read

$$A^*_{a\mu}=-\partial_\mu\bar{C}_a\,,\quad C^*_a=0\,,\quad \bar{C}^{*a}=\bar{\pi}^a/2\xi+\partial^\mu A^a_\mu\,,\quad \bar{\pi}^*_a=\bar{C}_a/2\xi\,.$$

The gauge-fixed action is thus

$$W_\psi=\int d^nx\,[-\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a-\partial_\mu\bar{C}_aD^{\mu a}{}_bC^b-(\bar{\pi}^a/2\xi+\partial^\mu A^a_\mu)\bar{\pi}_a]\,.$$

It is invariant under the gauge-fixed BRST transformation

$$s_\psi A^{a\mu}=D^{\mu a}{}_bC^b\,,\quad s_\psi C^a=\frac{1}{2}f_{bc}{}^aC^cC^b\,,\quad s_\psi\bar{C}_a=\bar{\pi}_a\,,\quad s_\psi\bar{\pi}_a=0\,.$$

One can check that this transformation squares to zero. This is a consequence of the fact that the algebra is closed, which implies the vanishing of the right-hand side of Eq.(1.5).

Notice that the field $\bar{\pi}_a$ is an auxiliary field: one can solve its equations of motion to eliminate it. Indeed, the latter read

$$-\bar{\pi}^a/\xi - \partial^\mu A^a_\mu = 0\,,$$

or equivalently

$$\bar{\pi}^a = -\xi \partial^\mu A^a_\mu\,.$$

Inserting this expression for $\bar{\pi}^a$ into the action yields

$$W_\psi = \int d^n x\,[-\tfrac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a - \partial_\mu \bar{C}_a D^{\mu a}{}_b C^b + \tfrac{1}{2}\,\xi \partial^\mu A^a_\mu \partial^\nu A_{a\nu}]\,.$$

The latter action is known as the Faddeev-Popov action for Yang-Mills theories. It is invariant under the BRST symmetry

$$s_\psi A^{a\mu} = D^{\mu a}{}_b C^b\ , \quad s_\psi C^a = \frac{1}{2} f_{bc}{}^a C^c C^b\ , \quad s_\psi \bar{C}^a = -\xi \partial^\mu A^a_\mu\ ,$$

which is obtained from the previous one by using the equations of motion for $\bar{\pi}_a$ to eliminate the latter field. This symmetry is nilpotent only on-shell:

$$s^2_\psi \bar{C}^a = -\xi \partial_\mu D^{\mu a}{}_b C^b \propto \frac{\delta S_\psi}{\delta \bar{C}^a}\,.$$

1.5 Is the path integral independent of the gauge fixing?

We now want to use the gauge-fixed action to quantize our gauge theory in the path integral approach. There are no more problems of infinities due to infinite gauge orbits since there is no gauge invariance left. However, one must check that the computations do not depend on the gauge-fixing fermion one chose. That is the point of this section.

The quantum quantities computed in the path integral approach are

$$I_\psi(X) = \int \mathcal{D}\Phi \mathcal{D}\Phi^*\, \delta(\Phi^*_A - \frac{\delta\psi}{\delta\Phi^A})\, e^{\frac{i}{\hbar}W^q[\Phi,\Phi^*]}\, X[\Phi,\Phi^*]\,,$$

where X is the correlation function of interest. Note that W^q is not the extended action W obtained in the previous sections, but its quantum generalization[4]. It coincides with W in the limit $\hbar \to 0$:

$$W^q = W + \sum_{n=1}^{\infty} \hbar^n W_n\ .$$

Note that there is no general existence theorem for this quantum generalisation, and we will see that it might not exist for certain theories.

Let us define $\mathcal{I}_\psi$ as the integrand:

$$\mathcal{I}[\Phi,\Phi^*] \equiv e^{\frac{i}{\hbar}W^q[\Phi,\Phi^*]}\, X[\Phi,\Phi^*]\,.$$

The variation of $I_\psi(X)$ under an infinitesimal change of ψ , is

$$\begin{aligned}
I_{\psi+\delta\psi}(X) - I_\psi(X) &= \int \mathcal{D}\Phi\, \Big(\mathcal{I}[\Phi, \frac{\delta(\psi+\delta\psi)}{\delta\Phi}] - \mathcal{I}[\Phi, \frac{\delta\psi}{\delta\Phi}]\Big) \\
&= \int \mathcal{D}\Phi\, \Big(\frac{\delta^L(\delta\psi)}{\delta\Phi^A}\frac{\delta^L\mathcal{I}}{\delta\Phi^*_A} + \mathcal{O}((\delta\psi)^2)\Big) \\
&= \int \mathcal{D}\Phi\, \Big(\delta\psi\, \Delta\mathcal{I} + \mathcal{O}((\delta\psi)^2)\Big)\,,
\end{aligned}$$

where

$$\Delta \equiv (-)^{\epsilon_A} \frac{\delta^L}{\delta\Phi^A}\frac{\delta^L}{\delta\Phi^*_A}\ . \tag{1.7}$$

(Some properties of this operator are given in the appendix B .) So the integral $I_\psi(X)$ is infinitesimally independent of ψ iff

$$\Delta\mathcal{I}(X) = 0\,.$$

[4]It contains additionnal terms the presence of which can be interpreted as a requirement to make the measure BRST-invariant [2].

This equation is not only a condition on the operators X of the correlation functions that one wants to compute, but also a condition on the quantum action W^q . Indeed, let us consider the simplest case, $X = 1$. The above equation reads

$$0 = \Delta e^{\frac{i}{\hbar}W^q} = e^{\frac{i}{\hbar}W^q}\Big(\frac{i}{\hbar}\Delta W^q - \frac{1}{2\hbar^2}(W^q, W^q)\Big) .$$

It thus implies

$$\frac{1}{2}(W^q, W^q) = i\hbar\Delta W^q ,$$

which is called the "quantum master equation". This equation is a consistency equation for the quantum extension of the extended action, which ensures that the path integral measure is independent of the choice of the gauge-fixing fermion.

Decomposing it in powers of $\hbar$, the quantum master equation reads

$$(W, W) = 0 , \quad (W_1, W) = i\Delta W ,$$
$$(W_p, W) = i\Delta W_{p-1} - \frac{1}{2}\sum_{n=1}^{p-1}(W_n, W_{p-n}) , \; p \geq 2 .$$

The first of these equations is satisfied automatically, it is the classical master equation (0.1).

The second equation can be seen as an equation to determine W_1. It has a solution iff ΔW is s-exact, $\Delta W = sX$, for some local X. This might be the case, or not. Can one somehow control the existence of such solutions? The answer is yes, through the BRST cohomological group $H^1(s)$ of s-cocycles modulo s-exact terms in ghost number one. Indeed, acting with the BRST differential on the equation $\Delta W = sW_1$, one gets $s\Delta W = s^2 W_1 = 0$. One thus sees that ΔW must be an s-cocycle if the theory is to be consistent, and it is then in some class of $H^1(s)$ since it has ghost number one. If the latter group is trivial, then ΔW is automatically in the trivial class and one can always find W_1. Up to this point the theory is consistent. If $H^1(s)$ is not trivial, then there can be obstructions to the existence of W_1, and one must compute ΔW explicitely to check in which equivalence class it is. If ΔW is in a nontrivial class of $H^1(s)$, then W_1 (and thus the quantum completion of W) does not exist. The theory is inconsistent and one says that it has an anomaly.

This point concludes the part of the lectures devoted to the quantization of gauge theories.

2 Consistent interactions

Let us now turn to the second topic of these lectures, i.e., we want to answer the following question: given a theory (that is, a set of fields and an action to describe their motion), can one introduce consistent local interactions into this theory?

2.1 Basic setting

Consider the "free" action $S_0[\phi^i]$ with "free" gauge symmetries

$$\delta_\varepsilon \phi^i = \overset{(0)i}{R}_\alpha \varepsilon^\alpha,$$

$$\overset{(0)i}{R}_\alpha \frac{\delta S_0}{\delta \phi^i} = 0 .$$

This can be any theory, it is just called "free" because it is our starting point. One wishes to introduce consistent interactions perturbatively, i.e. to modify S_0 as

$$S_0 \longrightarrow S = S_0 + gS_1 + g^2 S_2 + ... , \tag{2.1}$$

without changing the number of physical degrees of freedom.

The latter statement means that there should exist deformed gauge transformations $\delta_\varepsilon \phi^i = R^i_\alpha \varepsilon^\alpha$, where

$$R^i_\alpha = \overset{(0)i}{R}_\alpha + g \overset{(1)i}{R}_\alpha + g^2 \overset{(2)i}{R}_\alpha + ... , \tag{2.2}$$

that are gauge symmetries of the full action (2.1):

$$(\overset{(0)i}{R}_\alpha + g \overset{(1)i}{R}_\alpha + g^2 \overset{(2)i}{R}_\alpha + ...)\frac{\delta(S_0 + gS_1 + g^2 S_2 + ...)}{\delta\phi^i} = 0 \quad . \tag{2.3}$$

If the original gauge transformations are reducible, one should also demand that (2.2) remain reducible, with the same reducibility level.

The deformed theory then possesses the same number of (possibly deformed) independent gauge symmetries, reducibility identities, *etc.*, as the system one started with, so that the number of physical degrees of freedom is unchanged. Equation (2.3) is the key equation to be satisfied.

Interactions obtained by making *field redefinitions* are considered as trivial. Take such a transformation $\phi^i \longrightarrow \bar{\phi}^i = \phi^i + gF^i + ...$. One has

$$\begin{aligned} S_0[\phi^i] \longrightarrow S[\bar{\phi}^i] \equiv S_0[\phi^i[\bar{\phi}^i]] &= S_0[\bar{\phi}^i - gF^i + ...] \\ &= S_0[\bar{\phi}^i] - g\frac{\delta S_0}{\delta\phi^i}F^i + ... \ . \end{aligned} \tag{2.4}$$

We thus see that these transformations are characterized by the fact that the first-order deformation is proportional to the left-hand side of the equations of motion. In the sequel, we will have to keep in mind that we want to mod them out.

2.2 Antifield reformulation

Let us now reformulate the problem in the antifield framework.

If a consistent interacting theory exists, then one can construct the corresponding solution W of the master equation

$$(W, W) = 0 \ .$$

The latter equation guaranties the consistency of the interaction.

To make contact with the free theory, the generator W should admit the generator W_0 of the free theory as its vanishing coupling limit,

$$W = W_0 + gW_1 + g^2 W_2 + ... \ .$$

Inserting this expression into the master equation and splitting it according to the power in the coupling g yields

$$(W_0, W_0) = 0 \tag{2.5}$$
$$2(W_0, W_1) = 0 \tag{2.6}$$
$$2(W_0, W_2) + (W_1, W_1) = 0 \tag{2.7}$$
$$\vdots$$

The first equation is satisfied by assumption, while the second implies that W_1 is a cocycle for the free BRST differential $s \equiv (W_0, \cdot)$.

Can we somehow single out the interactions coming from *trivial field redefinitions*? Actually yes, in a very neet way: they correspond to the case where the first-order deformations are coboundaries: $W_1 = sT = (W_0, T)$. Indeed, S_0 then gets modified as in (2.4)

$$\begin{aligned} S_0 \longrightarrow S_0 + g\ [(W_0, T)]_{\Phi^*=0} &= S_0 + g\left[\frac{\delta^R W_0}{\delta\Phi^A}\frac{\delta^L T}{\delta\Phi^*_A} - \frac{\delta^R W_0}{\delta\Phi^*_A}\frac{\delta^L T}{\delta\Phi^A}\right]_{\Phi^*=0} \\ &= S_0 + g\ \frac{\delta^R S_0}{\delta\phi^i}\left[\frac{\delta^L T}{\delta\phi^*_i}\right]_{\Phi^*=0} \end{aligned} \tag{2.8}$$

The trivial deformations thus correspond to the s-exact quantities.

Since the consistent deformations must be s-cocycles, and that s-exact quantities correspond to trivial deformations, the **nontrivial deformations are thus determined by the equivalence classes of** $H^0(s)$, the cohomology group of the BRST differential of the undeformed theory in ghost number zero.

The next equation, Eq.(2.7), implies that W_1 should be such that the quantity (W_1, W_1) is trivial in $H(s)$ in ghost number one.

In practice, one prefers to work with Lagrangians. So, let $W_k = \int \mathcal{L}_k$ where $\mathcal{L}_k$ is a local d-form, which thus depends on the field variables and only a finite number of their derivatives. In terms of the integrands $\mathcal{L}_k$, the equations (2.6-2.7) for W_k read [5]

$$\begin{aligned} 2s\mathcal{L}_1 &= dj_1 && (2.10)\\ s\mathcal{L}_2 + (\mathcal{L}_1, \mathcal{L}_1) &= dj_2 && (2.11)\\ &\vdots \quad . \end{aligned}$$

The equation (2.10) expresses that $\mathcal{L}_1$ should be BRST-closed modulo d and again, it is easy to see that a BRST-exact term modulo d corresponds to trivial deformations. **Nontrivial local deformations of the master equation are thus determined by** $H^{n,0}(s|d)$, the cohomology of the BRST differential s modulo the total derivative d, in maximal form-degree n and in ghost number 0.

2.3 Example: Yang-Mills

Let us illustrate the above procedure with electromagnetism. We will consider N independent vector fields and sketch the computations to find all their consistent local Poincaré invariant interactions with at most two derivatives.

Basic setting

The set of fields and antifields is composed of

- the vector fields A^a_μ, of vanishing parity and ghost number;
- the ghosts C^a, one for each gauge transformation, of odd parity and ghost number one;
- the antifields $A^{*\mu}_a$, of odd parity and ghost number -1;
- the antifields C^*_a, of even parity and ghost number -2.

Defining the field strength $F^a_{\mu\nu}$ by

$$F^a_{\mu\nu} \equiv \partial_\mu A^a_\nu - \partial_\nu A^a_\mu \; ,$$

the extended action reads

$$W = \int d^n x \Big(-\frac{1}{4} F^a_{\mu\nu} F^{\mu\nu}_a + A^*_{a\mu} \partial^\mu C^a \Big) \; ,$$

and the BRST differential is $s = \gamma + \delta$, where the action of γ and δ is zero on any (anti)field except in the following cases:

$$\gamma A^{a\mu} = \partial^\mu C^b \; , \qquad \delta A^*_{a\mu} = \partial^\nu F_{a\nu\mu} \; , \qquad \delta C^*_a = \partial^\mu A^*_{a\mu} \; .$$

One easily checks that $\gamma^2 = 0$, $\delta^2 = 0$ and $\gamma\delta + \delta\gamma = 0$.

First-order deformation

To find all first-order deformations of the Lagrangian, we must compute $H(s|d)$, i.e. determine the general solution of the cocycle condition

$$s a^{n,0} + d b^{n-1,1} = 0, \tag{2.12}$$

[5] We also denote by (a, b) the antibracket for n-forms, i.e.,

$$(A, B) = \int (a, b) \tag{2.9}$$

if $A = \int a$ and $B = \int b$. Because (A, B) is a local functional, there exists (a, b) such that Eq.(2.9) holds, but (a, b) is defined only up to d-exact terms.

where $a^{n,0}$ is a topform of ghost number zero and $b^{n-1,1}$ a $(n-1)$-form of ghost number one, with the understanding that two solutions of Eq.(2.12) that differ by a trivial solution should be identified

$$a^{n,0} \sim a^{n,0} + sm^{n,-1} + dn^{n-1,0} .$$

The cocycles and coboundaries $a, b, m, n, \ldots$ are local forms of the field variables (including ghosts and antifields)

We will not go through the whole computation here. The main idea is to expand a, b and the equation (2.12) according to the antighost number. So $a = a_0 + a_1 + \ldots + a_k$ where a_i has antifield number i, a similar expansion holds for b, and they have to satisfy

$$\begin{aligned} \delta a_1 + \gamma a_0 + db_0 &= 0 , \\ \delta a_2 + \gamma a_1 + db_1 &= 0 , \\ &\vdots \\ \delta a_k + \gamma a_{k-1} + db_{k-1} &= 0 , \qquad (2.13) \\ \gamma a_k &= 0 . \qquad (2.14) \end{aligned}$$

One calls such a system of equations a "descent". One starts to solve it from the bottom, using the knowledge of cohomological groups like $H(\gamma)$, $H(\delta|d)$ in antighost number k, $H(\gamma|d)$, etc.

With the assumption that the deformation of the Lagrangian should have at most two derivatives, one finds [4] that a_k vanishes (or can be removed by trivial redefinitions) for any $k > 2$, and that the only nontrivial a_2 must be of the form

$$a_2 = f^a_{bc} C^*_a C^b C^c d^n x ,$$

where f^a_{bc} is a constant. Note that f^a_{bc} is antisymmetric in its indices b and c because the ghosts are anticommuting. We will not prove this statement here because the proof, though very interesting, is quite long and technical.

One then tries to "lift" a_2 , i.e. to find a_1 and a_0 that satisfy the other equations of the descent. Inserting a_2 into Eq.(2.13), one finds

$$\begin{aligned} \gamma a_1 &= -f^a_{bc} \partial^\mu A^*_{a\mu} C^b C^c d^n x - db_1 \\ &= 2\, f^a_{bc} A^*_{a\mu} \partial^\mu (C^b) C^c d^n x + d\tilde{b}_1 \\ &= 2\, f^a_{bc} A^*_{a\mu} \gamma(A^{b\mu}) C^c d^n x + d\tilde{b}_1 \\ &= \gamma(-2\, f^a_{bc} A^*_{a\mu} A^{b\mu} C^c d^n x) + d\tilde{b}_1 , \end{aligned}$$

so one can take

$$a_1 = -2\, f^a_{bc} A^*_{a\mu} A^{b\mu} C^c d^n x .$$

Inserting this expression into the top equation of the descent, one finds

$$\begin{aligned} \gamma a_0 &= 2\, f^a_{bc} \partial^\nu F_{a\nu\mu} A^{b\mu} C^c d^n x - db_0 \\ &= -2\, f^a_{bc} F_{a\nu\mu} \partial^\nu A^{b\mu} C^c d^n x - 2\, f^a_{bc} F_{a\nu\mu} A^{b\mu} \partial^\nu C^c d^n x - d\tilde{b}_0 \\ &= -f^a_{bc} F_{a\nu\mu} F^{b\nu\mu} C^c d^n x + \gamma(-2\, f^a_{bc} F_{a\nu\mu} A^{b\mu} A^{c\nu} d^n x) - d\tilde{b}_0 . \end{aligned}$$

It can be shown that the first term on the right-hand side is not γ-exact modulo d, so this equation can be solved only if it vanishes. This implies that the structure functions f^a_{bc} are antisymmetric in a and b, i.e. they must be completely antisymmetric.

We have now obtained a deformation of the free theory that is consistent at first-order in the coupling constant:

$$\mathcal{L}_1 = a_0 + a_1 + a_2 . \qquad (2.15)$$

This deformation reproduces the first-order term of a Yang-Mills theory, except that in the Yang-Mills case the structure functions have to verify one more condition: the Jacobi identity.

Second-order deformation

The next step is to find the second-order deformation of the theory. We will see that it exists only if f^a_{bc} satisfies the Jacobi identity. The equation to be solved is

$$s\mathcal{L}_2 + (\mathcal{L}_1, \mathcal{L}_1) = dj_2 \; .$$

A first consistency check is that $(\mathcal{L}_1, \mathcal{L}_1)$ must be s-exact modulo d. Using some theorems, one shows that this implies that the term with highest antighost number in $(\mathcal{L}_1, \mathcal{L}_1)$ must be γ-exact. However, this term is

$$(a_2, a_2) \propto f^a_{bc} C^b C^c f^e_{af} C^*_e C^f d^n x$$

and this expression is nontrivial in $H(\gamma)$. It must thus vanish. Because the product $C^b C^c C^f$ is antisymmetric in its indices, it is enough that the part of $f^a_{bc} f^e_{fa}$ antisymmetric in the indices b, c and f vanishes, i.e. that the structure functions f^a_{bc} obey the Jacobi identity.

End of the computation

The end of the computation is straightforward. We know that the only consistent $\mathcal{L}_1$ is given by Eq.(2.15) plus the Jacobi identity. We know that Yang-Mills theories are consistent and have exactly this first-order deformation. Putting the two facts together, we see that the only consistent interaction with at most two derivatives in the Lagrangian must be Yang-Mills.

We hope that this example, though incompletely treated, has given a taste of what the computations to find consistent interactions look like. The interested reader might want to have a look at a similar (and complete!) computations that has been done to prove the unicity of the Einstein action (under certain simple assumptions) [5]. The technique has been used in many other instances which will not be listed here. A more complete review can be found in the thesis [6].

Appendix

A Left and right derivatives

The following formulae involving left and right derivatives are very useful in the calculations.

The link between left and right derivatives of a function X is

$$\frac{\delta^R X}{\delta\phi} = (-)^{\epsilon_\phi(\epsilon_X+1)} \frac{\delta^L X}{\delta\phi} \; .$$

The commutation relations between derivatives are given by

$$\frac{\delta^R \delta^R X}{\delta\phi^A \delta\phi^B} = (-)^{\epsilon_A \epsilon_B} \frac{\delta^R \delta^R X}{\delta\phi^B \delta\phi^A} \tag{A1}$$

$$\frac{\delta^L \delta^R X}{\delta\phi^A \delta\phi^B} = \frac{\delta^R \delta^L X}{\delta\phi^B \delta\phi^A} \tag{A2}$$

$$\frac{\delta^L \delta^L X}{\delta\phi^A \delta\phi^B} = (-)^{\epsilon_A \epsilon_B} \frac{\delta^L \delta^L X}{\delta\phi^B \delta\phi^A} \tag{A3}$$

The derivation chain rule reads

$$\frac{\delta^R X(Y(\phi))}{\delta\phi} = \frac{\delta^R X}{\delta Y} \frac{\delta^R Y}{\delta\phi} \; , \qquad \frac{\delta^L X(Y(\phi))}{\delta\phi} = \frac{\delta^L Y}{\delta\phi} \frac{\delta^L X}{\delta Y} \tag{A4}$$

B Some properties of the operator Δ

The operator Δ defined by Eq.(1.7) is of odd grassmannian parity and its ghost number is one,

$$\epsilon(\Delta) = 1 \; , \quad gh(\Delta) = 1 \; .$$

It is nilpotent

$$\Delta^2 = 0 \; .$$

Its action on an antibracket and a pointwise product yields

$$\begin{aligned} \Delta(\alpha,\beta) &= (\Delta\alpha,\beta) - (-)^{\epsilon_\alpha}(\alpha,\Delta\beta)\ , \\ \Delta(\alpha\beta) &= (\Delta\alpha)\beta + (-)^{\epsilon_\alpha}\alpha(\Delta\beta) + (-)^{\epsilon_\alpha}(\alpha,\beta)\ . \end{aligned} \tag{B1}$$

The proofs of these equations are straightforward by using the definitions and the rules of Appendix A .

Acknowledgements

I of course want to thank the other organisers of the Modave Summer School. I hope this is the second milesstone of a long series. I would also like to thank the people who have been teaching me (also through their writings!) or working on the topic of these lectures with me, that is, Glenn Barnich, Xavier Bekaert, Nicolas Boulanger, Bernard Knaepen, Serge Leclercq, Marc Henneaux and Christiane Schomblond.

My work is supported in part by the “Actions de Recherche Concertées” of the “Direction de la Recherche Scientifique - Communauté Française de Belgique”, by a “Pôle d’Attraction Interuniversitaire” (Belgium), by IISN-Belgium (convention 4.4505.86) and by the European Commission RTN programme HPRN-CT-00131, in which I am associated with K. U. Leuven.

References

[1] J. Gomis, J. Paris and S. Samuel, “Antibracket, antifields and gauge theory quantization,” Phys. Rept. **259** (1995) 1 [arXiv:hep-th/9412228].

[2] M. Henneaux and C. Teitelboim, “Quantization of gauge systems,” Princeton, USA: Univ. Pr. (1992) 520 p.

[3] A. Fuster, M. Henneaux and A. Maas, “BRST quantization: A short review,” Int. J. Geom. Meth. Mod. Phys. **2** (2005) 939 [arXiv:hep-th/0506098].

[4] G. Barnich, M. Henneaux and R. Tatar, “Consistent interactions between gauge fields and the local BRST cohomology: Int. J. Mod. Phys. D **3** (1994) 139 [arXiv:hep-th/9307155].

[5] N. Boulanger, T. Damour, L. Gualtieri and M. Henneaux, “Inconsistency of interacting, multigraviton theories,” Nucl. Phys. B **597** (2001) 127 [arXiv:hep-th/0007220].

[6] S. Cnockaert, “Higher spin gauge field theories: Aspects of dualities and interactions,” arXiv:hep-th/0606121.

Topology of fibre bundles and global aspects of gauge theories

Andrés Collinucci[a,1] and Alexander Wijns[b,2]

[a] Instituut voor Theoretische Fysica, Katholieke Universiteit Leuven
Celestijnenlaan 200D, B-3001 Leuven, Belgium

[b] Theoretische Natuurkunde, Vrije Universitiet Brussel &
International Solvay Institutes,
Pleinlaan 2, B-1050 Brussels, Belgium

Abstract. In these lecture notes we will try to give an introduction to the use of the mathematics of fibre bundles in the understanding of some global aspects of gauge theories, such as monopoles and instantons. They are primarily aimed at beginning PhD students. First, we will briefly review the concept of a fibre bundle and define the notion of a connection and its curvature on a principal bundle. Then we will introduce some ideas from (algebraic and differential) topology such as homotopy, topological degree and characteristic classes. Finally, we will apply these notions to the bundle setup corresponding to monopoles and instantons. We will end with some remarks on index theorems and their applications and some hints towards a bigger picture.

[1] email: andres.collinucci@fys.kuleuven.be
[2] Aspirant FWO; e-mail: awijns@tena4.vub.ac.be

1 Fibre Bundles

A fibre bundle is a manifold[3] that looks locally like a product of two manifolds, but isn't necessarily a product globally. Because of their importance in modern theoretical physics, many introductory expositions of fibre bundles for physicists exist. We give a far from exhaustive list in the references. This section is mainly inspired by [1] and [2]. To get some intuition for the bundle concept, let us start off with the easiest possible example.

1.1 Invitation: the Möbius strip

Consider a rectangular strip. This can of course be seen as the product of two line segments. If now one wants to join two opposite edges of the strip to turn one of the line segments into a circle, there are two ways to go about this. The first possibility is to join the two edges in a straightforward way to form a cylinder C, as on the left hand side of figure 4.1. It should be intuitively clear that the cylinder is not only locally a product, but also globally; namely $C = S^1 \times L$, where L is a line segment. A fibre bundle will be called trivial if it can be described as a global product. The cylinder is trivial in this sense, because it is not very difficult to find a global diffeomorphism from $S^1 \times L$ to C.

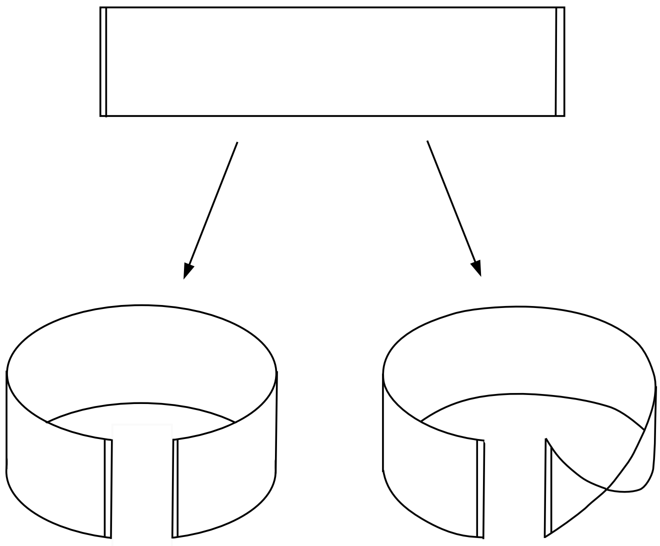


Figure 4.1: The cylinder and the Möbius strip.

The second way to join the edges of the strip is of course the more interesting one. Before gluing the edges together, perform a twist on one of them to arrive at a Möbius strip Mo, as shown on the right hand side of figure 4.1. Locally, along each open subset U of the S^1, the Möbius strip still looks like a product, $Mo = U \times L$. Globally, however, there is no unambiguous and continuous way to write a point m of Mo as a cartesian pair $(s, t) \in S^1 \times L$. The Möbius strip is therefore an example of a manifold that is not a global product, that is, of a non-trivial fibre bundle.

Since Mo is still locally a product, we can try to use this 'local triviality' to our advantage to find a useful way to describe it. Although we cannot write Mo as $S^1 \times L$, we can still project down any point m of Mo onto the circle, i.e. there is a projection π:

$$\pi : Mo \to S^1 \,, \tag{1.1}$$

so that, for every $x \in S^1$, its inverse image is isomorphic to the line segment, $\pi^{-1}(x) \cong L$. This leads to a natural way to define local coordinates on Mo, namely for every open subset U of S^1,

[3]More generally, one can define a fibre bundle as being any topological space. We will use the term fibre bundle in the more restrictive sense of manifolds throughout these notes. Whenever we will say manifold, we will mean differentiable manifold.

we can define a diffeomorphism

$$\phi : U \times L \to \pi^{-1}(U)\,. \tag{1.2}$$

This means that to every element p of $\pi^{-1}(U) \subset Mo$, we can assign local coordinates $\phi^{-1}(p) = (x,\, t)$, where $x = \pi(p) \in U$ by definition and $t \in L$. Now, how can we quantify the non-triviality of the Möbius strip? For this, cover the circle by two open sets, U_1 and U_2, which overlap on two disjoint open intervals, A and B. We also have the diffeomorphisms

$$\begin{aligned} \phi_1 &: U_1 \times L \to \pi^{-1}(U_1)\,, \\ \phi_2 &: U_2 \times L \to \pi^{-1}(U_2)\,. \end{aligned} \tag{1.3}$$

It is clear that the non-triviality of Mo will reside in the way in which the different copies of L will be mapped to each other on A and B. To this end, we need an automorphism of L over the region $A \cup B = U_1 \cap U_2$. This is provided by ϕ_1 and ϕ_2 of eq. (1.3). For every $x \in U_1 \cap U_2$, we can define

$$\phi_1^{-1} \circ \phi_2 : (A \cup B) \times L \to (A \cup B) \times L\,. \tag{1.4}$$

This induces a diffeomorphism g_{12} from L to L in such a way that

$$\phi_1^{-1}\phi_2(x,\, t) = (x,\, g_{12}(t))\,. \tag{1.5}$$

Since the only linear diffeomorphisms of L are the identity e and the sign-flip, $g(t) = -t$, we necessarily have that $g_{12} \in \{e,\, g\}$. We can always choose $g_{12} = e$ on A, so that for the Möbius strip g_{12} will have to equal g on B. We see that the non-triviality of the Möbius strip is encoded in the non-triviality of these 'transition functions'. We can now also understand the difference with the cylinder in another way. The same construction for the cylinder would lead to the identity on both A and B. So we see that the triviality of the cylinder is reflected in the triviality of the transition functions and that these functions encode the non-triviality of the Möbius strip. Since $g^2 = e$, in this example the transition functions form the group $\mathbb{Z}_2$. In general this group will be called the structure group of the bundle and it will turn out to be an ingredient of utmost importance in the description of bundles.

1.2 Definition of a bundle

Let us now turn to the formal definition of a fibre bundle. Most of the ingredients should now be intuitively clear from the previous example.

Definition 1. A (differentiable) **fibre bundle** $(E,\, \pi,\, M,\, F,\, G)$ consists of the following elements:

(i) A differentiable manifold E called the **total space**.

(ii) A differentiable manifold M called the **base space**.

(iii) A differentiable manifold F called the (typical) **fibre**.

(iv) A surjection $\pi : E \to M$ called the **projection**. For $x \in M$, the inverse image $\pi^{-1}(x) \equiv F_x \cong F$ is called the fibre at x.

(v) A (Lie) Group G called the **structure group**, which acts on the fibre on the left.

(vi) An open covering $\{U_i\}$ of M and a set of diffeomorphisms $\phi_i : U_i \times F \to \pi^{-1}(U_i)$ such that $\pi\phi_i(x,\, t) = x$. The map ϕ_i is called a **local trivialization**.

(vii) At each point $x \in M$, $\phi_{i,x}(t) \equiv \phi_i(x,\, t)$ is a diffeomorphism, $\phi_{i,x} : F \to F_x$. On each overlap $U_i \cap U_j \neq \emptyset$, we require $g_{ij} = \phi_{i,x}^{-1}\phi_{j,x} : F \to F$ to be an element of G, i.e. we have a smooth map $g_{ij} : U_i \cap U_j \to G$ such that

$$\phi_j(x,\, t) = \phi_i(x,\, g_{ij}(x)t).$$

In the mathematical literature this defines a coordinate bundle. Of course the properties of a bundle should not depend on the specific covering of the base manifold or choice of local trivialisations. A bundle is therefore defined as an equivalence class of coordinate bundles[4]. Since in practical applications physicists always work with an explicit choice of covering and trivialisations, we will not bother to make this distinction here.

Intuitively, one can view a fibre bundle as a manifold M with a copy of the fibre F at every point of M. The main difference with a product manifold (trivial bundle) is that the fibres can be 'twisted' so that the global structure becomes more intricate than a product. This 'twisting' is basically encoded in the transition functions which glue the fibres together in a non-trivial way. For the Möbius strip in the previous section, Mo was the total space, the base space was a circle and the fibre was the line segment L. In that example the structure group turned out to be the discrete group $\mathbb{Z}_2$. In that respect this was not a typical example, because in what follows all other examples will involve continuous structure groups. For convenience we will sometimes use $E \xrightarrow{\pi} M$ or simply E, to denote (E, π, M, F, G).

Let us look at some of the consequences of the above definition. From (vi) it follows that $\pi^{-1}(U_i)$ is diffeomorphic to a product, the diffeomorphism given by $\phi_i^{-1} : \pi^{-1}(U_i) \to U_i \times F$. It is in this sense that E is locally a product. From their definition (vii), it is clear that on triple overlaps, the transition functions obey

$$g_{ij} g_{jk} = g_{ik} \,, \qquad \text{on} \quad U_i \cap U_j \cap U_k \neq \emptyset. \tag{1.6}$$

Taking $i = k$ in the above equation shows that

$$g_{ij}^{-1} = g_{ji} \,, \qquad \text{on} \quad U_i \cap U_j \neq \emptyset. \tag{1.7}$$

These conditions evidently have to be fulfilled to be able to glue all local pieces of the bundle together in a consistent way. A fibre bundle is trivial if and only if all transition functions can be chosen to be identity maps. Since a choice of local trivialization ϕ_i results in a choice of local coordinates, the transition functions are nothing but a transformation of 'coordinates' in going from one open subset to another. When we will discuss gauge theories they will represent gauge transformations in going from one patch to another.

Of course, one should be able to change the choice of local trivializations (coordinates) within one patch. Say that for an open covering U_i of M we have two sets of trivializations $\{\phi_i\}$ and $\{\widetilde{\phi}_i\}$ of the same fibre bundle. Define a map $f_i : F \to F$ at each point $x \in U_i$

$$f_i(x) = \phi_{i,x}^{-1} \widetilde{\phi}_{i,x}. \tag{1.8}$$

It is easy to show that the transition functions corresponding to both trivializations are related by

$$\widetilde{g}_{ij}(x) = f_i(x)^{-1} g_{ij}(x) f_j(x). \tag{1.9}$$

While the g_{ij} will be gauge transformations for gluing patches together, the f_i will be gauge transformations within a patch. From eq. (1.9) it's clear that in general the transition functions of a trivial bundle will have the factorized form

$$g_{ij}(x) = f_i(x)^{-1} f_j(x). \tag{1.10}$$

1.3 More examples: vector and principal bundles

The prototype of a fibre bundle is the tangent bundle of a differentiable manifold. As described in subsection A , the collection of all tangent vectors to a manifold M at a point x is a vector space called the tangent space $T_x M$. The collection $\{T_x M | x \in M\}$ of all tangent spaces of M is called the tangent bundle TM. Its base manifold is M and fibre $\mathbb{R}^m$, where m is the dimension of M. Its structure group is a subgroup of $GL(m, \mathbb{R})$. Let us look at some examples of tangent bundles.

$T\mathbb{R}^n$. If M is $\mathbb{R}^n$, the tangent space to every point is isomorphic to M itself. Its tangent bundle $T\mathbb{R}^n$ is clearly trivial and equal to $\mathbb{R}^n \times \mathbb{R}^n \cong \mathbb{R}^{2n}$. It can be proven that every bundle over a manifold that is contractible to a point is trivial.

[4] For more details, see for example [8].

TS^1**.** The circle is not contractible, yet its tangent bundle TS^1 is trivial. The reason is that since one can globally define a (unit) vector along the circle in an unambiguous and smooth way, it is easy to find a diffeomorphism from TS^1 to $S^1 \times \mathbb{R}$.

TS^2**.** The tangent bundle of the 2-sphere TS^2 is our second example of a non-trivial bundle. There is no global diffeomorphism between TS^2 and $S^2 \times \mathbb{R}^2$, since to establish this one would have to be able to define two linearly independent vectors at every point of the sphere in a smooth fashion. (This is needed to be able to define coordinates on the tangent plane in a smooth way along the sphere.) In fact it's even worse for the 2-sphere, since in this case one cannot even find a single global non-vanishing vector field. The fact that this isn't possible has become known as the expression: "You cannot comb hair on a sphere."

FS^2**.** A set of pointwise linearly independent vectors over an open set of the base manifold of a tangent bundle is called a frame. So in the example above, the non-triviality of TS^2 was a consequence of not being able to find a frame over the entire sphere in a consistent way. At each point one can of course construct many different sets of linearly independent vectors. These are all related to each other by a transformation in the structure group, $GL(2,\mathbb{R})$ in this case. Since the action of $GL(2,\mathbb{R})$ on the set of frames is free (no fixed points for $g \neq e$) and transitive (every frame can be obtained from a fixed reference frame by a group element), the set of all possible frames over an open set U of S^2 is diffeomorphic to $U \times GL(2,\mathbb{R})$. Globally this becomes a bundle over S^2 with fibre $GL(2,\mathbb{R})$ and is called the frame bundle FS^2 of S^2. Note that for this bundle the fibre equals the structure group!

The first three examples above are examples of vector bundles, so let us define these properly.

Definition 2. A **vector bundle** $E \xrightarrow{\pi} M$ is a fibre bundle whose fibre is a vector space. If $F = \mathbb{R}^n$ it is common to call n the fibre dimension and denote it by $\dim E$ (although the total dimension of the bundle is $\dim M + n$). The transition functions belong to $GL(n,\mathbb{R})$.

Clearly a tangent bundle is always a vector bundle. Once one defines a frame $\{e_a\}$, $a \in \{1,...,n\}$, over a patch $U \subset M$, one can expand any vector field $V : U \to \mathbb{R}^n$ over U in terms of this frame, $V = V^a e_a$. A possible local trivialization would then become

$$\phi^{-1}(p) = (\pi(p), \{V^a\}), \qquad p \in E \tag{1.11}$$

Consider two coordinate frames, associated to a set of coordinates $\{x^a\}$ on U_x and $\{y^a\}$ on U_y respectively. A vector field V on the overlap $U_x \cap U_y$ can be expanded using either frame (this notation is discussed in more detail in subsection A),

$$V = V^a \frac{\partial}{\partial x^a} = \widetilde{V}^a \frac{\partial}{\partial y^a} \tag{1.12}$$

The resulting trivializations are related as follows (in a hopefully obvious notation),

$$\phi_y^{-1}\phi_x(\pi(p), \{V^a\}) = (\pi(p), \{\widetilde{V}^a\}), \tag{1.13}$$

where

$$V^a = \frac{\partial x^a}{\partial y^b} \widetilde{V}^b. \tag{1.14}$$

This relation is of course well known from basic tensor calculus. In the language we developed in the previous section this would be written as

$$(g_{yx})^a{}_b = \frac{\partial x^a}{\partial y^b}. \tag{1.15}$$

In the frame bundle example above, the fibre was not a vector space, but a Lie group. More importantly, the fibre equalled the structure group. This is an example of a principal bundle, the most important kind of bundle for understanding the topology of gauge theories. We will need them a lot in this set of lectures, so let us look at them in a little more detail.

Definition 3. A **principal bundle** has a fibre which is identical to the structure group G. It is usually denoted by $P(M,\, G)$ and called a G-bundle over M.

Most of the time, G will be a Lie group. The only example we will encounter where this is not the case, is the Möbius strip. The action of the structure group on the fibre now simply becomes left multiplication within G. In addition, we can also define right multiplication (on an element p of P) as follows. Let $\phi_i : U_i \times G \to \pi^{-1}(U_i)$ be a local trivialization,

$$\phi_i^{-1}(p) = (x,\, g_i), \qquad x = \pi(p). \tag{1.16}$$

Right multiplication by an element a of G is defined by

$$p\,a = \phi_i(x,\, g_i a). \tag{1.17}$$

Since left and right multiplications commute (associativity of the group), this action is independent of the choice of local coordinates. Let $x \in U_i \cap U_j$, then

$$p\,a = \phi_j(x,\, g_j a) = \phi_j(x,\, g_{ji}(x) g_i a) = \phi_i(x,\, g_i a). \tag{1.18}$$

We can thus just as well write the action as $P \times G \to P : (p,\, a) \mapsto p\,a$, without reference to local choices. One can show that this action is transitive and free on $\pi^{-1}(x)$ for each $x \in M$.

1.4 Triviality of a bundle

Later on, we will discuss a number of ways in which one can quantify the non-triviality of a given bundle. Of course before we try to quantify how much it deviates from triviality, it is interesting to know wether it is non-trivial at all. We will now discuss some equivalent ways to see wether a bundle is a global product or not. Before we do that, we define one more important notion

Definition 4. A **section** s is a smooth map $s : M \to E$ such that $\pi s(x) = x$ for all $x \in M$. This is sometimes also referred to as a global section. If a section can only be defined on an open set U of M, it is called a **local section** and one only has the smooth map $s : U \to E$. The set of all sections of E is called $\Gamma(M,\, E)$, while the set of all local sections over U will be called $\Gamma(U,\, E)$.

The best known example of this is a vector field over a manifold M, which is a section of the tangent bundle TM. Clearly, it is not a great challenge to construct a local section over some open subset $U \subset M$. Being able to construct a (global) section over M will place much stronger requirements on the topology of the bundle and it will have a lot to say about the non-triviality of a bundle. This is reflected in the following theorem:

Theorem 1. A vector bundle of rank n is trivial if and only if it admits n point-wise linearly independent sections, i.e. a global frame.

This is precisely why the tangent bundle over the 2-sphere is non-trivial. On the other hand, for a vector bundle there always exists at least one global section, namely the so called zero section, i.e. the section which maps every point $x \in M$ to $(x,\, O)$, where O is the origin of the vector space. This is always possible irrespective of the non-triviality of the bundle, since we do not need a frame to characterize the origin uniquely. This is not so for a principal bundle and the group structure of the fibre allows for the following very powerful theorem:

Theorem 2. A principal bundle is trivial if and only if it allows a global section.

To illustrate the above theorems, we return to the discussion of the Möbius strip Mo. In the first section we saw that the non-triviality of the Möbius strip had to do with the fact that one could not find a global trivialization. We now understand that this is the case because one cannot define a linearly independent (which in one dimension means everywhere nonzero) section on Mo[5]. It is not very difficult to see (figure 4.2) that every smooth section would have to take the value zero over at least one point of the circle. According to Theorem 1 this is equivalent to the bundle being non-trivial.

To illustrate the second theorem, we would like to associate a principal bundle $P(S^1, \mathbb{Z}_2)$ to Mo. To accomplish this, replace the fibre L by the structure group $\mathbb{Z}_2$. Take the same open

[5]The reader might be bothered by the fact that the Möbius strip is not a vector bundle. One can however replace the line segment L by the real line $\mathbb{R}$ and all the arguments used in the text will still apply. In this case one would speak of a line bundle over S^1

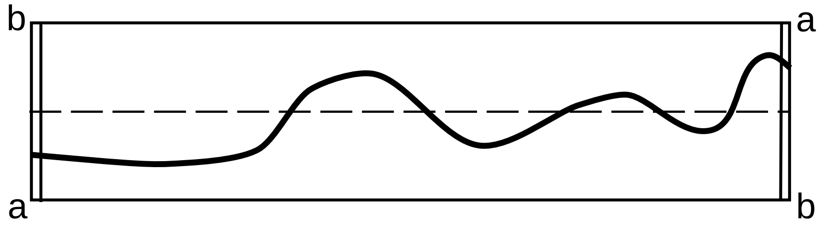


Figure 4.2: A section of the bundle corresponding to the Möbius strip. The points marked with the same letter (a or b) should be identified.

covering of the circle as in subsection 1.1 and use the same transition functions on the overlap $A \cup B$, where now instead of acting on the line segment L, they act by left multiplication within $\mathbb{Z}_2$. What one gets is a double cover of the circle as depicted in figure 4.3. To get a global section one has to go around the circle once and it is clear that this will always have a discontinuity somewhere. Thus we cannot find a section of this principal bundle, so that according to Theorem 2 it has to be non-trivial.

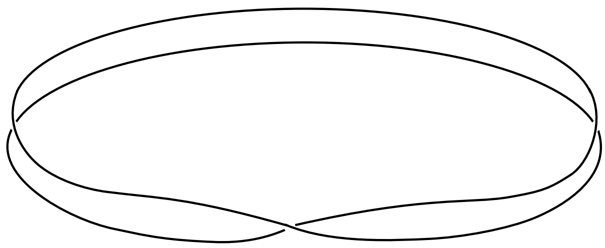


Figure 4.3: The principal bundle $P(S^1, \mathbb{Z}_2)$ associated to the Möbius strip is a double cover of the circle.

The above construction is more general: starting from a fibre bundle (not necessarily a vector bundle), one can always construct the associated principal bundle by replacing the fibre with the structure group and keeping the transition functions. Note that the frame bundle FS^2 over S^2 from subsection 1.3 was nothing but the principal bundle associated to the tangent bundle TS^2. More generally two bundles with same base space and structure group are called associated if their respective associated principal bundles are equivalent. In gauge theories it is important to be able to associate a principal bundle to a vector bundle and vice versa.

Definition 5. Start from a principal bundle $P(M, G)$ and an n-dimensional faithful representation $\rho : G \to GL(n, \mathbb{R})$ which acts on $V = \mathbb{R}^n$ from the left. Consider the product $P \times V$. Define the equivalence relation $(p, v) \sim (p\, g^{-1}, \rho(g)v)$, where $p \in P$, $v \in V$ and $g \in G$. The vector bundle E_ρ associated to P via the representation ρ is defined as

$$E_\rho = P \times_\rho V \equiv P \times V/\sim \tag{1.19}$$

This is basically a complicated way of saying that one changes the fibre from G to V and use as transition function $\rho(g_{ij})$ instead of g_{ij}. Since every element of P over a certain point $x \in M$ can be obtained from (x, e) (where e is the identity of G) by an element of G, the equivalence relation $(p\, g, v) \sim (p, \rho(g)v)$ effectively replaces the fibre over x with V, thus replacing a principal bundle by a vector bundle. If we define the new projection by

$$\pi_E[(p, v)] = \pi(p), \tag{1.20}$$

this is well defined under the equivalence relation since $\pi(p\, g) = \pi(p)$. If $\phi_P(\pi(p), g) = p$, $p \in P$ is a local trivialization on $U \subset M$, we define for E_ρ

$$\phi_E^{-1} : E_\rho \to U \times V : [(p, v)] \mapsto (\pi(p), \rho(g)v). \tag{1.21}$$

This definition is independent of the representative of the equivalence class. To see this, take two different representatives:

$$[(p,\, v)] = [(p\, h^{-1},\, \rho(h)v)]. \tag{1.22}$$

To find their trivialization corresponding to a trivialization over U in the associated principal bundle P, we note that

$$p = \phi_P(x, g), \qquad x = \pi(p); \tag{1.23}$$
$$p\, h^{-1} = \phi_P(x, g\, h^{-1}), \qquad x = \pi(p\, h^{-1}) = \pi(p). \tag{1.24}$$

From this, we find

$$[(p\, h^{-1},\, \rho(h)v)] = \phi_E(x,\, \rho(g\, h^{-1})\rho(h)v) = \phi_E(x,\, \rho(g)v). \tag{1.25}$$

To see what the transition functions are, consider a point $[(p,\, v)] \in E_\rho$. If we choose a trivialization of U_i such that $p = \phi^i_P(x,\, e)$, then on U_j there is a trivialization such that on $U_i \cap U_j \neq \emptyset$, we have $p = \phi^i_P(x,\, g_{ji})$. For the corresponding trivializations of E_ρ, this means

$$[(p,\, v)] = \phi^i_E(x, v) = \phi^j_E(x, \rho(g_{ji})v) = \phi^j_E(x, \rho_{ji}v). \tag{1.26}$$

This shows that the new transition functions ρ_{ij} are just $\rho(g_{ij})$.

Now that we have defined all necessary ingredients, we can formulate the main theorem of this subsection[6]

Theorem 3. A vector bundle is trivial if and only if its associated principal bundle is trivial.

This means that for establishing the (non-)triviality of a vector bundle, we only need to study its associated principal bundle. More concretely:

Corollary 1. A vector bundle is trivial if an only if its associated principal bundle admits a section.

Sometimes the converse is more useful:

Corollary 2. A principal bundle is trivial if an only if its associated vector bundle of rank n admits n point-wise linearly independent sections.

This is exactly why both the bundle Mo and its associated principal bundle were non-trivial. We were basically looking at the same issue from two different points of view.

2 Connections on fibre bundles

Now that we have gained some feeling for the concept of a bundle, we want to define some extra structure on it. We are all familiar with the notion of parallel transport of a tangent vector on a manifold in General Relativity. Given a curve in space-time, there are many possible choices to transport a given vector along this curve, which are all equally valid as a choice for what 'parallel' might mean[7]. Translated into the language of fibre bundles, we want to, given a curve γ in the base (space-time) M, define a corresponding section of the tangent bundle $s^\gamma \in TM$, in such a way that $\pi(s^\gamma) = \gamma$ (otherwise it would not be a section).

The question we want to ask ourselves now is basically a generalization of this. Given a certain motion in the base manifold, how can we define a corresponding motion in the fibre? Since, given a bundle, there is no a priori notion of what 'parallel' should mean, we need some additional structure to give meaning to the notion of 'parallel motion'. This will be provided by the choice of a connection on the bundle. This section follows parts of [1] and [4] closely, although some specific points are more indebted to [3]. We will start by defining parallel transport on a principal bundle and later sketch how this connection can be used to provide a connection on an associated vector bundle. But first of all, let us pause and recall some facts about Lie groups.

[6] There is actually a more general theorem which states that a bundle is trivial iff all its associated bundles are trivial.

[7] The Levi-Civita connection corresponds to a choice of parallel transport such that if a curve is the shortest path between two points, a tangent vector to this curve stays tangent to the curve under parallel transport.

2.1 Lie groups and algebras

From now on the structure group G will be a Lie group, i.e. a differential manifold with a group structure, where the group operations (multiplication, inverse) are differentiable. Given an element $g \in G$, we can define the left- and right-translation of an element $h \in G$ by g,

$$L_g(h) = gh; \tag{2.1}$$
$$R_g(h) = hg. \tag{2.2}$$

These induce differential maps in the tangent space (see subsection C for a review of differential maps)

$$L_{g*}: \quad T_hG \to T_{gh}G; \tag{2.3}$$
$$R_{g*}: \quad T_hG \to T_{hg}G. \tag{2.4}$$

We say that a vector field X is left-invariant if it satisfies

$$L_{g*}\left(X|_h\right) = X|_{gh}. \tag{2.5}$$

One can show that if X and Y are left-invariant vector fields, their Lie bracket $[X, Y]$ is also left-invariant. The algebra formed in this way by the left-invariant vector fields of G, is called the Lie algebra $\mathfrak{g}$. Because of (2.5), a vector A in T_eG (e is the unit element of G) uniquely defines a left-invariant vector field X_A over G (that is, a section of TG). This establishes an isomorphism between T_eG and $\mathfrak{g}$. From now on we will not make the distinction between the two anymore and say that the Lie algebra $\mathfrak{g}$ is the tangent space to the identity in G. The generators T_a, $a \in \{1, ..., r\}$ ($r = \dim\mathfrak{g} = \dim G$) of $\mathfrak{g}$ satisfy the well known relations

$$[T_a, T_b] = f_{ab}{}^c\, T_c\,, \tag{2.6}$$

where $f_{ab}{}^c$ are the structure constants of $\mathfrak{g}$ (and can be shown to really be constant).

By Lie group we will always mean a matrix group (subgroup of $GL(n)$ with matrix multiplication as group operation), although most of what we will discuss can be proven in a more general context. From our experience with matrix groups we know that exponentiation maps elements of the Lie algebra to elements of the Lie group. Consequently, $A \in \mathfrak{g}$ generates a curve (one-parameter subgroup) through g in G by

$$\sigma_t(g) = g\exp(tA) = R_{\exp(tA)}(g). \tag{2.7}$$

The corresponding flow equation is (in matrix notation)

$$\left.\frac{d\sigma_t(g)}{dt}\right|_{t=0} = gA = L_{g*}A = X_A|_g\,, \tag{2.8}$$

where we used that in matrix notation $L_{g*}A = gA$ (exercise) and that A generates a left-invariant vector field X_A. This shows that the tangent vector to the curve through g is nothing but the left-translation of A by L_{g*} (or the left-invariant vector field generated by A evaluated at g, $X_A|_g$). More formally one would define the tangent vector to the one-parameter flow by making use of a function $f : G \to \mathbb{R}$,

$$X_A(f(g)) = \left.\frac{d}{dt}f(\sigma_t(g))\right|_{t=0}. \tag{2.9}$$

Given a basis $\{T_a\}$ of $\mathfrak{g}$, one defines a corresponding basis $\{X_a\}$ of T_gG by left translation to $g \in G$, $X_a = L_{g*}T_a$. This means that for an element $A \in \mathfrak{g}$, if $A = A^aT_a$, the left-invariant vector field corresponding to A can be expanded as $X_A = A^aX_a$. One can also define a basis for left-invariant 1-forms $\{\eta^a\}$ dual to to $\{X_a\}$ by, $\eta^a(X_b) = \delta^a_b$. The Maurer-Cartan form Θ is then defined by

$$\Theta = T_a \otimes \eta^a. \tag{2.10}$$

This is a $\mathfrak{g}$-valued 1-form, which takes a left-invariant vector field $X|_g$ at g and pulls it back to the identity, giving back the Lie algebra element $X|_e$. To see this, evaluate the Maurer-Cartan form on X_A defined above,

$$\Theta(X_A)|_g = T_a \otimes \eta^a(A^bX_b) = A^aT_a = A. \tag{2.11}$$

In this way, the Maurer-Cartan form establishes an explicit isomorphism between $\mathfrak{g}$ and T_eG. More generally, it takes any vector at g and returns a Lie algebra element, thus establishing the decomposition of the vector in terms of a basis of left-invariant vector fields at g.

To make contact with widely used notation in physics, we will now be a bit less formal. Choosing coordinates $\{g^i\}$ on a patch of G, a coordinate basis at g can be written as $\{\partial/\partial g^i\}$. A coordinate basis at the identity e would in this notation be written as

$$\left.\frac{\partial}{\partial g^i}\right|_e = L_{g^{-1}*}\frac{\partial}{\partial g^i}. \tag{2.12}$$

This would mean that (2.10) can be rewritten as

$$\Theta = L_{g^{-1}*}\frac{\partial}{\partial g^i}\otimes dg^i = g^{-1}\frac{\partial}{\partial g^i}\otimes dg^i, \tag{2.13}$$

where the last equality is for a matrix group. Physicists write this as

$$\Theta = g^{-1}dg, \tag{2.14}$$

where dg should be interpreted as the identity operator at g,

$$dg = \frac{\partial}{\partial g^i}\otimes dg^i. \tag{2.15}$$

Since for a matrix group $X_A|_g = gA$, we get indeed that

$$\Theta(X_A) = g^{-1}dg(X_A) = g^{-1}gA = A. \tag{2.16}$$

The reason why this notation makes sense is because if A is tangent to a flow $\sigma_t(e)$, $X_A|_g = gA$ is tangent to the flow $\sigma_t(g)$. Concretely, we have

$$A = \left.\frac{d\sigma_t(e)}{dt}\right|_{t=0}, \tag{2.17}$$

so that,

$$X_A|_g = gA = \left.\frac{d\sigma_t(g)}{dt}\right|_{t=0} \equiv X_A(g) = dg(X_A). \tag{2.18}$$

We see that if we interpret $X_A(g)$ as $X_A|_g$ (for a matrix group in matrix notation), we can almost act as if $dg(X_A)$ has the usual meaning. In a lot of practical calculations (2.14) is used in an even more direct sense: if $g = \exp tA$ for some $A \in G$ and some t,

$$\Theta|_g = h^{-1}dh|_g \equiv e^{-tA}\frac{d}{dt'}e^{t'A}\Big|_{t'=t} = A, \tag{2.19}$$

where $\Theta|_g$ should now be thought of as simply the Lie algebra element associated to g. What we are doing here is exactly the same as above, because

$$\frac{d}{dt'}e^{t'A}\Big|_{t'=t} = e^{tA}A = gA = g\,\frac{d}{dt'}e^{t'A}\Big|_{t'=0}, \tag{2.20}$$

so that

$$\Theta|_g = e^{-tA}\frac{d}{dt'}e^{t'A}\Big|_{t'=t} = g^{-1}\,\frac{d}{dt'}ge^{t'A}\Big|_{t'=0} = \Theta(X_A), \tag{2.21}$$

where X_A is the left invariant vector field associated to A.

Finally the adjoint map

$$ad_g : G \to G : h \mapsto ghg^{-1}, \tag{2.22}$$

induces a map in the tangent space

$$ad_{g*} : T_hG \to T_{ghg^{-1}}G. \tag{2.23}$$

When we restrict this map to $h = e$, we get the adjoint representation of $T_eG \cong \mathfrak{g}$,

$$Ad_g : \mathfrak{g} \to \mathfrak{g} : V \mapsto gVg^{-1}. \tag{2.24}$$

2.2 Parallel transport in a principal bundle

Consider a principal bundle $P(M, G)$. Given a curve γ on the base manifold M, to define parallel transport, we want to define a corresponding choice of curve γ_P in the total space P. There are of course many possible choices. To characterize our choice, we will look at vectors tangent to this curve. At every point along γ, we can define a lift of the tangent vector to γ to an element in TP, the tangent vectors to P. This will define an integral curve γ_P in P. See figure 4.4 for an illustration

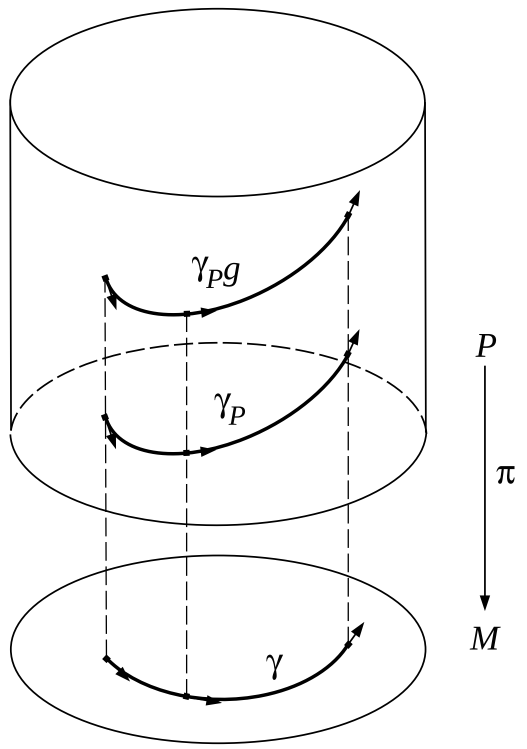


Figure 4.4: Illustration of horizontal lift.

The question is then how to lift a vector in $T_\gamma M$ to TP. At every point $p \in P$, we can decompose T_pP into a subspace of vectors tangent to the fibre G, called the vertical subspace V_pP and a complement H_pP, called the horizontal subspace, such that $T_pP = V_pP \oplus H_pP$. Since V_pP corresponds to motion along the fibres and is essentially fixed, a choice of H_pP is the crucial ingredient in the definition of parallel transport. We will require the vectors tangent to γ_P to lie in H_pP.

A choice of connection now essentially boils down to a choice of horizontal subspace. Let us be a bit more precise.

Definition 6. A **connection** on P is a smooth and unique separation of the tangent space T_pP at each p into a vertical subspace V_pP and a horizontal subspace H_pP such that

(i) $T_pP = V_pP \oplus H_pP$;

(ii) $H_{pg}P = R_{g*}H_pP$ for every $g \in G$.

Condition (i) just means that every $X \in T_pP$ can be written in a unique way as a sum $X = X^V + X^H$, where $X^V \in V_pP$ and $X^H \in H_pP$. The equivariance condition (ii) means that the choice of horizontal subspace at p determines all the horizontal subspaces at points pg. This roughly means that all points above the same point $x = \pi(p)$ in the base space will be parallel transported in the same way (recall that $\pi(p) = \pi(pg)$).

Parallel transport can now immediately be defined by what is called a horizontal lift.

Definition 7. Let $\gamma : [0, 1] \to M$ be a curve in the base manifold (a base curve). A curve $\gamma_P : [0, 1] \to P$ is called the **horizontal lift** of γ if

(i) $\pi(\gamma_P) = \gamma$;

(ii) All tangent vectors X_P to γ_P are horizontal: $X_P \in H_{\gamma_P}P$.

Theorem 4. Let $\gamma : [0, 1] \to M$ be a base curve and let $p \in \pi^{-1}(\gamma(0))$. Given a connection, there exists a unique horizontal lift γ_P such that $\gamma_P(0) = p$.

This means that we can (given a connection) uniquely define the parallel transport of a point p in P along a curve γ in M by moving it along the unique horizontal lift of γ through p.

A loop in M is defined as a curve γ with $\gamma(0) = \gamma(1)$. It is interesting to see what happens to a horizontal lift of this loop. In other words, what happens if we parallel transport an element of P along a closed loop? Starting from a point p and moving it along a horizontal lift of a loop, there is no guarantee that we will end up at the same point. In general we will obtain a different point p_γ, which depends on the loop γ. Since

$$\pi(\gamma_P(0)) = \gamma(0) = \gamma(1) = \pi(\gamma_P(1)), \tag{2.25}$$

we know that both points will belong to the same fibre, $\pi(p) = \pi(p_\gamma)$. This means that $p_\gamma = pg$ for some $g \in G$. If we vary the loop γ, but keep the base point p fixed, we generate a group called the holonomy group $Hol_p(P)$ of P at p, which by definition is a subgroup of G. This group of course also depends on the connection, so $Hol_p(P)$ is a characteristic not only of P, but also of the connection. If M is connected the holonomy group at all points of M are isomorphic and we can speak of the holonomy group of P, $Hol(P)$.

Given a notion of parallel transport in a principal bundle P, one can easily define parallel transport in an associated bundle E_ρ by

Definition 8. If $\gamma_P(t)$ is a horizontal lift of $\gamma(t) \in M$ in P, then $\gamma_E(t)$ is defined to be the horizontal lift of $\gamma(t)$ in E_ρ if

$$\gamma_E(t) = [(\gamma_P(t), v)], \tag{2.26}$$

where v is a constant element of V.

This is independent of the lift chosen in P, since (because of equivariance) another lift would be related to γ_P by $\gamma'_P(t) = \gamma_P(t)a$, with a constant element $a \in G$, so that

$$\gamma'_E(t) = [(\gamma'_P(t), v)] = [(\gamma_P(t), \rho(a)^{-1}v)] \tag{2.27}$$

where $\rho(a)^{-1}v$ is still a constant element. So this would still be a horizontal lift, albeit through a different element of V. Choosing a trivialization for γ_P, $\gamma_P(t) = \phi_P(\gamma(t), g(t))$, leads to the corresponding trivialization for γ_E

$$\gamma_E(t) = \phi_E(\gamma(t), \rho(g(t))v). \tag{2.28}$$

We see that if parallel transport in P is described by $g(t)$, then parallel transport in E_ρ is defined by $\rho(g(t))$.

2.3 Connection one-form on a principal bundle

Up to this point, the reader might be confused as to what this all has to do with gauge theories and the usual definition of a connection in physics. To establish the link with physics, we now introduce the connection one-form and clarify its relation to the gauge potential in Yang-Mills theories.

To define the connection one-form properly, we need a more specific construction of the vertical subspace V_pP. Let $A \in \mathfrak{g} = T_eG$ be an element of the Lie algebra of G. We saw in section 2.1 that A generates a one-parameter flow $\sigma_t(g)$ through g in G. A slight modification of this construction shows that A will generate a flow in P along the fibre at each point of M by the right action of G on P:

$$\sigma_t(p) = R_{\exp(tA)}p = p \exp(tA). \tag{2.29}$$

Note that $\pi(p) = \pi(\sigma_t(p))$, so that indeed vectors tangent to the curves are elements of V_pP. We now define a map $\mathfrak{g} \to V_pP$ which maps A to the vector tangent to $\sigma_t(p)$ for $t = 0$, which we will call (with slight abuse of notation with respect to equation (2.8)) $X_A \in V_pP$. The equivalent of the flow equation (2.9) now becomes

$$X_A(f(p)) = \frac{d}{dt} f(\sigma_t(p))\bigg|_{t=0}. \tag{2.30}$$

X_A is called the fundamental vector field associated with A. The fundamental vector fields associated to a basis of the Lie algebra form a basis of the vertical subspace. A connection one-form is now defined as follows.

Definition 9. A **connection one-form** $\omega \in \Lambda P \otimes \mathfrak{g}$ (where $\Lambda P \equiv \Gamma(P, T^*P)$) is a Lie algebra valued one-form defined by a projection of the tangent space T_pP onto the vertical subspace V_pP satisfying

(i) $\omega(X_A) = A$ for every $A \in \mathfrak{g}$;

(ii) $R_g^*\omega = Ad_{g^{-1}}\omega$ for every $g \in G$.

More concretely, (i) means that ω acts as a Maurer-Cartan form on the vertical subspace and (ii) means that for $X \in T_pP$,

$$R_g{}^*\omega|_p(X) = \omega|_{pg}(R_{g*}X) = g^{-1}\omega(X)g. \tag{2.31}$$

The horizontal subspace is then defined as the set

$$H_pP = \{X \in T_pP |\, \omega(X) = 0\}. \tag{2.32}$$

When defined in this way, H_pP still satisfies the equivariance condition. To see this, take $X \in H_pP$ and construct $R_{g*}X \in T_{pg}P$. This is an element of H_{pg} because

$$\omega(R_{g*}X) = R_g{}^*\omega(X) = g^{-1}\omega(X)g = 0. \tag{2.33}$$

So both definitions of a connection are equivalent. This connection is defined over all of P. To connect to physics, we have to relate this to a one-form in the base M. It turns out that this can only be done locally (when P is non-trivial).

Definition 10. Let $\{U_i\}$ be an open covering of M. Choose a local section s_i on U_i

$$s_i : U_i \to \pi^{-1}(U_i). \tag{2.34}$$

The local connection one-form or **gauge potential** is now defined as

$$A_i \equiv s_i{}^*\omega \in \Gamma(U_i, T^*M,) \otimes \mathfrak{g} = \Lambda U_i \otimes \mathfrak{g} \tag{2.35}$$

Also the converse is true.

Theorem 5. Given a local connection one-form A_i and a section s_i on an open subset $U_i \subset M$, there is a unique connection one-form $\omega \in \pi^{-1}(U_i)$ such that $A_i = s_i{}^*\omega$.

We will prove this theorem rather explicitly, since this will give better insight into the emergence of a gauge potential on M from the connection one-form on P. First of all, we introduce the notion of a canonical local trivialization with respect to a section. Given a section s_i on U_i and a $p \in \pi^{-1}(U_i)$, there always exists a $g_i \in G$ such that $p = s_i(x)g_i$, where $x = \pi(p)$. This means that we can define a local trivialization by

$$\phi^{-1} : \pi^{-1}(U_i) \to U_i \times G : p \mapsto (x,\, g_i). \tag{2.36}$$

This means that the section itself is represented as $s_i(x) = (x,\, e)$. On an overlap $U_i \cap U_j \neq \emptyset$ two sections are related by

$$\begin{aligned} s_i(x) &= \phi_i(x,\, e) = \phi_j(x,\, g_{ji}(x)e) = \phi_j(x,\, g_{ji}(x)) \\ &= \phi_j(x,\, e)g_{ji}(x) = s_j(x)g_{ji}(x). \end{aligned} \tag{2.37}$$

First we proof that an ω exists, then we will sketch a proof of its uniqueness.

Proof (existence):
Given a section s_i and a gauge potential A_i on U_i, we propose the following form of the connection one-form:

$$\omega|_{U_i} = g_i^{-1}\pi^* A_i g_i + g_i^{-1} d_P g_i\,, \tag{2.38}$$

where d_P is the exterior derivative on P and g_i is the group element which appears in the definition of the canonical local trivialization with respect to s_i.

(i) First of all, we have to show that pulling back (2.38) with the section results in A_i. Note that since $\pi \circ s_i = \mathrm{Id}_{U_i}$, we have that $\pi_* s_{i*} = \mathrm{Id}_{TU_i}$ and that $g_i = e$ on s_i. For a $X \in T_x M$ we have

$$\begin{aligned} {s_i}^*\omega|_{U_i}(X) &= \omega(s_{i*}X) = \pi^* A_i(s_{i*}X) + d_P e(s_{i*}X) \\ &= A_i(\pi_* s_{i*}X) = A_i(X). \end{aligned} \tag{2.39}$$

(ii) Now we need to establish that (2.38) satisfies the conditions from definition 9. A fundamental vector field V_A satisfies $\pi_* X_A = 0$ so that only the second term from (2.38) contributes. We need to evaluate this in the sense of a Maurer-Cartan form as discussed in subsection (2.1),

$$\begin{aligned} w|_{U_i}(X_A) &= g_i^{-1} d_P g_i(X_A) = g_i^{-1} X_A(g_i) = g_i^{-1} X_A|_{g_i} \\ &= g_i^{-1}(p)\, \frac{d}{dt} g_i(\sigma_t(p))\Big|_{t=0} = g_i^{-1}(p)\, \frac{d}{dt} g_i(p\,\exp(tA))\Big|_{t=0} \\ &= g_i^{-1}(p) g_i(p)\, \frac{d}{dt}\exp(tA)\Big|_{t=0} = \frac{d\sigma_t(e)}{dt}\Big|_{t=0} = A, \end{aligned} \tag{2.40}$$

where we used both the definition of the fundamental vector field (2.30) and the flow equation (2.8). This proves the first condition for being a connection. To prove the second condition, take an $X \in T_p P$. Note next that $g_i(p\,h) = g_i(p)h$ and that since $\pi \circ R_h = \pi$, we have that $\pi_* R_{h*} = \pi_*$. We find

$$\begin{aligned} {R_h}^*\omega(X) &= \omega(R_{h*}X) = h^{-1} g_i^{-1} A_i(\pi_* X) g_i h + h^{-1} g_i^{-1} d_P g_i(X) h \\ &= h^{-1}\omega(X) h = Ad_{h^{-1}}\omega(X). \end{aligned} \tag{2.41}$$

□

We will now sketch the proof of the uniqueness of the connection one-form. For this we need to see what happens on overlaps $U_i \cap U_j \neq \emptyset$.

Proof (uniqueness):
The two definitions of the connection on each patch have to agree on the intersection, $\omega|_{U_i} = \omega|_{U_j}$ on $U_i \cap U_j \neq \emptyset$. Writing this out, we find the following condition:

$$g_i^{-1}\pi^* A_i g_i + g_i^{-1} d_P g_i = g_j^{-1}\pi^* A_j g_j + g_j^{-1} d_P g_j \tag{2.42}$$

Noting that on the intersection we have $g_j = g_{ji} g_i$ ($s_j = s_i g_{ij}$), a small calculation shows that

$$\pi^* A_j = g_{ij}^{-1}\pi^* A_i g_{ij} + g_{ij}^{-1} d_P g_{ij}. \tag{2.43}$$

We can use either one of the sections $s_{i,j}$ on $U_i \cap U_j$ to pull this back to a local statement (note that $s_i^*\pi^* = \mathrm{Id}_M$ and that the pull-back and the exterior derivative commute),

$$A_j = g_{ij}^{-1} A_i g_{ij} + g_{ij}^{-1} d g_{ij}. \tag{2.44}$$

So we see that both definitions of ω agree on $U_i \cap U_j \neq \emptyset$ if the local gauge potentials are related in the above way.

□

This is the important point we wanted to reach. Note that the connection one-form ω on P is defined globally; it contains global information on the non-triviality of P. The gauge potentials $\{A_i\}$ are defined locally and we have just seen that if the fibres over two intersecting open sets on the base space have to be identified in a non-trivial way, the two gauge potentials defined on the overlap are necessarily different. This means that on a non-trivial bundle one local gauge potential has no global information, only the collection of all locally defined gauge potentials knows about the global topology. This means that gauge freedom is sometimes not just a matter of choice, but more of necessity!

Of course also in this language gauge freedom is a reflection of choice. Say that on an open set U, two sections are related by $s'(x) = s(x)g(x)$. We can choose either section to define a local gauge potential and almost the same reasoning as above shows that both are related as follows

$$A'(x) = g(x)^{-1}A(x)g(x) + g(x)^{-1}dg(x). \tag{2.45}$$

We see that in the bundle language, gauge freedom is equivalent to the freedom to choose local coordinates on a principal bundle!

2.4 Curvature of a connection

A very important notion is of course the curvature of a connection. To introduce this, we first define the concept of covariant derivative on a principal bundle.

Definition 11. Consider a Lie algebra valued p-form, $\alpha \in \Lambda^n P \otimes \mathfrak{g}$. This can be decomposed as $\alpha = \alpha^a \otimes T_a$, where α^a is an ordinary p-form and T_a is a basis of $\mathfrak{g}$. Let $X_1, ..., X_{p+1} \in T_pP$ be $p+1$ tangent vectors on P. The **exterior covariant derivative** of α is defined by

$$D\alpha(X_1, ..., X_{p+1}) = d_P\alpha(X_1^H, ..., X_{p+1}^H), \tag{2.46}$$

where $X_i^H \in H_pP$ is the horizontal component of $X_i \in T_pP$ and $d_P\alpha = (d_P\alpha^a) \otimes T_a$.

The curvature is now readily defined.

Definition 12. The **curvature 2-form** Ω of a connection ω on P is defined as

$$\Omega = D\omega \in \Lambda^2 P \otimes \mathfrak{g} \tag{2.47}$$

At every point $p \in P$, the horizontal vectors define a subspace of T_pP. The assignment of such a subspace at every point of P is called a distribution. Since in this case the distribution is defined by the equation $\omega = 0$, the Frobenius condition for integrability of the submanifold of P tangent to this distribution is exactly the vanishing of $d\omega$ along the distribution, that is, the vanishing of the curvature. As such, the curvature is an obstruction to finding a submanifold of P that is 'completely horizontal'.

Theorem 6. The curvature 2-form has the property

$$R_g{}^*\Omega = Ad_{g^{-1}}\Omega = g^{-1}\Omega g, \tag{2.48}$$

where g is a constant element of G.

Proof:
Recall that, because of the equivariance property of horizontal subspaces

$$\begin{aligned}(R_{g*}X)^H &= R_{g*}X^H;\\ R_g{}^*\omega &= Ad_{g^{-1}}\omega = g^{-1}\omega g.\end{aligned}$$

Also recall that pull-backs and exterior derivatives commute, $d_P R_g{}^* = R_g{}^* d_P$ and because g is constant, $d_P g = 0$. For $X, Y \in T_pP$ we then have

$$\begin{aligned}R_g{}^*\Omega(X,Y) &= \Omega(R_{g*}X, R_{g*}Y) = d_P\omega(R_{g*}X^H, R_{g*}Y^H)\\ &= R_g{}^* d_P\omega(X^H, Y^H) = d_P R_g{}^*\omega(X^H, Y^H)\\ &= d_P(g^{-1}\omega g)(X^H, Y^H) = g^{-1}d_P\omega(X^H, Y^H)g\\ &= g^{-1}\Omega(X,Y)g\end{aligned} \tag{2.49}$$

□

Let $\alpha \in \Lambda^p \otimes \mathfrak{g}$ and $\beta \in \Lambda^q \otimes \mathfrak{g}$ be two Lie algebra valued forms. The Lie bracket (commutator) between the two is defined as

$$\begin{aligned} [\alpha, \beta] &\equiv \alpha \wedge \beta - (-)^{pq} \beta \wedge \alpha \\ &= T_a T_b \alpha^a \wedge \beta^b - (-)^{pq} T_b T_a \beta^b \wedge \alpha^a \\ &= [T_a, T_b] \alpha^a \wedge \beta^b = f_{ab}{}^c T_c \, \alpha^a \wedge \beta^b . \end{aligned} \tag{2.50}$$

Note that for odd p this means that

$$[\alpha, \alpha] = 2\alpha \wedge \alpha \neq 0. \tag{2.51}$$

For even p, $[\alpha, \alpha] = 0$.

Let us now proof the following important theorem:

Theorem 7. Ω and ω satisfy the Cartan structure equations (X, $Y \in T_pP$)

$$\Omega(X, Y) = d_P\omega(X, Y) + [\omega(X), \omega(Y)], \tag{2.52}$$

or

$$\Omega = d_P\omega + \omega \wedge \omega = d_P\omega + \frac{1}{2}[\omega, \omega]. \tag{2.53}$$

To see the relation between the two forms of the theorem note that

$$\begin{aligned} [\omega, \omega](X, Y) &= [T_a, T_b] \omega^a \wedge \omega^b (X, Y) \\ &= [T_a, T_b](\omega^a(X)\omega^b(Y) - \omega^a(Y)\omega^b(X)) \\ &= [\omega(X), \omega(Y)] - [\omega(Y), \omega(X)] = 2[\omega(X), \omega(Y)]. \end{aligned}$$

Proof:
We consider three cases separately

(i) Let X, $Y \in H_pP$. Then by definition $\omega(X) = \omega(Y) = 0$. Then (2.52) follows trivially since by definition

$$\Omega(X, Y) = d_P\omega(X^H, Y^H) = d_P\omega(X, Y).$$

(ii) Let $X \in H_pP$ and $Y \in V_pP$. Since $Y^H = 0$, by definition $\Omega(X, Y) = 0$. We still have that $\omega(X) = 0$, so we still have to prove that $d_pP\omega(X, Y) = 0$. To do this, we use the following identity:

$$\begin{aligned} d_P\omega(X, Y) &= X\omega(Y) - Y\omega(X) - \omega([X, Y]) \\ &= X\omega(Y) - \omega([X, Y]). \end{aligned}$$

Since $Y \in V_pP$, it is a fundamental vector field[8] for some $V \in \mathfrak{g}$. This means that $\omega(Y) = V$ is a constant, so $X\omega(Y) = XV = 0$. One can show that $[X, Y] \in H_pP$, so that also $\omega([X, Y]) = 0$.

(iii) Let X, $Y \in V_pP$. Then again $\Omega(X, Y) = 0$ and this time we have

$$d_P\omega(X, Y) = -\omega([X, Y]).$$

So we still have to prove $\omega([X, Y]) = [\omega(X), \omega(Y)]$. Since also $[X, Y] \in V_pP$, every vector $X = X_B$, $Y = X_C$ and $[X, Y] = X_A$ are fundamental vector fields associated to Lie algebra elements B, C and A respectively. One can prove that necessarily $A = [B, C]$, which completes the proof.

□

[8]Or a linear combination of fundamental vector fields, in which case the result follows by linearity

We will now again use a section to pull back this globally defined object on P to a local object defined on a patch on M.

Definition 13. Given a section s_i on U_i, the local (Yang-Mills) **field strength** is defined by

$$F_i = {s_i}^*\Omega \in \Lambda^2 U_i \otimes \mathfrak{g}. \tag{2.54}$$

The relation with the gauge potential is now easily obtained

$$\begin{aligned} F_i &= {s_i}^* d_P\omega + {s_i}^*(\omega\wedge\omega) = d({s_i}^*\omega) + {s_i}^*\omega \wedge {s_i}^*\omega \\ &= dA_i + A_i \wedge A_i. \end{aligned} \tag{2.55}$$

Writing $A = A_a dx^a$ and $F = \frac{1}{2}F_{ab}dx^a \wedge dx^b$ (we dropped the subscript i for convenience), we find the usual expression,

$$F_{ab} = \partial_a A_b - \partial_b A_a + [A_a, A_b]. \tag{2.56}$$

The effect of a coordinate change on the field strength 2-form can be deduced (as usual) from the transformation properties of the gauge potential 1-form (2.45). More specifically, if two sections are related by $s'(x) = s(x)g(x)$, the corresponding field strengths are related by

$$F'(x) = g(x)^{-1}F(x)g(x). \tag{2.57}$$

We will need an important identity involving the curvature. Since, $\omega(X) = 0$ for $X \in H_pP$, we find that for X, Y, $Z \in T_pP$

$$D\Omega(X,Y,Z) = d_P\Omega(X^H,Y^H,Z^H) = (d_P\omega\wedge\omega - \omega\wedge d_P\omega)(X^H,Y^H,Z^H) = 0.$$

This proves the **Bianchi identity**

$$D\Omega = 0. \tag{2.58}$$

To find the local form of this identity we use a section s_i to pull back the relation

$$d_P\Omega = d_P\omega\wedge\omega - \omega\wedge d_P\omega. \tag{2.59}$$

This results in

$$\begin{aligned} dF_i &= d{s_i}^*\Omega = {s_i}^* d_P\Omega = {s_i}^*(d_P\omega\wedge\omega - \omega\wedge d_P\omega) \\ &= d{s_i}^*\omega \wedge {s_i}^*\omega - {s_i}^*\omega\wedge d{s_i}^*\omega = dA_i\wedge A_i - A_i \wedge dA_i \\ &= F_i \wedge A_i - A_i \wedge F_i = -[A_i, F_i]. \end{aligned}$$

So we find the local identity

$$\mathcal{D}_i F_i = dF_i + [A_i, F_i] = 0, \tag{2.60}$$

where we defined the **covariant derivative**

$$\mathcal{D}_i = d + [A_i, \ \ ]. \tag{2.61}$$

3 The topology of principal bundles

We will now discuss some aspects of the topology of gauge bundles. As we have seen, pure Yang-Mills theory can be described using only principal bundles, so we restrict ourselves to a discussion of the topology of principal bundles. Again, this section is mainly influenced by [1], [4] and [3]. For more advanced treatments and different perspectives, we recommend [8], [9] and [7].

3.1 Aspects of homotopy theory

We start by making some simple remarks about the classification of topological spaces. Usually in topology, two spaces are considered equivalent if they can continuously be deformed into each other. In other words, they are considered topologically the same, if there exists a homeomorphism between them (for differentiable manifolds this would have to be diffeomorphism). Classifying spaces up to homeomorphism is a difficult thing to do. The idea is to find as much topological invariants (in general, numbers that do not depend on continuous parameters) of a type of space as possible. Finding a full set of invariants that completely classifies a space is rather difficult. The converse is however easily stated:

> If two topological spaces have different topological invariants, they are not homeomorphic, hence not topologically equivalent.

Since classification up to homeomorphism is such a difficult task, one can try to answer somewhat easier problems. For instance, one can try to classify spaces up to homotopy. Two spaces are said to be homotopic to each other if one can be mapped to the other in a continuous way, but this map need not have an inverse. For instance a circle and a cylinder are homotopic (one can continuously shrink the cylinder until its length disappears), but clearly not homeomorphic. In an intuitive sense, homotopic equivalence occurs when in the process of deforming one space to another, a part of the space is 'lost', so that it becomes impossible to define the reverse process.

For the moment, we are interested in homotopic equivalence classes of loops on a differentiable manifold M. It turns out that these reveal a very interesting group structure. We will only sketch this construction and state the results that we need for the remainder of these notes. By a (based) loop we will mean a map $\alpha : I = [0,1] \to M : t \mapsto \alpha(t)$, such that $\alpha(0) = \alpha(1) = x \in M$, the base point of the loop. It turns out that not these loops by themselves exhibit a group structure, but rather their equivalence classes under homopoty.

Definition 14. Two loops α and β based at the point $x \in M$ are **homotopic** to each other if there exists a continuous map

$$H : I \times I \to: M : (s,t) \mapsto H(s,t) \tag{3.1}$$

such that

$$\begin{array}{ll} H(s,0) = \alpha(s) \text{ and } H(s,1) = \beta(s), & \forall s \in I, \\ H(0,t) = H(1,t) = x, & \forall t \in I. \end{array}$$

$H(s,t)$ is called the **homotopy** between α and β.

One can show that this is an equivalence relation $\sim$ (reflexive, symmetric, transitive) between based loops. We will denote the equivalence class or homotopy class by $[\alpha]$. So we have that

$$\alpha \sim \beta \Rightarrow [\alpha] = [\beta]. \tag{3.2}$$

In other words, two loops are considered the same if one can continuously deform one into the other. One can now define the 'product' $\alpha \circ \beta$ of two loops as the result of first going through α and then through β. The inverse α^{-1} of a loop is then just going through the loop α in reverse order and the unit element is the constant loop $\alpha(t) = x$, $\forall t \in I$. It is clear that $\alpha \circ \alpha^{-1}$ are not equal, but homotopic to the identity. This is why only the homotopy classes of loops exhibit a group structure.

Definition 15. The group formed by the homotopy classes of loops based at x on a manifold M is called the **fundamental group** or first homotopy group $\Pi_1(M,x)$.

One can show that, if the manifold is arc-wise connected, the fundamental groups at different points are isomorphic. In that case we just refer to the fundamental group of the manifold $\Pi_1(M)$. If two manifolds are homotopic (of the same homotopy type) one can show that their fundamental groups are isomorphic. Since homotopy is a weaker than homeomorphy, if the fundamental group is invariant under homotopy, it must certainly be under homeomorphy. So we arrive at the following conclusion.

Theorem 8. The fundamental group is invariant under homeomorphisms and hence is a topological invariant.

As an example, let us look at $\Pi_1(S^1) = \Pi_1(U(1))$. This basically means that we want to classify maps from one circle to another. Intuitively, we know that one circle can wind the other an integer n times. This is accomplished by functions of the form

$$g_{n,a} : I \to S^1 : t \mapsto g_{n,a}(t) = e^{i(nt+a)}, \quad a \in \mathbb{R}. \tag{3.3}$$

It's easily shown that two maps $g_{n,a}$ and $g_{m,b}$ are homotopic to each other for any a, $b \in \mathbb{R}$ if $n = m$, but that for $n \neq m$ they are homotopically distinct. This means that we have homotopy classes

$$[n] \equiv [g_n], \tag{3.4}$$

where $g_n \equiv g_{n,0} = e^{int}$ is a good representative for each equivalence class. In this easy case one calls the integer n the degree or winding number of the map and we find that homotopy classes are characterized by their winding number. This winding number can be represented by the integral

$$n = \frac{1}{2\pi i} \int_0^{2\pi} dt\, g_n(t)^{-1} \frac{d}{dt} g_n(t). \tag{3.5}$$

From this it is clear that if $f_1(t)$ represents a map of winding number one, a map of winding number n is represented by $f_n(t) = f_1(t)^n$. It is also clear that the product of $[g_n]$ and $[g_m]$ is nothing but $[g_{n+m}]$, which means that the fundamental group is isomorphic to the additive group $\mathbb{Z}$. So we get the well known result $\Pi_1(S^1) = \mathbb{Z}$. In general $\Pi_1(M)$ can be non-Abelian.

One can also define the higher homotopy groups by looking at homotopy classes of maps from higher dimensional spheres to a manifold. More concretely, one looks at maps $\alpha : I^n \to M$, such that the entire boundary of the n-dimensional cube, ∂I^n maps to a single point $x \in M$. Going through the same procedure as for $\Pi_1(M)$ one arrives at the higher homotopy groups $\Pi_n(M)$, which are always Abelian. In general these homotopy groups are quite hard to calculate. However, in the examples we will be studying, the maps we want to classify are always between two spaces of the same dimension. In that case, there is the notion of the (Brouwer) degree of a map, which is somewhat easier to handle. In the end, it will give us the same information as the related homotopy groups would provide.

Definition 16. Consider a map $\phi : M \to N$, where $\dim M = \dim N = n$ and let Ω be a normalized volume form on N. The **Brouwer degree** of this map is defined as

$$\deg(\phi) = \int_M \phi^* \Omega, \quad \int_N \Omega = 1 \tag{3.6}$$

This definition does not depend on the volume form chosen, since the difference of two normalized volume forms has to be exact (its integral over N has to vanish) and the pull-back commutes with the exterior derivative. In addition one can show that the degree is an integer and hence has to be a topological invariant. We will show below that the degree we defined in (3.5) for the circle can interpreted exactly in this way.

We would like to study the topology of the simplest non-trivial bundles. Since, as we already mentioned, a bundle over a contractible space is always trivial, bundles over $\mathbb{R}^n$ are always trivial. The next simplest thing to study are bundles over n-spheres S^n and since these bundles are also very relevant and interesting for physics, we will mainly focus on these. One can always cover an n-sphere by two patches, say the north and the south hemisphere. The intersection of these two patches is homotopic to the equator, an $(n-1)$-sphere. This means that to classify a principal G-bundle over S^n, one would have to classify the transition functions on S^{n-1}, that is all maps from S^{n-1} to G. As we already discussed, a very interesting object to study in this regard is the homotopy group $\Pi_{n-1}(G)$. Later on we will look at $U(1)$-bundles over S^2 (Dirac monopoles) and $SU(2)$-bundles over S^4 (instantons). This requires knowledge of the groups $\Pi_1(U(1))$ and $\Pi_3(SU(2))$, respectively. $\Pi_1(U(1)) = \mathbb{Z}$ was already considered in the previous example, so let's now focus on $\Pi_3(SU(2))$.

We will again use the fact that homotopy classes of maps from S^3 to S^3 are characterized by their topological degree. First of all, we have to find a well defined volume form on $SU(2)$. For a general compact (matrix) group manifold, this is done as follows. At every point $g \in G$, one can define the left invariant $\mathfrak{g}$-valued Maurer-Cartan form $\Theta = g^{-1}dg$, as discussed in subsection (2.1). From this one can define a well-defined bi-invariant (left- and right-invariant) volume form on G. For instance, for $SU(2)$ a normalized volume form is given by

$$\Omega = \frac{1}{24\pi^2} \operatorname{Tr}\left(g^{-1}dg \wedge g^{-1}dg \wedge g^{-1}dg\right), \tag{3.7}$$

where,

$$g = c_0 1_2 + c_i \tau_i\,, \qquad c_0^2 + c_i c_i = 1, \tag{3.8}$$

and τ_i, $i \in \{1,2,3\}$ are the Pauli matrices. Let (with a slight abuse of notation) $g : S^3 \to SU(2) : x \mapsto g(x)$. Then the degree of this map, according to equation (3.6), is given by

$$\deg(g) = \frac{1}{24\pi^2} \int_{S^3} \operatorname{Tr}\left(g^{-1}dg \wedge g^{-1}dg \wedge g^{-1}dg\right), \tag{3.9}$$

where the integrand should now be interpreted as the pull-back $g^*\Omega$. In other words, dg should now be interpreted as $\partial_i g dx^i$. Again, equation (3.9) will always give an integer n and since this degree fully characterizes elements of $\Pi_3(SU(2))$, we find that $\Pi_3(SU(2)) = \mathbb{Z}$.

In the case $G = U(1)$, the Maurer-Cartan form itself is a bi-invariant volume form, so we can take

$$\Omega = \frac{1}{2\pi i}\, g^{-1}dg\,, \qquad g \in U(1) \tag{3.10}$$

For the map $g_n : S^1 \to U(1) : t \mapsto g_n(t)$ we considered above we get

$${g_n}^*\Omega = \frac{1}{2\pi i}\, g_n(t)^{-1} \frac{dg_n(t)}{dt}\, dt\,, \tag{3.11}$$

so that the integer n we defined for the circle above, can rightfully be called the degree of the map g_n, $\deg(g_n) = n$.

3.2 Characteristic classes

Besides Homotopy there are of course many different ways to construct topological invariants. An important example are groups generated by (co)homology classes of a manifold. We will now focus on certain integer cohomology classes constructed from polynomials in the field strength of a bundle, called characteristic classes. First we define invariant polynomials.

Definition 17. Let $\mathfrak{g}$ be the Lie algebra of some G. A totally symmetric and n-linear polynomial

$$P(X_1, ..., X_i, ..., X_j, ..., X_n) = P(X_1, ..., X_j, ..., X_i, ..., X_n), \quad \forall\, i, j, \tag{3.12}$$

where X_i, $i \in \{1, ..., n\}$ are elements of $\mathfrak{g}$, is called a **symmetric invariant** (or characteristic) **polynomial** if

$$P(g^{-1}X_1 g, ..., g^{-1}X_n g) = P(X_1, ..., X_n), \quad g \in G. \tag{3.13}$$

An immediate consequence of this definition is that (take g close to the identity, $g = 1 + tY$, and expand (3.13) to first order in t),

$$\sum_{i=1}^{n} P(X_1, ..., X_{i-1}, [Y, X_i], X_{i+1}, ..., X_n) = 0. \tag{3.14}$$

This will be of use to us later.

Definition 18. An **invariant polynomial** of degree n is defined as a symmetric invariant polynomial with all its entries equal,

$$P_n(X) \equiv P(\underbrace{X, ..., X}_{n}) \equiv P(X^n) \tag{3.15}$$

Now we want to extend this definition to $\mathfrak{g}$-valued differential forms on a manifold. A $\mathfrak{g}$-valued p-form, α_i, we write as (no sum over i) $\alpha_i = \eta_i X_i$, where η_i is an ordinary p-form and X_i again an element of $\mathfrak{g}$. We then extend the previous definition as follows:

Definition 19. An invariant polynomial for $\mathfrak{g}$-valued forms is defined as

$$P(\alpha_1, ..., \alpha_n) = P(X_1, ..., X_n)\, \eta_1 \wedge ... \wedge \eta_n. \tag{3.16}$$

The diagonal combination is again called an invariant polynomial of degree n,

$$P_n(\alpha) = P(\alpha^n) = P(X^n)\, \eta \wedge ... \wedge \eta. \tag{3.17}$$

Let β be a $\mathfrak{g}$-valued 1-form. From (3.14) we find that

$$\sum_{i=1}^{n} (-)^{p_1+...+p_{i-1}} P(\alpha_1, ..., \alpha_{i-1}, [\beta, \alpha_i], \alpha_{i+1}, ..., \alpha_n) = 0, \tag{3.18}$$

where p_i is the degree of α_i and the minus signs arise from pulling the 1-form to the front each time. Equally easy to compute is

$$dP(\alpha_1, ..., \alpha_n) = \sum_{i=1}^{n} (-)^{p_1+...+p_{i-1}} P(\alpha_1, ..., \alpha_{i-1}, d\alpha_i, \alpha_{i+1}, ..., \alpha_n). \tag{3.19}$$

Adding both equations for the specific case $\beta = A$, where A is the local gauge potential associated to a connection on the bundle (we drop the index referring to the specific local patch for convenience) and recalling the expression for the covariant derivative $\mathcal{D} = d + [A, \ \ ]$, we find the important expression

$$dP(\alpha_1, ..., \alpha_n) = \sum_{i=1}^{n} (-)^{p_1+...+p_{i-1}} P(\alpha_1, ..., \alpha_{i-1}, \mathcal{D}\alpha_i, \alpha_{i+1}, ..., \alpha_n). \tag{3.20}$$

The objects we want to study are invariant polynomials in the field strength 2-form F, $P_n(F)$, because these turn out to have very interesting properties. We will work with a connection on a principal bundle, although the following results equally hold for its associated vector bundle. We are now ready to prove the following important theorem:

Theorem 9. Let $P_n(F)$ be an invariant polynomial, then

(i) $P_n(F)$ is closed, $dP_n(F) = 0$.

(ii) Let F and F' be local curvature 2-forms corresponding to two different connections on the same bundle. Then the difference $P_n(F) - P_n(F')$ is exact.

Note that, since $P_n(F)$ is closed, it can always locally be written as the d of something (Poincaré's lemma). The important point about this theorem is that the difference $P_n(F) - P_n(F') = dQ$ is exact in a global sense (which is of course the meaning of exact), meaning that we have to prove that Q is globally defined! (a point which is usually ignored in physics textbooks)

Proof:

(i) The first part of the theorem follows immediately from (3.20), because of the Bianchi identity $\mathcal{D}F = 0$, see (2.60).

(ii) To prove the second part, consider two 1-form gauge potentials A and A', both referring to the same system of local trivializations, and their respective 2-form field strengths F and F'. We define the homotopic connection[9]

$$A_t = A + t\theta, \quad \theta = A' - A, \tag{3.21}$$

[9]We are a bit sloppy here, because we should define this homotopy locally on a patch, where it is clear that this can be done. However, since both connections are defined on the same bundle (same set of transition functions) this turns out to be possible globally.

so that $A_0 = A$ and $A_1 = A'$, and its field strength

$$\begin{aligned} F_t &= dA_t + A_t \wedge A_t = F + t(d\theta + A \wedge \theta + \theta \wedge A) + t^2 \theta \wedge \theta \\ &= F + t\mathcal{D}\theta + t^2\theta \wedge \theta, \end{aligned} \tag{3.22}$$

where $\mathcal{D} = d + [A, \ ]$ (note the sign convention in the definition of the commutator). We now differentiate F_t with respect to t,

$$\frac{d}{dt}F_t = \mathcal{D}\theta + 2t\theta \wedge \theta = d\theta + A_t \wedge \theta + \theta \wedge A_t = \mathcal{D}_t\theta \tag{3.23}$$

with the obvious notation $\mathcal{D}_t = d + [A_t, \ ]$. Considering then the invariant polynomial $P_n(F_t)$, we get

$$\frac{d}{dt}P_n(F_t) = nP(\frac{d}{dt}F_t, \underbrace{F_t, ..., F_t}_{n-1}) = nP(\mathcal{D}_t\theta, F_t^{n-1}). \tag{3.24}$$

From equation (3.20) and $\mathcal{D}_t F_t = 0$, we know that $dP(\theta, F_t^{n-1}) = P(\mathcal{D}_t\theta, F_t^{n-1})$, so that we find that

$$\frac{d}{dt}P_n(F_t) = ndP(\theta, F_t^{n-1}). \tag{3.25}$$

Integrating this from $t = 0$ to $t = 1$, we find

$$P_n(F') - P_n(F) = dQ_{2n-1}(A', A), \tag{3.26}$$

where we defined the **transgression** $Q_{2n-1}(A', A)$ as

$$Q_{2n-1}(A', A) = n\int_0^1 dt\, P(A' - A, F_t^{n-1}). \tag{3.27}$$

Note that $Q_{2n-1}(A', A)$ is indeed a gauge invariant and hence globally defined object, since under a gauge transformation (the inhomogeneous term cancels) $\theta' = g^{-1}\theta g$ and P is invariant.

□

Equation (3.26) is called a transgression formula and is quite important in the study of anomalies. We use it here to define the **Chern-Simons form**. Say that one can define a trivial connection $A' = 0$ on a bundle. This means that either the bundle is trivial or that we are working on a local patch. We know that since $P_n(F)$ is closed, it is locally exact; it can locally be written as the d of a Chern-Simons form. The transgression formula provides a means for calculating this Chern-Simons form. Indeed, from (3.26) we find

$$P_n(F) = dQ_{2n-1}(A), \tag{3.28}$$

where we defined the Chern-Simons form

$$Q_{2n-1}(A) \equiv Q_{2n-1}(A, 0) = n\int_0^1 dt\, P(A, F_t^{n-1}), \tag{3.29}$$

and now,

$$A_t = tA, \quad F_t = t\,dA + t^2 A \wedge A = tF + (t^2 - t)A \wedge A \tag{3.30}$$

We see that, given an invariant polynomial $P_n(F)$, we can always construct the associated Chern-Simons form $Q_{2n-1}(A)$ from (3.29). Note that $P_n(F)$ is a $2n$-form, while Q_{2n-1} is a $(2n-1)$-form.

Since an invariant polynomial in F, $P_n(F)$, is closed and generically non-trivial, it represents a non-trivial (de Rham) cohomology class $[P_n(F)] \in H^{2n}(M, \mathbb{R})$, which is called a **characteristic**

class. Since we have shown that the difference of two invariant polynomials defined with respect to two different connections is exact, we have by Stoke's theorem and for a manifold M without boundary, $\partial M = 0$,

$$\int_M P_n(F') - \int_M P_n(F) = \int_M dQ_{2n-1}(A', A) = \int_{\partial M} Q_{2n-1}(A', A) = 0. \tag{3.31}$$

This means that the integrals or periods $([P_n(F)], M)$ of these classes, usually called characteristic numbers, do not depend on the connection chosen, in other words, they are characteristic of the bundle itself (transition functions)! This makes characteristic classes very interesting objects to study the topology of fibre bundles. In contrast, Chern-Simons forms obviously do depend on the connection chosen, they are not even gauge invariant, but will prove to be very useful nonetheless. We now go on to define some examples of characteristic classes and Chern-Simons forms which are useful in the study of gauge theories.

3.3 Chern classes and Chern characters

Let P be a principal bundle, with structure group $G = GL(k, \mathbb{C})$ or a subgroup thereof ($U(k)$, SU(k),...)[10]. The **total Chern class** is defined by (the normalization of F is for later convenience)

$$c(F) = \det\left(1 + \frac{i}{2\pi}F\right). \tag{3.32}$$

Since F is a 2-form, $c(F)$ is a sum of forms of even degrees,

$$c(F) = 1 + c_1(F) + c_2(F) + ... \tag{3.33}$$

where $c_n(F) \in \Lambda^{2n}M$ is called the n-th **Chern class**[11]. If $\dim M = m$, all Chern classes, $c_n(F)$ of degree $2n > m$, vanish. In general it can be quite cumbersome to compute this determinant for higher dimensional manifolds. Therefore we will diagonalize the matrix $\frac{i}{2\pi}F$ (if for instance $G = SU(k)$, F is anti-hermitian, so iF is hermitian and can be diagonalized by an $SU(k)$ rotation g) to a matrix $\widetilde{F}$, with 2-forms x_i on the diagonal. This leads to

$$\begin{aligned}
\det(1 + \widetilde{F}) &= \det[\mathrm{diag}(1 + x_1, ..., 1 + x_k)] = \prod_{i=1}^{k}(1 + x_i) \\
&= 1 + (x_1 + ... + x_k) + (x_1x_2 + ... + x_{k-1}x_k) + ... + (x_1x_2...x_k) \\
&= 1 + \mathrm{Tr}\,\widetilde{F} + \frac{1}{2}\left[(\mathrm{Tr}\,\widetilde{F})^2 - \mathrm{Tr}\,\widetilde{F}^2\right] + ... + \det\widetilde{F}.
\end{aligned} \tag{3.34}$$

Note that in the second line we encounter the elementary symmetric functions of $\{x_i\}$ and that all manipulations are well defined since the x_i are 2-forms and thus commute (the wedge product is always understood). For an invariant polynomial $P_n(F) = P_n(g^{-1}Fg) = P_n(2\pi\widetilde{F}/i)$ (note that the trace always guaranties invariance), so we find the following expressions for the Chern classes:

$$c_1(F) = \mathrm{Tr}\,\widetilde{F} = \frac{i}{2\pi}\mathrm{Tr}\,F \tag{3.35}$$

$$c_2(F) = \frac{1}{2}\left[(\mathrm{Tr}\,\widetilde{F})^2 - \mathrm{Tr}\,\widetilde{F}^2\right] = \frac{1}{8\pi^2}\left[\mathrm{Tr}(F \wedge F) - \mathrm{Tr}\,F \wedge \mathrm{Tr}\,F\right] \tag{3.36}$$

$$\vdots$$

$$c_k(F) = \det\widetilde{F} = \left(\frac{i}{2\pi}\right)^k \det F \tag{3.37}$$

[10]One can equally well take the bundle to be an associated complex vector bundle E with one of the mentioned structure groups.

[11]Strictly speaking, this is a representative of the n-th Chern class, but we will follow the rest of the world in calling these Chern classes by themselves.

To show how the computation of Chern-Simons forms goes about, let's start with a ridiculously easy example. Consider a $U(1)$-bundle over some 2-dimensional manifold. The only Chern class which can be defined is $c_1(F)$ and obviously, since locally $F = dA$, we find

$$c_1(F) = d\left(\frac{i}{2\pi}A\right), \quad \text{so that} \quad Q_1(A) = \frac{i}{2\pi}A. \tag{3.38}$$

Killing a fly with a jackhammer, we now use formula (3.29) to compute the same thing

$$Q_1(A) = \int_0^1 dt\, P(A) = \int_0^1 dt\, c_1(A) = \int_0^1 dt\, \frac{i}{2\pi}A = \frac{i}{2\pi}A. \tag{3.39}$$

Now that we have earned some trust in (3.29), we compute something less trivial. Consider an $SU(2)$-bundle over a 4-dimensional manifold. Since for $SU(k)$ we have that $\mathrm{Tr}\, F = 0$, the first Chern class vanishes. Let's try to compute the Chern-Simons form related to the second Chern class,

$$\begin{aligned} Q_3(A) &= 2\int_0^1 dt\, P(A, F_t) = \frac{1}{4\pi^2}\int_0^1 dt\ \mathrm{Tr}(A \wedge F_t) \\ &= \frac{1}{4\pi^2}\int_0^1 dt\ \mathrm{Tr}(tA \wedge dA + t^2 A \wedge A \wedge A) \\ &= \frac{1}{8\pi^2}\mathrm{Tr}\left(A \wedge dA + \frac{2}{3}A \wedge A \wedge A\right). \end{aligned} \tag{3.40}$$

which is of course the most famous example of a Chern-Simons form in physics.

Since the periods of Chern classes are independent of the connection, these numbers, called **Chern numbers**, are denoted as

$$c_n \equiv ([c_n(F)], M) = \int_M c_n(F). \tag{3.41}$$

One can show that on a compact manifold, these numbers are always integers, $c_n = k$, a phenomenon called topological quantization. We will see instances of this where the Chern numbers compute the monopole charge or instanton number later.

We will now briefly discuss another characteristic class called the Chern character, because it has some properties which make it easier to compute than Chern classes (one can afterwards compute the Chern classes from the Chern characters) and because it appears in the Atiyah-Singer index theorem. The **total Chern character** (again for $G \subseteq GL(k, \mathbb{C})$) is defined by

$$ch(F) = \mathrm{Tr}\exp\left(\frac{i}{2\pi}F\right) = \sum_n \frac{1}{n!}\mathrm{Tr}\left(\frac{i}{2\pi}F\right)^n. \tag{3.42}$$

This is again a sum over even forms, the **Chern characters**

$$ch_n(F) = \frac{1}{n!}\mathrm{Tr}\left(\frac{i}{2\pi}F\right)^n \tag{3.43}$$

By again diagonalizing iF to a diagonal matrix $\widetilde{F}$ with eigenvalues $\{x_i\}$, using

$$\mathrm{Tr}\exp(\widetilde{F}) = \sum_{i=1}^k \exp(x_i) = \sum_{i=1}^k (1 + x_i + \frac{1}{2}x_i^2 + ...), \tag{3.44}$$

and expressing the result in terms of elementary symmetric functions of $\{x_i\}$, one can relate the Chern characters to the Chern classes. A few examples are,

$$ch_0(F) = k \tag{3.45}$$

$$ch_1(F) = c_1(F) \tag{3.46}$$

$$ch_2(F) = -c_2(F) + \frac{1}{2}c_1(F) \wedge c_1(F). \tag{3.47}$$

Since for a Dirac monopole we only need $c_1(F) = ch_1(F)$ and for $SU(2)$-instantons $c_1(F) = 0$, so that $ch_2(F) = -c_2(F)$, we will not really see the difference between the two. We will mostly refer to the Chern class, when speaking about either of the two.

4 Some applications

We will now apply the formalism we developed to two standard examples of the topology of gauge bundles, Dirac monopoles and instantons. We will not talk about (the more interesting) non-Abelian 't Hooft-Polyakov monopoles, because these do not appear in pure gauge theory, but require a Higgs field, which is a section of an associated vector bundle and are not as such 'pure' examples of the topology of gauge bundles. For more on these and other applications of bundles to physics, see [1] - [7].

4.1 Dirac monopoles

Consider a magnetic monopole in Maxwell theory (Abelian) at the origin of 3-dimensional Euclidean space, $\mathbb{R}^3$. If q is the magnetic charge of the monopole, we can take the magnetic charge distribution to be $\rho(x) = 4\pi q\delta(x)$. Let B^i be the components of the magnetic field (a vector field on $\mathbb{R}^3$). From Maxwell's equations, we know that $\partial_i B_i(x) = q\delta(x)$, which has the spherically symmetric solution

$$B = \frac{q}{r^2}\frac{\partial}{\partial r}. \tag{4.1}$$

This expression becomes infinite at the origin, so that strictly speaking it is only a physically relevant solution on $\mathbb{R}^3_0 = \mathbb{R}^3 \backslash \{(0,0,0)\}$. There are no non-trivial bundles over $\mathbb{R}^3$, but for $\mathbb{R}^3_0$ there is the possibility that there exists a fibre bundle description of this kind of monopole. Since $\mathbb{R}^3_0$ is homotopic to a sphere and we are considering a pure $U(1)$-gauge theory, the proper bundle setting is that of a principal $U(1)$-bundle over S^2, $P(U(1), S^2)$.

We described two ways for classifying this bundle. One was the characterization of $\Pi_1(U(1))$ by the topological degree or winding number of the map from the equator S^1 of S^2 to the fibre $U(1)$. The other was by computing the first Chern class $c_1(F)$ of a $U(1)$-connection on the sphere S^2.

Let's look at the second approach. The sphere can be covered by two patches U_N and U_S, corresponding to the northern and southern hemisphere respectively, with $U_N \cap U_S = S^1$. On each patch we have a local 1-form gauge potential, A^N on U_N and A^S on U_S. Since we are dealing with an Abelian structure group, the 2-form field strength F is gauge invariant. This means that on the equator

$$F|_{S^1} = dA^N = dA^S. \tag{4.2}$$

The first Chern number is computed as follows:

$$c_1 = \int_{S^2} c_1(F) = \frac{i}{2\pi}\int_{S^2} F = \frac{i}{2\pi}\left(\int_{U_N} F + \int_{U_S} F\right). \tag{4.3}$$

The field strength is locally exact on both hemispheres, so by Stoke's theorem we find,

$$c_1 = \frac{i}{2\pi}\left(\int_{\partial U_N} A^N + \int_{\partial U_S} A^S\right) = \frac{i}{2\pi}\int_{S^1}\left(A^N - A^S\right). \tag{4.4}$$

On the equator both potential 1-forms are related by a transition function $g \in U(1)$,

$$A^S = A^N + g^{-1}dg. \tag{4.5}$$

This leads to

$$c_1 = \frac{1}{2\pi i}\int_{S^1} g^{-1}dg, \tag{4.6}$$

which we recognize as the winding number of the map $g : S^1 \to U(1)$. We see that for a map g_n of winding number n, we find that $c_1 = \deg(g_n) = n$.

To connect to physics, we now write down an explicit solution. The 1-form β associated to the vector field B in (4.1) is (we use spherical coordinates with metric components $\eta_{rr} = 1$, $\eta_{\theta\theta} = r$ and $\eta_{\varphi\varphi} = r\sin\theta$),

$$\beta = \frac{q}{r^2} dr. \tag{4.7}$$

The field strength 2-form F is the Hodge dual to this (to find complete agreement with the general theory, we need to include the Lie algebra factor i),

$$F = i * \beta = i\sqrt{\det\eta}\, B_r\, \varepsilon^r{}_{\theta\varphi}\, d\theta \wedge d\varphi = iq\sin\theta\, d\theta \wedge d\varphi. \tag{4.8}$$

This is the field strength 2-form which represents a Dirac monopole of charge q. A possible gauge potential 1-form that leads to this field strength is $A = -i\cos\theta\, d\varphi$. Since spherical coordinates are badly behaved along the entire z-axis ($\theta = 0,\ \pi$), we can't use this potential for either hemisphere. We can however define,

$$A^N = iq(1-\cos\theta)d\varphi \quad \text{on} \quad U_N \tag{4.9}$$
$$A^S = -iq(1+\cos\theta)d\varphi \quad \text{on} \quad U_S, \tag{4.10}$$

which are well defined on their respective patches and lead to the same F. On the equator, we find

$$A^N = A^S + 2iqd\varphi. \tag{4.11}$$

This completely agrees with equation (4.5) for

$$g_n = e^{in\varphi}, \quad n = 2q, \tag{4.12}$$

so that a monopole of charge q corresponds to a $U(1)$-bundle over S^2 with winding number $n = 2q$, or, to put it differently, corresponds to an element $[2q]$ of $\Pi_1(U(1))$.

We conclude this subsection with an important note. The above reasoning might falsely cause one to believe that one does not need quantum mechanics to prove the quantization of magnetic charge. The main assumption we used, however, is that one can use a bundle to describe the magnetic field on a sphere in the first place. This manifests itself in a gauge transformation by $g^{-1}dg$ in eq. (4.5), instead of by just a general closed one-form on the equator. This is equivalent to the assumption that the magnetic field is described by an integral 2-form (a first Chern class) as opposed to a generic 2-form, which is not integral. The integrality of the magnetic field of a monopole is usually proved by considering the wave function of a quantum mechanical particle in the neighborhood of the monopole. To summarize: classically, the magnetic field is just a (non-integral) 2-form (which consequently is not the curvature of a bundle), while quantum mechanically, it is an integral 2-form, which means that it can be seen as the curvature (first Chern class) of a bundle.

4.2 Holonomy and the Aharonov-Bohm effect

We already briefly discussed holonomy in subsection 2.2, but let us come back to it in little more detail. Consider a principal bundle $P(M, G)$ with a connection 1-form ω. Let γ be a curve on M and γ_P be a horizontal lift. Suppose for the moment that γ is contained within a single patch U and let $s : U \to P$ be a section on U. This means that $\gamma_P(t) = s(\gamma(t))g(t)$. The aim is to compute $g(t)$ to have a local description of parallel transport (local because this description depends on s).

If $X \in T\gamma M$ is the tangent to γ, the tangent to γ_P is $X_P = \gamma_{P*}X \in T_{\gamma_P}P$. Since γ_P is horizonal, we have that $X_P \in H_{\gamma_P}P$ or $\omega(X_P) = 0$. According to (2.38), this means

$$g^{-1}\pi^*A(X_P)g + g^{-1}d_P g(X_P) = 0. \tag{4.13}$$

Using that $\pi_* X_P = \pi_*\gamma_{P*}X = X$, we find on M,

$$g^{-1}A(X)g + g^{-1}dg(X) = 0, \tag{4.14}$$

or

$$dg(X) = X(g) = -A(X)g. \tag{4.15}$$

Since X is tangent to γ, we have

$$X(g) = \frac{d}{dt}g(\gamma(t)) \quad \text{and} \quad A(X) = A_a X^a = A_a \frac{d}{dt}x^a(\gamma(t)), \tag{4.16}$$

where $A = A_a dx^a$. Writing $g(\gamma(t)) = g(t)$ and $x^a(\gamma(t)) = x^a(t)$, this leads to

$$\frac{dg(t)}{dt}g(t)^{-1} = -A_a \frac{dx^a(t)}{dt}. \tag{4.17}$$

For $G = U(1)$ this has the solution (suppose that $g(0) = e$)

$$g(t) = \exp\left[-\int_0^t dt\, A_a \frac{dx^a(t)}{dt}\right] = \exp\left[-\int_{\gamma(0)}^{\gamma(t)} A_a dx^a\right] \tag{4.18}$$

For short, we can write

$$g(t) = \exp\left[-\int_\gamma A\right]. \tag{4.19}$$

For a non-Abelian structure (gauge) group, this is modified to

$$g(t) = P\exp\left[-\int_\gamma A\right]. \tag{4.20}$$

where P indicates that the exponential is defined by its power series expansion and that the matrix-valued forms should always be path ordered. In physics this is called a Wilson line.

If $s' = sh$ is another section on U (or on an overlap with another patch U'), related to s by a group element $h \in G$, one can show that if $\gamma_P(t) = s'(\gamma(t))g'(t)$, we find

$$g'(t) = h^{-1}(t)g(t)h(0). \tag{4.21}$$

This shows that if $G = U(1)$ and if $\gamma(0) = \gamma(1)$, so that $h(1) \equiv h(\gamma(1)) = h(\gamma(0)) = h(0)$, then

$$g_\gamma \equiv \exp\left[-\oint_\gamma A\right], \tag{4.22}$$

called a Wilson loop, is gauge invariant. g_γ is nothing but an element of the holonomy group $Hol(P)$ we discussed in subsection 2.2. We see that in the non-Abelian case this procedure doesn't lead to a gauge invariant quantity, but

$$g'_\gamma = h^{-1}g_\gamma h, \quad h = h(0) = h(1). \tag{4.23}$$

If we take the trace of this though, we do get a gauge invariant quantity. In non-Abelian gauge theories the Wilson loop is thus defined as the trace of the Wilson line around a closed loop,

$$W_\gamma = \mathrm{Tr}\, g_\gamma = \mathrm{Tr}\, P\exp\left[-\oint_\gamma A\right], \tag{4.24}$$

and equals the trace of the holonomy at $p = \gamma_P(0)$.

To illustrate this, consider a solenoid along the x_3-axis in $\mathbb{R}^3$. The $U(1)$ magnetic field is uniform in the interior of the solenoid and practically vanishing outside of it. In the limit of infinitely thin and long solenoid, the magnetic field is strictly 0 in the exterior region, which is $\mathbb{R}^3_0$, but there still is a non-zero flux Φ associated to it. By Stoke's theorem, this means that the gauge potential A cannot be zero in the exterior region, since for any curve γ encircling the solenoid, which spans a surface A_γ

$$\oint_\gamma A = \int_{A_\gamma} dA = \int_{A_\gamma} F = \Phi. \tag{4.25}$$

Since $F = 0$ outside of the solenoid, there is no classical effect on a particle moving alongside it. Quantum mechanically however, it is known that A can have a physical meaning. To see this, consider a wave function ψ of a particle moving in the x_1x_2-plane perpendicular to the solenoid. This is described by a section of a complex line bundle over $\mathbb{R}_0^2$ associated to the principal bundle $P(\mathbb{R}_0^2, U(1))$ by the obvious representation

$$\psi \to \rho(e^\alpha)\psi = e^\alpha \psi. \tag{4.26}$$

In a path integral approach, every path γ is weighed by a factor

$$g(t) = e^{iS_\gamma}, \quad S_\gamma = \int_\gamma dt\, L, \tag{4.27}$$

where the important part of the Lagrangian L is the part involving the gauge potential,

$$L = A_i(x)\frac{dx^i}{dt}. \tag{4.28}$$

In other words, every path is weighed by a factor (note that we absorbed the Lie algebra factor into A to make contact with our formalism),

$$g(t) = \exp \int_\gamma A. \tag{4.29}$$

In a double slit experiment, with the solenoid placed between the slits, part of the wave function ψ_1 will move along path γ_1 above the solenoid and part ψ_2 will move along a path γ_2 underneath it. The total wave function is[12]

$$\begin{aligned} \psi &= \exp\left[\int_{\gamma_1} A\right]\psi_1 + \exp\left[\int_{\gamma_2} A\right]\psi_2 \\ &= \exp\left[\int_{\gamma_1} A\right]\left\{\psi_1 + \exp\left[\int_{\gamma_2} A - \int_{\gamma_1} A\right]\psi_2\right\}. \end{aligned} \tag{4.30}$$

What is important of course, is the phase difference,

$$\int_{\gamma_2} A - \int_{\gamma_1} A = \oint_\gamma A, \quad \gamma = \gamma_2 - \gamma_1. \tag{4.31}$$

We see that the probability to find a particle at a certain point on the screen is influenced by the gauge potential through the holonomy of the connection

$$|\psi|^2 = |\psi_1 + g_\gamma \psi_2|^2 = |\psi_1 + e^\Phi \psi_2|^2. \tag{4.32}$$

Note that the only reason why $\oint A$ can have physical meaning is because it is gauge invariant (independent of local trivializations).

G-bundles over a circle are classified by $\Pi_0(G)$, implying that a $U(1)$-bundle over S^1 is necessarily trivial. Hence, this example shows that a trivial bundle can contain non-trivial physics, i.e. the Aharonov-Bohm effect.

[12] We are mixing path integral and ordinary QM arguments. Since the only real importance is wether the path passes underneath or above the solenoid, the path integral essentially reduces to the sum of 2 paths, so that both viewpoints essentially give the same information.

4.3 Instantons

Instantons are traditionally defined as smooth finite action solutions of Yang-Mills theory on 4-dimensional Euclidian space $\mathbb{R}^4$. We will only consider the case of $SU(2)$. There exist no non-trivial bundles over $\mathbb{R}^4$, but the finiteness of the action imposes boundary conditions at infinity, which allow for the existence of topologically non-trivial solutions of the field equations. This is seen as follows: To get a finite action, the field strength has to go to zero (fast enough) at infinity. This means that along a sufficiently large 3-sphere, S^3_∞, the gauge potential has to be pure gauge

$$A|_{S^3_\infty} = g^{-1}dg \implies F|_{S^3_\infty} = 0, \tag{4.33}$$

so these solutions are classified by maps $g : S^3_\infty \to SU(2)$. The reason for their stability is the fact that one cannot change the homotopy class of this map while keeping the total action finite. Computing the second Chern number, using that $c_2(F) = dQ_3(A)$ (and assuming that there is no contribution outside of S^3_∞),

$$\begin{aligned} c_2 &= \int_{\mathbb{R}^4} c_2(F) = \int_{S^3_\infty} Q_3(A) = \frac{1}{8\pi^2}\int_{S^3_\infty} \mathrm{Tr}\left(A\wedge dA + \frac{2}{3}A\wedge A\wedge A\right) \\ &= \frac{1}{8\pi^2}\int_{S^3_\infty} \mathrm{Tr}\left(F\wedge A - \frac{1}{3}A\wedge A\wedge A\right) = -\frac{1}{24\pi^2}\int_{S^3_\infty}\mathrm{Tr}\left(A\wedge A\wedge A\right) \\ &= -\frac{1}{24\pi^2}\int_{S^3_\infty}\mathrm{Tr}\left(g^{-1}dg\wedge g^{-1}dg\wedge g^{-1}dg\right). \end{aligned} \tag{4.34}$$

Which is (up to a sign) exactly the topological degree of the map $g : S^3_\infty \to SU(2)$.

To gain more control over the situation and allow for a bundle description of instantons, we consider a one-point compactification of $\mathbb{R}^4$ to S^4, by adding to it the point at infinity, $\mathbb{R}^4\cup\{\infty\} = S^4$. This means that we want to look at principal $SU(2)$-bundles over S^4, $P(SU(2), S^4)$. As we saw in subsection 3.1, this is classified by $\Pi_3(SU(2))$, while on the other hand one can compute the second Chern number of an $SU(2)$-connection on S^4.

Again, we can cover the 4-sphere by two open sets, the northern and southern hemisphere, U_N and U_S respectively. This time, the field strength 2-form is not invariant, but

$$F^N = dA^N + A^N\wedge A^N, \quad F^S = dA^S + A^S\wedge A^S, \tag{4.35}$$

with

$$A^N = g^{-1}A^S g + g^{-1}dg \Rightarrow F^N = g^{-1}F^S g. \tag{4.36}$$

Let's try to compute the second Chern number

$$\begin{aligned} c_2 &= \int_{S^4} c_2(F) = \int_{U_N} dQ_3(A^N) + \int_{U_S} dQ_3(A^S) = \int_{S^3}\left(Q_3(A^N) - Q_3(A^S)\right) \\ &= \frac{1}{8\pi^2}\int_{S^3}\mathrm{Tr}\left(F^N\wedge A^N - \frac{1}{3}A^N\wedge A^N\wedge A^N - F^S\wedge A^S + \frac{1}{3}A^S\wedge A^S\wedge A^S\right) \\ &= -\frac{1}{8\pi^2}\int_{S^3}\mathrm{Tr}\left(\frac{1}{3}g^{-1}dg\wedge g^{-1}dg\wedge g^{-1}dg - d(g^{-1}A^N\wedge dg)\right) \end{aligned} \tag{4.37}$$

where in the last step we used the gauge transformations (4.36), the identity

$$dg^{-1} = -g^{-1}dgg^{-1}, \tag{4.38}$$

and the fact that for three $\mathfrak{g}$-valued 1-forms

$$\mathrm{Tr}(\alpha\wedge\beta\wedge\gamma) = \mathrm{Tr}(\gamma\wedge\alpha\wedge\beta). \tag{4.39}$$

Since S^3 has no boundary, we arrive at the expected result

$$c_2 = \frac{1}{24\pi^2} \int_{S^3} \mathrm{Tr}\left(g^{-1}dg \wedge g^{-1}dg \wedge g^{-1}dg\right). \tag{4.40}$$

Again we find that $c_2 = \deg(g) = k$, as described in (3.9). We conclude that the classification by the second Chern class $[c_2(F)]$ is equivalent to the one given by $\Pi_3(SU(2))$. Both are characterized by the degree of the transition function of the bundle.

What does this have to do with instantons? As we already stated, we have to find finite action solutions of the action for $F = F_{ij}dx^i \wedge dx^j$, $F_{ij} = F_{ij}{}^a\, T_a$

$$S_E = \frac{1}{4}\int dx^4\, F_{ij}{}^a F^{ij}_{\ a} = -\frac{1}{2}\int dx^4\ \mathrm{Tr}\left(F_{ij}F^{ij}\right) = -\int \mathrm{Tr}(F \wedge *F), \tag{4.41}$$

where $*F$ is the Hodge dual of F, in flat space

$$* F_{ij} = \frac{1}{2}\varepsilon_{ijkl}F^{kl}. \tag{4.42}$$

For $\mathfrak{su}(2)$, the Lie algebra of $SU(2)$, we take the generators (in the defining representation) to be,

$$T_a = \frac{1}{2i}\,\tau_a, \tag{4.43}$$

where τ_a, $a \in \{1,2,3\}$, are the Pauli matrices. The generators have the following properties

$$\mathrm{Tr}(\tau_a\tau_b) = 2\delta_{ab} \quad \Rightarrow \quad \mathrm{Tr}(T_aT_b) = -\frac{1}{2}\,\delta_{ab}, \tag{4.44}$$
$$[\tau_a,\tau_b] = 2i\,\varepsilon_{ab}{}^c\,\tau_c \quad \Rightarrow \quad [T_a,T_b] = \varepsilon_{ab}{}^c\,T_c. \tag{4.45}$$

To write the action in a convenient way, we calculate the following positive definite object

$$\begin{aligned}\frac{1}{4}\int (F_{ij}{}^a \pm *F_{ij}{}^a)(F^{ij}_{\ a} \pm *F^{ij}_{\ a}) &= -\int \mathrm{Tr}(F \pm *F) \wedge *(F \pm *F) \\ &= -2\int \mathrm{Tr}(F \wedge *F) \mp 2\int \mathrm{Tr}(F \wedge F). \end{aligned}\tag{4.46}$$

From this, we find

$$\begin{aligned} S_E &= -\frac{1}{2}\int \mathrm{Tr}(F \pm *F) \wedge *(F \pm *F) \pm \int \mathrm{Tr}(F \wedge F) \\ &\geq 8\pi^2|k|. \end{aligned}\tag{4.47}$$

Where we defined the **instanton number** k to be

$$k = -c_2 = ch_2. \tag{4.48}$$

We see that the positive definite action S_E is bounded from below and that a minimum is attained when

$$F = \pm * F, \tag{4.49}$$

called self-dual (SD) and anti-self-dual (ASD) instantons respectively. We have chosen the instanton number k in such a way that it is positive (negative) for (A)SD instantons, as one can see from (4.41). This shows that (4.47) does establish a lower bound. Since (A)SD instantons are minima of the action, they are solutions to the equations of motion.

4.4 Further applications and remarks

Characteristic classes are a very important ingredient for the Atiyah-Singer index theorem. It would take us too far to go into a lot of details, but we will try to sketch some aspects of this application of characteristic classes to the study of anomalies and moduli spaces. For more information on the use of fibre bundles in the study of anomalies, see [1] and [4]

The idea is to compute an analytic index of certain differential operators defined on a bundle by computing topological quantities of the bundle expressed as integrals of characteristic classes of the bundle. By an analytic index of an operator D, we mean

$$\mathrm{Ind}D = \dim\ \ker D - \dim\ \ker D^\dagger \tag{4.50}$$

where $D^\dagger$ is the adjoint of D with respect to an inner product,

$$(u, Dv) = (D^\dagger u, v). \tag{4.51}$$

For the analytic index to be well defined, we need that both ker D and ker $D^\dagger$ are finite dimensional. A (bounded) operator which satisfies these conditions is called a Fredholm operator. One can show that certain differential operators on compact boundaryless manifolds, called elliptic operators, are always Fredholm operators.

It is for these elliptic differential operators that Atiyah and Singer found another way to express the index, namely in terms of characteristic classes. One important example of where this is relevant in physics, is for the chiral anomaly. Consider a massless Dirac spinor coupled to an $SU(2)$ gauge theory on a 4-dimensional manifold M. Classically, there is a global chiral symmetry

$$\psi' = e^{i\gamma_5\alpha}\psi, \quad \bar{\psi}' = \bar{\psi}e^{i\gamma_5\alpha}, \tag{4.52}$$

which leads to a conserved current

$$\partial_a j_5^a = 0. \tag{4.53}$$

Quantum mechanically this symmetry is broken, so that it is anomalous. One can show that the right hand side of (4.53) is no longer zero, but that its integral over M is given by the index of the Dirac operator

$$\mathrm{ind}(i\nabla\!\!\!/_+) = \dim\ \ker(i\nabla\!\!\!/_+) - \dim\ \ker(i\nabla\!\!\!/_-) = n_+ - n_- \tag{4.54}$$

where

$$i\nabla\!\!\!/_\pm = i\gamma^a\nabla_a P_+ = i\gamma^a(\partial_a + A_a)\frac{1}{2}(1 \pm \gamma_5), \tag{4.55}$$

and $n_\pm$ are the number of positive and negative chirality zero modes of the Dirac operator, respectively. It is clear that if the index of this operator is nonzero, the chiral symmetry is broken. For this example, the Atiyah-Singer index theorem states that (if all relevant characteristic classes of the tangent bundle TM are zero),

$$\mathrm{ind}(i\nabla\!\!\!/_+) = \int_M ch_2(F) = ch_2, \tag{4.56}$$

so the index is given by the second Chern character of P. Since for $SU(2)$ this is equal to the instanton number k, the statement becomes,

$$\int_M dx^4\, \partial_a j_5^a = \mathrm{ind}(i\nabla\!\!\!/_+) = n_+ - n_- = ch_2 = k. \tag{4.57}$$

We see that the index computes the anomaly of the theory and shows the obstruction for the classical symmetry to become a quantum symmetry. Moreover, (4.57) shows that the instanton background breaks the symmetry. On the other hand if there are no instantons, the chiral symmetry is a symmetry in the full quantum theory. Because the current in this case carries no group index, this is also called the Abelian anomaly. A similar discussion can be given for a non-Abelian anomaly.

The index theorem can for instance also be used to compute the dimension of the moduli space of instantons in $SU(N)$. One can show that the number of parameters to describe a general instanton (for a given winding number k) is related to the number of zero modes of a kind of

Dirac operator. It would take us too far to discuss this in detail, but an index slightly different from the above one computes the number of zero modes of this operator. Using this, one can show that the moduli space of $SU(N)$ instantons with winding number k is $4kN$-dimensional [10].

We have seen two examples of integral cohomology classes at work, $[c_1(F)]$ for the monopole and $[c_2(F)]$ for the instanton. Both are elements of $H^p(M,\mathbb{Z})$, where $\dim M = p$ ($p = 2$ for the monopole and $p = 4$ for the instanton). They both represent an obstruction for the principal bundle P to be trivial. In this sense it's clear that they both represent an obstruction for P to have a section, this can only happen if P is trivial. Note also that in both cases, $\mathbb{Z}$ turns out to be $\Pi_{p-1}(G)$, namely $\Pi_1(U(1)) = \Pi_3(SU(2)) = \mathbb{Z}$. These turn out to be special cases of a far more general result.

Let $P(M,G)$ be a principal bundle, and let M_p be the p-dimensional skeleton of M. Define s_p to be a section of the bundle P defined only over the skeleton, i.e. $s_p : M_p \to \pi^{-1}(M_p)$. Generically, any given s_p can be extended to a section s_{p+1} over M_{p+1}. However, if $P(M,G)$ is not trivial, there will be obstructions to continuing such a chain of extensions. Without intending to be fully rigorous, the general result found in [8] can be stated as follows:

> Let p be the first dimension such that sections over M_{p-1} cannot be extended to sections over M_p. Then, the obstruction to building a section of P over an p-dimensional skeleton of M lies in the cohomology group $H^p(M,\Pi_{p-1}(G))$.

We will not try to delve much deeper into these very interesting matters, but note that $\Pi_{p-1}(G)$ need not be $\mathbb{Z}$. So the relevant cohomology classes need not be integer valued, but might be, for instance, $\mathbb{Z}_2$ valued, like Stiefel-Whitney classes. For instance, the non-triviality of the Möbius strip is measured by the first Stiefel-Whitney class, which belongs to the cohomology group

$$H^1(S^1,\Pi_0(\mathbb{Z}_2)) = H^1(S^1,\mathbb{Z}_2) = \mathbb{Z}_2. \tag{4.58}$$

This tells us that there are two ways to make a $\mathbb{Z}_2$ bundle over S^1.

The two physics examples we studied can be interpreted as follows: For the monopole, we had

$$H^2(S^2,\Pi_1(U(1))) = H^2(S^2,\mathbb{Z}) = \mathbb{Z}. \tag{4.59}$$

This means that, if the class is not the trivial element in cohomology, there is an obstruction to finding a section over a 2-dimensional skeleton of S^2. In this case, because this is incidentally the top class of the manifold, this means indeed that one cannot construct a section of $P(S^2,U(1))$, so that the bundle is non-trivial. A charge n monopole corresponds to a connection on a bundle which is in the class $[n]$ of $H^2(S^2,\Pi_1(U(1)))$, where n is the degree of the appropriate map or is the first Chern number c_1 (see subsection 4.1).

For instantons, the story is analogous. In this case the relevant object is,

$$H^4(S^4,\Pi_3(SU(2))) = H^4(S^4,\mathbb{Z}) = \mathbb{Z}. \tag{4.60}$$

This again describes the possible obstruction to finding a section of $P(S^4,SU(2))$, because this is again the top class. The instanton with instanton number k corresponds to a connection on a bundle in class $[k]$, where $k = -c_2 = ch_2$ or is again the degree of the appropriate map (see subsection 4.3). Note that for both the monopole and instanton, the gauge potential describes one possible connection on the bundle. There are many possible connection on a bundle in a certain class $[n]$, but they will all compute the same Chern class and be characterized by the same topological degree. In this sense, the bundle itself is more fundamental than the specific solution of the gauge theory we construct. Of course, this does not take away the meaning of these specific solutions. For instance, an instanton is still a minimum of the action, while other gauge configurations in the same topological class are generically not.

Notice that a characteristic class can only be viewed as an element of the group $H^p(M,\Pi_{p-1}(G))$ if it represents the *first* obstruction to the extendibility of sections over skeletons. Higher obstructions require a different interpretation. For example, if $H^2(M,\mathbb{Z})$ is not trivial for some four-dimensional M, then it is possible to have a non-trivial $U(1)$ bundle over M with non-zero first Chern class F. Although the second Chern class of a rank one bundle is always trivial, it is possible in this case to have a non-trivial second Chern character. This means that one can have a $U(1)$ instanton on a four-dimensional manifold even though $\Pi_3(U(1)) = 0$. This is possible because the second Chern character does not represent the first obstruction of the $U(1)$ bundle, and hence does not lie in $H^4(M, \Pi_3(U(1))$. Instead, in this case it lies in $H^4(M, \Pi_1(U(1))$. Although such instantons are not wide-spread objects in quantum field theory, they certainly make their appearance in string theory. For those who are familiar with string theory, the example in question is a D4 brane wrapped on a manifold with $H^2 \neq 0$, that carries lower-dimensional D2 and D0 charge. The D2 brane can be viewed as vortex or string living on the D4, whose charge is given by the first Chern class of the $U(1)$ bundle on the D4. The D0 brane can be viewed as a $U(1)$ instanton on the D4 (if one ignores the time direction), such that its charge is given by the second Chern character of the $U(1)$ bundle.

Acknowledgments

We thank, of course, the organizers of the Second Modave School on Mathematical Physics for providing us with the opportunity to give this set of lectures and for creating the perfect environment for doing so. We also thank the participants of the Modave School for their useful questions and comments, in particular Jarah Evslin and Laurent Claessens. Both authors are supported in part by the Belgian Federal Science Policy Office through the Interuniversity Attraction Pole P5/27 and in part by the European Commission FP6 RTN programme MRTN-CT-2004-005104. AW is supported in part by the "FWO-Vlaanderen" through project G.0428.06.

Appendix: Some differential geometry

In this appendix we will quickly review some of the concepts of the theory of differentiable manifolds that we will use in the rest of these notes. We will also establish a lot of notation used throughout the main text.

A Manifolds and tangent spaces

First of all, a differentiable manifold is roughly a smooth topological space, which locally looks like $\mathbb{R}^n$. By identifying an open subset of a manifold with an open subset of $\mathbb{R}^n$, the notion of differentiability of a function from $\mathbb{R}^n$ to $\mathbb{R}^m$ is passed on to one of a function from one manifold to another. This means that one can compare manifolds as smooth spaces. More importantly for these lectures, it allows for doing physics on them, much like on $\mathbb{R}^n$. Let's be a bit more concrete.

Definition 20. We call M a differentiable manifold if the following conditions are satisfied

(i) M is a topological space

(ii) M is equipped with a set of pairs $\{(U_i, \varphi_i)\}$, where $\{U_i\}$ is an open cover of M (all U_i are open sets and $M = \bigcup_i U_i$) and φ_i is a homeomorphism from U_i to an open subset of $\mathbb{R}^n$. n is called the dimension of M

(iii) On an overlap $U_i \cap U_j \neq \emptyset$, the map $\varphi_j \circ \varphi_i^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable.

Again, given a local chart (U_i, φ_i) one can define physical quantities on U_i much like one would do on $\mathbb{R}^n$, where the φ_i define coordinates on U_i. The different patches U_i can however be glued together in a nontrivial way by the transition functions $\varphi_j\,\varphi_i^{-1}$, so that globally a manifold is a generalization of $\mathbb{R}^n$. We give this definition for completeness, but to keep the notation tractable we will be a bit sloppy throughout the text and keep the φ_i implicit. For instance, to define the derivative of a function $f : U_i \to \mathbb{R}$ at a point $x \in U_i$, one would have to consider instead $f \circ \varphi_i^{-1} : \mathbb{R}^n \to \mathbb{R}$ and use the definition of a derivative of a functional on $\mathbb{R}^n$. We will omit φ from this expression and treat f as though it were a function on $\mathbb{R}^n$.

First of all, we define the tangent space to $x \in M$.

Definition 21. Let $\gamma : [0, 1] \to U_i : t \mapsto \gamma(t)$ be a curve on a chart of M, such that $\gamma(0) = x$. A vector X at x tangent to the curve γ is called a tangent vector to M at x. If $\{x^a\}$ are a set of coordinates on U_i, X can be represented by the components

$$X^a = \frac{d}{dt} x^a(\gamma(t))\Big|_{t=0} . \tag{A1}$$

The collection of the vectors at x tangent to al curves that go through x is called the tangent space T_xM at x.

In a lot of practical situations (and to have an explicit representation) it can be convenient to define a tangent vector by using a function on M. Let f be a function from M to $\mathbb{R}$. One defines a tangent vector by

$$X(f) = \frac{d}{dt} f(\gamma(t))\Big|_{t=0}, \tag{A2}$$

where now X is represented by a differential operator

$$X = X^a \frac{\partial}{\partial x^a} \equiv X^a \partial_a, \tag{A3}$$

so that the set $\{\partial_a\}$ can be considered as a basis for T_xM. Note that this definition is consistent with the first one, since if we take the function f to be the coordinate map x^a that maps every point to its a-th coordinate, we get from (A2) that $X(x^a) = X^a$. A smooth assignment of a tangent vector at every point $x \in M$ is called a vector field $X(x)$ on M. In subsection 1.4 this is called a section of TM, $X(x) \in \Gamma(M, TM)$.

B Differential forms

Definition 22. The space dual to the tangent space T_xM is the cotangent space T^*_xM, that is, T^*_xM is the space of all functionals from T_xM to $\mathbb{R}$. An element of T^*_xM is called a cotangent vector or a 1-form α.

The dual basis to $\{\partial_a\}$ is denoted by $\{dx^a\}$, so that we have that $dx^a(\partial_b) = \delta^a_b$. More generally, this leads to

$$\alpha(X) = \alpha_a dx^a(X^b \partial_b) = \alpha_a X^a. \tag{B1}$$

In general a differential form of degree p is an element of the totally anti-symmetric tensor product of p copies of T^*_xM. This is accomplished by introducing the wedge product

$$dx^a \wedge dx^b = -dx^b \wedge dx^a. \tag{B2}$$

Continuing this process, one can build a basis for the space of all differential forms. The space of p-forms at x is denoted by $\Lambda^p_x M$. Again a smooth assignment of a 1-form at every point of M, is called a 1-form $\alpha(x)$ on M and is a section of T^*M, $\alpha(x) \in \Gamma(M, T^*M) \equiv \Lambda M$. More generally, a p-form on M is an element of $\Lambda^p M$. A general $\alpha_p \in \Lambda^p M$ can be expanded as

$$\alpha_p = \frac{1}{p!}\, \alpha_{a_1 \ldots a_p}(x)\, dx^{a_1} \wedge \ldots \wedge dx^{a_p}. \tag{B3}$$

In this way the product of two differential forms is defined, with the property

$$\alpha_p \wedge \beta_q = (-)^{pq} \beta_q \wedge \alpha_p, \tag{B4}$$

as can be seen from the expansion (B3). The exterior differential is defined by

$$d = \partial_a\, dx^a \wedge . \tag{B5}$$

which is a symbolic notation for

$$d\alpha_p = \frac{1}{p!}\, \partial_a \alpha_{a_1 \ldots a_p} dx^a \wedge dx^{a_1} \wedge \ldots \wedge dx^{a_p}. \tag{B6}$$

This means that d sends p-forms to $(p+1)$-forms, that it is nilpotent, $d^2 = 0$, and that it is an anti-derivation

$$d(\alpha_p \wedge \beta_q) = d\alpha_p \wedge \beta_q + (-)^p \alpha_p \wedge d\beta_q \tag{B7}$$

Note the very useful identity $X(f) = df(X)$, as is easily seen by expanding both expressions. A similar expression for a 2-form α is

$$d\alpha(X,Y) = X(\alpha(Y)) - X(\alpha(Y)) - \alpha([X,Y]) \tag{B8}$$

where X, $Y \in T_xM$ and $[X,Y] = [X,Y]^a\partial_a$ is the Lie bracket,

$$[X,Y](f) = X(Y(f)) - Y(X(f)). \tag{B9}$$

Differential forms are very important when it comes to defining integration on a general manifold. On an n-dimensional manifold M an n-form α_n transforms as a volume element because of the wedge product, so that the integral

$$\int_M \alpha_n \equiv \int_M \alpha(x)\, dx^1 ...\, dx^n, \tag{B10}$$

is well defined. Here we used that a top form (of maximal dimension) is always characterized by a single function,

$$\alpha_n = \frac{1}{n!}\,\alpha_{a_1...a_n}(x) dx^{a_1} \wedge ... \wedge dx^{a_n} \Rightarrow \alpha_{a_1...a_n}(x) = \alpha(x)\varepsilon_{a_1...a_n}, \tag{B11}$$

and

$$\frac{1}{n!}\,\varepsilon_{a_1...a_n} dx^{a_1} \wedge ... \wedge dx^{a_n} = dx^1 \wedge ... \wedge dx^n \tag{B12}$$

Since we always work in Euclidean signature, the totally ant-symmetric tensor is defined by $\varepsilon_{1...n} = 1$ and raising indices does not affect the sign. The most important tool we will use, is Stoke's theorem,

$$\int_M d\alpha = \int_{\partial M} \alpha, \tag{B13}$$

where ∂M is the boundary of M ($\partial\partial M = 0$).

A form α for which $d\alpha = 0$ is called closed, and if $\alpha = d\beta$ for some form β, α is called exact. A closed form is also called a cocycle and an exact form a coboundary. It is clear that all exact forms are closed (because $d^2 = 0$), but the reverse is not necessarily true on a non-trivial manifold. An important theorem is Poincaré's lemma, which states that locally (on an open set which is contractible) every closed form is exact. The fact that on a general manifold this is not the case globally means that the cohomology defined by closed forms that are not exact contains a lot of information about the topology of a manifold. To be more precise, one defines the group of p-cocycles $Z^p(M,\mathbb{R})$ and the group of p-coboundaries $B^p(M,\mathbb{R})$ as follows

$$Z^p(M,\mathbb{R}) = \{\alpha \in \Lambda^p M | d\alpha = 0\}, \tag{B14}$$
$$B^p(M,\mathbb{R}) = \{\alpha \in \Lambda^p M | \alpha = d\beta,\, \beta \in \Lambda^{p-1} M\}. \tag{B15}$$

The p-th de Rham cohomology class is then defined as the quotient,

$$H^p(M,\mathbb{R}) = Z^p(M,\mathbb{R})/B^p(M,\mathbb{R}), \tag{B16}$$

so that a general element of $H^p(M,\mathbb{R})$ is an equivalence class under the equivalence relation,

$$\alpha_p \sim \beta_p \quad \text{iff} \quad \alpha_p = \beta_p + d\gamma_{p-1}, \tag{B17}$$

and we denote their common equivalence class by $[\alpha_p] = [\beta_p] \in H^p(M,\mathbb{R})$. A period of an element $[\alpha] \in H^p(M,\mathbb{R})$ over a boundaryless submanifold $C \in M$ is defined by

$$([\alpha], C) = \int_C \alpha, \tag{B18}$$

and because of Stoke's theorem this is independent of the choice of representative of the equivalence class. Some cohomology classes, like Chern classes (see section 3), are known to have integral periods, $([c_n(F)], M) \in \mathbb{Z}$. We denote these integral cohomology groups by $H^p(M,\mathbb{Z})$.

Note that if dim$M = m$, $\Lambda^p M$ and $\Lambda^{m-p} M$ have the same dimension. Given a metric g_{ab} on M, one can define an isomorphism between the two called Hodge duality. When α_p is given by (B3), its Hodge dual $*\alpha_{m-p}$ is defined by

$$* \, \alpha_{m-p} = \frac{1}{p!(m-p)!} \sqrt{g} \, \alpha_{a_1 \dots a_p} \varepsilon^{a_1 \dots a_p}{}_{a_{p+1} \dots a_m} \, dx^{a_{p+1}} \wedge \dots \wedge dx^{a_m} . \tag{B19}$$

For Euclidean signature spaces, one finds $** = (-)^{p(m-p)}$. For example, for a 2-form in 4-dimensional Euclidean space, we have $** = 1$. This means that a self-duality condition is well defined.

C Push-forwards and pull-backs

Definition 23. Given a map $f : M \to N$, there is always an induced map $f_* : T_x M \to T_{f(x)} N$ called the push-forward (or differential map) of f. This sends a vector X tangent to a curve γ at $x = \gamma(0) \in M$ to a vector $f_* X$ tangent to the curve $f \circ \gamma$ at the point $y = f(x) = f(\gamma(0)) \in N$.

Concretely, this means that, if $\{y^a\}$ are a set of coordinates for $y \in V \subset N$, where V contains $f(x)$ the components of $f_* X$ are

$$(f_* X)^a = \frac{d}{dt} y^a(f(\gamma(t))) \bigg|_{t=0} . \tag{C1}$$

Again, sometimes it is convenient to define $f_* X$ by using an auxiliary function $g : N \to \mathbb{R}$. A definition equivalent to the previous one is

$$f_* X(g) = X(g \circ f). \tag{C2}$$

From this definition it is easy to express $f_* X$ in terms of X. First of all, using coordinate bases for the coordinates $\{x^a\}$ on $U \subset M$, where U contains x, and the $\{y^a\}$ defined previously, we find

$$(f_* X)^b \frac{\partial}{\partial y^b} g(y) = X^b \frac{\partial}{\partial x^b} (g(f(x))). \tag{C3}$$

Note that on the left hand side, $g(y)$ means that g is expressed in the coordinates $\{y^a\}$, while on the right hand side, $g(f(x))$ means that g is now expressed in the coordinates $\{x^a\}$ by means of the function f. The latter is usually simply denoted by $g(f(x)) \equiv g(x)$. Choosing now $g = y^a$, the a-th coordinate map on $V \subset N$, we find

$$(f_* X)^a = X^b \frac{\partial y^a(x)}{\partial x^b}. \tag{C4}$$

We find that the push-forward of X under the map f is simply expressed in terms of the Jacobian of $f = y(x)$. An important property of the push-forward is that for $f : M \to N$ and $h : N \to P$,

$$(h \circ f)_* = h_* f_* \, . \tag{C5}$$

Definition 24. Given a map $f : M \to N$, there always exists an induced map $f^* : T^*_{f(x)} N \to T^*_x M$, called the pull-back of f. For an $X \in T_x M$, the pull-back of a 1-form α is given by

$$f^* \alpha(X) = \alpha(f_* X). \tag{C6}$$

Using (B1) and (C4), we find,

$$(f^* \alpha)_b X^b = \alpha_b (f_* X)^b = \alpha_b X^c \frac{\partial y^b(x)}{\partial x^c}. \tag{C7}$$

Choosing now, $X = \partial / \partial x^a$, so that $X^b = \delta^b_a$, leads to

$$(f^* \alpha)_a = \alpha_b \frac{\partial y^b(x)}{\partial x^a}. \tag{C8}$$

Again, the pull-back of α under the map f is simply expressed in terms of the Jacobian of $f = y(x)$.

One can easily extend this to a definition of the pull-back of a p-form α. For $X_i \in T_xM$,

$$f^*\alpha(X_1, X_2, ..., X_p) = \alpha(f_*X_1, f_*X_2, ..., f_*X_p) \tag{C9}$$

Now the induced map is $f^* : \Lambda^p_{f(x)}N \to \Lambda^p_x M$ and in component form we find,

$$f^*\alpha_{a_1 \dots a_p}(x) = \alpha_{b_1 \dots b_p}(y(x)) \frac{\partial y^{b_1}}{\partial x^{a_1}} \cdots \frac{\partial y^{b_p}}{\partial x^{a_p}}. \tag{C10}$$

The most important properties of the pull-back of a p-form we will use are,

$$\begin{aligned} d(f^*\alpha) &= f^*d\alpha, &\text{(C11)}\\ (h \circ f)^* &= f^*h^*, &\text{(C12)}\\ f^*(\alpha \wedge \beta) &= f^*\alpha \wedge f^*\beta, &\text{(C13)} \end{aligned}$$

with $f : M \to N$, $h : N \to P$, $\alpha \in \Lambda^p N$ and $\beta \in \Lambda^q N$. These identities are not too difficult to prove using their component form.

References

[1] M. Nakahara, *Geometry, Topology and Physics*, Institute of Physics Publishing, London (1990);

[2] C. Nash and S. Sen, *Topology and Geometry for Physicists*, Academic Press, San Diego (1983);

[3] T. Frankel, *The Geometry of Physics: an Introduction*, Cambridge University Press, Cambridge (1999);

[4] R. Bertlmann, *Anomalies in Quantum Field Theory*, Oxford University Press, Oxford (2000);

[5] T. Eguchi, P. Gilkey and A. Hanson, *Gravitation, Gauge Theories and Differential Geometry*, Physics Reports 66, 213-393 (1980);

[6] C. J. Isham, *Modern Differential Geometry for Physicists*, World Scientific Publishing, Singapore (2001);

[7] C. Nash, *Differential Topology and Quantum Field Theory*, Academic Press, San Diego (1991);

[8] N. Steenrod, *The Topology of Fibre Bundles*, Princeton University Press, New Jersey (1951);

[9] J. Milnor, J. D. Stasheff, *Characteristic Classes*, Princeton University Press, New Jersey (1974);

[10] A.V. Belitsky, S. Vandoren, P. van Nieuwenhuizen, *Yang-Mills- and D-instantons*, Class.Quant.Grav. **17** (2000) 3521-3570, `hep-th/0004186`.

A mini-introduction to topological K-theory

Wen Jiang[1]

Mathematical Institute, Oxford

[1]email: jiang@cantab.net

1 Introduction

K-theory has become a very important tool in many branches of mathematics and, more recently, in theoretical physics. Despite its seeming esotericism, topological K-theory may be introduced quite naturally via vector bundles over a topological space. It will turn out to have very nice properties, such as periodicity and similarities to a cohomology theory. We will content ourselves with statements of relevant results and not delve deeper into the mathematical details.

For proofs missing in this talk, the reader is advised to consult [1] and [2], while the paper [3] contains a good overview placing K-theory in the context of topology in general.

Some materials in the following two chapters are not strictly necessary for our purpose, but serve to introduce related concepts important in their own right.

2 Complex Vector Bundles

2.1 Generalities

Definitions

For the purpose of this talk, all the spaces involved will be compact topological manifolds, all the maps continuous and all the vector spaces complex, unless otherwise stated. Almost all the results hold for smooth vector bundles, but it must be noted that holomorphic vector bundles are completely different.

Vector bundles are familiar objects to most at this school. They can be visualised as families of vector spaces parametrised by points in a manifold and homomorphisms between them must be vector space homomorphisms when restricted to the fibres. A vector bundle $p : E \to X$ with fibre V must be locally trivial in the sense that around any point $x \in X$ is an open neighbourhood U such that E restricted to U is isomorphic to the Cartesian product $U \times V$. Such a U is called a trivialising neighbourhood.

Given that X is compact, we can find a finite open cover U_i of X such that over each U_i there is an isomorphism $\phi_i : E|_{U_i} \to U_i \times V$. Then for any $x \in U_i \cap U_j$, the map $\phi_i \phi_j^{-1}$ restricted to $\{x\} \times V$ gives an isomorphism $g_{ij}(x) : V \to V$. The function $g_{ij} : U_i \cap U_j \to GL(V)$ is often called a transition function. We may work backwards and glue together E from the data $\{U_i\}$ and $\{g_{ij}\}$. In fact, thinking about which sets of such data define equivalent bundles will lead to Čech cohomology, of which we will speak no further.

Note that any vector bundle admits a hermitian metric. The transition functions will consequently take values in $U(V)$. Thus complex line bundles are just $U(1)$ or circle bundles.

Constructions

Let us look at various ways of obtaining new bundles from old. Natural operations on vector spaces, such as taking the dual, quotient, direct sum and tensor product, all carry over to vector bundles. Thus, the set of (isomorphism classes of) vector bundles over X almost forms a ring (like $\mathbb{Z}$), except that we cannot subtract bundles. Attempting to define additive inverses will lead one directly to topological K-theory.

Another method of producing bundles is the pull-back. Suppose we have a map $f : Y \to X$, we can define the 'pull-back bundle' f^*E as a subset of $E \times Y$ by

$$f^*E = \{(e, y) \in E \times Y \ : \ p(e) = f(y)\}.$$

In words, a point $y \in Y$ is associated with the vector space over the point $f(y)$. If $E|_U \cong U \times V$, then $W = f^{-1}(U)$ is open and $(f^*E)|_W \cong W \times V$. Hence, f^*E is locally trivial. The properties of the pull-back bundle is summarised by the following commutative diagram.

$$\begin{array}{ccc} f^*E & \xrightarrow{\tilde{f}} & E \\ {\scriptstyle q}\downarrow & & \downarrow{\scriptstyle p} \\ Y & \xrightarrow{f} & X \end{array}$$

In terms of transition functions, those for f^*E are simply $g_{ij} \circ f$ over the intersections $W_i \cap W_j$.

Fact: If $f_0, f_1 : Y \to X$ are homotopic maps, then f_0^*E and f_1^*E are isomorphic. It follows that homotopy-equivalent spaces have the same isomorphism classes of vector bundles over them. In particular, any bundle over a contractible space is trivial.

Example (Bundles over S^n): The n-sphere S^n may be covered by two open disks U_1 and U_2, the intersection of which is homotopy-equivalent to S^{n-1}. Any bundle $V \to E \to S^n$ is trivial when restricted to U_1 and U_2. From the gluing point of view, E is determined by the homotopy class of the transition function $g : S^{n-1} \to GL(V)$. Thus, isomorphism classes of vector bundles over S^n correspond to $[S^{n-1},\ GL(V)]$, the set of homotopy classes of maps from $S^{n-1} \to GL(V)$.

For $n = 1$, this set contains a single element, as $GL(V)$ is path-connected. Therefore, any complex vector bundle over S^1 is trivial – this is easy to see directly anyway.

When $n = 2$, $[S^1,\ GL(V)]$ is the same as the fundamental group $\pi_1(GL(V))$, which is $\mathbb{Z}$ for any V of dimension> 0. This integer invariant can be identified with the first Chern class of the bundle. Hence, two vector bundles over S^2 are isomorphic if and only if they have the same rank and first Chern class.

Example (Tautological bundles): A projective space is naturally equipped with a line bundle. Each point $[l]$ in the space $\mathbb{CP}^n$ by definition labels a 1-dimensional subspace l of $V = \mathbb{C}^{n+1}$. The bundle $H^* \to \mathbb{CP}^n$ simply assigns l as the fibre over $[l]$. Its dual is denoted by H and the direct sum (resp. tensor product) of m copies of H is written as mH (resp. H^m). Since $\mathbb{CP}^1 = S^2$, from the previous example, we deduce that any line bundle over S^2 is isomorphic to H^m for some $m \in \mathbb{Z}$.

A slight generalisation of a projective space is a Grassmanian $G_k(V)$, the set of subspaces of V of dimension k. So $\mathbb{CP}^n$ is simply $G_1(V)$. As above, the tautological bundle Q assigns the vector space $u \subset V$ to the point $[u] \in G_k(V)$. It is also called the classifying bundle over $G_k(V)$.

Now any map $f : Y \to G_k(V)$ gives a vector bundle f^*Q. We will see later that any rank k bundle over Y is of this form for some $V = \mathbb{C}^N$.

Complementary bundles

Certainly not every vector bundle $V \to E \to X$ is trivial. However, provided that the base space X is compact, as we assume, it can be shown that there always exists a bundle $V' \to E' \to X$ such that

$$E \oplus E' \cong X \times \mathbb{C}^N,$$

where $N = \mathrm{dim}V + \mathrm{dim}V'$. Such an E' is called complementary to E.

The existence of complementary bundles is due to the fact that any (compact) manifold admits a partition of unity. Given any (finite) open cover $\{U_i\}$ of X, a partition of unity subordinate to $\{U_i\}$ is a collection of continuous functions $f_i : X \to \mathbb{R}$ such that

- $f_i(x) \geq 0$, for any $x \in X$ and any i;
- the support of each f_i lies in U_i;
- $\sum_i f_i(x) = 1$, for any $x \in X$.

Partitions of unity are used to patch up local definitions into a global one. Suppose that the triviality of $E|_{U_i}$ allows us to produce a map h_i on U_i, then the object $\sum_i f_i h_i$ is well-defined over the whole manifold.

Example (Hermitian metrics): If we take h_i to be a hermitian metric on $E|_{U_i}$, we obtain a hermitian metric on the whole of E. As a consequence, any short exact sequence of vector bundles

$$0 \to E_1 \to E \to E_2 \to 0$$

splits, that is to say, $E \cong E_1 \oplus E_2$. To see this, note that $E \cong E_1 \oplus E_1^{\perp}$ given any metric. The exact sequence, however, implies that $E_2 \cong E/E_1$, which is in turn isomorphic to $E_1^{\perp}$.

2.2 Classifying space for vector bundles

The importance of complementary bundles will become apparent in this subsection, as we find a homotopy characterisation of $\mathrm{Vect}_n(X)$, the set of isomorphic classes of rank n bundles on X.

Let $E \to X$ be a rank n vector bundle as before. Take a complementary bundle $E' \to X$, so that $E \oplus E' \cong X \times \mathbb{C}^N$ for some N. Then we have an inclusion $i : E \to X \times \mathbb{C}^N$, a injective bundle morphism. Over each point $x \in X$, the map $i_x : E_x \to \mathbb{C}^N$ is an injective vector space homomorphism and the image $\mathrm{Im}\, i_x$ is of codimension n. We can then define a map

$$f : X \to G_n(\mathbb{C}^N); \quad x \mapsto \mathrm{Im}\, i_x.$$

Recall that the Grassmanian $G_n(\mathbb{C}^N)$ has a classifying bundle Q as described in subsection 2.1. To see that the pull-back f^*Q is isomorphic to E, note that the fibre of f^*Q over $x \in X$ is the fibre of Q over $f(x)$, which is tautologically $\mathrm{Im} i_x$, naturally isomorphic to E_x. Hence, $f^*Q \cong E$.

It can be shown that different choices of E' lead to homotopic maps f, and thus isomorphic bundles. Denote by $[X,\ G_n(\mathbb{C}^N)]$ the set of homotopy classes of maps $f : X \to G_n(\mathbb{C}^N)$.

Theorem: The map $f \mapsto f^*Q$ induces an isomorphism

$$[X,\ G_n(\mathbb{C}^N)] \to \mathrm{Vect}_n(X),$$

where the natural number N only depends on the compact space X.

The number N increases when X becomes more complicated, that is, if a large number of trivialising neighbourhoods are required to cover X. The theorem above holds for any X if $G_n(\mathbb{C}^N)$ is replaced by a construction called the direct limit, $G_n(\mathbb{C}^\infty) = \varinjlim_N G_n(\mathbb{C}^N)$. It is a 'classifying space' for complex rank n bundles. We will not elaborate on this. For more details, see [6] or [9].

3 Characteristic Classes

The aim of this section is to make a detour into the theory of characteristic classes, so that one could introduce those the Chern character and Todd class, which are often encountered in K-theory. We begin with Chern classes, not only because they are the most commonly encountered characteristic classes, but also because they form the building blocks of others, as we shall see.

The distinction between topological and smooth vector bundles will be blurred in this section. This poses no real ambiguity as any topological vector bundle we are considering could be given differential structures and any continuous maps between them can be approximated by smooth ones.

3.1 Chern classes

Topological approach

There are a number of definitions of Chern classes for complex vector bundles. They are usually introduced to physicists through the Chern-Weil homomorphism: given a hermitian metric and a connection on a complex bundle $E \to X$, the total Chern class $c(E) \in H^*(M, \mathbb{R})$ is represented by the closed form

$$\det(I + \frac{i}{2\pi} F),$$

where F is the curvature of the connection. It then has to be shown that this definition is independent of the metric and connection.

Having introduced the classifying bundle, we can give an alternative, purely topological, definition. Given a bundle $E \to X$, represent it as f^*Q for some map $f : X \to G_n(\mathbb{C}^\infty)$, as in the theorem in the last subsection. The cohomology ring of $G_n(\mathbb{C}^\infty)$ is a free polynomial ring on n generators,

$$\mathbb{R}[c_1,\ \ldots,\ c_n].$$

Then set the kth Chern class of E to be $c_k(E) = f^*c_k$. This definition can be verified to agree with the previous one, if the c_k's are normalised appropriately.

The definition is not as important as the properties. As a matter of fact, Chern classes are uniquely determined by the following of its properties.

- For any bundle $E \to X$ and continuous map $f : Y \to X$, $c(f^*E) = f^*(c(E))$, where f^*E is the pull-back bundle and $f^*(c(E))$ is the pull-back of the cohomology class on X.
- If $E \cong E_1 \oplus E_2$ on X, then $c(E) = c(E_1)\, c(E_2)$.
- The Chern class of a complex line bundle $L \to X$ is $c(L) = 1 + e(L_{\mathbb{R}})$, where $e(L_{\mathbb{R}})$ is the Euler class of L regarded as a rank 2 real vector bundle.

The last is the key, as it indicates that vector bundles can be built from line bundles as far as Chern classes are concerned. To explain this, we need to learn about the splitting principle, the subject of the next subsection. If the reader is not familiar with the Euler class, he could think of it as $iF/2\pi$, in terms of a $U(1)$ connection on L, or consult [4] for an elementary exposition.

Before carrying on, let's note a few results. If L, L' are line bundles, then

$$c_1(L \oplus L') = c_1(L)\, c_1(L') \quad \text{and} \quad c_1(L \otimes L') = c_1(L) + c_1(L').$$

Let L^* be the dual line bundle of L. Since $L \otimes L^* = X \times \mathbb{C}$ and c_1 of a trivial bundle is 0, we have $c_1(L^*) = -c_1(L)$.

Over $S^2 = \mathbb{CP}^1$, denote by 1 the trivial line bundle and by x the first Chern class $c_1(H)$ of the tautological line bundle H we introduced earlier. Then the total Chern classes of the two bundles $H^2 \oplus 1$ and $H \oplus H$ are

$$c(H^2 \oplus 1) = c(H^2)\, c(1) = c(H^2) = 1 + c_1(H^2) = 1 + 2x;$$

and

$$c(H \oplus H) = c(H)\, c(H) = (1 + x)^2 = 1 + 2x,$$

since x^2 is a degree 4 form, which vanishes on S^2. Now the rank and the first Chern class are the only topological invariants over S^2, as we saw in subsection 2.1. Therefore,

$$H^2 \oplus 1 \cong H \oplus H.$$

One could also see this directly, by noting that the two bundles have transition functions $\begin{pmatrix} 1/z^2 & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1/z & 0 \\ 0 & 1/z \end{pmatrix}$ respectively. The following formula gives an explicit homotopy between the two. It is equal to the first matrix when $t = 0$ and to the second when $t = \pi/2$.

$$\begin{pmatrix} 1/z & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} \begin{pmatrix} 1/z & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$$

The splitting principle

Without going into details, we simply state that, given a vector bundle $E \to X$, there exists a compact space $F(E)$ and a map $\sigma : F(E) \to X$ such that

- $\sigma^*E \to F(E)$ splits into a sum of line bundles over $F(E)$,
$$\sigma^*E = L_1 \oplus \cdots \oplus L_n;$$
- The pull-back of cohomology $\sigma^* : H^*(X) \to H^*(F(E))$ is injective.

More generally, for two bundles E, E' on X, one can apply the above twice to split both E and E' on $F(E, E') := F(\sigma_E^* E')$.

To illustrate the use of this construction, suppose that we wish to prove a *polynomial* relation $P(c(E), c(E')) = 0$ between two vector bundles on X and assume that this relation holds when E and E' are sums of line bundles on any space. Pulling back by σ to $F(E, E')$ gives

$$\sigma^* P(c(E), c(E')) = P(\sigma^* c(E), \sigma^* c(E')) = P(c(\sigma^* E), c(\sigma^* E')) = 0,$$

since σ^*E and σ^*E' are sums of line bundles on $F(E, E')$. But the injectivity of σ^* implies that $P(c(E), c(E')) = 0$. (This argument of course goes unchanged if more than two bundles appear in P.) Thus, we have the following

Splitting Principle: To prove polynomial identities of Chern classes, it suffices to assume that the bundles involved are sums of line bundles.

This principle can in fact be used to *define* other characteristic classes in terms of Chern classes, as we will explain next.

3.2 More characteristic classes

The splitting principle allows us to write

$$\sigma^* c(E) = \prod_{i=1}^{n}(1 + x_i),$$

for certain $x_i \in H^2(F(E))$, the first Chern class of the line bundle L_i. Since the whole expression is symmetric in x_i, its degree k part, $s_k(x_1, \ldots, x_n)$, is also. It is called the kth elementary symmetric polynomial in x_i. For example,

$$s_1 = \sum_i x_i, \quad s_2 = \sum_{i<j} x_i x_j, \quad \ldots, \quad s_n = \prod_i x_i.$$

Fact: Any symmetric polynomial in x_i may be written as a polynomial in s_i.

Note that $s_k = \sigma^* c_k(E)$ by construction. Therefore, any symmetric polynomial $P(x_1, \ldots, x_n)$ is equal to a polynomial

$$P'(s_1, \ldots, s_n) = \sigma^* P'(c_1(E), \ldots, c_n(E)) \equiv \sigma^* \tau_P(E).$$

Since σ^* is injective, P uniquely determines a class $\tau_P(E) \in H^*(X)$ in this fashion. Such a 'characteristic' class evidently possesses the same functorial property as the Chern classes. One can in fact allow P to be a power series expression, since $H^m(X) = 0$ for $m > \dim X$ and so higher products of the x_i's eventually vanish.

We may now introduce the Chern character. Let P be the power series

$$f_{\text{ch}} = \sum_{i=1}^{n} e^{x_i} = n + \sum x_i + \frac{1}{2}\sum x_i^2 + \cdots .$$

The class it determines is the total Chern character $\text{ch}(E)$. In terms of the Chern classes,

$$\text{ch}(E) = \text{rank}\, E + c_1(E) + \left(\frac{1}{2}c_1(E)^2 - c_2(E)\right) + \cdots .$$

The importance of the Chern character is that, for any bundles $E, F \to X$,

$$\text{ch}(E \oplus F) = \text{ch}(E) + \text{ch}(F) \quad \text{and} \quad \text{ch}(E \otimes F) = \text{ch}(E)\,\text{ch}(F).$$

To prove this, we use the splitting principle and assume that E and F are sums of line bundles, L_i and L'_j, with c_1 being x_i and y_j respectively. Then

$$c(E \oplus F) = c(\bigoplus_i L_i \oplus \bigoplus_j L'_j) = \prod_i (1 + x_i) \prod_j (1 + y_j),$$

so

$$\text{ch}(E \oplus F) = \sum_i e^{x_i} + \sum_j e^{y_j} = \text{ch}(E) + \text{ch}(F).$$

For the second equality, we go through the same procedure:

$$c(E \otimes F) = c(\sum_{i,j} L_i \otimes L'_j) = \prod_{i,j}(1 + x_i + y_j)$$

implies that

$$\text{ch}(E \otimes F) = \sum_{i,j} e^{x_i + y_j} = \sum_i e^{x_i} \sum_j e^{y_j} = \text{ch}(E)\,\text{ch}(F).$$

In the Chern-Weil theory, the Chern character may be represented by the following closed form

$$\text{Tr}\,\exp\left(\frac{F}{4\pi i}\right).$$

More characteristic classes may be obtained by considering other choices of P. Of course, the only interesting ones are those which arise naturally from real problems. For example, in the Atiyah-Singer Index theorem, one encounters the formal object

$$\mathrm{ch}\,(\bigwedge^{\text{even}} E - \bigwedge^{\text{odd}} E) = \prod_i (1 - e^{x_i}),$$

where the $\bigwedge$ is the exterior product and we will worry about the '$-$' sign in the next subsection. Dividing the top Chern class $c_n(E) = \prod x_i$ by this leads one to define the Todd class through the symmetric power series

$$f_{\mathrm{Td}} = \prod_{i=1}^{n} \frac{x_i}{1 - e^{-x_i}} = \prod_{i=1}^{n} \left(1 + \frac{1}{2}x_i + \frac{1}{12}x_i^2 + \cdots\right).$$

In terms of Chern classes, $\mathrm{Td}_1 = c_1/2$, $\mathrm{Td}_2 = (c_1^2 + c_2)/12$, etc.

There is a similar story for real vector bundles. One uses the Pontryagin classes, which are Chern classes of the complexification, to build the $\hat{A}$-classes through the power series $\prod \dfrac{x_i/2}{\sinh(x_i/2)}$, and also the $\hat{L}$-classes where 'sinh' is replaced by 'tanh'.

3.3 Other cohomological invariants?

In the previous subsection, we constructed some characteristic classes as polynomials in the Chern classes. Are there any other characteristic classes not expressible in this way?

What do we mean then by a characteristic class? One may define it as a natural transformation τ between $\mathrm{Vect}_n(\cdot)$ and $H^*(\,\cdot\,;\mathbb{R})$, both contravariant functors from the category of compact manifolds to the category of sets. By this 'functorial statement', we simply mean that, for any bundle $E \to X$ and map $f : Y \to X$, $\tau(E)$ is a cohomology class in $H^*(X)$ and that

$$\tau(f^*E) = f^*\tau(E).$$

This is of course in complete analogy with the first property that Chern classes satisfy, as stated in the last subsection. From this point of view, the following theorem shows that Chern classes are the only cohomological invariants of complex vector bundles in this sense.

Theorem: Every natural transformation τ as above can be expressed as a polynomial in the Chern classes.

Proof. Given a bundle $E \to X$, let $f : X \to G_n(\mathbb{C}^\infty)$ be a classifying map, so that $E \cong f^*Q$. Now $\tau(Q) \in H^*(G_n(\mathbb{C}^\infty))$, which is a polynomial ring generated by c_k, as we said. (It does not matter here that $G_n(\mathbb{C}^\infty)$ is infinite-dimensional.) Hence, $\tau(Q) = P(c_1, \ldots, c_n)$ for some polynomial. The 'functoriality' of τ then implies that

$$\tau(E) = \tau(f^*Q) = f^*\tau(Q) = P(f^*c_1, \ldots, f^*c_n) = P(c_1(E), \ldots, c_n(E)),$$

as we claimed. □

4 K-theory

As appropriate to this school, we will concentrate on the basic ideas, which are fairly natural, even though proofs of results are generally technical and will be omitted here. Those who wish may find them in the copious literature, for example, [1], [5] and [7].

4.1 Basic objects

The group $K(X)$

We saw in subsection 2.1 that vector bundles on a space X can be added and multiplied. To make the set of vector bundles $\mathrm{Vect}(X)$ into a ring, it would require one to define a 'subtraction' or, equivalently, additive inverses. This can be done formally. Let's recall the construction of negative numbers from natural numbers. Define a relation $\sim$ on $\mathbb{N} \times \mathbb{N}$ by

$$(a, b) \sim (c, d) \quad \text{if} \quad a + d = c + b.$$

Here (a, b) is to be thought of as '$a - b$'. This relation is evidently reflexive and symmetric. It is also transitive, since $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$ implies that

$$a + f + c + d = b + e + c + d$$

and so cancelling $(c + d)$ gives $(a, b) \sim (e, f)$. Equivalence classes under $\sim$ then form the ring of integers $\mathbb{Z}$.

One can attempt the same procedure with $\mathrm{Vect}(X) \times \mathrm{Vect}(X)$, but one immediately runs into difficulty, as the 'cancellation' above does not hold for vector bundles. (A real bundle example is the tangent bundle of $S^2 \subset \mathbb{R}^3$. Its direct sum with the normal bundle $S^2 \times \mathbb{R}$ is $S^2 \times \mathbb{R}^3$.) Therefore, one is led to define $(E, F) \sim (E', F')$ if there exists G such that

$$E \oplus F' \oplus G \cong E' \oplus F \oplus G.$$

Transitivity holds for this relation. The equivalence classes of these pairs $[(E, F)]$ form the additive abelian group $K(X)$. Note that indeed $[E \oplus F] = [E] + [F]$

Each vector bundle $E \to X$ determines an element $[E] \equiv [(E, 0)]$ of $K(X)$ and since $(E, 0) + (0, E) \sim (0, 0)$, we have $-[E] = [(0, E)]$. Thus, every element of $K(X)$ is of the form $[E] - [F]$. Moreover, there is a complementary bundle F' such that $F \oplus F' \cong n$, the trivial rank n bundle. Consequently

$$[E] - [F] = [E \oplus F'] - [F \oplus F'] = [E \oplus F'] - [n],$$

and so all the elements of $K(X)$ are in fact of the form $[E] - [n]$. It follows that $[E] = [F]$ in $K(X)$ if and only if $E \oplus n = F \oplus n$ for some n. Such bundles are called stably isomorphic.

Any map $f : Y \to X$ enables one to pull back bundles from X to Y and so induces a map $f^* : K(X) \to K(Y)$, which only depends on the homotopy class of f. One can view $K(\,\cdot\,)$ as a contravariant functor from the category of compact topological spaces to the category of abelian groups. Furthermore, the tensor product between vector bundles on X induces a multiplication in $K(X)$, which makes $K(X)$ into a ring.

One can remove the appearance of an arbitrary natural number from elements $[E] - [n]$ of $K(X)$ as follows. Take any point $x_0 \in X$, the inclusion $i : \{x_0\} \to X$ induces a map $i^* : K(X) \to K(x_0) \cong \mathbb{Z}$. Define $\widetilde{K}(X)$ to be the kernel of i^*. Hence the elements of $\widetilde{K}(X)$, if X is connected, are of the form $[E] - [\mathrm{rk}E]$ and $[E] = [F]$ in $\widetilde{K}(X)$ if and only if $E \oplus m \cong F \oplus n$ for some m and n. Such bundles are called stably equivalent.

The collapsing map $c : X \to \{x_0\}$ induces an splitting and

$$K(X) \cong \widetilde{K}(X) \oplus K(x_0) \cong \widetilde{K}(X) \oplus \mathbb{Z}.$$

Example ($K(S^k)$): Because of its properties mentioned above, $K(X)$ is a homotopy invariant, to calculate which is generally very difficult. However, Bott periodicity, to be introduced later, makes it possible for certain spaces. From our gluing point of view in subsection 2.1, we see that $\mathrm{Vect}(S^1) = \mathbb{Z}$ and $\mathrm{Vect}(S^2) = \mathbb{Z} \times \mathbb{Z}$ as sets. If S^0 means the '0-sphere' $\{1, -1\}$, then $\mathrm{Vect}(S^0) = \mathbb{Z} \times \mathbb{Z}$. Evidently $K(X) = \mathrm{Vect}(X)$ for $X = S^0$ or S^1, since all the bundles are trivial over these spaces. We assert that this is also true when $X = S^2$. We see, therefore, that

$$K(S^0) = \mathbb{Z} \times \mathbb{Z}, \quad K(S^1) = \mathbb{Z}, \quad \text{and} \quad K(S^2) = \mathbb{Z} \times \mathbb{Z}.$$

The equality of $K(S^0)$ and $K(S^2)$ is no coincidence and is an example of Bott periodicity, which implies that, for any $k \in \mathbb{N}$,

$$\begin{aligned} K(S^{2k}) &= K(S^0), \\ K(S^{2k+1}) &= K(S^1). \end{aligned}$$

The ring, i.e. multiplicative, structure will be described later.

Suspension and $K^*(X)$

The group $K(X)$ is really $K^0(X)$ in a infinite sequence of K-groups. To define these, one has to introduce several topological constructions. For a space X with a distinguished base point x_0, let the suspension of X be

$$SX = ((X \times [0,1])/(X \times \{0\}))\,/(X \times \{1\}),$$

i.e. a cylinder with each end pinched to a point. The reduced suspension is

$$\Sigma X = SX/(\{x_0\} \times [0,1]),$$

i.e. the line segment through x_0 is further shrunk to a point. For example, for a sphere, $SS^n \cong \Sigma S^n \cong S^{n+1}$. Note that SX and ΣX are homotopy equivalent and so have identical K-groups.

When $Y \subset X$ is non-empty and compact, X/Y is the quotient space with Y/Y as the base point. For $Y = \emptyset$, let $X/\emptyset$ be the space $X_+ = X \sqcup \{*\}$, the disjoint union of X with a point.

Definition: For $n \geq 0$, let

$$\begin{aligned} \widetilde{K}^{-n}(X) &= \widetilde{K}(\Sigma^n X) \\ K^{-n}(X,Y) &= \widetilde{K}^{-n}(X/Y) \\ K^{-n}(X) &= K^{-n}(X,\emptyset) = \widetilde{K}(\Sigma^n(X_+)). \end{aligned}$$

The minus sign in the exponents was chosen so that the index increases in the following exact sequence (see [1] for proof)

$$\begin{gathered} \cdots \to K^{-2}(Y) \xrightarrow{\delta} K^{-1}(X,Y) \xrightarrow{j^*} K^{-1}(X) \xrightarrow{i^*} \\ K^{-1}(Y) \xrightarrow{\delta} K^0(X,Y) \xrightarrow{j^*} K^0(X) \xrightarrow{i^*} K^0(Y), \end{gathered}$$

where $Y \subset X$ is compact and $i : Y \to X$ and $j : (X,\emptyset) \to (X,Y)$ are the inclusions. Please consult [1] for the description of the map δ.

Some may have noticed that this is similar in form to the long exact sequence in relative cohomology. In fact, $K*$ satisfies all the axioms of Eilenberg and Steenrod, except the dimension axiom, and is therefore an 'extraordinary cohomology theory'. The Chern character, constructed in subsection 3.2, gives rise to a ring homomorphism

$$\mathrm{ch} : K^0(X) \to H^{2*}(X;\mathbb{Q}),$$

which turns out to induce an isomorphism $K^0(X) \otimes \mathbb{Q} \to H^{2*}(X;\mathbb{Q})$ for 'nice enough' spaces.

4.2 Bott Periodicity

Bott originally discovered the periodicity named after him when he applied Morse theory to loop spaces of (compact) classical groups. This translates into a natural isomorphism of the K-groups: $K^{-n}(X) \xrightarrow{\sim} K^{-n-2}(X)$, for all $n \geq 0$. One could read Bott's original approach in Milnor's beautiful book [8], or a direct proof for K-theory in [1].

Unfortunately, there is no time to explain the precise nature of the isomorphism. We will take account of the periodicity and set, for $n > 0$, $K^n = K^{-n}$. Note that the long exact sequence in the previous subsection reduces to a six-term exact sequence.

$$\begin{tikzcd} K^0(X,Y) \arrow[r, "j^*"] & K^0(X) \arrow[r, "i^*"] & K^0(Y) \arrow[d, "\delta"] \\ K^1(Y) \arrow[u, "\delta"] & K^1(X) \arrow[l, "i^*"'] & K^1(X,Y) \arrow[l, "j^*"'] \end{tikzcd}$$

Example ($K^*(S^n)$): In the case of a point,

$$K^{-n}(\mathrm{pt}) = \widetilde{K}(\Sigma^n(\mathrm{pt}_+)) = \widetilde{K}(S^n).$$

Therefore, the periodicity $K^{-n}(\mathrm{pt}) = K^{-n-2}(\mathrm{pt})$ implies that the K-groups of spheres are periodic, as we claimed before. The ring structure can be described as follows. We will look at the simpler reduced version, $\widetilde{K}^*(S^n) = \widetilde{K}^0(S^n) \oplus \widetilde{K}^1(S^n)$. There are only two distinct cases: $n = 1$ or 2.

The 0-th group $\widetilde{K}^0(S^2) \cong \mathbb{Z}$ is generated additively by $[H] - [1]$. We saw in subsection 3.1 that $H^2 \oplus 1 \cong 2H$, which implies that

$$([H] - [1])^2 = 0$$

in K-theory. On the other hand, $\widetilde{K}^1(S^2) = \widetilde{K}(\Sigma S^2) = \widetilde{K}(S^3) \cong \widetilde{K}(S^1) = 0$. Thus, as rings,

$$\widetilde{K}^*(S^2) = \mathbb{Z}t \oplus 0, \quad \text{where } t^2 = 0.$$

The ring $\widetilde{K}^*(S^1)$ is the same except that the even-odd grading is reversed.

Aside. One could give some motivation to the use of suspension in defining K^{-n}. It could be seen as directly taking advantage of Bott periodicity for the classical groups. Let $U = \varinjlim U(n)$ and $BU = \varinjlim BU(n)$. Bott periodicity for $U(n)$ (deformation retract of $GL_n(\mathbb{C})$) says that the loop space $\Omega(U)$ is homotopy-equivalent to $BU \times \mathbb{Z}$. For X compact and connected, it is not difficult to see that $[X, BU]$, the set of (free) homotopy classes of maps $X \to BU$, corresponds to $\widetilde{K}(X)$, the set of classes of stably equivalent bundles on X. We need to consider based maps, so noting that $[X_+, BU]_{\text{based}} = [X, BU]_{\text{free}}$ and $K(X) = \widetilde{K}(X) \oplus \mathbb{Z}$, we have

$$\begin{aligned} K(X) &= [X_+, BU \times \mathbb{Z}] && \text{since } X \text{ is connected,} \\ &= [X_+, \Omega(U)] && \text{by Bott periodicity,} \\ &= [\Sigma(X_+), U] && \text{as } \Sigma \text{ and } \Omega \text{ are adjoint,} \\ &= \widetilde{K}(\Sigma^2(X_+)) && \text{by the gluing construction,} \\ &= K^{-2}(X)\,. \end{aligned}$$

Thus we have periodicity in K^*, too. (Here all maps are based.)

4.3 Appearance in the Index theorem

The first major application of K-theory was in Atiyah and Singer's proof of their Index theorem. We cannot go into any detail except to describe briefly the role of K-theory in this context.

Let X be a compact smooth manifold of dimension n and E, F be smooth vector bundles over X. An elliptic differential operator $P : \Gamma(E) \to \Gamma(F)$ determines a principal symbol $\sigma(P)$, a topological object that can be regarded as an element of $K^0_c(T^*M)$. Here the cotangent bundle T^*M is of course non-compact and $K^0_c(T^*M)$ is K-group with compact support. One then defines an 'analytical index', which is basically the index of the differential operator P:

$$\operatorname{Ind} P = \dim \ker P - \dim\,(\operatorname{Im} P)^{\perp}.$$

This quantity is defined through the analytical properties of P, but may in fact be expressed using purely topological data of the bundles. One could see this by defining a 'topological index' Ind_T and then showing that the two indices are equal.

The number Ind_T can be expressed explicitly as the integral over X of a cohomological class and the Atiyah-Singer Index theorem is

$$\operatorname{Ind} P = (-1)^{\frac{1}{2}n(n+1)} \int_X \pi_! \operatorname{ch}(\sigma(P)) \wedge \operatorname{Td}(TM \otimes \mathbb{C}).$$

For more details, please consult [6] or any of the many books written on this beautiful result.

Acknowledgements

The author would like to thank the organisers of Modave Summer School 2006 for inviting him to speak on this subject and to enjoy the wonderful village of Modave.

References

[1] M. F. Atiyah, *K-theory*, W.A. Benjamin 1967

[2] R. Bott, Lectures on $K(X)$, *Collected Papers*, Vol. 2, Birkhäuser 1994

[3] R. Bott, The Periodicity Theorem for the Classical Groups and Some of its Applications, *Collected Papers*, Vol. 1, Birkhäuser 1994

[4] R. Bott, L. W. Tu, *Differential Forms in Algebraic Topology*, Springer-Verlag 1982

[5] M. Karoubi, *K-theory*, Springer-Verlag 1978

[6] H. B. Lawson, M. L. Michelsohn, *Spin geometry*, Princeton Univ. Press 1989

[7] Peter May, Online notes on K-theory, `http://www.math.uchicago.edu/~fowler/k-theory/`

[8] J. Milnor, *Morse Theory*, Ann. of Math. Studies 51, Princeton Univ. Press 1963

[9] J. Milnor, J. Stasheff, *Characteristic Classes*, Ann. of Math. Studies 76, Princeton Univ. Press 1974

What does(n't) K-theory classify?

Jarah Evslin[1]

International Solvay Institutes,
Physique Théorique et Mathématique,
Statistical and Plasma Physics C.P. 231,
Université Libre de Bruxelles,
B-1050, Bruxelles, Belgium

Abstract. We review various K-theory classification conjectures in string theory. Sen conjecture based proposals classify D-brane trajectories in backgrounds with no H flux, while Freed-Witten anomaly based proposals classify conserved RR charges and magnetic RR fluxes in topologically time-independent backgrounds. In exactly solvable CFTs a classification of well-defined boundary states implies that there are branes representing every twisted K-theory class. Some of these proposals fail to respect the self-duality of the RR fields in the democratic formulation of type II supergravity and none respect S-duality in type IIB string theory. We discuss two applications. The twisted K-theory classification has led to a conjecture for the topology of the T-dual of any configuration. In the Klebanov-Strassler geometry twisted K-theory classifies universality classes of baryonic vacua.

[1]email: jevslin@ulb.ac.be

1 Introduction

In 1997 Minasian and Moore suggested that Ramond-Ramond charges in type II string theories are classified by K-theory [1], which is a group that characterizes the complex vector bundles on a given spacetime. Ramond-Ramond (RR) fluxes are sourced by RR charges, in the same way that the magnetic field is sourced by magnetic monopoles in QED. In QED the well-definedness of the partition function of the dual electric charges implies that the magnetic field is quantized. Combining this quantization condition with Gauss' Law one finds that magnetic charge cannot be smeared at will but must instead be localized on a codimension three surface in spacetime, which is the trajectory of a magnetic monopole. Similarly in type II string theory the well-definedness of the partition function of dual RR charges, which are also RR charges, imposes that the RR field strengths are quantized. This in turn implies that RR sources are localized on submanifolds, which are called D-branes. And so Minasian and Moore's conjecture is that D-brane configurations are classified by K-theory.

Since 1997 many versions of this conjecture have appeared, classifying various features of D-brane configurations in terms of different K-theories on distinct subspaces. Some use the K-theory of the spacetime, some use the K-theory of a spatial slice, some use relative K-theory. In these lectures we will review several of these proposals, their physical motivations, and their ranges of validity.

We begin in Sec. 2 with the older, homology classification of D-branes. In Sec. 3 we provide some background material on charges and fluxes in type II supergravity theories and explain why the homology classification fails.

Finally we come to K-theory in Sec. 4. We first describe Witten's proposal [2] for a classification of D-brane trajectories in type IIB string theory. This classification scheme is a consequence of the Sen conjecture [3], which states that all D-brane configurations in type IIB can be realized as gauge field configurations on a stack of spacefilling branes. Distinct D-brane trajectories are identified if they are related by tachyon condensation. Here tachyon condensation is not interpreted as a physical process, as it relates different trajectories and not states on different time slices. Instead in concrete realizations, configurations before and after tachyon condensation are related by an RG flow. Thus K-theory classifies universality classes of the theories which are described by various vacua of the UV theory. While this physical interpretation of the K-theory classification conjecture is perhaps the most popular to date, its range of validity is limited. For example, spacefilling branes are inconsistent in backgrounds with nontrivial H flux [4]. In addition the spacetime is taken to be compact, but then in the case of nontorsion K-theory classes, which are the branes that continue to be stable in the classical limit, there is the usual problem that the sourced fluxes have nowhere to go and so no corresponding supergravity solution exists.

Instead of classifying D-brane trajectories, many authors use K-theory to classify D-brane charges at a fixed moment in time. The first were Moore and Witten [5], who used relative K-theory to classify the charges on a noncompact spatial slice. Diaconescu, Moore and Witten [6] found that some D-branes which wrap nontrivial homology cycles but carry no K-theory charge can decay. This violation of D-brane charge conservation is a result of the Freed-Witten anomaly, which we describe in Sec. 5. As we will review in Sec. 6, using this idea Maldacena, Moore and Seiberg [7] (MMS) were able to interpret the K-theory classification as a classification of stable D-branes, in other words, they argue that K-theory classifies the set of conserved RR charges. The naive group of conserved charges, the homology of a spatial slice, contains charges whose conservation is violated by the Freed-Witten anomaly. MMS were therefore able to use the Freed-Witten anomaly to test the K-theory classification in several cases, some of which will be reviewed in Sec. 7. This strategy has an advantage over the tachyon condensation strategy in that it can be extended to configurations with nontrivial H flux, where one finds twisted K-theory [8]. Ramond-Ramond fluxes are sourced by D-branes, and so Moore and Witten argued [5] that these fluxes should also be classified by twisted K-theory.

The above classification schemes suffer from two major shortcomings, which are the subject of section 8. The first is that in the democratic formulation of type II supergravity [9, 10, 11] the set of Ramond-Ramond field strengths is self-dual. More precisely, the p-form RR improved field strength is the Hodge dual of the $(10-p)$-form improved field strength. The Hodge duality operator, which is called the Hodge star, is a matrix which depends continuously on the metric and so its components are in general irrational. However according to the K-theory classification, the RR field strengths are Chern characters which are rational. Therefore the Hodge dual of a rational field strength component is an irrational number, which is in contradiction with the fact that RR field strengths should be rational. The solution to this problem is to choose half of the field strengths to be interpreted by K-theory, and to let the other half not be quantized. This is the way in which, for example, the chiral scalar field theory is defined. When the topology of each spatial slice of spacetime is time-independent there is a natural choice of a half of the fields to be classified by K-theory, one can choose the magnetic half. The cure to the second shortcoming is still unknown. We will see that all of the schemes relevant to type IIB superstrings suffer from a lack S-duality covariance, and so must at some level fail.

Twisted K-theory also classifies branes in some string theories besides type II. The twisted K-theory classification of D-branes in WZW models is already well established. Recently Moore and Segal [12] have also found a K-theory classification of branes in the target space of some simple 2-dimensional topological gauge theories, although the conjecture for A and B model topological string theory branes and pure spinors has not yet formally appeared.

Finally in section 9 we will describe two applications of the K-theory classification. First, we will see how it led to a formula for the topology of the T-dual of any spacetime. Next we will calculate the twisted K-theory of the Klebanov-Strassler geometry and find that the S-dual twisted K-classes correspond to universality classes of worldvolume gauge theories on D-branes.

2 Warmup: The Homology Classification

In these lectures we will review some of the most common K-theory classification schemes. We begin with an older classification scheme, the homology classification of D-branes, which is still widely used in the literature despite the fact that it includes some unphysical and some unstable branes. The failure of this scheme to reproduce the known gauge theory operators in the case of D-branes on $\mathrm{AdS}^5 \times \mathbb{RP}^5$ was observed by Witten in [13] and led to the discovery of the Freed-Witten anomaly [14], which is the basis of the modern understanding of the K-theory classification scheme that will be discussed in Sec. 6.

Like the K-theory classification of D-branes, there are many variations of the homology classification scheme, some of which classify D-brane charges, some of which classify D-brane trajectories and some of which classify RR fluxes, which are the fields sourced by D-branes. For concreteness we will restrict our attention to the homology classification of D-brane charges. In particular, we will consider type II string theory on a spacetime whose topology is of the form $\mathbb{R} \times M$ where $\mathbb{R}$ is the time direction and M is a compact 9-manifold. Notice that this choice implies that the topology of each spatial slice M is time-independent, and so we are not allowing processes in which the topology of spacetime changes. However the metric of M is allowed to change, so the universe can be expanding for example. Thus eternal inflation may be allowed, but a big bang which starts with the universe at a point is not contained in our ansatz. A big bang would be allowed in a classification of D-brane trajectories, for example in the K-theoretic classification scheme presented in Sec. 4.

We will not consider S-branes, as the full spectrum of D-brane charges is carried by ordinary D-branes. It will instead suffice to consider Dp-branes that extend along the time direction and also wrap a p-dimensional submanifold of the 9-manifold M. Each Dp-brane will then carry a charge which depends on the cycle that is wrapped.

2.1 The Homotopy Classification

Following the basic strategy of [7], we are interested in classifying charges that satisfy two conditions. First, we impose that some Dp-brane that carries the charge be consistent. In the case of the homology classification it suffices to impose that the brane has no boundary, and so the wrapped p-submanifold has no boundary. p-submanifolds without boundaries are called p-cycles. Second, we impose that the charge is conserved. This implies, for example, that two p-cycles which are related by a small, continuous deformation should correspond to the same charge, since a D-brane wrapping one could move to the other and so while the number of branes wrapping either cycle individually is not conserved by this process the sum of these numbers is conserved. Thus Dp-brane charges appear to be classified by p-cycles where two p-cycles correspond to the same charge if one can be deformed to another. Such cycles are said to be homotopic, and this set of charges is called the set of homotopy classes of p-cycles.

In fact we know that the charges corresponding to homotopy classes of p-cycles are not always conserved. Consider a simplified case in which M is the product of a 7-manifold and a

2-manifold. Let the 2-manifold be the Riemann surface Σ_2 of genus 2 and ignore the 7-manifold. We can wrap a D1-brane around any loop on the Riemann surface. In particular, we can wrap a D1-brane around the red loop in the center of Figure 6.5. The homotopy classes of loops on the genus 2 Riemann surface are elements of a nonabelian group called the fundamental group of the surface. The fundamental group is generated by four elements A_1, A_2, B_1 and B_2 which satisfy a single relation: the product $A_1B_1A_1^{-1}B_1^{-1}A_2B_2A_2^{-1}B_2^{-1}$ is equal to the identity element

$$\pi_1(\Sigma_2) =< A_1, A_2, B_1, B_2 | A_1B_1A_1^{-1}B_1^{-1}A_2B_2A_2^{-1}B_2^{-1} = 1 > . \tag{2.1}$$

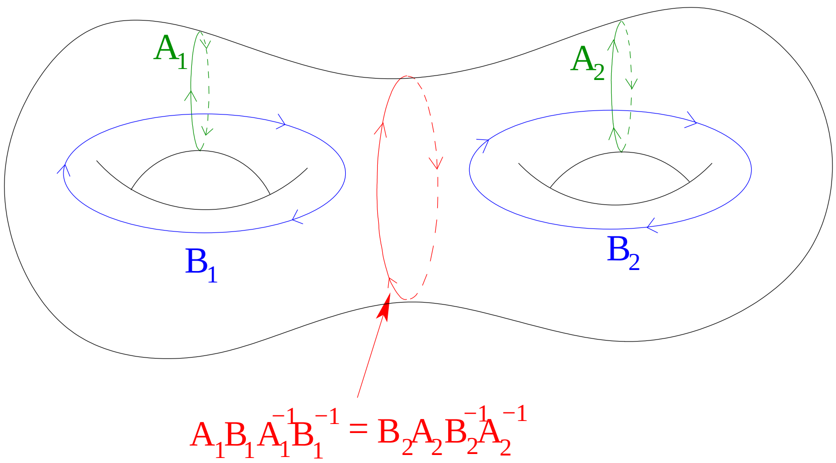


Figure 6.5: Loops on the genus 2 Riemann surface. A D1-brane wrapping the red loop in the middle can decay by moving steadily to the right, but it will need to split into two loops wrapping A_2 with opposite orientations to pass the handle, which then rejoin once the handle has been passed. After the rejoining the brane wraps a contractible cycle and it can shrink into nothingness. The nontrivial homotopy charge of the brane is not conserved by this process.

The loop in Figure 6.5 represents the nontrivial element $A_1B_1A_1^{-1}B_1^{-1}$ of the fundamental group. Nonetheless, a brane wrapping this loop does not carry a conserved charge because it can decay via the following process. At each moment in time, the brane can move a little to the right. Once it reaches the handle it will need to divide into two loops, one wrapping A_2 with each orientation. Then, further into the future, once they have passed the handle completely, the two loops will coalesce and the brane will wrap a contractible loop on the right hand side. This loop will then shrink into nothingness and the brane will decay.

Thus branes wrapping some nontrivial homotopy classes do not carry conserved charges because the branes can decay after some processes in which they change their topology, like the splitting of one loop into two which then coalesce that occurs in this example. To obtain a classification of conserved charges, one must identify the unstable brane charges with the zero element of the charge group. Unstable branes, at least Dp-branes that can decay via the above process, are those which wrap p-cycles that are boundaries of $(p+1)$-dimensional submanifolds of M. The $(p+1)$-dimensional submanifold is the space swept out by the decaying brane, for example the right hand side of the Riemann surface M is a 2-dimensional submanifold of M. The group of p-cycles modulo cycles which are boundaries of $(p+1)$-submanifolds is called the pth homology group of M, and is often denoted $\mathrm{H}_p(M)$.

2.2 The Homology Classification

In the above example, D1-branes are classified by the first homology group of the genus 2 Riemann surface, which is

$$\mathrm{H}_1(\Sigma_2) = \mathbb{Z}^4 \tag{2.2}$$

the additive group of quadruples of integers. This is an abelian group which is generated by the same generators A_1, A_2, B_1 and B_2 as the homotopy group. In the abelian group, the element $A_1B_1A_1^{-1}B_1^{-1}$ is trivial, as the A_1 and B_1 commute and so the A_1 can be moved one space to the right where it annihilates its inverse. Thus the D1-brane wrapping the cycle in Figure 6.5

represents the trivial element of $H_1(M)$ and so carries no conserved homology charge, which is consistent with the fact that it can decay.

On any manifold, homology groups are finitely generated abelian groups. This means that they can always be written as the sum of copies of the integers and copies of the integers modulo powers of a prime number

$$H_p(M) = \mathbb{Z}^{b_p} \oplus_i \mathbb{Z}_{p_i^{k_i}}. \tag{2.3}$$

The number b_p of powers of the integers that appears in the pth homology class is called the pth Betti number. The second term, which is a sum of finite order cyclic groups, is said to be the torsion part of the homology group, while the $\mathbb{Z}^{b_p}$ term is called the free part. In the case of the genus 2 Riemann surface we have seen that the 1st Betti number is equal to 4, meaning that D1-branes on this surface can carry four distinct kinds of conserved charges, all of which are quantized.

None of the homology groups of a Riemann surface contains a torsion part. One example of a manifold with a torsion part is the three-dimensional real projective space $\mathbb{RP}^3$, which is the quotient of the 3-sphere S^3 by the antipodal map, which generates a $\mathbb{Z}_2$ symmetry. The homology groups of $\mathbb{RP}^3$ are

$$H_0(\mathbb{RP}^3) = \mathbb{Z}, \quad H_1(\mathbb{RP}^3) = \mathbb{Z}_2, \quad H_2(\mathbb{RP}^3) = 0, \quad H_3(\mathbb{RP}^3) = \mathbb{Z} \tag{2.4}$$

where 0 is the trivial group which only contains the identity element. Now we can use the homology classification to classify conserved charges of D-branes wrapping a p-cycle in M where M is the product of $\mathbb{RP}^3$ and an irrelevant 6-manifold. As above we will restrict our attention to branes that are a single point on the 6-manifold.

The above homology classes (2.4) tell us that in type IIA, where there are only even-dimensional D-branes, the only available conserved charge is the D0-brane charge, which is an integer. The fact that the second homology group is the trivial group means that all D2-branes on $\mathbb{RP}^3$ can decay. On the other hand in type IIB there are odd-dimensional branes. D3-branes can wrap the whole of $\mathbb{RP}^3$ and these carry a conserved charge which measures the number of times that the $\mathbb{RP}^3$ is wrapped. The new ingredient in this example is $H_1 = \mathbb{Z}_2$, which contains the torsion term $\mathbb{Z}_2$. This means that D1-branes on $\mathbb{RP}^3$ carry a $\mathbb{Z}_2$ torsion-valued conserved charge. Physically, $\mathbb{RP}^3$ contains a single nontrivial loop which can be represented, for example, by a meridian that extended from the north pole to the south pole in the S^3 before it was quotiented by $\mathbb{Z}_2$. After the quotient the north and south poles are identified and so the meridian becomes a loop. A D1-brane wrapping this loop a single time is stable. However if it wraps the loop twice, the brane is the image of a brane that extends from the north pole to the south pole and back on the original S^3. Like any loop on a sphere, this loop can be contracted out of existence, and so the brane can decay. A brane that wraps the loop twice is the same as, or can turn into, two branes that wrap the loop once. Therefore a single brane wrapping the nontrivial loop in $\mathbb{RP}^3$ is stable, but if two come into contact then they can annihilate.

More generally, if the p-th homology group contains a $\mathbb{Z}_n$ factor then there is an interaction in which n coincident Dp-branes annihilate. Such $\mathbb{Z}_n$-charged objects are familiar in gauge theories, for example in $U(n)$ pure Yang-Mills there are $\mathbb{Z}_n$-charged strings, and in $\mathcal{N} = 1$ supersymmetric pure Yang-Mills there are $\mathbb{Z}_n$-charged Douglas-Shenker strings [15]. In examples these strings attach to each other and form bound states with a binding energy that increases for each string that is bound. When n strings are bound then the binding energy is equal to the total energy and all of the strings decay into radiation. For this reason torsion charged D-branes can never preserve any supersymmetry. More generally, degree n torsion cycles can never be calibrated because n times a calibrating cycle is a boundary and so is homologous to the empty set, and so if a cycle minimizes the integral of the calibrating form then n times the integral of the calibrating form is at most equal to the integral over the empty set which is equal to zero. However the homology classification indicates that D-branes wrapping torsion cycles may nonetheless be stable. Thus the homology classification is in a sense more powerful than a classification of conserved charges based on supersymmetry, because there exist conserved homology charges which are carried by objects which are stable but are never BPS.

2.3 Generalizations: Other Coefficients and Cohomology

The homology that we have discussed so far is a particular kind of homology, known as homology with integer coefficients or simply as integral homology. Representatives of integral homology classes are not precisely submanifolds for two reasons. First, they are sometimes singular, as we will discuss further in subsection 5.1. A homology class whose representatives are all singular is said to be nonrepresentable. On a manifold of dimension nine or less every homology cycle is representable, and so in the classification of conserved D-brane charges we can think of homology classes as nonsingular submanifolds. However in 10-dimensions there can be nonrepresentable 7-dimension homology classes, although some multiple of the cycle is always representable. Thus later when we classify D-brane trajectories, in particular the 7-dimensional surfaces swept out by D6-branes in IIA, and also when we look at WZW models, which can have arbitrarily large dimensions, this distinction will play a role. D-brane physics on nonrepresentable cycles is poorly understood, and we will see that the K-theory classification makes concrete predictions as to which singular submanifolds can be wrapped and which cannot.

The second difference between an integral homology class and a submanifold is that integral homology classes come with weights, which are integers. When the weights are positive they can be thought of intuitively as winding numbers of the submanifold around some cycle, but the weights can also be negative. These weights are the charges, thus negative weights can be thought of as winding numbers for anti D-branes. So a homology class contains a little more information than just the submanifold which is wrapped, it also contains integers which are the charges of the objects doing the wrapping.

In quantum gauge theories the charges are integers because of the Dirac quantization condition. However there are other physical theories in which the charges are not integers. For example, in the classical Maxwell theory electric charges are real numbers. Thus charges in classical electrodynamics are not of the form (2.3) and so are not classified by integral homology. More generally, in the classical supergravity limit of a superstring theory there is no Dirac quantization condition and so charges are not classified by integral homology.

There is another kind of homology, called real homology or homology with real coefficients, that does classify p-branes in supergravity. The term "real coefficients" means that real homology classes are submanifolds weighted by real numbers. In general one can define homology with coefficients valued in any abelian group G. When we want the choice of abelian group to be explicit, we will write

$$\mathrm{H}_p(M;G) \tag{2.5}$$

for the pth homology of M with coefficients in G. This is always the quotient of the kernel of the boundary map with coefficients in G by its image with coefficients in G. Thus it will always be a quotient of a subgroup of some copies of G by another subgroup of some copies of G.

One can obtain real homology from integral homology by tensoring with the real numbers, which kills the torsion part but leaves the Betti numbers b_p invariant

$$\mathrm{H}_p(M;\mathbb{R}) = \mathrm{H}_p(M;\mathbb{Z}) \otimes \mathbb{R} = (\mathbb{Z}^{b_p} \oplus_i \mathbb{Z}_{p_i^{k_i}}) \otimes \mathbb{R} = \mathbb{R}^{b_p}. \tag{2.6}$$

Thus torsion charged D-branes are unstable in the classical limit. Physically their decay process is easy to understand. If a Dp-brane is of charge one under a group

$$\mathrm{H}_p(M;\mathbb{Z}) = \mathbb{Z}_n \tag{2.7}$$

then in the quantum theory it is absolutely stable, as it wraps a noncontractible cycle and it cannot split into smaller pieces as its charge, one, is defined to be the lowest charge compatible with the Dirac quantization condition.

However in the classical theory there is no Dirac quantization condition and so without even moving it can divide into n branes of charge $1/n$ that wrap the same cycle. These branes can then reattach in a different way, again without moving, to form one brane of charge $1/n$ that wraps the cycle n times. However the fact that the homology is $\mathbb{Z}_n$ means that a submanifold which wraps the generator n times is the trivial element of homology and so can be deformed, maybe after nucleating some other branes as we saw on the Riemann surface, into nothingness. Thus the long charge $1/n$ classical brane may decay. The fact that all p-branes are unstable in this background reflects the triviality of the real homology group

$$\mathrm{H}_p(M;\mathbb{R}) = \mathrm{H}_p(M;\mathbb{Z}) \otimes \mathbb{R} = \mathbb{Z}_n \otimes \mathbb{R} = 0 \tag{2.8}$$

where 0 is the group consisting of only the identity element, which corresponds to zero charge.

Similarly in the classical theory every p-brane can be considered to wrap a representable cycle. This is because for every nonrepresentable cycle there is some multiple N of the cycle which is representable. One can then divide the charge of the classical p-brane by N and declare that it wraps the cycle N times. A small deformation then makes the brane honestly wrap a nonsingular cycle.

In general the relationship between homology groups with different coefficients is not so simple as in Eq. (2.6). Homology groups with different coefficients are related by a mathematical structure which is called an exact sequence of abelian groups, or simply an exact sequence. An exact sequence consists of a series of abelian groups G_i indexed by an integer $i \in \mathbb{Z}$ with homomorphisms

$$f_i : G_i \longrightarrow G_{i+1} \tag{2.9}$$

such that the image of each homomorphism is the kernel of the next

$$\mathrm{Im}(f_i : G_i \longrightarrow G_{i+1}) = \mathrm{Ker}(f_{i+1} : G_{i+1} \longrightarrow G_{i+2}). \tag{2.10}$$

An exact sequence is usually represented by a list of groups separated by arrows with the functions over the arrows

$$... \xrightarrow{f_{i-1}} G_i \xrightarrow{f_i} G_{i+1} \xrightarrow{f_{i+1}} G_{i+2} \xrightarrow{f_{i+2}} .... \tag{2.11}$$

If all of the abelian groups are the trivial group 0 except for three consecutive groups

$$0 \xrightarrow{0} G_0 \xrightarrow{f_0} G_1 \xrightarrow{f_1} G_2 \xrightarrow{0} 0 \tag{2.12}$$

then one calls the sequence a short exact sequence. In this case 0 is the zero map, whose image is 0. Exactness implies that f_0 is into and that f_1 is onto. Thus elements of G_1 are roughly pairs of elements from G_0 and G_2, but the additive structures are mixed. An example of a short exact sequence is

$$0 \xrightarrow{0} \mathbb{Z} \xrightarrow{\times 2\pi} \mathbb{R} \xrightarrow{exp} U(1) \xrightarrow{0} 0 \tag{2.13}$$

where $\times 2\pi : \mathbb{Z} \longrightarrow \mathbb{R}$ is multiplication by 2π and $exp : \mathbb{R} \longrightarrow U(1)$ is the exponential map $r \mapsto e^{ir}$. The kernel of the exponential map consists of all multiples of 2π, which is the image of the previous map and so the sequence is exact.

Short exact sequences will be useful for us because given any short exact sequence of abelian groups (2.12) one can construct a long exact sequence of homology groups which have those groups as coefficients. In what follows we will be interested less in homology groups than in cohomology groups, which are like homology groups but one replaces the boundary map with its transpose which is called the coboundary map. Weighted submanifolds which are closed under the coboundary are called cocycles while those in the image of the coboundary map are called cochains. The pth cohomology group H^p is then defined to be the quotient of the p-cocycles by the p-cochains. When the n-dimensional space M is orientable, as it always is in type II string theory compactifications, there is a theorem called Poincaré duality which states that the pth homology and $(n-p)$th cohomology groups are isomorphic

$$\mathrm{H}_p(M;G) = \mathrm{H}^{n-p}(M;G). \tag{2.14}$$

A short exact sequence of abelian groups also induces a long exact sequence of cohomology groups, which we will use in Subsec. 5.5 when discussing the Freed-Witten anomaly.

Like homology groups, cohomology groups also may be defined using any abelian group of coefficients. When one uses real coefficients the elements may be represented by differential forms and they generate a ring whose product is the wedge product $\wedge$ of differential forms.

In a quantized quantum theory instead one uses integral cohomology classes, which in general cannot be represented by differential forms of a single degree. Instead a degree p class can be represented by a collection of differential forms, one in each degree between 0 and p, with a series of gauge equivalence relations. In the mathematics literature this collection is called a Deligne cohomology class, named after a ULB graduate. Physically the collection of forms corresponds to the tower of ghosts of ghosts familiar in the quantization of p-form electrodynamics. Integral cohomology is also a ring and the product is known as the cup product $\cup$. However for simplicity in the sequel we will often abuse the notation and pretend that integral classes are represented by differential forms and write the wedge product $\wedge$ instead of the cup product.

3 Fluxes and Worldvolume Actions

3.1 The Failure of the Homology Classification

The homology classification of D-branes has met with a great deal of success. However in general it suffers from a number of shortcomings. First, some homology cycles cannot be represented by any smooth submanifold and so any brane carrying such a homology charge would necessarily have a singularity, even if the spacetime is nonsingular. D-branes wrapped on such cycles were first considered in Ref. [16] in a study of the Freed-Witten anomaly and later in Ref. [17] in the context of the K-theory classification. It was argued that sometimes, but not always, such wrappings are inconsistent in Ref. [18].

When M is 9-dimensional or less, as is the case in the classification of conserved charges in critical string theories considered above, all homology classes can be represented by smooth submanifolds. But it may still be that a brane wrapping such a submanifold is necessarily anomalous. Freed and Witten [14] have found a topological expression for this anomaly, which will be the subject of Sec. 5. Examples of nontrivial homology classes such that any brane wrapping any representative of the class will be anomalous have appeared in, for example, Ref. [7]. Thus homology is too big to be the class of conserved charges, because in general it contains charges that cannot be realized by any physical branes. Such classes need to be removed.

In addition to containing unrealizable charges, homology also contains charges that are not conserved. In Ref. [6] it was argued that sometimes a brane wrapping a nontrivial homology cycle can nonetheless decay. Such classes of unstable branes need to be quotiented out of homology to arrive at a classification of conserved charges. Thus the real group of conserved charges is equal to homology, minus the unphysical branes, quotiented by the unstable branes. In Section 6 we will review the argument of Ref. [7] that this prescription leads one precisely to twisted K-theory. In Section 8 we will then argue that yet more unphysical branes need to be removed and more unstable branes need to be quotiented, which leaves one with an as of yet unidentified mathematical structure.

One may now wonder why the homology classification has been so successful. Part of the reason is that, in the absence of the topologically nontrivial fluxes which lead to twisting, homology and K-theory and the unidentified structure are still unequal. However they can all be expressed as the sum of a free part, which is a sum of copies of the integers, and a torsion part, which is a sum of finite cyclic groups. Homology is so successful because the free parts of all of these structures are isomorphic, only the torsion parts differ. Recall that the charges of BPS states are always elements of the free part of the charge group, this is true not only for homology but for K-theory as well. Therefore in the absence of nontrivial fluxes the homology classification correctly identifies the free conserved charges, and so classifies all of the BPS states.

The first hint of the failure of the homology classification scheme came from the study of the charges of D-branes embedded in the worldvolumes of higher dimensional D-branes, which were calculated by demanding a cancellation of worldvolume anomalies on intersecting branes. To explain this development, we will begin with a short review of D-brane charges.

3.2 D-brane Charges and RR Fields

D-brane charges, like the charges of electrons in QED, measure the coupling of a D-brane to a gauge potential. Field configurations in QED are described by a 2-form field strength F, which in the absence of magnetic monopoles is closed. The closure of F ensures that locally one can introduce a one-form A, called the vector potential, such that F is the exterior derivative of A. The electric coupling e of an electron to the field F is defined by the topological term of the worldline action of an electron

$$S = e\int_\gamma A, \quad F = dA \tag{3.1}$$

where γ is the electron's worldline. This coupling leads to the celebrated Berry's phase in the electron's wave function $e^{iS/\hbar}$.

In the democratic formulation [9, 10, 11] of type II supergravity, Dp-branes couple not to two-form field strengths, but to $(p+2)$-form field strengths G_{p+2} which are called Ramond-Ramond (RR) field strengths or sometimes improved RR field strengths. In the case $p=0$ the

Dp-brane is a particle and, like an electron, it couples to a two-form G_2. When the NSNS 3-form field strength, which is called the H flux, vanishes, the RR field strengths are closed. Again this implies that locally one can introduce a gauge potential, this time a $(p+1)$-form C_{p+1}, whose exterior derivative is the field strength. One can then define the D-brane's coupling to the RR field strength to be the coefficient k in the following term in its worldvolume action

$$S \supset k \int_N C_{p+1}, \quad G_{p+2} = dC_{p+1} \tag{3.2}$$

where N is the $(p+1)$-dimensional worldvolume of the Dp-brane.

While only the above term calculates the Dp-brane's RR charge k, there are three physical effects in type II string theories with no analogue in QED which add extra couplings to the worldvolume action beyond those in Eq. (3.2). There may be a nontrivial H field, D-branes support gauge theories whose gauge fields couple to the RR fields and also D-branes couple to gravity which couples to both the RR fields and the gauge fields. We will consider each of these corrections in turn.

An H flux is a 3-form NSNS field strength H. In the absence of NSNS magnetic monopoles, which are called NS5-branes, H is closed $dH = 0$. In the presence of a nontrivial H flux, the RR field strengths are no longer closed, but instead they satisfy a twisted Bianchi identity

$$dG_{p+2} + H \wedge G_p = 0. \tag{3.3}$$

Notice that the twisted Bianchi identity mixes RR field strengths which are differential forms of different degrees. It will be convenient to combine all of the RR fields of different degrees into a single polyform. In type IIA supergravity, the classical limit of IIA string theory, all of the RR field strengths are of even degree, whereas they are all of odd degree in type IIB. Therefore in type IIA or IIB we can add all of the RR field strengths together in one even or odd polyform G, which satisfies the Bianchi identity

$$(d + H\wedge)G = 0. \tag{3.4}$$

As in QED, one may introduce magnetic monopoles that are defined to be sources for violations of the Bianchi identity, in type II supergravity these sources are p-branes.

The inclusion of an H flux had the effect of changing the operator in the Bianchi identity from the exterior derivative d to the twisted exterior derivative $(d + H\wedge)$. The closedness of H guarantees that the twisted exterior derivative is still nilpotent. To see this, note that for any polyform ω

$$\begin{aligned}(d + H\wedge)(d + H\wedge)\omega &= d^2\omega + d(H\wedge\omega) + H\wedge d\omega + H\wedge H\wedge\omega \\ &= 0 + (dH)\wedge\omega - H\wedge d\omega + H\wedge d\omega + 0 = 0 + 0 + 0 = 0\end{aligned} \tag{3.5}$$

where $d^2\omega$ vanishes because d is nilpotent, dH vanishes because H is closed and $H \wedge H$ vanishes because H is an odd form. The nilpotence of $(d + H\wedge)$ means that any $(d + H\wedge)$-closed form is locally $(d + H\wedge)$-exact. In other words, we can introduce a polyform C on each open patch such that

$$G = (d + H\wedge)C. \tag{3.6}$$

The terms of C are the RR gauge potentials C_{p+1}, which are odd differential forms in IIA and even forms in IIB. It is these gauge potentials which define the Dp-brane charge Eq. (3.2) in the presence of H flux.

The condition Eq. (3.6) does not define the gauge potentials C uniquely. Instead, analogously to the condition $F = dA$ in QED, the gauge potentials are only determined up to gauge transformations. In this case one may add any $(d + H\wedge)$-closed polyform Φ to C and obtain another solution to (3.6)

$$C \longrightarrow C + \Phi, \quad (d + H\wedge)\Phi = 0. \tag{3.7}$$

This means that the D-brane worldvolume action is in turn not well-defined. Fortunately, the action of a configuration is not an observable, the quantity which needs to be well-defined is the path integral measure e^{iS}, which implies that the action must be well-defined up to a shift of an integral multiple of 2π.

The fact that C is not uniquely defined is not so surprising. After all, the closedness of G only implied that C exists on contractible patches. Thus we may consider different solutions C of (3.6) to correspond to values of C on different contractible patches. Consider now a $(p+1)$-dimensional worldvolume N that lies in a single contractible patch. Then inside the patch there exists a $(p+2)$-manifold X whose boundary is N. We may then use Stoke's theorem to recast (3.2) as an integral on X

$$S \supset k \int_X dC_{p+1}. \tag{3.8}$$

The polyform dC is sometimes called the unimproved field strength.

When the H flux is nontrivial it may not seem like this expression is much of an improvement on (3.2), because dC_{p+1} is still not gauge-invariant. Using Eq. (3.7) we see that it is subject to the gauge transformations

$$dC_{p+1} \longrightarrow dC_{p+1} + d\Phi_p = dC_{p+1} - H \wedge \Phi_{p-2} \tag{3.9}$$

where we have used the $(d + H\wedge)$ closure of Φ. Notice that the shift in dC_{p+1} is a closed form because

$$-d(H \wedge \Phi_{p-2}) = H \wedge d\Phi_{p-2} = -H \wedge H \wedge \Phi_{p-4} = 0 \wedge \Phi_{p-4} = 0 \tag{3.10}$$

so after the gauge transformation (3.7) dC_{p+1} remains closed.

3.3 The Dirac Quantization Condition

Now we are ready to ask whether the path integral measure e^{iS} is really invariant under the gauge transformations (3.7). The answer is no. Even if all of spacetime is contractible, so that there is only a single patch and even if $H = 0$ then if the integral of Φ over N is not a multiple of 2π it will change the phase of the measure. This phase generalizes the Wilson loop in QED. It is an observable in the sense that two different trajectories can interfere with each other, and the total wave function depends on the relative phases of the wave functions at coincident points. This relative phase is the Berry's phase, and it is a new observable in the quantum theory that is not captured by G and so it does not exist in the classical supergravity. Thus the path integral measure need not be entirely determined by G as there are other observables available.

However all of the observables are invariant under integer shifts in the integral of $k\Phi$ over a closed surface, which are called large gauge transformations. Thus we still need to check that e^{iS} is invariant under large gauge transformations. Large gauge transformations may be transition functions between patches and so it remains to check that the measure is the same when calculated on two different patches, in other words, it must be independent of the choice of X.

To this end, let us choose a different manifold, Y, such that the boundary of Y is again N, but this time with the opposite orientation. In particular let us assume that the union of X and Y is a smooth manifold without boundary. Then we can use Stoke's theorem to rewrite the action (3.2) as

$$S \supset -k \int_Y dC_{p+1} \tag{3.11}$$

and demand that this expression for S give the same measure as (3.8). The measures e^{iS} will be equal if the actions differ by 2π times an integer. Therefore we demand

$$k \int_X dC_{p+1} - (-k \int_Y dC_{p+1}) = k \int_{X\cup Y} dC_{p+1} \in 2\pi\mathbb{Z}. \tag{3.12}$$

$X \cup Y$ can be any closed submanifold, and so we need to demand that the product of the D-brane charge k and the flux dC_{p+1} over any closed manifold be integral.

One might worry that this condition is ill-defined, because dC_{p+1} is only defined patchwise, as it enjoys the gauge transformations (3.9). One would be right. Fortunately branes may only wrap those cycles on which the integral of H is zero, so that H is exact

$$H = dB. \tag{3.13}$$

On these cycles the transition functions are topologically trivial and so one may simultaneously define dC on X and Y with no transition function and thus render (3.12) well-defined. Now that the term (3.12) is well-defined, we need to understand when it is integral.

There is only one way to ensure that $k\int dC_{p+1}/2\pi$ is always integral, one needs to quantize the RR charge k. Let us define our units such that the smallest quantity of charge allowed is a single unit $k=1$. Then the well-definedness of the path integral measure implies that

$$\frac{1}{2\pi}\int dC_{p+1}\in\mathbb{Z} \tag{3.14}$$

integrated over any cycle. This is the Dirac quantization condition for RR fluxes, and we have seen that we also need to quantize the RR charges k carried by D-branes. Roughly speaking the quantization condition implies that RR field strengths, in particular the unimproved RR field strengths dC, are classified by integral cohomology. However one needs to bear in mind the caveat that we have only been able to define dC on those cycles on which the pullback of H is exact. This is a major shortcoming of the cohomology classification of RR fields, but is automatic in the K-theoretic interpretation.

The quantization condition was already implied in the previous section when we classified D-brane charges by finitely generated abelian groups. In classical supergravity, where there is no quantization condition, RR charges instead are classified by n-tuples of real numbers called real homology classes. In the quantum theory instead only integral classes are allowed. However this does not mean that for each real-valued charge in the classical theory there is a single integer-valued charge in the quantum theory. In fact, we have seen that the integral homology classes that classify D-branes also contain extra torsion charges, which were nontrivial in the example of string theory on $\mathbb{RP}^3$.

3.4 Gauge and Gravitational Couplings

Finally we are ready to discuss the worldvolume gauge bundle, which is often called the Chan-Paton gauge bundle. A D-brane of charge k, which may alternately be thought of as a stack of k coincident D-branes, is a place where open strings can end. Quantizing these open strings one finds, among other massless modes, gluons that transform under a $U(k)$ gauge symmetry. Notice that the quantization of D-brane charge in the quantum theory is critical, as the rank of the gauge group must be integral. Configurations of the gluon field in a $U(k)$ gauge theory describe a geometrical object which is called a $U(k)$ gauge bundle E. The gauge bundle E is a vector bundle which consists of a copy of the complex vector space $\mathbb{C}^k$ at each point on the worldvolume N of a stack of k Dp-branes.

While a Dp-brane couples to the $(p+1)$-form RR gauge potential C_{p+1}, the gauge fields on the D-brane's worldvolume couple to the lower dimensional RR forms. Given any gauge field configuration one can determine a gauge bundle E, and the information of the bundle is partially characterized by an even-dimensional polyform called the total Chern character $ch(E)$. Different degree forms in the Chern character have different physical interpretations. For example, the 0-form part, which is called the 0th Chern character $ch_0(E)$ is a constant integer which is just the rank of the gauge group k. The next lowest dimensional component, the 2-form which is called the 1st Chern character $ch_1(E)$ is the trace of the gauge field strength F. It measures, for example, the flux of a magnetic vortex in the gauge theory. The next component is a 4-form called the 2nd Chern character $ch_2(E)$, which is the trace of the gauge field strength squared. It roughly measures the instanton number of a gauge field configuration.

The coupling to the RR fields is now easy to describe, the jth Chern character is a $2j$-form characterizing the gauge field configuration and it couples to the RR gauge connection C_{p+1-2j} via

$$S\supset\int_N ch(E)\wedge C\supset\int_N\left(kC_{p+1}+\frac{\mathbf{Tr}(F)}{2\pi}\wedge C_{p-1}+\frac{\mathbf{Tr}(F\wedge F)}{8\pi^2}\wedge C_{p-3}\right). \tag{3.15}$$

Notice that the first term on the right hand side, the $ch_0(E)=Tr(1)=k$ term, is the term (3.2) which measures the Dp-brane charge k, which is defined to be the strength of the coupling to C_{p+1}. The second term $F\wedge C_{p-1}$ instead is a coupling to C_{p-1} so it measures D$(p-2)$-brane charge. Thus we conclude that a Dp-brane may carry D$(p-2)$ brane charge, and that this charge is equal to the trace of its gauge field strength divided by 2π. Intuitively, magnetic flux tubes in a Dp-brane carry D$(p-2)$-brane charge. In fact if one dissolves a parallel D$(p-2)$-brane inside of a Dp-brane one finds that the gauge field configuration of the worldvolume theory on the Dp-brane changes, a magnetic flux tube appears where the D$(p-2)$ was, with a flux equal to the D$(p-2)$-brane's RR charge. Similarly D$(p-4)$-branes carry instanton charge in the Dp-brane's worldvolume theory.

A simple dimension-counting argument using the gauge theory kinetic term $F \wedge \star F$ shows that while one gains energy by smearing out a D$(p-2)$ in a Dp, and the size of a D$(p-4)$ is a modulus, it costs energy to dissolve a yet lower dimensional brane and so in general these tend to be ejected by the dynamics. However in this review we will be interested in a classification of conserved charges and rarely discuss their dynamics. The total charge of two D-branes is independent of whether they are separated or whether one is dissolved in the other, describing some nontrivial gauge configuration.

There are yet more couplings in the D-brane worldvolume action. Usually one includes the coupling of NSNS fields. Recall that the 3-form H field is everywhere closed, and on a D-brane worldvolume it is exact. Therefore on a worldvolume it can be written as the exterior derivative of a 2-form which is called the B field. The B field, like the other gauge potentials in string theory, is not quantized. One often includes the factor e^B in the worldvolume action. This may seem disastrous, as its couplings appear to lead to nonquantized lower dimensional D-brane charges [19]. However these charges, or at least the nonintegral parts, are canceled by contributions from the action of the bulk supergravity [20] and so we will neglect them for now. They will play a role later when we note that, in the case of a single D-brane $k=1$, the action is invariant under gauge transformations that leave $B+F$ constant even if F is not constant.

The final couplings that we need to consider are those of the D-brane to gravity. Gravity couples similarly to gauge fields. Instead of using a gauge field configuration to determine a gauge bundle, one uses the configuration of gravitons, which determines the topology of the spacetime itself, to determine another vector bundle which is called the tangent bundle **TM**. This vector bundle is the direct sum of two subbundles, the tangent bundle to the brane **TN** and the normal bundle to the brane inside of M, which we will denote **NN**. Unlike the $U(k)$ gauge bundle, whose fibers were complex k-dimensional vector spaces, the fibers of the tangent bundle are 10-dimensional real vector spaces $\mathbb{R}^{10}$ while those of **TN** and **NN** are $\mathbb{R}^{p+1}$ and $\mathbb{R}^{9-p}$ respectively. One rarely considers the Chern characters of a real vector bundle, because the odd Chern characters vanish being traces of odd numbers of the antisymmetric matrices that generate the Lie algebra of SO(N). Thus instead of defining a form of each even degree $2k$, real vector bundles determine a form in each degree of the form $4k$. One common basis of these forms, which plays the role played by Chern characters in the case of complex vector bundles, is the set of Pontrjagin classes p_k which are traces of degree $2k$ homogeneous polynomials in the curvature tensor. Just as the Chern character differential forms can be assembled into distinct polyforms, the total Chern class and the total Chern character, by weighting them differently, the Pontrjagin classes can be assembled into a number of distinct polyforms. Gravitational anomaly cancellation then determines how these polyforms couple to RR fields and to the worldvolume gauge field.

It turns out that this coupling is most simply expressed in terms of a polyform called the A-roof genus $\hat{A}$. The lowest terms in the A-roof genus are

$$\hat{A} = 1 - \frac{p_1}{24} + \frac{7p_1^2 - 4p_2}{5760} + ... \tag{3.16}$$

where the higher order terms are forms of degree at least 12, and so will vanish in the 10-dimensional world of critical superstrings. In fact, crucially for the success of the K-theory interpretation, the RR fields and gauge fields will couple not to the A-roof genus but to its square root

$$\sqrt{\hat{A}} = 1 - \frac{p_1}{48} + \frac{9p_1^2 - 8p_2}{23040} + ... \tag{3.17}$$

whose wedge product with itself is the A-roof genus.

With these ingredients, the unique coupling of RR fields and gauge fields to gravity which renders the measure of the path integral of the chiral fermions at the intersection of two branes well-defined was calculated in [1], generalizing the result in [21] which applied when **NN** is the trivial bundle. At the level of differential forms they found

$$S \supset \int_N C \wedge ch(E) \wedge \frac{\sqrt{\hat{A}(\mathrm{TN})}}{\sqrt{\hat{A}(\mathrm{NN})}} \tag{3.18}$$

where division is the inverse of the wedge product. The integral of all forms of degree not equal to $p+1$ is defined to be zero. This is our final answer for the worldvolume action of a D-brane. In fact this is not the complete action, it is a collection of terms known as the Wess-Zumino terms, but it will suffice to determine the lower-dimensional D-brane charges as the other terms do not contain couplings to RR fields.

4 K-Theory from the Sen Conjecture

4.1 Charges and K-Theory's Inner Product

As we did with Eq. (3.15), one may analyze the action (3.18) and read the charges of the various D-branes off of the coefficients of each RR field. This gives an expression for each lower-dimensional D-brane charge in terms of the gauge field of the original D-brane and the topology of its embedding. However, to classify D-brane charges in all of spacetime one would like to discuss some object which lives not on the worldvolume of a particular brane like the gauge bundle E and the bundles **TN** and **NN**, but rather an object that lives in the bulk. In Ref. [1] the authors reexpress these lower D-brane charges in terms of the cohomology of the bulk spacetime, using the pushforward map $f_!$ which takes the worldvolume gauge bundle to a bundle on all of spacetime. Here f is an embedding of the Dp-brane's worldvolume N into the spacetime M.

They found that the charges with respect to all RR fields may be summarized by the following simple expression

$$Q = ch(f_! E)\sqrt{\hat{A}(\mathrm{TM})} \in \mathrm{H}^*(M). \tag{4.1}$$

E is a gauge bundle on the D-brane worldvolume N and so its Chern character $ch(\mathrm{E})$ is a polyform which represents a class in the de Rham cohomology of N. However $f_!$ E is a bundle on the entire spacetime M, and so its Chern character $ch(f_!\mathrm{E})$ represents a class in the cohomology of M. Also the Pontrjagin classes of the tangent bundle, which are the summands of the square root of the A-roof genus of **TM**, represent cohomology classes in M. Therefore the D-brane charge Q in (4.1) is an element of the cohomology of M.

To relate the charge (4.1) to the earlier worldvolume expression (3.18) Minasian and Moore used a version of the Grothendieck-Riemann-Roch theorem that was proven by Atiyah and Hirzebruch. This theorem states that the Chern character of the bundle $f_!$E multiplied by the A-roof genus of the normal bundle is the class in $\mathrm{H}^*(M)$ which is Poincaré dual, in M, to the pushforward f_* of the Poincaré dual in N of the element $ch(\mathrm{E})$ of $\mathrm{H}^*(N)$

$$ch(f_!\mathrm{E})\hat{A}(NN) = \mathrm{PD}|_M(f_*(\mathrm{PD}|_N(ch(\mathrm{E})))), \qquad f_* : \mathrm{H}_*(N) \longrightarrow \mathrm{H}_*(M). \tag{4.2}$$

These Poincaré dualities are necessary because, unlike cohomology classes, homology classes pushforward naturally. The $\hat{A}(NN)$ correction is the price one must pay for the unnatural pushforward. It kills the denominator of Eq. (3.18) and combines with the worldvolume tangent bundle's A-roof genus to form the spacetime tangent bundle's A-roof genus in Eq. (4.1).

It may seem that this argument demonstrates that D-branes are classified by the de Rham cohomology of M. However in [1], Minasian and Moore argued that instead (4.1) suggests that D-brane charges are more naturally characterized by the K-theory classes $f_!$E. Before explaining why this seemed natural to Minasian and Moore, we will digress to describe what K-theory is and why $f_!$E represents a K-theory class.

4.2 What is K-Theory?

Recall that a complex vector bundle E over M consists of a complex vector space $\mathbb{C}^k$ fibered over every point in M such that, on each contractible neighborhood $U \subset M$, the total space of the vector bundle is topologically $U \times \mathbb{C}^k$. These neighborhoods are then glued together via transition functions in $U(k)$. Given two vector bundles E and F of rank j and k over the same base M, there is an easy way to add them, called the direct sum E$\oplus$F. One simply takes the direct sum fiber by fiber of the vector spaces, so that the total space above U is $U \times \mathbb{C}^{j+k}$ and the transition function in $U(j+k)$ is block diagonal with the $j \times j$ and $k \times k$ blocks being the original transition functions in E and F.

While the direct sum provides an easy way to add vector bundles, subtraction is more difficult. If E and F are vector bundles, one can define their difference E-F to be the pair (E,F). Subtraction should in some sense be the inverse of addition, which motivates the following equivalence relation. If E, F and G are vector bundles then one identifies

$$(\mathrm{E}, \mathrm{F}) = (\mathrm{E}\oplus\mathrm{G}, \mathrm{F}\oplus\mathrm{G}). \tag{4.3}$$

This intuitively corresponds to subtraction because E-E= 0

$$\mathrm{E} - \mathrm{E} = (\mathrm{E}, \mathrm{E}) = (0 \oplus \mathrm{E}, 0 \oplus \mathrm{E}) = 0 - 0 = 0 \tag{4.4}$$

where 0 is the trivial bundle of rank 0. One can now define addition and subtraction for a pair of vector bundles

$$(\mathrm{E}, \mathrm{F}) + (\mathrm{E}', \mathrm{F}') = (\mathrm{E} \oplus \mathrm{E}', \mathrm{F} \oplus \mathrm{F}'), \quad (\mathrm{E}, \mathrm{F}) - (\mathrm{E}', \mathrm{F}') = (\mathrm{E} \oplus \mathrm{F}', \mathrm{F} \oplus \mathrm{E}'). \tag{4.5}$$

With this definition of addition and subtraction, the space of pairs of bundles (E,F) is a group. The inverse of any element is

$$-(\mathrm{E}, \mathrm{F}) = (\mathrm{F}, \mathrm{E}) \tag{4.6}$$

and the identity is (0,0). This group is called the *K-theory* of M, and is denoted $\mathrm{K}^0(M)$.

Now we can identify $f_!\mathrm{E}$ with an element of the K-theory of M. It is a complex vector bundle on M, so it can be identified with the pair $(f_!\mathrm{E}, 0)$. The Chern character of a bundle does not contain all of the information about a bundle, and so one may worry that assigning charges to K-group elements really uses more information than appears in the physical coupling. However the Chern character is the coupling to the differential form representing the RR fields. Moore and Witten have conjectured [5] that RR fields should also be classified by twisted K-theory, as we will review in Subsec. 6.5. Thus one may suspect that since the Chern character of $f_!E$ is a differential form approximation of a K-class, and since the RR fields are differential form approximations of a K-class, perhaps the coupling expressed in Ref. [1] in terms of cohomology is just an approximation of a direct coupling of the K-theory. In this interpretation, it appears that the charge Q only knows about the Chern character, and not about the whole K-theory class, because of the approximation used in the calculation of D-brane charges which is inherent in the use of cohomology classes. If one had a formulation which used only K-theoretic operations from the beginning perhaps the charge would depend upon all of the information in the K-theory class.

One may wonder about the presence of the A-roof genus term in the charge (4.1). This term is independent of the choice of gauge bundle. Minasian and Moore explain that this term has a natural interpretation in K-theory. It ensures that the assignment of a charge in cohomology to a K-class is an isometry. More explicitly, as we will review momentarily, there is a natural inner product in both cohomology and in K-theory, and the A-roof genus term ensures that Q, which maps a K-theory class to a cohomology class, preserves these inner products.

A natural inner product between any two elements $[\eta]$ and $[\omega]$ of de Rham cohomology is as follows. Choose two differential forms η and ω that represent the two classes, the choice of representative will not matter. The inner product of the two classes is just the integral of the wedge product of the forms over the spacetime

$$([\eta], [\omega]) = \int_M \eta \wedge \omega. \tag{4.7}$$

The inner product of two K-theory classes E and F is slightly more complicated. Given two vector bundles E and F of rank j and k we have seen that one may form their sum by taking the direct sum of each complex vector space fiber. We will be interested in a different operation. One may also form the product of E and F by taking the tensor product of each complex vector space fiber. The product is a new vector bundle E$\otimes$F whose fibers are jk dimensional. This product makes $K^0(M)$ into a ring. Consider a Dirac operator $\not{D}$ acting on sections of the bundle E⊗F, which intuitively are just jk-tuples of functions on each patch with transition functions between patches which are the same as the transition functions that define the bundle E⊗F. The Dirac operator is just a differential operator times a gamma matrix, so it takes vectors in $\mathbb{C}^{jk}$ to vectors in $\mathbb{C}^{jk}$ and sections of the $\mathbb{C}^{jk}$ bundle to other sections. In particular, it has a kernel which consists of all of the sections that it takes to the zero section, which is the section that consists of the origin of $\mathbb{C}^{jk}$ above each point in M. Also it has a cokernel, which is the kernel of its transpose.

While neither the dimension of the kernel nor the dimension of the cokernel is a topological invariant of the bundle E⊗F, the difference between these dimensions is an invariant. This difference is called the index of $\not{D}$ and its value is determined from the topology of the bundle via the Atiyah-Singer Index Theorem

$$\mathrm{ind}\not{D}_{E\otimes F}=\int_M ch(E)\wedge ch(F)\wedge\hat{A}(\mathrm{TM}). \tag{4.8}$$

Now we can define the K-theoretic inner product $(\mathrm{E},\mathrm{F})_K$ of the K-classes corresponding to the vector bundles E and F to be the index of the Dirac operator on their tensor product

$$(\mathrm{E},\mathrm{F})_K=\mathrm{ind}\not{D}_{\mathrm{E}\otimes\mathrm{F}}. \tag{4.9}$$

Minasian and Moore noticed that this K-theoretic inner product is precisely equal to the de Rham inner product of the corresponding D-brane charges Q. Concretely, given two D-branes with worldvolume gauge bundles E and F and embedding maps e and f of their worldvolumes into M, the de Rham inner product of their charges from Eq. (4.1) equals the K-theory inner product of their gauge bundles pushed forward onto the full spacetime M via the maps $e_!$ and $f_!$

$$\begin{aligned}(Q(\mathrm{E}),Q(\mathrm{F})) &= \left(ch(e_!(\mathrm{E}))\wedge\sqrt{\hat{A}(\mathrm{TM})},ch(f_!(\mathrm{F}))\wedge\sqrt{\hat{A}(\mathrm{TM})}\right) \\ &= \int_M ch(e_!(\mathrm{E}))\wedge ch(f_!(\mathrm{F}))\wedge\hat{A}(\mathrm{TM})=\mathrm{ind}\not{D}_{e_!\mathrm{E}\otimes f_!\mathrm{F}}=(e_!\mathrm{E},f_!\mathrm{F})_K.\end{aligned} \tag{4.10}$$

Therefore the square root of the A-roof genus in Eq. (4.1) guarantees that the charge map Q is an isometry of K-theory onto de Rham cohomology.

While this evidence for the K-theory classification of D-branes was quite circumstantial, resting on the compatibility of the inner products in the known cohomological expressions for D-brane charge and in a conjectured K-theory framework, it suggested that the older homological classification might be incorrect. Witten later demonstrated in [2] that in the case of type I string theory the K-theory classification, or more precisely a variation of K-theory using real bundles, successfully predicts the existence of nontrivial $SO(32)$ gauge configurations missed by the homological classification. In fact several months earlier he had already, in the context of the AdS/CFT correspondence, presented an example of a failure of the homological classification in type II with an orientifold 3-plane in [13]. Thus it was eventually irrefutable that the homology classification of branes was incorrect, it contains unphysical branes and also unstable branes, and it needed to be replaced.

However the K-theory description of Minasian and Moore was mysterious as it described D-brane configurations in terms of a gauge bundle on the entire spacetime. The original gauge bundle on a D-brane was easy to interpret physically. Open strings end on the D-brane, when quantized some of their modes correspond to gluons and nontrivial configurations of the gluon field are described by gauge bundles on the D-brane. But there were no gluons in the bulk, open strings must end on branes, and so the presence of a gauge field in the bulk was a mystery.

4.3 The Sen Conjecture

An interpretation for this apparent bulk gauge bundle that classifies all D-brane configurations came from a parallel development in string field theory. In quantum field theories, a particle with a negative mass squared, called a tachyon, is a sign of instability. These particles will be spontaneously created, reducing the energy of the system and changing the vacuum itself via a process called tachyon condensation. Sometimes, as in the case of the standard model's Brout-Englert-Higgs mechanism, this process will end when the system arrives in a new vacuum with no tachyonic excitations available. One example of such a process in string theory is the closed string tachyon condensation when spacetime is a discrete quotient of the complex plane [22] with a conical singularity, which decays to a stable vacuum in which spacetime is the full complex plane. Sometimes the field theory is hopelessly unstable and the decay will never end or will end once all of spacetime has been destroyed, as has been conjectured to occur in the wrong sign heterotic theory of Ref. [23]. More often, no one knows the end point of this process, as is the case for closed string tachyon condensation in critical bosonic string theory.

Consider a D-brane and an anti D-brane in type II string theory. Open strings may extend from one to the other, and quantizing these strings one finds a tachyonic mode. Tachyonic modes correspond to instabilities, and in this case Ashoke Sen argued in 1998 [3] that the condensation of the tachyon corresponds to a dynamical process in which the brane and antibrane annihilate. We have seen that Dp-branes can carry the charges of lower dimensional D-branes, encoded in the topology of their worldvolume gauge bundles. Therefore if the Dp and anti Dp carried topologically distinct gauge bundles, while the total Dp-brane RR charge is zero there is still a net RR charge of lower-dimensional D-branes. As net RR charge is conserved, this means that lower dimensional D-branes will be present after the annihilation of the Dp-branes is complete. In the case $p = 9$ the D9-brane and anti D9-brane fill all of space, and so one might suspect that any configuration of lower dimensional Dp-branes could be encoded in the gauge bundle on this pair.

This is not possible, as these branes each carry a single $U(1)$ gauge bundle and abelian gauge theories have, for example, no instantons and so there would be no D5-branes. However if one allows the gauge group to be large enough then there is no such obstruction. This led Sen to conjecture that any configuration of D-brane charges can be realized as a gauge field configuration on a stack of sufficiently many D9-branes and anti D9-branes. Then, after these D9-branes and anti D9-branes have decayed, or equivalently after the open string tachyons have condensed, the arbitrary configuration of lower dimensional Dp-branes will remain. For example if the D9-branes contained a magnetic flux tube with k units of magnetic flux then after the annihilation k D7-branes will remain. If instead the anti D9-branes carry this flux tube then k units of anti D7 charge will remain. If the D9 worldvolume contained an instanton then a D5-brane will remain wrapped around the same cycle as the instanton, and so on.

Now we can interpret the classification of D-brane charges (4.1) by bundles $f_!\mathrm{E}$ on the bulk, these are the gauge bundles on the D9-branes. We can also interpret the K-theory equivalence relation (4.3). Consider a brane anti-brane system that represents the element $(\mathbf{E}\oplus\mathbf{G},\mathbf{F}\oplus\mathbf{G})$ in K-theory. If a brane and antibrane with the same gauge bundle, which pushes forward to $\mathbf{G}$, annihilate, then the equivalence condition states that the conserved charges $(\mathbf{E},\mathbf{F})$ are conserved, as they must be. Thus Witten has argued [2] that Sen's conjecture implies that D-branes are classified by K-theory.

These vector bundles over all of space time are not conserved charges in the usual sense. Usually, when one speaks of the conserved charges of a configuration, one integrates a conserved current over a timeslice and then the conservation implies that the choice of timeslice was irrelevant. If instead the charge is different at two timeslices then a process has occurred which can destroy this charge. Here instead one is considering a D-brane charge on all of the spacetime, not just on a timeslice. The process of tachyon condensation is not being treated as a process relating configurations at two moments in time, but rather it relates two different trajectories on the entire spacetime. This is not to say that tachyon condensation is not a real dynamical process that can occur over time, but it is merely to say that in the viewpoint taken in this subsection, tachyon condensation is something which relates two entire trajectories.

Instead of a dynamical process which proceeds in the time direction, closed string tachyon condensation is treated in Ref. [22] as an RG flow that proceeds along an internal direction of configuration space. Similarly the deformations in Ref. [2] relating different representatives of the same K-theory class identify configurations related by deformations in the configuration space. Thus the K-theory of the entire spacetime classifies not charges that are invariant under time-dependent physical processes, but rather charges associated with entire trajectories which are invariant under deformations such as renormalization group flow.

While the K-theoretic classification of D-brane charges using the Sen conjecture was revolutionary, it was somewhat limited. First, it did not apply to type IIA string theory which has no spacefilling D-branes. Hořava later claimed [24] that one can make the same arguments in IIA string theory using unstable D9 sphalerons, and one arrives not at K^0 but at a similar object named K^1. Also Kapustin has argued [4] that D9-branes are inconsistent in the presence of H-flux, and so the K-theory classification does not apply in these cases. As we will discuss presently, in these cases one needs to use a variation of K-theory known as twisted K-theory. Finally, as was emphasized in the last two paragraphs, this classification scheme does not classify charges that are preserved under dynamical processes, but rather a kind of property of trajectories that is preserved under deformations. But the question of what structure classifies conserved RR charges was left open. These shortcomings were all addressed by Maldacena, Moore and Seiberg [7] using a new approach to understand K-theory based not on tachyon condensation, but based on the Freed-Witten anomaly. This approach will be the subject of Section 6.

5 The Freed-Witten Anomaly

5.1 How Are Homology and K-theory Different?

The above K-theory classification of D-brane charges in the absence of H flux is not equivalent to the homology classification. There are three differences:

- Some homology classes do not correspond to any K-theory class.
- Some K-theory classes correspond to multiple homology classes.
- While homology and K-theory are both abelian groups, their addition rules are not compatible.

Each of these differences has an interpretation in terms of D-brane physics, and in each case K-theory gets it right while homology gets it wrong. As we will argue shortly, the homology classes that do not correspond to K-theory classes correspond to inconsistent D-branes that lead to worldsheet global anomalies on the strings that end on them. The multiple homology classes that correspond to the same K-theory class correspond to distinct brane configurations such that processes exist which transform one configuration into another, so that the conserved charges corresponding to all of these configurations must be the same. The transformations of D-branes that create homology charge can be thought of as a violation of the conservation of a RR current due to the aforementioned anomalies [25]. Thus the first two differences are the consequences of the same anomalies.

The difference in the addition rules of homology and K-theory corresponds to the fact that a bound state of two D-branes should carry the sum of the conserved charges of the constituents. In particular the homology classification would imply that the sum of two Dp-branes is again a Dp-brane, whereas K-theory successfully predicts that a Dp-brane can, for example, carry a half unit of D$(p-2)$-brane charge so that two Dp-branes can combine to yield a D$(p-2)$. For brevity we will have little to say about this additive structure, but the literature is filled with examples. A particularly rich example involving D-branes of several dimensions is described in Ref. [26].

Despite all of these differences, homology and K-theory, in the absence of H flux, are remarkably similar. Classically they are identical. Mathematically one says that they are isomorphic when tensored with the real numbers

$$\mathrm{H}^{even}\otimes\mathbb{R}\cong\mathrm{K}^0\otimes\mathbb{R},\quad \mathrm{H}^{odd}\otimes\mathbb{R}\cong\mathrm{K}^1\otimes\mathbb{R} \tag{5.1}$$

which yields homology and K-theory with real coefficients. In physics we saw in Subsec. 2.3 that this tensoring eliminates the Dirac quantization condition and it is interpreted as a classical limit. One finds that branes in supergravity are classified by homology with real coefficients which is isomorphic to K-theory with real coefficients. Intuitively, the inconsistencies which lead some D-branes to be anomalous all disappear when there are enough D-branes, or more precisely when the number of D-branes has the right prime factors, as occurs in the classical limit.

There appear to be two things that can go wrong with a D-brane wrapped on a homology cycle that imply that the homology cycle does not correspond to any K-theory cycle. One is that the D-brane suffers from the Freed-Witten anomaly [14], which implies that worldsheets of open strings ending on the D-brane are afflicted with a global anomaly that does not allow one to globally define the phases of the measures of their path integrals. No D-branes with Freed-Witten anomalies carry K-theory charges, and we will argue that they are all inconsistent. If a D-brane wraps a nonsingular submanifold and is not Freed-Witten anomalous then it automatically carries a K-theory charge. This does not mean that all homology classes that can be wrapped by Freed-Witten anomaly-free branes correspond to K-classes, because some homology classes can not be represented by nonsingular submanifolds [27, 28]. Thus D-branes can fail to carry K-theory charge if they are Freed-Witten anomalous or if they wrap a homology cycle which cannot be represented by any nonsingular submanifold [18].

It may seem odd that there are homology classes that are not representable by any nonsingular submanifold. In fact, homology classes in a space M of dimension 9 or less can always be represented by nonsingular submanifolds, but precisely in dimension 10 there appear 7-dimensional homology cycles which cannot be represented by any nonsingular submanifold, meaning that the representability of cycles is only an issue for D6-branes in IIA string theory and generically for branes in noncritical string theories like WZW models. When one says that a class is not representable by a nonsingular submanifold, this means that it can still be represented by a subset, but that this subset is not a nonsingular manifold because it contains at least one point, called a singularity, where the tangent space is not of the form $\mathbb{R}^n$. For example, the real cone over the complex projective space $\mathbb{CP}^3$ has a singularity at the tip of the cone.

The singularities of these cycles are different from the singularities usually encountered in string theory because the ambient spacetime is nonsingular, it is only the wrapped cycle which is singular. Not only is the cycle singular, but every cycle in its homology class is singular and so there is no way that the singularity may be blown up or deformed away. The nonsingularity of the spacetime means that there are no strong gravitational effects that can change the geometry, the physics of the singularity is governed by open strings. However some nonrepresentable homology cycles nonetheless lift to K-theory, and so the corresponding branes carry K-theory charges. In Ref. [18] the authors gave an example of a nonrepresentable homology cycle that lifts to a K-theory class, and one that does not. If one believes the type IIA version of the Sen conjecture, and therefore that type IIA branes are classified by K-theory, then the first representable cycle can be consistently wrapped by a D6-brane and the second cannot. However, unlike the Freed-Witten case, the global anomalies of strings ending on these two D-branes have not yet been analyzed. Such an analysis would be a strong test of the K-theory classification hypothesis.

Summarizing, a D-brane wrapping a homology cycle is inconsistent if it suffers from a Freed-Witten anomaly, and is sometimes inconsistent if the homology cycle cannot be represented by any nonsingular submanifold. This brings us to the question, what is a Freed-Witten anomaly? To answer this question we will first need some mathematical background on *spin* structures and *spin*^{c} structures.

5.2 *Spin* Structures

The remainder of this section is technically more difficult than the rest of the lectures. The uninterested reader can skip to Sec. 6. To understand the following sections he will only need the conclusion that on a D-brane worldvolume there is an anomaly called the Freed-Witten anomaly which is equal to $W_3 + H$. Here W_3 is a class in the third integral cohomology of the worldvolume called the third Stiefel-Whitney class. It is determined by the topology of the D-brane's worldvolume. H is also an element of the third integral cohomology of the worldvolume, it is equal to the pullback of the spacetime NSNS field strength. A D-brane can only wrap cycles N such that

$$W_3 + H = 0 \in \mathrm{H}^3(N; \mathbb{Z}). \tag{5.2}$$

The Freed-Witten anomaly is an obstruction to the existence of certain representations of the Lorentz group in the spectrum of a theory on spacetimes with some topological properties. To understand it, we will first need to review the action of the Lorentz group on fields defined on a topologically nontrivial space.

The wavefunctions of fermions do not transform under any representation of the Lorentz group. To see this, consider a 360 degree rotation on your favorite plane. This is the identity rotation of the Lorentz group. Representations are, by definition, homomorphisms of groups to matrices which means among other things that they map the identity group element to the identity matrix. However a 360 degree rotation is the identity element of the Lorentz group but it changes the sign of the wavefunction of a fermion, and so is not represented by the identity matrix. If U is the map from the Lorentz group to the group action on the spinor ψ then

$$\psi = U(1)\psi \neq U(e^{2\pi i})\psi = -\psi \tag{5.3}$$

implies that U is not a representation. However U is a representation up to a phase, which in this case is a choice of sign. Maps like U that are representations up to a choice of phase are called projective representations, of which representations are examples in which the phase factor is one. Symmetries in quantum theories always act via projective representations.

In a topologically trivial spacetime one can always define an action on states via a projective representation. This is because a physical state does not correspond to a vector in a Hilbert space, but to a ray in a Hilbert space, which is the set consisting of a vector multiplied by any phase. In particular, the multiplication of a fermion wave function by minus one negates the corresponding vector in the Hilbert space, but it is still described by the same ray and so it still corresponds to the same physical state. Thus the action of the Lorentz group on rays, that is to say on physical states, is well-defined.

One might then be tempted to do away with the Hilbert space and just deal with the rays directly, but perhaps unfortunately the known formalisms for doing calculations in quantum field theory require a choice of a representative vector in each ray. In the case of the Lorentz group, this means that given any action of the Lorentz group, one must also determine a phase. A choice of an element of the Lorentz group plus a phase is the same as a choice of an element of the nontrivial central extension of the Lorentz group which is called the *Spin* group. An element of the *Spin* group is an element of the corresponding Lorentz group plus a $\mathbb{Z}_2$ choice of sign, which topologically means that the group manifold of a *Spin* group is a certain $\mathbb{Z}_2$ bundle over the group manifold of a Lorentz group $SO(N)$

$$\begin{array}{ccc} \mathbb{Z}_2 & \longrightarrow & Spin(N) \\ & & \downarrow \pi \\ & & SO(N) \end{array} \tag{5.4}$$

where π is the projection map from the *Spin* group to the Lorentz group.

The simplest example of the relationship between the Lorentz and Spin groups occurs in the 2-dimensional Euclidean space $\mathbb{R}^2$. Here the Lorentz group is $U(1)$, and the *Spin* group is another copy of $U(1)$ which is twice as big. π is the identification of antipodal points on the big $U(1)$. In particular a rotation by α in the Lorentz group is only a rotation by $\alpha/2$ in the *Spin* group. Vectors, like the gauge connection A in electrodynamics, are sections of the tangent bundle of 2-dimensional Euclidean space which is a trivial vector bundle whose fibers are $\mathbb{R}^2$ and transform in the 2-dimensional $SO(2)$ representation of the Lie algebra $u(1)$, which topologically means that the structure group of the bundle is the Lorentz group $SO(2)$. Spinors are also sections of a 2-dimensional trivial bundle, which again is subject to transition functions in $SO(2)$, which is the *Spin* group of $SO(2)$. These are the same transition functions as were used for the tangent bundle, but with a projective representation in which all rotations are divided by two. However, as the bundle is trivial, these transition functions can be taken to be trivial and so the tangent and spin bundles are equivalent in this case.

If instead one considers QED on a topologically nontrivial 2-dimensional Euclidean space Σ then the two bundles are different. Now the photon wavefunction is a section of the tangent bundle $\mathbf{T\Sigma}$ while the electron wavefunction is a section of the square root of the tangent bundle $\sqrt{\mathbf{T\Sigma}}$, which is called the *Spin* bundle. Both bundles may be constructed from local trivializations and transition functions. The transition functions that create both bundles are the same, but the transition functions act on the square root bundle using a different representation, in which the rotation angles are all halved. If one considers the fiber to be $\mathbb{C}$ and the transition functions to be 1-dimensional complex representations of $U(1)$ then the *Spin* bundle transition functions act on the *Spin* bundle via the projective representation

$$U(e^{i\alpha}) = e^{i\alpha/2} \tag{5.5}$$

where $e^{i\alpha}$ is the action on the tangent bundle.

Notice that the map (5.5) is not well-defined, as, given a transition function $e^{i\alpha}$ one may interpret α as $\alpha + 2\pi$ by choosing a different fundamental domain for the angles and one would find a different sign for the right hand side, which translates into a minus sign in the transition functions for the spinors. To define the spin bundle one needs to choose all of these signs in the transition functions. The choice of these signs is called the *spin structure*.

The choice of spin structure is a physical choice, it can determine, for example, whether a compactification is supersymmetric or not. The choice of spin structure on the worldsheet of a closed string determines whether one is describing NS or R modes, which in turn determines the spin of the spacetime field corresponding to an excitation of the worldsheet spinor. Sometimes different choices of spin structure are physically equivalent. Two choices are only inequivalent if, after a spinor moves around some closed loop, the two spin structures disagree on the spinor's sign. Thus there is a $\mathbb{Z}_2$ choice of *spin* structure for each inequivalent closed loop of the theory. For example, the fundamental group of the genus g Riemann surface Σ_g has $2g$ generators and its group of spin structures is

$$\mathrm{H}^1(\Sigma_g,\mathbb{Z}_2)=\mathbb{Z}_2^{2g}. \tag{5.6}$$

In this example one can see that a choice of *spin* structure is just a choice of whether a spinor will have periodic or antiperiodic boundary conditions on each of the $2g$ inequivalent loops.

We have seen that a choice of *spin* structure is just a choice of an element of $\mathbb{Z}_2$ at each transition function between patches. This information defines a $\mathbb{Z}_2$ bundle. Actually, not every choice of $\mathbb{Z}_2$ on overlaps gives a bundle. For example, consider three patches U_i, U_j and U_k which are discs and whose double and triple overlaps are all discs. When all possible intersections of a set of patches covering a space are topologically trivial, one says that the patches form a *good cover*. We first trivialize the *Spin* bundle on each patch. To define the *Spin* bundle globally we must define transition functions which are $+1$ or -1 on each of the three overlaps $U_i\cap U_j$, $U_j\cap U_k$ and $U_k\cap U_i$. Name these transition functions x_{ij}, x_{jk} and x_{ki} respectively. The $\mathbb{Z}_2$ bundle is only well-defined if these patches satisfy the triple overlap condition on the triple overlap $U_i\cap U_j\cap U_k$

$$x_{ij}x_{jk}x_{ki}=1. \tag{5.7}$$

In fact, the *spinor* is also only well-defined when we impose this condition. This is because, given the value of a spinor on the overlap written using the trivialization of U_i, one may switch to the U_j trivialization and then to the U_k trivialization and finally back to the U_i trivialization without ever moving, therefore the value of the spinor should not change. This means that we must impose the condition (5.7) at every triple overlap in our definition of *spin* structures. In the example (5.6) we have already done this.

Mathematically the group of choices of elements of a group G on 2-way overlaps of an atlas of M subject to the 3-way overlap condition (5.7) has a name. It is called the first Čech cohomology group with G coefficients, and is denoted

$$\check{\mathrm{H}}^1(M;G). \tag{5.8}$$

When one uses a good cover Čech cohomology is isomorphic to the cohomology $\mathrm{H}^*(M;G)$ that we have been using all along. From now on we will restrict our attention to good covers. Therefore the group of spin structures on M is $\mathrm{H}^1(M;\mathbb{Z}_2)$.

The classification of $\mathbb{Z}_2$ bundles by characteristic classes in the first cohomology group with $\mathbb{Z}_2$ coefficients is an example of a more general feature in the classification of bundles. As we will later be interested in a different example, $Spin^c$ bundles, we will now summarize the main features of this structure. Consider a bundle whose fiber F is $(p-1)$-connected, meaning that it $k<p$ then the kth homotopy group of F vanishes. Imagine that the pth homotopy group is nontrivial, and is equal to the group G

$$\pi_{k<p}(F)=0, \qquad \pi_p(F)=G. \tag{5.9}$$

Then F bundles will be characterized entirely by a degree $p+1$ characteristic class in the cohomology with G coefficients

$$\omega_{p+1}\in\mathrm{H}^{p+1}(M;G) \tag{5.10}$$

and also some characteristic classes of higher degree. If furthermore all of the homotopy classes of F of degree higher than p vanish then the bundles are entirely characterized by just ω_{p+1}.

In our case we are interested in $\mathbb{Z}_2$ bundles, so $p=0$ and $G=\mathbb{Z}_2$. The characteristic class $\omega_1\in\mathrm{H}^1(M;\mathbb{Z}_2)$ is the *spin* structure. We will also be interested in circle bundles, for which $p=1$ and $G=\mathbb{Z}$. Circle bundles are completely characterized by a single characteristic class

$$c_1=\omega_2\in\mathrm{H}^2(M;\mathbb{Z}) \tag{5.11}$$

which is called the first Chern class.

5.3 When is a Manifold *Spin*?

Beyond two dimensions a new complication will arise, due to the fact that we are not really classifying $\mathbb{Z}_2$ bundles over spacetime, but rather *Spin* bundles over spacetime which are $\mathbb{Z}_2$ bundles over Lorentz bundles over spacetime. $\mathbb{Z}_2$ bundles over any space always exist, for example the trivial bundle always exists. However, given a particular Lorentz bundle over spacetime, which is the principal bundle associated to the tangent bundle **TM**, it is not always possible to define a *Spin* bundle whose transition functions are those of the tangent bundle but in a *Spin* representation. The problem is that when the Lorentz bundle is nontrivial, sometimes no choice of signs for the *Spin* lift satisfies the triple overlap condition (5.7).

The triple overlaps are determined entirely by the topology of the tangent bundle, and so provide a characteristic class of the tangent bundle. While transition functions of our $\mathbb{Z}_2$ bundle are choices of elements of $\mathbb{Z}_2$ on each 2-way overlap and define elements of the first Čech cohomology group $\check{H}^1(M;\mathbb{Z}_2)$, the obstruction to (5.7) is a choice of element $x_{ij}x_{jk}x_{ki}$ in $\mathbb{Z}_2$ on each 3-way overlap and defines an element of the second Čech cohomology group $\check{H}^2(M;\mathbb{Z}_2)$.

This is part of the above general structure. F bundles are classified by elements ω_{p+1} of $\mathrm{H}^{p+1}(M;G)$ but if G is the extension of a group H by F then, given an H bundle, one can define another characteristic class

$$\omega_{p+2} \in \mathrm{H}^{p+2}(M;G) \tag{5.12}$$

which is the obstruction to lifting the H bundle P over M to a G bundle P' over M

$$\begin{array}{ccc} F & \rightarrow & G \\ & & \downarrow \pi \\ & & H \end{array}, \quad \begin{array}{ccc} H & \rightarrow & P \\ & & \downarrow p \\ & & M \end{array} \quad \overset{\text{Lift}}{\Longrightarrow} \quad \begin{array}{ccc} F & \rightarrow & P' \\ & & \downarrow \pi \\ & & P \end{array}, \quad \begin{array}{ccc} G & \rightarrow & P' \\ & & \downarrow p' \\ & & M \end{array}. \tag{5.13}$$

The H bundle P can be lifted to a G bundle P' only when $\omega_{p+2}=0$. In our case F is $\mathbb{Z}_2$, H is the Lorentz group, G is the *Spin* group and ω_{p+2} is called the second Stiefel-Whitney class and is denoted w_2. The bundles P and P' are the tangent and *Spin* bundles respectively, or more precisely the associated principle bundles. When w_2 is nonzero the triple overlap condition is never satisfied and spinors cannot be globally defined. If

$$w_2(\mathbf{TM}) = 0 \tag{5.14}$$

then the spacetime M is said to be *spin*. More generally w_2 can be defined for any SO bundle P, not just the tangent bundle, and it always measures the obstruction to lifting the SO bundle P to a *Spin* bundle P'.

This discussion has been somewhat abstract. We will now describe a concrete example of the obstruction to the lift of an SO bundle to a *Spin* bundle. We will consider a higher rank bundle, although already in two dimensions the Hopf bundle has no *spin* lift. In general the topologies of the Lorentz group and the corresponding *Spin* group are different. The simplest case in which the topologies are different is the 3-dimensional Euclidean space $\mathbb{R}^3$. Now the Lorentz group is $SO(3)$ and its *spin* cover is $Spin(3) \cong SU(2)$. If spacetime is a Euclidean 3-manifold M, then vector fields like the photon are sections of the tangent bundle **TM** and spinor fields are sections of its square root, which is the spin bundle $\sqrt{\mathbf{TM}}$. Again, both of these bundles are built using the same trivializations and transition functions, but while one uses the 3 real dimensional $SO(3)$ representation of the Lie algebra $su(2)$ in the definition of the action of the transition functions of the rank 3 tangent bundle **TM**, one uses the 2 complex dimensional representation $SU(2)$ for the transition functions of the rank 2 spin bundle $\sqrt{\mathbf{TM}}$.

We now ask the question, given an $SO(3)$ bundle P on a manifold M, is it always possible to lift P to an $SU(2)$ bundle P' on M? Recall that the group manifold $SU(2)$ is a $\mathbb{Z}_2$ bundle over the group manifold $SO(3)$ and so there is a projection map

$$\pi : SU(2) \longrightarrow SO(3) \tag{5.15}$$

with kernel $\mathbb{Z}_2$ that maps $SU(2)$ to $SO(3)$. The $SU(2)$ bundle P' is then said to be the lift of the $SO(3)$ bundle P if π acting fiberwise on P' takes it to P. If P is the tangent bundle **TM** then P' exists if and only if M is *spin*. All manifolds of dimensional less than 4 are *spin*, an example of a 4-dimensional manifold that is not *spin* is the complex projective space $\mathbb{CP}^2$. However manifolds of dimension four or more are difficult to picture, so we will consider a 2-dimensional example.

As all 2-dimensional manifolds are *spin*, the tangent bundle of a 2-manifold always lifts to a *Spin* bundle. Thus to find an example of a bundle that does not lift, we will have to consider a bundle which is not the tangent bundle.

When discussing the worldvolume gauge theories on D-branes one is often interested not in the tangent bundle but in the normal bundle **NN** of the submanifold $N \subset M$ wrapped by a D-brane. This is because worldvolume scalars are often sections of the normal bundle and worldvolume fermions are often sections of the *spin* lift of the normal bundle.

When classifying topologically nontrivial configurations in gauge theories one is often interested instead in the topology of the gauge bundle itself. For example, consider an $SU(2)$ gauge theory with no matter or with only adjoint matter. Gluons always transform in the adjoint representation of the gauge group, and by hypothesis any matter in this theory also transforms in the adjoint. The only generators of the gauge group that appear in this theory are then those of the adjoint representation, which are three-dimensional and generate the Lie group $SO(3)$. Therefore this theory really only exhibits an $SO(3)$ gauge symmetry. The difference between $SU(2)$ and $SO(3)$ is that $SU(2)$ contains an element which is minus the identity, which gets identified with the zero element in the quotient π to $SO(3)$. However conjugating any field by minus the identity leaves the field invariant and so this element acts trivially on our field configurations, only $SO(3)$ can act faithfully.

The fact that this gauge theory is really an $SO(3)$ gauge theory means that gluon configurations are classified by $SO(3)$ bundles and not by $SU(2)$ bundles. This is a physically important distinction because there are more $SO(3)$ bundles than there are $SU(2)$ bundles. This is true in part because only $SU(2)$ is simply-connected

$$\pi_1(SU(2)) = 0, \qquad \pi_1(SO(3)) = \mathbb{Z}_2. \tag{5.16}$$

The nontriviality of the fundamental group of $SO(3)$ implies that, unlike $SU(2)$ gauge theories, $SO(3)$ gauge theories admit $\mathbb{Z}_2$-charged magnetic monopoles and Dirac strings, which in the Higgs phase become physical $\mathbb{Z}_2$ vortices.

For example, consider an $SO(3)$ gauge theory on S^2. We can trivialize the $SO(3)$ gauge bundle on the northern and southern hemispheres. The bundle is then topologically classified by the transition function from the equator to $SO(3)$, which is a map in $\pi_1(SO(3)) = \mathbb{Z}_2$. The nontrivial map leads to the $\mathbb{Z}_2$-charged vortex. If two of these vortices collide they may annihilate, but a lone vortex is topologically stable. The Douglas-Shenker strings [15] are examples of these stable vortices in supersymmetric gauge theories with only adjoint matter.

The key observation is that there are no vortices in the $SU(2)$ gauge theory, and so this $SO(3)$ bundle has no *spin* lift. In fact, using the general characterization of bundles in Eq. (5.10), the first nontrivial homotopy class of $SO(3)$ is at $p = 1$ and is given by (5.16). Therefore $SO(3)$ bundles are partially classified by a characteristic class

$$\omega_2 \in \mathrm{H}^2(S^2; \mathbb{Z}_2) = \mathbb{Z}_2. \tag{5.17}$$

This characteristic class is precisely the second Stiefel-Whitney class w_2. Therefore the nontrivial vortex configuration of the $SO(3)$ bundle has nonvanishing w_2 and so does not have a *spin* lift to an $SU(2)$ bundle.

The absence of a *spin* lift means that one cannot define spinors. One spinor representation of $SO(3)$ is the two-dimensional fundamental representation of $SU(2)$. Therefore $SU(2)$ fundamental matter cannot be defined in the presence of a $\mathbb{Z}_2$ vortex, as its wavefunction would be ill-defined. Instead, in the presence of $SU(2)$ fundamental matter one may only consider $SU(2)$ gauge bundles, which contain no vortices.

Similarly, when spacetime is not *spin* or equivalently when w_2 of the tangent bundle is nonzero, the wavefunctions of Lorentz spinors are ill-defined and so spinors cannot be consistently incorporated in the theory. In the next subsection we will see that sometimes spinors can be included if they are charged under a $U(1)$ gauge symmetry.

5.4 $Spin^c$ Structures

We have seen that when spacetime is not a *spin* manifold there is a nontrivial class w_2 and that this implies that the transition functions $x_{ij} = \pm 1$ of the sign of a spinor do not satisfy the triple overlap condition (5.7). Instead at some triple overlaps the product of the transition functions is

$$x_{ij} x_{jk} x_{ki} = -1. \tag{5.18}$$

The fact that the product is not equal to one means that the *Spin* bundle is not really a bundle. Therefore spinors cannot be sections of the *Spin* bundle and therefore cannot appear in the spectrum. In general a particle can only appear in a theory if its wavefunction is a section of a legitimate bundle that satisfies the triple overlap condition.

While the *Spin* bundle does not exist in this case, it is sometimes possible to construct a different bundle that does exist from the inconsistent *Spin* bundle. The strategy will be to consider a $U(1)$ bundle Q whose second Stiefel-Whitney class is equal to that of P. The tensor product bundle $P \otimes Q$ will then have a *spin* lift and the sections of its associated vector bundle will be the wavefunctions of $U(1)$-charged spinors. So while uncharged spinors are still inconsistent, charged spinors may exist if we can define Q.

If there is a $U(1)$ gauge group, then there is also a $U(1)$ gauge bundle Q. $U(1)$ bundles are entirely characterized by a degree two integral cohomology class called the Chern class

$$c_1(Q) \in \mathrm{H}^2(M;\mathbb{Z}). \tag{5.19}$$

Any element of $\mathrm{H}^2(M;\mathbb{Z})$ defines a $U(1)$ bundle Q, and therefore one can construct trivializations and $U(1)$ transition functions and the transition functions always satisfy the triple overlap condition (5.7).

Recall that $U(1)$ is isomorphic to $SO(2)$, and so one may also try to define a *spin* lift $\sqrt{Q}$ of Q whose transition functions are the square roots of those of Q. The Chern class of the square root bundle $\sqrt{Q}$ will satisfy

$$2c_1(\sqrt{Q}) = c_1(Q). \tag{5.20}$$

This implies that $\sqrt{Q}$ will only be a bundle if $c_1(Q)$ is divisible by two. We have another criterion for when $\sqrt{Q}$ exists. The lift $\sqrt{Q}$ of Q is a bundle if and only if its transition functions satisfy the triple overlap condition (5.7). These triple overlaps are measured by the Čech class

$$w_2(Q) \in \check{H}^2(M;\mathbb{Z}_2) \tag{5.21}$$

which consists of the triple products $x_{ij}x_{jk}x_{ki}$ of the transition functions. The triple products vanish, or more precisely are equal to the identity, precisely when $w_2(Q)$ is equal to zero, but we have seen that the bundle also exists precisely when $c_1(Q)$ is even. This suggests that $w_2(Q)$ is the mod 2 reduction of $c_1(Q)$

$$w_2(Q) = c_1(Q) \bmod 2 \tag{5.22}$$

and in fact it is.

Our strategy to make a bundle from the *Spin* nonbundle $\sqrt{P}$ is to construct another nonbundle $\sqrt{Q}$ that fails the triple overlap conditions at just the same triple overlaps as $\sqrt{P}$. The tensor product of these two nonbundles $\sqrt{P} \otimes \sqrt{Q}$ will then satisfy the triple overlap condition because the $U(1)$ commutes with everything, so the products of the triple overlaps just multiply. At each triple overlap the products of the transition functions for the two bundles are either both $+1$ or both -1, in either case the products of the transition functions of the tensor bundle is $+1^2 = 1$ or $-1^2 = 1$ and so the tensor bundle satisfies the triple overlap condition and is a legitimate bundle.

To complete the construction, we need only choose Q such that

$$c_1(Q) \bmod 2 = w_2(Q) = w_2(P), \qquad c_1(Q) \in \mathrm{H}^2(M;\mathbb{Z}). \tag{5.23}$$

While there exists a $U(1)$ bundle for any class $c_1(Q)$ in the integral cohomology group $\mathrm{H}^2(\mathbb{Z})$, not every class $w_2 \in \mathrm{H}^2(M;\mathbb{Z}_2)$ may be obtained as the mod 2 reduction of an integral class. As we will see in the next subsection, such an integral class exists if and only if a certain 2-torsion integral 3-class called the third Stiefel-Whitney class vanishes

$$W_3(P) = 0 \in \mathrm{H}^3(M;\mathbb{Z}). \tag{5.24}$$

For the moment we will assume that W_3 vanishes and so a U(1) bundle Q exists satisfying (5.23). In this case the tensor product bundle $P \otimes Q$ has a *spin* lift because

$$w_2(P \otimes Q) = w_2(P) + w_2(Q) = 2w_2(P) = 0 \tag{5.25}$$

where we have used the fact that w_2 is $\mathbb{Z}_2$ torsion. Now we can define our charged spinor wavefunction. If the spinor has an odd $U(1)$ charge $2q + 1$ then its wavefunction is a section of the associated vector bundle to the principal bundle $\sqrt{P} \otimes \sqrt{Q} \otimes Q^q$.

The fiber of the tensor bundle is the product of the groups *Spin* and $U(1)$. Both of these groups share the element which is equal to -1 times the identity, so to avoid double-counting the fiber of the tensor bundle is the group

$$Spin^c(N) = \frac{Spin(N) \times U(1)}{\mathbb{Z}_2}. \tag{5.26}$$

The tensor bundle is called a $Spin^c$ bundle. For example if $N = 2$ then

$$Spin^c(3) = U(2) = \frac{SU(2) \times U(1)}{\mathbb{Z}_2} = \frac{Spin(3) \times U(1)}{\mathbb{Z}_2}. \tag{5.27}$$

$U(2)$ gauge theories admit Dirac strings, which are the $spin^c$ lifts of the $\mathbb{Z}_2$ vortices in the $SO(3)$ gauge theory.

Using the above general classification of bundles combined with the fact that

$$\pi_1(Spin^c) = \mathbb{Z} \oplus \mathbb{Z}_2 \tag{5.28}$$

we find that $Spin^c$ bundles are partially characterized by two degree two cohomology classes. There is an element of integral cohomology which is just the magnetic flux of the $U(1)$ gauge theory, and there is an element of cohomology with $\mathbb{Z}_2$ coefficients $\mathrm{H}^2(M;\mathbb{Z}_2)$ which is called the $spin^c$ structure.

Recall that for every degree $p+1$ characteristic class there is a degree $p+2$ obstruction to a lift. In this case the obstruction to the existence of a $spin^c$ structure is the third Stiefel-Whitney class

$$w_3 \in \mathrm{H}^3(M;\mathbb{Z}_2). \tag{5.29}$$

While w_3 is always $\mathbb{Z}_2$ torsion, unlike w_2 it always has a lift to cohomology with integral coefficients, which will be denoted

$$W_3 \in \mathrm{H}^3(M;\mathbb{Z}) \tag{5.30}$$

and is also always $\mathbb{Z}_2$ torsion. When the third Stiefel-Whitney class of the tangent bundle **TM** is equal to zero, M is said to be a $spin^c$ manifold and a $spin^c$ lift of the tangent bundle exists. On such manifolds one can define $U(1)$-charged spinors. However recall that the $U(1)$ bundles $\sqrt{Q}$ themselves may fail the triple overlap condition by a sign, and in fact they must fail if M is not *spin* in order to cancel the failure of the *spin* bundle. In this case the field strength $c_1(\sqrt{Q})$ defined in (5.20) fails the Dirac quantization condition by a half-integer. These shifted quantization conditions are often responsible for fractional brane charges in string theory.

5.5 The Freed-Witten Anomaly

We will be interested in fermions not in the bulk spacetime, but in the spectra of the open strings that end on a particular Dp-brane. While Freed and Witten have demonstrated the existence of their anomaly with and also without the need of the following assumption, we will assume, in line with Sen's conjecture, that it is always possible to nucleate a spacefilling D9 and anti D9 pair from nothing. Then we may consider the open strings that stretch from any fixed Dp-brane to the D9. The ground state Ramond sector open strings that stretch from the Dp to the D9 are fermions whose wavefunctions ψ are sections of the tensor product of the spin bundle over the Dp-brane worldvolume of N and the $U(1)$ line bundle which is the gauge bundle of the $U(1)$ worldvolume gauge field on the Dp-brane. This bundle exists precisely when the Dp-brane's worldvolume is $spin^c$, and so the wavefunctions of the strings between the Dp and the D9 only exist if the Dp wraps a $spin^c$ submanifold N of the spacetime M. If it does not, the Dp-brane is said to suffer from a Freed-Witten anomaly which is equal to W_3, the failure of N to be $spin^c$.

Fundamental strings couple electrically to the NSNS B field. Therefore if one introduces a nontrivial B field, the path integral measure of a fundamental string trajectory will change. In particular, the wave functions of the fermions found in the Ramond groundstate of our p-9 strings will be shifted by a B field. The phase of the wavefunction will increase by the integral of the B field over the worldsheet. However, if the Dp-brane was $spin^c$ then the wavefunction was well-defined, and so after multiplying it by a constant shift it will continue to be well defined. Conversely, if the Dp-brane wrapped a non-$spin^c$ submanifold then the constant shift provided by the B field will not render the wave function well-defined, in this case the wavefunction is a section of a $spin^c$ bundle which cannot be globally extended over the entire Dp worldvolume. Below we will argue that a nontrivial H flux, which corresponds to a B flux that cannot be globally defined in exactly the same way as the fermion wavefunction, can render the wave function well-defined if and only if

$$W_3 + H = 0. \tag{5.31}$$

To arrive at (5.31) we will need to understand more precisely just what W_3 measures. We will now introduce an exact sequence based characterization of W_3 that generalizes to the case with H flux.

Remember that when a *Spin* bundle does not exist, as in the case of $\mathbb{CP}^2$ which is $spin^c$ but is not a *spin* manifold, one may not define uncharged fermions but one may define $U(1)$ charged fermions if the gauge bundle fails to be a gauge bundle in just the right places.

If $W_3 \neq 0$, so that the Dp-brane's worldvolume is not $spin^c$, then no choice of gauge bundle Q, even one that fails the triple overlap condition, can render the $spin^c$ bundle well-defined. We have already noted that this is a consequence of the condition that $w_2(\mathrm{TN})$ be the mod 2 reduction of the integral class $c_1(Q)$, as no such integral class exists when $W_3(\mathrm{TN})$ is nonzero. Now we will see this more explicitly. First we will need to understand how the gauge bundle's field strength $F = c_1(\sqrt{Q})$, the second Stiefel-Whitney class w_2 and the third Stiefel-Whitney class W_3 are related. They are related by a long exact sequence which arises from the short exact sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{\times 2} \mathbb{Z} \xrightarrow{mod\ 2} \mathbb{Z}_2 \longrightarrow 0 \tag{5.32}$$

where the first map is multiplication by two and the second is reduction modulo two. This short exact sequence leads to a long exact sequence in cohomology with $\mathbb{Z}$ and $\mathbb{Z}_2$ coefficients.

Given a short exact sequence (5.32) one can always form a long exact sequence of cohomology groups with coefficients given by the elements of the short sequence. In this case one finds

$$... \xrightarrow{\beta} \mathrm{H}^2(M;\mathbb{Z}) \xrightarrow{\times 2} \mathrm{H}^2(M;\mathbb{Z}) \xrightarrow{mod\ 2} \mathrm{H}^2(M;\mathbb{Z}_2) \xrightarrow{\beta} \mathrm{H}^3(M;\mathbb{Z}) \xrightarrow{\times 2} ... \tag{5.33}$$

where β is a map called the Bockstein homomorphism, which is a close relative of the exterior derivative of differential forms. All of the elements that we have been discussing, the $U(1)$ field strength $F = c_1(\sqrt{Q}) \in \mathrm{H}^2(M;\mathbb{Z})$, the second Stiefel-Whitney class $w_2 \in \mathrm{H}^2(M;\mathbb{Z}_2)$ and the third Stiefel-Whitney class $W_3 \in \mathrm{H}^3(M;\mathbb{Z})$ fit into this long exact sequence

$$... \xrightarrow{\beta} c_1(\sqrt{Q}) \xrightarrow{\times 2} c_1(Q) \xrightarrow{mod\ 2} w_2 \xrightarrow{\beta} W_3 \xrightarrow{\times 2} ... \tag{5.34}$$

In particular we see that W_3 is defined to be the Bockstein of w_2

$$W_3 = \beta w_2. \tag{5.35}$$

If and only if the Dp-brane wraps a $spin^c$ submanifold, W_3 will be zero. W_3 is the image of w_2 under the Bockstein, and so W_3 is zero if and only if w_2 is in the kernel of the Bockstein homomorphism. The above sequence is exact, which means that the kernel of the Bockstein map β is the image of the preceding mod two reduction map. Thus w_2 is in the kernel of the Bockstein map if and only if there exists an element $c_1(Q)$ of integral cohomology such that the mod two reduction of $c_1(Q)$ is equal to w_2. Remember that this was the condition such that a $U(1)$ bundle Q exists satisfying

$$w_2(Q) = w_2(\mathrm{TM}). \tag{5.36}$$

Thus the existence of $c_1(Q)$ which reduces modulo two to w_2 is equivalent to the condition that the D-brane wrap a $spin^c$ cycle, as desired.

If the D-brane wraps a $spin^c$ cycle which is not *spin* then w_2 is not equal to zero and so $c_1(Q)$ is not in the kernel of mod two reduction. By exactness of the sequence, this means that $c_1(Q)$ is not in the image of multiplication by two, and so $c_1(\sqrt{Q})$ does not really exist as an integral class. This corresponds to the fact that the compensating $U(1)$ bundle $\sqrt{Q}$ fails the triple overlap condition. Circle bundles that fail triple overlap by a sign do not have integral Chern classes, instead they are classified by the integral class $c_1(Q)$ which is intuitively twice their Chern class.

$c_1(Q)$ is the degree 2 characteristic class which partially characterizes the topology of the $Spin^c$ bundle. When it fails to exist, because w_2 is not in the kernel of β, then the $Spin^c$ lift of the tangent bundle does not exist. The fermion wavefunction is a section of the *Spin*^{c} bundle, so it does not exist either.

What changes when we add a nontrivial H field? The fermion wavefunction no longer is a section of just the *Spin*^{c} bundle, but of the tensor product of the *Spin*^{c} bundle with another bundle that has characteristic class B. Recall that the sum of the $U(1)$ gauge field F and the B is gauge invariant, and so F and B always appear together in gauge-invariant expressions. The B field, like w_2, is naturally an element of cohomology with torsion coefficients. This is because it enters in the string partition function via the factor $e^{\int_\Sigma B}$, and so adding an integer to the B field has no observable effect.

Thus one may add B to w_2 in the long exact sequence (5.34). In fact it is the sum that appears in the string partition function. The failure of B to lift to an integer class in $\mathrm{H}^2(M;\mathbb{Z})$ is, by exactness, measured by its image under the Bockstein homomorphism. We refer to its image as the H flux, even though in general it may contain torsion components. Thus we have the modified long exact sequence

$$... \xrightarrow{\times 2} c_1(Q) + B \xrightarrow{mod\ 2} w_2 + B \xrightarrow{\beta} W_3 + H \xrightarrow{\times 2} ... \tag{5.37}$$

where by abuse of notation we refer to both the B field and its integer lift as B.

Now $c_1(Q) + B$ is the characteristic class for the combined bundle, whose sections are the fermion wavefunction. The class $c_1(Q) + B$ is defined to be any class such that its mod two reduction is $w_2 + B$, thus the combined bundle exists if and only if $w_2 + B$ is in the image of the mod 2 reduction map. By exactness, this is the case if and only if $w_2 + B$ is in the kernel of the Bockstein, which implies (5.31)

$$0 = \beta(w_2 + B) = W_3 + H. \tag{5.38}$$

Therefore a D-brane can sometimes consistently wrap a non-*spin*^{c} cycle, but this can happen precisely when the H flux cancels W_3, the obstruction to the cycle being *spin*^{c}. Summarizing, the Freed-Witten anomaly condition states that a D-brane may wrap a cycle if and only if $W_3 + H$ of that cycle vanishes as an integral cohomology class.

6 Twisted K-Theory from the Freed-Witten Anomaly

6.1 Brane Insertions and MMS Instantons

So what happens if you ignore the Freed-Witten anomaly and wrap a Dp-brane on a cycle with $W_3 + H$ flux anyway? For simplicity, we will just consider the free part of $W_3 + H$, the $\mathbb{Z}^k$ terms that survive the supergravity limit, but the same conclusions apply to the torsion part. We will now provide a heuristic argument for the claim of Ref. [7] that this anomaly may be canceled by a D$(p-2)$-brane insertion.

Recall that there are large gauge transformations that change the B field and the $U(1)$ gauge field strength F but preserve their sum. Thus one may intuitively try to gauge transform $H = dB$ into dF. Concretely this means the following. Consider a small, contractible sphere inside of the Dp-brane. The integral of the B field over the sphere is not well-defined, as the B field is not gauge invariant. One may define it via Stoke's theorem as the integral of H, either over the inside or the outside of the sphere. In the first case, one finds that the integral of B is about zero because the sphere is small and H is well-defined. In the second case, one finds that the integral of B is about equal to the integral of the H flux on some three-cycle.

These two cases are related by a large gauge transformation, and so F must also differ in the two cases by the same amount, the integral of H over a three-cycle. By Stoke's theorem, we can then conclude, using either gauge, that the integral of dF over the union of these two three-cycles is equal to the integral of H. dF, by the Bianchi identity, is equal to the magnetic monopole charge that intersects the 3-cycle. Thus for every unit of H flux on a given cycle there is a single unit of magnetic monopole charge on that cycle. Considering again the full integral cohomology we conclude that there is a magnetic monopole which is Poincaré dual to $W_3 + H$

$$Q_{\mathbf{monopole}} = \mathrm{PD}(W_3 + H). \tag{6.1}$$

Recall that magnetic flux F on a Dp-brane worldvolume couples to the RR field C_{p-1} and so carries D$(p-2)$-brane charge. A magnetic monopole is a codimension three surface that is a source for magnetic flux, so it is the endpoint of a D$(p-2)$-brane. Thus a Dp-brane is always free to wrap a cycle on which $W_3 + H$ is nonvanishing, but the price is that there must be a D$(p-2)$ brane which ends on the Dp-brane on a cycle dual to the offending $W_3 + H$. The D$(p-2)$ extends away from the brane until it finds another place to end, or, if there is no other place to end, it is semi-infinite. Such configurations of branes ending on branes were referred to as baryons in Ref. [13] and, for reasons that will be apparent momentarily, as (MMS) instantons in Ref. [7].

Depending on the compactification in question, baryons may have infinite masses because the D$(p-2)$-branes may be forced to extend to infinity. If the transverse dimensions are compact then the D$(p-2)$-branes may have no place for the other endpoint, and so the baryons configurations may be inconsistent. Of course, the D$(p-2)$-branes may end on an anti Dp-brane wrapping the same cycle, but the Dp and anti Dp may decay so this system carries no conserved charges. Thus baryonic configurations are often dropped in classifications of D-brane charges. As we have mentioned, while baryonic configurations carry homology charges, they do not carry K-theory charges. Thus K-theory does *not* classify baryons.

The Freed-Witten anomaly not only eliminates certain configurations from our classification, but also it leads to the nonconservation of a charge. To see this, consider again a baryonic configuration in which a Dp-brane wraps a cycle with nonvanishing W_3+H and a D$(p-2)$ ends on the Dp and extends away in another direction. If this other direction is a spatial direction, then one arrives at the high-mass spider-like configuration described above. But the other direction could also be timelike. This corresponds to D$(p-2)$-branes being created or destroyed. Thus the nonconserved charge is a D$(p-2)$-brane charge, it is the charge of the homology $(p-2)$-cycle wrapped by the created or destroyed branes. This charge maps to the trivial K-theoretic charge, as K-theory is the quotient of the legal homology charges by a set of nonconserved charges. These timelike configurations are called MMS instantons because the Dp-brane may live for only a short period of time as it sweeps out a $(p+1)$-cycle and its trajectory may be a solution to the Euclidean equations of motion.

When $g_s \neq 0$ the intersection of a D$(p-2)$-brane and a Dp-brane is not a sharp corner. Rather there is a continuous transition, in which one sees, if one looks closely at the D$(p-2)$, that it is always a thin tube of Dp-brane with S^2 cross-sections that carry D$(p-2)$-charge because the integral of the $U(1)$ field strength over these spheres is nonzero. The spheres may be thought of as Dp dipoles, as antipodal points carry opposite Dp charge, thus they carry no net Dp charge and appear to be D$(p-2)$-branes from far away. The actual monopole is at the end of the tube, at infinity, so in a sense these branes avoid the Freed-Witten anomaly by putting it all at a point and cutting out that point. If instead of extending away in a spatial direction the D$(p-2)$ extends in the time direction then the configuration is an MMS instanton. Smoothing out the corners at nonzero string coupling one finds that this corresponds to, for example, a homologically nontrivial stack of D$(p-2)$ branes that grows, via the Myers dielectric effect [29] for example, into a Dp-brane and sweeps out a $(p+1)$-cycle before eventually collapsing into heat as it finishes and closes up.

In conclusion, the Freed-Witten anomaly is responsible for two physical effects. First, D-branes carrying $W_3 + H$ not equal to zero are, without extra insertions, inconsistent. Secondly D-branes wrapping homology cycles which are Poincaré dual to W_3+H in some bigger cycle are unstable, and their corresponding homology charges are not preserved. The inconsistent charges do not lift to K-theory classes, while the unstable charges correspond to the zero K-theory class.

Now that we have turned on fluxes, the tachyon condensation argument for the K-theory classification no longer applies. In fact D-branes are not quite classified by K-theory when there is a topologically nontrivial H flux, instead they are classified by a variation of K-theory known as twisted K-theory [8] where the twist is given by H. Some details on the construction of twisted K-theory from cohomology will appear in the next subsection. This construction, known as the Atiyah-Hirzebruch spectral sequence, will make it apparent that the Freed-Witten anomaly indeed removes and quotients out the correct homology charges to arrive at twisted K-theory.

6.2 Atiyah-Hirzebruch Spectral Sequence

Dp-branes in type IIA string theory exist for every even p. The branes carrying conserved RR charges are extended in one time direction and p spatial dimensions, and so are classified by the even homology groups $\mathrm{H}_p(M)$. By Poincaré duality these are isomorphic to the odd cohomology groups

$$\mathrm{H}^{9-p}(M) = \mathrm{H}_p(M). \tag{6.2}$$

While Dp-brane charges for a fixed p are classified by the single odd cohomology group H^{9-p}, the collection of all Dp-brane charges in type IIA string theory is the direct sum of these groups for all even values of p. We refer to this direct sum as the odd cohomology and denote it $\mathrm{H}^{\mathrm{odd}}$. Similarly D-branes in IIB are classified by the even cohomology $\mathrm{H}^{\mathrm{even}}$. Summarizing, all D-branes in type II string theories carry charges in either $\mathrm{H}^{\mathrm{even}}$ or $\mathrm{H}^{\mathrm{odd}}$ which are defined by

$$\mathrm{H}^{\mathrm{even}}(M) = \oplus_k \mathrm{H}^{2k}(M), \qquad \mathrm{H}^{\mathrm{odd}}(M) = \oplus_k \mathrm{H}^{2k+1}(M). \tag{6.3}$$

However we know that these charge groups are too big, we really want the twisted K-groups K_H^0 and K_H^1 which will be quotients of subsets of $\mathrm{H}^{\mathrm{even}}$ and $\mathrm{H}^{\mathrm{odd}}$ respectively.

The Atiyah-Hirzebruch spectral sequence (AHSS) was introduced by Atiyah and Hirzebruch to compute untwisted K-groups, although it has since been generalized to the twisted case. It is an algorithm which begins with $\mathrm{H}^{\mathrm{even}}$ and $\mathrm{H}^{\mathrm{odd}}$. It arrives at twisted K-theory after a series of approximations, the nth of which are denoted E_n^0 and E_n^1. At each step some elements are removed and some are quotiented out. After a finite number N of steps the algorithm arrives at groups which as sets are equivalent to the twisted K-groups K_H^0 and K_H^1

$$E_1^0 = \mathrm{H}^{\mathrm{even}}(M), \qquad E_1^1 = \mathrm{H}^{\mathrm{odd}}(M), \qquad |E_N^i| = |\mathrm{K}_H^i(M)| \tag{6.4}$$

which classify D-branes in IIB and IIA string theories respectively.

To calculate the nth approximation E_n^i of K_H^i, one needs both of the $(n-1)$st approximations E_{n-1}^0 and E_{n-1}^1 and also a differential operator d_{2n-1} which maps one to the other

$$d_{2n-1} : E_{n-1}^0 \leftrightarrow E_{n-1}^1, \qquad d_{2n-1} d_{2n-1} = 0. \tag{6.5}$$

The subscript $2n+1$ on the differential reflects Atiyah and Hirzebruch's observation that each element of E_{n-1}^i can be represented by a class in the degree $2k+i$ cohomology group H^{2k+i} for some k, and d_{2n-1} increases the degree of this class by $2n-1$. However, in the cases $n \geq 2$, the action of d_{2n-1} on a given cohomology class is only well defined up to the quotients that occurred at previous steps in the spectral sequence, and so the action is only well defined on E_{n-1}^i itself.

Given E_{n-1}^i and d_{2n-1} one can determine the E_n^i. They are simply the cohomology of the E_{n-1}^i with respect to the differentials d_{2n-1}

$$E_n^i = \frac{\mathrm{Ker}(d_{2n-1} : E_{n-1}^i \to E_{n-1}^{i+1})}{\Im(d_{2n-1} : E_{n-1}^{i+1} \to E_{n-1}^i)}. \tag{6.6}$$

Only the first differential, d_3, has been explicitly computed in the literature. It is

$$d_3 x = Sq^3 x + H \cup x \tag{6.7}$$

where Sq^3 is a cohomology operation known as a Steenrod square which takes an integral class in the degree k cohomology to a class in the degree $k+3$ integral cohomology, as does the cup product with H. The image of Sq^3 is always a $\mathbb{Z}_2$ torsion component of $\mathrm{H}^{k+3}(M;\mathbb{Z})$. Recall that in the supergravity limit the torsion components disappear, as the finite cyclic group $\mathbb{Z}_2$ tensored with the real numbers is zero. Therefore in the classical theory the Sq^3 term vanishes and d_3 reduces to the wedge product with the H flux.

While limited results about the higher differentials have appeared, they have been computed in general in the case of real coefficients [30], in which they compute the tensor products of the K groups with the real numbers. The nth differential on a differential form x is

$$d_{2n+1} x = [H, ..., H, x] \tag{6.8}$$

which is the Massey product of n copies of a differential form representing H with x. Massey products are intuitively products of differential forms multiplied by inverses of the exterior derivative. These inverses in general are not well-defined on differential forms, or even on the de Rham cohomology of differential forms. However they are well-defined on E_{n-1}^i, as required. For this reason the Massey product is called a secondary cohomology operation. Note in particular that the real differential (6.8) disappears in the case $H = 0$. Recalling that the real case corresponds to the supergravity limit, we find that p-branes in supergravity are classified by de Rham cohomology if $H = 0$.

However the nontriviality of (6.8) when H is nontrivial implies that even in the supergravity limit the homology or equivalently the cohomology classification of D-branes fails. There is a an important exception to this last statement. A manifold is called *formal* if all Massey products with more than two terms vanish. Many of the kinds of manifolds that are of interest in compactifications are formal. For example, all Kähler manifolds are formal, as are all simply-connected six-manifolds. The differentials (6.8) at $n > 1$ vanish on a formal manifold and so branes in supergravity on a formal background are just classified by the $H\wedge$ cohomology of the de Rham cohomology, which is isomorphic to the $d + H\wedge$ twisted cohomology.

6.3 Freed-Witten from the AHSS

One may now ask what the AHSS has to do with the Freed-Witten anomaly. Recall that the Freed-Witten anomaly provided a necessary condition for the consistency of a homology-valued D-brane charge, and also yielded a set of D-brane charges that are not conserved. It did not provide a sufficient condition for consistency, because some nonrepresentable cycles that do not lead to Freed-Witten anomalies cannot be wrapped. On the other hand, the AHSS computes the twisted K-theory precisely, at least as a set, from the cohomology. Thus it must eliminate all unphysical branes, those afflicted with the Freed-Witten anomaly and those wrapping unwrappable singularities. It must also quotient by all nonconserved charges, those corresponding to branes that may decay via MMS instantons and those that may decay via trajectories in which the brane grows, wraps a nonrepresentable cycle and then is sucked into a worldvolume singularity that forms, eats its hosts and finally eats itself.

Thus the consistency of the AHSS and the Freed-Witten anomaly, which is critical for the K-theory classification, requires that the AHSS indeed eliminates all Freed-Witten anomalous D-branes, although it may eliminate others. We will now argue that this is the case. In particular, we will show that cohomology classes which are not in the kernel of d_3 are Poincaré dual to D-brane worldvolumes that suffer from the Freed-Witten anomaly, while its image consists of cohomology classes which are dual to the worldvolumes of unstable D-branes. However some Freed-Witten anomalous D-branes and Freed-Witten unstable D-branes will be missing from this classification, we will note that in the only known example of this phenomenon [7] these anomalies and instabilities are captured by d_5 and correspond to MMS instantons in which a D$(p-4)$-brane ends on a Dp-brane.

We will need the definition of the Steenrod square Sq^3, which appears in the differential d_3 in (6.7). Consider an integral cohomology class x^p of degree p. This class is Poincaré dual, in the 9-dimensional timeslice M^9, to a homology class x_{9-p} of degree $9-p$

$$\mathrm{PD}(x^p) = x_{9-p} \in \mathrm{H}_{9-p}(M;\mathbb{Z}) \tag{6.9}$$

which corresponds to a spatial slice of the worldvolume of a D$(9-p)$-brane. In particular, a spatial slice of the worldvolume is a $(9-p)$-cycle $N \subset M$ that represents the homology class dual to x^p. The submanifold $N \subset M$ has a normal bundle $\mathbf{NN}$, which is a p-dimensional real vector bundle on N.

Recall that $W_3(\mathbf{NN})$ is the integral cohomology 3-class which measures the obstruction to the existence of a *spin*^{c} lift of the normal bundle $\mathbf{NN}$. The Steenrod squares are defined to be the pushforwards i_* of the Stiefel-Whitney classes of the normal bundle of N onto M using the inclusion map $i : N \hookrightarrow M$

$$sq^n x = i_* w_n(\mathrm{NN}), \qquad i_* : \mathrm{H}^n(N) \longrightarrow \mathrm{H}^{p+n}(M). \tag{6.10}$$

Recall that homology classes pushforward naturally and cohomology classes pullback naturally. The pushforward of a cohomology class is defined to be the Poincaré dual in M of the pushforward of its Poincaré dual in N

$$sq^n x = i_* w_n(\mathrm{NN}) = \mathrm{PD}|_M(i_* \mathrm{PD}|_N(w_n(\mathrm{NN}))). \tag{6.11}$$

The unnaturalness of the pushforward of a cohomology class is responsible for the fact that we need an infinite tower of differentials on the cohomology of M to capture a single anomaly on the worldvolume of N, as well as the fact that the pushforward of a cohomology class is a class of a different degree. We use the notation Sq^3 to denote the integral lift of sq^3, which always exists. Equation (6.10) combined with the linearity of the pushforward map implies that Sq^3 will be trivial when $W_3(\mathbf{NN})$ is trivial, but it does not imply the inverse.

This is not quite the Freed-Witten condition for two reasons. First $W_3(\mathbf{NN})$ is not quite the Freed-Witten anomaly, the Freed-Witten anomaly was the the third Stiefel-Whitney class of the tangent bundle. Second, i_* can have a nontrivial kernel, so some D-branes afflicted with the Freed-Witten anomaly and so having a nontrivial W_3 can nonetheless lead to a trivial d_3. We will now resolve the first issue and speculate about the second, then we will extend to the case with nontrivial H flux.

To relate the differential defined in (6.10) to the Stiefel-Whitney class of the tangent bundle, we first observe that the direct sum of the tangent and normal bundles to the submanifold $N \subset M$ is the tangent bundle of M, restricted to N

$$\mathrm{TN} \oplus \mathrm{NN} = \mathrm{TM}|_N. \tag{6.12}$$

We can now insert this relation into the addition formula for Stiefel-Whitney classes

$$w(A \oplus B) = w(A) \wedge w(B) \tag{6.13}$$

where $w(A)$ is the sum of the Stiefel-Whitney classes of the bundle A. As usual a lower case w denotes a Stiefel-Whitney class in the cohomology with $\mathbb{Z}_2$ coefficients. Only the odd degree Stiefel-Whitney classes w_{2k+1} can be lifted to integral cohomology classes W_{2k+1}, and even these will continue to be $\mathbb{Z}_2$ torsion

$$2W_{2k+1} = 0. \tag{6.14}$$

Expanding the addition law (6.13) in components, in the case A=**TN**, B=**NN**, $A \oplus B$=$\mathbf{TM}|_N$, and using the identity $w_0 = 1$ we find at degree one

$$w_1(\mathrm{TM}|_N) = w_0(\mathrm{TN}) \wedge w_1(\mathrm{NN}) + w_1(\mathrm{TN}) \wedge w_0(\mathrm{NN}) = w_1(\mathrm{NN}) + w_1(\mathrm{TN}). \tag{6.15}$$

We recall that the first Stiefel-Whitney class of a manifold is 0 if and only if the manifold is orientable. D-branes in type II string theory are always orientable, as are compactification manifolds. Therefore

$$w_1(\mathrm{TM}) = w_1(\mathrm{TN}) = 0. \tag{6.16}$$

Let i be the inclusion map of the D-brane worldvolume N into M, then the Stiefel-Whitney classes of the tangent space **TM** of M restricted to N are just the pullbacks of those of **TM** using i^*

$$i : N \hookrightarrow M, \qquad w_k(\mathrm{TM}|_N) = i^* w_k(\mathrm{TM}). \tag{6.17}$$

In particular, the fact that the first Stiefel-Whitney class of the tangent space of M vanishes (6.16) implies that its image $w_1(\mathbf{TM}|_N)$ under the homomorphism i^* also vanishes. Thus the left hand side and the last term on the right hand side of (6.15) both vanish. This means that the other term on the right hand side of (6.15) must also vanish

$$w_1(\mathrm{NN}) = 0. \tag{6.18}$$

We will need this fact in the following argument.

Our goal is to relate the two third Stiefel-Whitney classes $W_3(\mathbf{TN})$ and $W_3(\mathbf{NN})$, the first of which appears in the Freed-Witten anomaly and the second of which appears in the AHSS. These classes both appear in the degree three part of (6.13)

$$\begin{aligned} w_3(\mathrm{TM}|_N) &= w_0(\mathrm{TN}) \wedge w_3(\mathrm{NN}) + w_1(\mathrm{TN}) \wedge w_2(\mathrm{NN}) \\ &\quad + w_2(\mathrm{TN}) \wedge w_1(\mathrm{NN}) + w_3(\mathrm{TN}) \wedge w_0(\mathrm{NN}) \\ &= 1 \wedge w_3(\mathrm{NN}) + 0 + 0 + w_3(\mathrm{TN}) \wedge 1 = w_3(\mathrm{NN}) + w_3(\mathrm{TN}). \end{aligned} \tag{6.19}$$

Compactification manifolds M in type II string theory are always $spin^c$. Therefore

$$w_3(\mathrm{TM}|_N) = i^* w_3(\mathrm{TM}) = i^* 0 = 0. \tag{6.20}$$

Thus (6.20) implies that the third Stiefel-Whitney classes of the normal and tangent bundles to N add to zero. Using the fact (6.14) that Stiefel-Whitney classes are mod 2 torsion, either of these classes may be moved to the other side of the equality to yield

$$w_3(\mathrm{TN}) = w_3(\mathrm{NN}). \tag{6.21}$$

This equality is preserved by their integral lifts.

We can now reformulate the Freed-Witten anomaly as follows. In the absence of H flux, Dp-branes must wrap cycles N such that the third Stiefel-Whitney class of the normal bundle of N is trivial. If $W_3(\mathbf{NN})$ is zero then we saw in (6.10) that square three of the Poincare dual of N vanishes. Thus the cohomology class x dual to a Dp-brane's worldvolume is annihilated by the first AHSS differential

$$d_3x = Sq^3x = i_*W_3(\mathrm{NN}) = 0 \tag{6.22}$$

if $W_3(\mathbf{NN}) = 0$ which in turn occurs if and only if $W_3(\mathbf{TN}) = 0$, which occurs if the Dp-brane wrapping is Freed-Witten anomaly free. Therefore consistent Dp-brane wrappings are in the kernel of d_3, as they must be in order to lift to K-theory classes.

This argument is not invertible. In fact the invertibility fails at two points. First, some inconsistent Dp-brane wrappings have trivial $W_3(\mathbf{TN})$, although only when N is nonrepresentable so this will not occur when M is 9-dimensional or less. It may also fail because sometimes $W_3(\mathbf{NN}) \neq 0$ but $Sq^3x^p = 0$ because W_3 is in the kernel of the pushforward map i_* in (6.10). In Ref. [7] it was shown that such a situation arises in the identification of the twisted K-theory of $SU(3)$ with D-brane charges in the SU(3) WZW model. As will be seen in Subsec. 7, in this case the unphysical branes are in the kernel of d_3 but are not in the kernel of d_5, and so they do not carry K-theory charges. Generalizing this result, in Ref. [31] a general form of the Freed-Witten part of the d_5 term was proposed.

The main advantage of the Freed-Witten perspective of the K-theory classification over the Sen conjecture perspective is that one can easily include a nontrivial H flux. We will now argue that, even in the presence of H flux, Dp-branes that are Freed-Witten anomaly free are Poincaré dual to cohomology classes that are in the kernel of the AHSS differential d_3, as they must be if their charges are to lift to K-theory classes.

The Freed-Witten anomaly on a Dp-brane's worldvolume is equal to the sum of the third Stiefel-Whitney class of its normal bundle plus the integral of the pullback of the H flux onto its worldvolume. More precisely, now using the language of integral cohomology and not that of differential forms, the H flux in the bulk defines an integral 3-class in $\mathrm{H}^3(M;\mathbb{Z})$ and its contribution to the anomaly is

$$H \cap N, \qquad \cap : \mathrm{H}^j \times \mathrm{H}_k \longrightarrow \mathrm{H}_{k-j} \tag{6.23}$$

where $\cap$ is an operation called the cap product. Roughly speaking the cap product takes a p-manifold N and a 3-class H and produces a $(p-3)$-manifold which is the Poincaré dual of H in N, in other words, it produces the magnetic monopole in the worldvolume gauge theory of the Dp-brane.

The cap product satisfies a useful relation. If x^{9-p} is the degree $9-p$ cohomology class which is Poincaré dual to a p-dimensional timeslice N of the worldvolume of our Dp-brane, then the magnetic monopole is the Poincare dual (PD) of $H \cup x^{9-p}$

$$H \cup x^{9-p} = H \cup \mathrm{PD}(N) = \mathrm{PD}(H \cap N). \tag{6.24}$$

where we recall that the cup product $\cup$ is the integral cohomology version of the wedge product of differential forms, which we have been using implicitly for most of this note. For example the H flux on a Dp-brane's worldvolume is topologically trivial if and only if $H \cup x^{9-p}$ is topologically trivial.

Adding the contribution of the H flux (6.24) to the Freed-Witten anomaly to the contribution of W_3 from (6.22) we find that the vanishing of the entire Freed-Witten anomaly $W_3 + H$ implies the vanishing of

$$d_3x = Sq^3x + H \cup x = i_*(W_3(\mathrm{NN}) + H) = 0. \tag{6.25}$$

Thus, as desired, Freed-Witten anomaly free configurations are in the kernel of d_3. The higher differentials have never been explicitly calculated, but it is known that Freed-Witten anomaly free Dp-branes which wrap representable cycles are in fact in the kernel of all of the differentials d_{2k+1} at least in the untwisted case, and so these D-branes all carry K-theory charges. Furthermore, the AHSS guarantees that these charges are trivial precisely when the D-brane wraps a cycle whose Poincaré dual is in the image of one of the differentials d_{2k+1}.

6.4 The Supergravity Limit and Twisted Homology

In the supergravity limit everything is tensored by the real numbers $\mathbb{R}$. This kills the W_3 term, as torsion groups tensored with the group of real numbers are trivial and W_3 is always $\mathbb{Z}_2$ torsion by (6.14). Similarly, Steenrod squares are all identically equal to zero after tensoring by the reals, as their images are also torsion. Thus in the supergravity limit the AHSS depends entirely on H. The first nontrivial differential is given by the wedge product with H, and more generally all of the differentials are given by the Massey products in (6.8).

Working with differential forms, which capture all of the information in the classical theory, one may begin the AHSS with the set of not just closed differential forms but all differential forms. Then one arrives at the set of closed differential forms by introducing the differential d_1 equal to the exterior derivative and taking its cohomology, which is de Rham cohomology. The forms in the image of the exterior derivative are exact. These do not correspond to trivial configurations, but rather to configurations that contain nontrivial RR fluxes and no D-branes.

Poincaré duality only applies to closed differential forms, and so general differential forms are difficult to relate to D-branes. However one may interpret every step in terms of D-branes if, instead of passing through differential forms, one works with submanifolds, or more precisely simplicial complexes, from the beginning. Then the first differential d_1 is the boundary operator, and the cohomology of the boundary operator consists of all D-branes which have no boundaries quotiented by those that wrap boundaries and so can decay, in other words, the homology group. The second differential d_3 can be defined to be the cap product with H, whose kernel consists of branes that contain no magnetic monopoles

$$d_1 = \partial, \qquad d_3 = H \cap . \tag{6.26}$$

In the full quantum theory in which one uses integral homology, one needs to include a term in d_3 which captures W_3. However in the classical theory, and also in the quantum theory in certain cases, W_3 vanishes. In particular if M is any orientable 6-manifold then W_3 vanishes, and so in orientifold-free flux compactifications of type II which yield 4-dimensional physics one can drop the W_3 term and use $d_3 = H\cap$.

In general one needs the entire series of differentials from (6.8), but when the compactification manifold possesses a special property called formality we have noted that they all vanish, at least rationally. Simply connected 6-manifolds are always formal, for example. Also all manifolds which are diffeomorphic to Kahler manifolds are formal. Generalized Kähler manifolds may also be formal, at least they share many of the nice properties of formal manifolds [32]. Thus, in flux compactifications one can generally compute the twisted K-theory using the simple differentials (6.26).

In fact, on formal manifolds the cohomology with respect to these two differentials is isomorphic to the cohomology with respect to the single differential [33]

$$\partial_H = \partial + H \cap . \tag{6.27}$$

The cohomology of ∂_H is called twisted homology, in the real case it was applied to the classification of D-branes on calibrated cycles in Refs. [34, 35]. Notice that the two summands are of different dimensions, the first reduces the dimension of a submanifold by 1 and the second by 3. Thus twisted homology classes contain networks of branes whose dimensions are all equal modulo 2. These are all baryons.

More generally, and perhaps more usefully for phenomenology, on all simply connected 6-manifolds twisted homology and twisted K-theory are equivalent as sets. This is a consequence of the fact that the obstructions to this equivalence, d_5 in twisted K-theory and the next element of the spectral sequence in the identification of $d + H$ with d and H cohomologies are both degree 5 operations. Degree 5 operations on a 6-manifold either have images in H^5 or operate on elements of H^1, both of which are trivial in the simply connected case, and so these two obstructions vanish.

The cap product with a differential form is only well defined on homology classes, but in (6.27) it is applied to a general submanifold which may have a boundary. To render it well defined, one needs to include two pieces of extra data on the submanifold. One needs a nonquantized 2-form called the B-field whose exterior derivative is the pullback of the H field, and one needs an integral two-class F, that turns out to be the field strength of the worldvolume gauge bundle.

Then the boundary map yields

$$\partial : N \mapsto N', \qquad \int_{N'} B + F = \int_{N} H. \tag{6.28}$$

This only defines the sum of $B+F$. There are large gauge transformations which change B and F, but these must always preserve both the sum and the quantization condition on F.

The extra data needed to define a twisted homology class not only matches with the physical expectations of a D-brane, but it also precisely matches the extra data needed to define a twisted generalized complex submanifold in Ref. [36]. This is because cycles in both cases can be thought of as sections of a bundle whose fibers are the Eilenberg-MacLane space $\mathrm{K}(\mathbb{Z},2)$ with 3-class H or alternately of its S^2 subbundle with Euler class H. We recall that an Eilenberg-MacLane space $\mathrm{K}(G,n)$ is defined to be any space whose only nontrivial homotopy group is

$$\pi_n(\mathrm{K}(G,n)) = G \tag{6.29}$$

so that, in particular, $\mathrm{K}(G,n)$ bundles are characterized completely by the characteristic class ω_{n+1} in $\mathrm{H}^{n+1}(M;G)$, which in the case of $\mathrm{K}(\mathbb{Z},2)$ is the H flux

$$H = \omega_3 \in \mathrm{H}^3(M;\mathbb{Z}). \tag{6.30}$$

The extra data is the 2-form connection B on this bundle and also its choice of trivialization F. The subbundle realization is particularly useful because the homology of the total space of the sphere subbundle is just an extension of the ordinary homology of M by its twisted homology, thus it provides a tool for the calculation and also the visualization of twisted homology classes.

Unlike K-theory, which topologically can only be twisted by a 3-form, homology can be twisted by any p-form G_p. One then needs a worldvolume $(p-1)$-form connection and trivialization, which describe the fibration of a $K(\mathbb{Z},p-1)$ bundle or its S^{p-1} subbundle with characteristic class or Euler class G_p. Thus while D-branes are sections of $K(\mathbb{Z},2)$ bundles, M-branes are sections of $K(\mathbb{Z},3)$ bundles, or equivalently are subsets of the total space of the bundle. The Freed-Witten anomaly is then interpreted as the topological obstruction to the existence of a section of the bundle restricted to the D-branes worldvolume. If one uses the approximation E_8 for $K(\mathbb{Z},3)$ and the based loopgroup of E_8 for $K(\mathbb{Z},2)$ one arrives at the characterization of D-branes in E_8 gauge theories in Refs. [37, 38, 39], which was based on the E_8 gauge theory of Ref. [6, 40]. While the existence of this bundle away from the boundaries is not evident in perturbative string theories, $PU(\infty)$, which is the quotient of $U(\infty)$ by its $U(1)$ center, is also a model for $K(\mathbb{Z},2)$ and the existence of a $PU(\infty)$ bundle in the bulk in string field theory is well-known, see for example Refs. [41, 42].

6.5 RR Magnetic Field Strengths

Ramond-Ramond fluxes are sourced by D-branes. Thus one might expect that inequivalent RR field strengths are in one to one correspondence with inequivalent D-branes and so should be classified by the same twisted K-group. Moore and Witten proposed just this in Ref. [5]. If one believes that any acceptable RR field can be sourced by some D-brane, in other words, that if a nontrivial flux is supported on some cycle then that cycle is free to degenerate and at the point where it degenerates one will find a D-brane sourcing the flux, then one version of the K-theory classification of RR field strengths is as follows.

Consider type II string theory on a spacetime which is topologically $\mathbb{R}\times M$ where $\mathbb{R}$ may be thought of as the time direction and M is compact. Recall that conserved Dp-brane charges are characterized by Dp-branes that extend along the time direction and also wrap a p-cycle $N\subset M$. If the spacetime metric were really the Cartesian product of $\mathbb{R}\times M$ then such an eternal configuration would be inconsistent because the sourced flux would have nowhere to go, however no such assumption is made on the metric. Thus, as in the familiar cases of D-branes on $AdS^p\times M$, configurations in which cycles of M are wrapped are consistent because M gets large sufficiently quickly far from the origin of AdS^p. So we will not worry about the problem of fluxes having no place to go, essentially because $\mathbb{R}\times M$ is noncompact. We have argued that in type IIA p is even and conserved D-brane charges are classified by $\mathrm{K}^1_H(M)$, whereas in type IIB p is odd and conserved D-brane charges are classified by $\mathrm{K}^0_H(M)$.

To extend this classification to fluxes one needs the above conjecture, that fluxes are in one to one correspondence with the branes that source them. More concretely, consider time-dependent Dp-brane configurations in which a Dp-brane nucleates, wrapping a contractible p-cycle, sweeps out a $(p+1)$-cycle N and then shrinks into oblivion on the other side of N. If N fails to lift to a K-theory class, for example if there is a Freed-Witten anomaly or if it is nonrepresentable, then when the Dp-brane disappears a lower-dimensional D-brane will remain. However if N does lift to a K-theory class then this process, while perhaps not energetically favorable, is possible.

Recall from Section 3 that Dp-branes are violations of the Bianchi identity

$$ddC_{7-p} = \mathrm{PD}(\mathrm{D}p\text{-brane}) = \mathrm{PD}(N). \tag{6.31}$$

Therefore one can use Stoke's theorem to write Gauss' Law (3.14) for the flux over any $(8-p)$-cycle N' that links the Dp-brane worldvolume N. As our Dp-brane only existed during a finite lifetime, we can choose the union of a timeslice before and after the D-brane came into existence. The difference between the unimproved field strength dC_{7-p} on the timeslice before and after will, using (6.31) and an integral cohomology version of Stoke's theorem, be the Poincaré dual of N. As N sweeps out a $(p+1)$-dimensional cycle in the 9-manifold M, the Poincaré dual will be a degree $(8-p)$ integral cohomology class of M

$$x = \mathrm{PD}(N) \in \mathrm{H}^{8-p}(M). \tag{6.32}$$

It was crucial that the slices which link the brane are compact. Technically this is necessary because otherwise the linking number would vanish and Stoke's theorem would not apply. Also noncompact slices would be disastrous as Poincaré duality is an isomorphism of the cohomology groups of compact spaces. In fact, the fluxes on a noncompact cycle are not quantized, as Dirac's argument requires a nontrivial linking number and so fluxes on noncompact cycles cannot be quantized by either integral cohomology or integral K-theory. Thus only the fluxes along M, corresponding in the language of differential forms to forms with all legs along M, will be quantized. In the language of electrodynamics these are the magnetic fluxes, which are related by Hodge duality to the electric fluxes, which by definition have one leg along the time direction. This distinction is important, as RR field strengths in the democratic formulation [10] of type II supergravity, which we are using, are self-dual

$$G_p = \star G_{10-p} \tag{6.33}$$

where $\star$ is the Hodge star. The self-duality implies that the magnetic fields completely determine the electric fields, and vice versa. Thus a classification of the magnetic fields is a complete classification of all RR fields.

In particular, if we start with the trivial flux $dC = 0$ then the conjectured relation between D-branes and RR field strengths implies that we can turn on any flux x by nucleating a D-brane which then sweeps out the dual cycle N. Conversely, given any consistent nucleation, sweeping and annihilation of a D-brane one creates some flux. Thus RR field strengths dC_{7-p} on M are conjectured to be classified by the group that classifies not Dp-brane charges but the surfaces that they sweep out in M. In the case of IIA this is the twisted K-group $\mathrm{K}^0_H(M)$ and in the case of type IIB it is the twisted K-group $\mathrm{K}^1_H(M)$.

Recall that all D-branes carry homology charges. Similarly all fluxes carry cohomology charges. The AHSS eliminates some of these charges and quotients others to arrive at the twisted K-theory classification. We identified the eliminated D-brane charges as those charges which are carried only by anomalous branes, and we noted that the anomalies may be canceled by inserting branes that end on certain submanifolds of the anomalous branes. We also identified quotiented D-brane charges as nonconserved, and saw that they are carried by the branes whose insertions cancel anomalies. Thus we provided a physical interpretation for the mismatch between the homology and the K-theory classification of branes.

A similar physical interpretation exists for the mismatch between the cohomology and K-theory classification of fluxes. We have seen that fluxes that do not lift to K-theory are created by Freed-Witten anomalous branes. These Dp-branes fail to decay, leaving behind lower dimensional D-branes. Thus RR field strengths whose cohomology classes fail to lift to K-theory can only exist in the presence of these lower-dimensional D-branes. This is similar to the observation that D-branes whose charges do not lift to K-theory are only consistent in the presence of lower dimensional D-brane insertions.

	RR Charges	**RR Field Strengths**
In Image of d_p:	Charge is not conserved	Pure gauge
Not In Kernel of d_p:	Baryons and MMS instantons	Contains D-branes

Table 6.1: Interpretation of those RR charges and field strengths that are in the image of the AHSS differentials and those that are not in the kernel

When discussing the worldvolume theories of branes in supergravity we argued that the field strengths dC are only well-defined up to large gauge transformations, which at the level of differential forms are the wedge products of H with another integral class. By comparison, RR field strengths that lift to the zero twisted K-theory class are those that are in the image of the AHSS differentials, which include the image of $H\wedge$ as well as torsion and Massey product corrections. This leads one to the conjecture that while the gauge symmetries of the classical supergravity (3.9) shift dC by forms in the image of $H\wedge$, the gauge symmetries of the full quantum string theory shift dC by the classes in the image of the AHSS differentials. Thus RR field strengths whose cohomology classes lift to the trivial K-class are conjectured to be pure gauge.

Summarizing, magnetic RR field strengths dC appear to be classified by twisted K-theory. While naively they are classified by cohomology, those whose cohomology class does not lift to K-theory can only exist in the presence of unstable D-branes and those whose cohomology class lifts to the trivial K-theory class are pure gauge.

One can simultaneously classify D-branes and RR fluxes in configurations that contain both, as the RR fluxes that lift to twisted K-theory do not require the insertion of any D-branes. However one must remember in such cases that the true RR flux is not just the corresponding K-theory element, but one must also add the RR flux sourced by the D-brane. If the D-brane is unstable then it is a trivial element in K-theory, and one must add an RR flux which is a cohomology class that does not lift to K-theory such that an AHSS differential of the cohomology class of the RR flux yields the cohomology class which is Poincaré dual to the unstable D-brane.

Notice that we have classified RR field strengths dC, but have not classified RR potentials C. Gauge potentials are continuously valued and so are not classified by ordinary twisted K-theory. It has been conjectured that they are classified by a generalization of twisted K-theory known as differential K-theory [43], which characterizes not just the topologies of bundles but also their Wilson loops.

7 Examples

The literature contains a modest number of examples of compactifications in which the twisted K-theory of the spacetime or of a spatial slice has been computed and has been successfully tested against an independent calculation of the D-brane spectrum. This is because the full spectrum of stable D-branes can only be computed in a few special classes of compactifications. More generally applicable D-brane classifying schemes, such as derived categories [44, 45], only classify BPS D-branes. D-branes which wrap torsion homology classes, for example, may be stable but are never BPS nor even calibrated and so are missed by schemes which are based on supersymmetry or calibrations.

However in some special classes of simple string models one can calculate the full D-brane spectrum. The simplest class of examples consists of two-dimensional topological field theories in which the target space is a finite set of points. In Ref. [12] it was proven, using sewing conditions for the field theories, that boundary conditions are indeed classified by K-theory. While this analysis relied on the fact that the spacetimes considered are *spec* of semisimple Frobenius algebras, the authors hint that the K-theory classification of D-branes may extend to general topological string theories.

More traditional tests of the K-theory classification are in the context of exactly solvable conformal field theories. One set of such exactly solvable conformal CFTs is the WZW model corresponding to a Lie group G, which describes strings propagating on the group manifold G. When G is semisimple and compact one can compute the D-brane spectrum directly from the admissible boundary conditions of the worldsheet [46], and one invariably finds twisted K-theory with a twist determined by the level of the affine Lie algebra corresponding to G. In this section we will review the first two cases in which the spectra derived from CFT boundary conditions and from twisted K-theory have been compared, the $SU(2)$ and $SU(3)$ WZW models.

7.1 The SU(2) WZW Model

The group manifold of $SU(2)$ is the three-sphere. Three-spheres appear in spacetime, for example, in type II string theory on a topologically trivial space with k flat, coincident NS5-branes. The NS5-branes are linked by a three-sphere and, although the dilaton diverges close to the NS5-brane, fundamental strings away from the brane are described by an exactly solvable CFT. The radial coordinate of the embedding of a fundamental string is described by a nonlinear dilaton model and the embedding on the linking three-sphere is described by an SU(2) WZW model at level $k-2$, which is unitary when $k>2$. Using Gauss' Law, the integral of the H flux on the linking 3-sphere is equal to k

$$\int_{S^3} H = k \tag{7.1}$$

and so we will expect our twisted K-theory to be twisted by the class

$$k \in \mathrm{H}^3(S^3) = \mathbb{Z}. \tag{7.2}$$

In the sequel we will not restrict our attention to this embedding of the CFT in type II superstring theory, but rather we will consider the WZW model on its own. This CFT has a 3-dimensional target space and so is not critical when coupled to 2-dimensional gravity, in particular it has a central charge of

$$c_{SU(2),k-2} = \frac{3k-6}{k-1}. \tag{7.3}$$

Topologically inequivalent boundary conditions of the CFT wrap conjugacy classes of the three-sphere, which are points or two-spheres.

Surprisingly only a discrete set of conjugacy classes is consistent [30]. This is because a boundary state on a two-sphere is a wavefunction on each choice of loop in the two-sphere on which a fundamental string may end. The phase of this wavefunction must be well-defined on the loopspace, or else the boundary state is ill-defined. In particular if one takes the loop and moves if off of the 2-sphere to the north and then pulls it back on from the south, dragging it to its original position, one needs to arrive at the same state. In general, however, the state shifts by the integral of the worldvolume field strength F over the 2-sphere, reflecting the fact that the loop on which the string ends is the trajectory of an electrically charged particle in the D-brane's worldvolume gauge theory. This shift is also easily calculable in the CFT. The authors of Ref. [47] found that, imposing a particular boundary condition on the string worldsheets, there are only k possible wrappings in which this shift is a multiple of 2π, and so only k conjugacy classes may be wrapped.

Today we know that any 2-dimensional submanifold may be wrapped, but in Ref. [19] it was shown that the the D-branes will feel a Myers dielectric force [29] that integrates to a potential energy function which has one of k minima, corresponding to the k special conjugacy classes found in Ref. [47]. The minimum applicable to a given brane depends on the integral of F on its worldvolume. Of these k conjugacy classes, $k-2$ are two-spheres at constant, evenly spaced latitudes and two are points, the north and south poles. The locations of the poles may be changed by applying an inner automorphism of $SU(2)$ to the ansatz of boundary conditions investigated in [47]. The existence of yet more general boundary conditions remains an open question.

Now we are ready to compute the twisted K-theory of the 3-sphere with k units of H flux and compare it to the conformal field theory result that there should be k stable even-dimensional D-brane configurations and no odd-dimensional D-brane configurations. The homology and cohomology of the three-sphere, like that of any sphere, is torsion-free and is generated by a single generator at the lowest and highest dimension

$$\begin{aligned} \mathrm{H}_0(S^3) &\cong H^3(S^3) = \mathbb{Z}, \qquad & \mathrm{H}_1(S^3) &\cong H^2(S^3) = 0 \\ \mathrm{H}_2(S^3) &\cong H^1(S^3) = 0, \qquad & \mathrm{H}_3(S^3) &\cong H^0(S^3) = \mathbb{Z} \end{aligned} \tag{7.4}$$

where $\cong$ is Poincaré duality. Recall that the H flux H is the element k of H^3.

As the cohomology ring is trivial beyond dimension three, all cohomology operations of dimension greater than three are trivial. Therefore only the dimension three AHSS differential d_3 may be nonzero. In addition the image of Sq^3 is always torsion but there are no torsion classes and so the Sq^3 term in d_3 is trivial. It is also trivial for dimensional reasons. Therefore only the $H\cup$ term in d_3 may be nonvanishing. This augments the degree of a cohomology class by 3, and so it kills all of the 3-classes since its image would be a six class but there are no nonzero six classes.

If we let a and b be the generators of H^0 and H^3 respectively, then we have found

$$d_3 : \mathrm{H}^0(S^3) \longrightarrow \mathrm{H}^3(S^3) : a \mapsto kb, \qquad d_3 : \mathrm{H}^3(S^3) \longrightarrow \mathrm{H}^6(S^3) : b \mapsto 0. \tag{7.5}$$

Therefore the kernel of d_3 consists of all elements in $\mathrm{H}^3(S^3)$ while the image consists of those elements of $\mathrm{H}^3(S^3)$ which are divisible by k. The fact that the image is a subset of the kernel was guaranteed by the nilpotency of d_3.

The twisted K-theory of S^3, as a set, is just the quotient of the kernel of d_3 by its image. In particular there are no even classes in the kernel and so $\mathrm{K}^0_H(S^3)$ is trivial

$$\mathrm{K}^0_H(S^3) = \frac{\mathrm{Ker}(d_3 : \mathrm{H}^{\mathrm{even}} \longrightarrow \mathrm{H}^{\mathrm{odd}})}{\Im(d_3 : \mathrm{H}^{\mathrm{odd}} \longrightarrow \mathrm{H}^{\mathrm{even}})} = \frac{0}{0} = 0. \tag{7.6}$$

On the other hand K^1_H is nontrivial. The kernel of d_3 acting on the odd cohomology, in particular acting on $\mathrm{H}^3(S^3)$, is all of $\mathrm{H}^3(S^3)$ and so is a copy of the integers $\mathbb{Z}$ which contains the image $k\mathbb{Z}$ as a proper subgroup

$$\mathrm{K}^1_H(S^3) = \frac{\mathrm{Ker}(d_3 : \mathrm{H}^{\mathrm{odd}} \longrightarrow \mathrm{H}^{\mathrm{even}})}{\Im(d_3 : \mathrm{H}^{\mathrm{even}} \longrightarrow \mathrm{H}^{\mathrm{odd}})} = \frac{\mathrm{Ker}(d_3 : \mathrm{H}^3 \longrightarrow \mathrm{H}^6)}{\Im(d_3 : \mathrm{H}^0 \longrightarrow \mathrm{H}^3)} = \frac{\mathbb{Z}}{k\mathbb{Z}} = \mathbb{Z}_k. \tag{7.7}$$

These may be identified with the K-homology groups that classify D-branes via Poincaré duality

$$\mathrm{K}^H_0(S^3) \cong \mathrm{K}^1_H(S^3) = \mathbb{Z}_k, \qquad \mathrm{K}^H_1(S^3) \cong \mathrm{K}^0_H(S^3) = 0. \tag{7.8}$$

The triviality of K^H_1 agrees with the CFT expectation that there are no odd-dimensional branes. K^H_0 contains k elements which again is in agreement with the CFT result that the level $k-2$ $SU(2)$ WZW model contains k inequivalent D-brane embeddings.

Physically the quotient by $k\mathbb{Z}$ corresponds to the fact that a D2-brane may sweep out the 3-sphere, and the cancellation of its Freed-Witten anomaly causes it to change the D0-brane charge by k units during this process. As a result D0 charge is only conserved modulo k. These D2-branes are the MMS instantons described in Subsec. 6.1. The fact that one restricts to the kernel of d_3, thus eliminating all of H^0, corresponds to the fact that all space-filling D-branes are FW anomalous since the integral of H over their worldvolumes is equal to k and not to zero.

7.2 The SU(3) WZW Model

The twisted K-theory of the group manifold $SU(3)$ was calculated in Ref. [7]. The authors used their result to classify conserved D-brane charges in type II string theory on $\mathbb{R}^{1,1} \times SU(3)$, where branes do not wrap the spatial $\mathbb{R}$, so as to avoid the tadpoles resulting from D9-brane charge. Again their results were successfully compared against the CFT expectations, and against the expectation that consistent branes on a manifold of dimension 9 or less are precisely those whose Freed-Witten anomaly vanishes. Here we will sketch their calculation of the K-groups.

We begin by reviewing the nontrivial homology and cohomology classes of the group manifold $SU(3)$, which are identical to that of the similar manifold $S^3 \times S^5$

$$\begin{aligned} &\mathrm{H}_0(SU(3)) \cong H^8(SU(3)) = \mathbb{Z}, \qquad &&\mathrm{H}_3(SU(3)) \cong H^5(SU(3)) = \mathbb{Z} \\ &\mathrm{H}_5(SU(3)) \cong H^3(SU(3)) = \mathbb{Z}, \qquad &&\mathrm{H}_8(SU(3)) \cong H^0(SU(3)) = \mathbb{Z}. \end{aligned} \tag{7.9}$$

Notice that again there are no torsion groups and so Sq^3 is trivial. Therefore d_3 is just the cup product with the H flux, which is k times the generator of H^3 when the WZW model is at level $k-3$.

As d_3 increases the degree by 3, it may only be nontrivial on H^0 and H^5, and in fact it is. Let x^i be the generator of H^i. Then the nontrivial actions of d_3 are just

$$d_3 : \mathrm{H}^0(S^3) \longrightarrow \mathrm{H}^3(S^3) : x^0 \mapsto kx^3, \qquad d_3 : \mathrm{H}^5(S^3) \longrightarrow \mathrm{H}^8(S^3) : x^5 \mapsto kx^8 \tag{7.10}$$

while d_3 annihilates x^3 and x^8. The second step in the AHSS is then the quotient of the kernel of d_3, consisting of $\mathrm{H}^3 \oplus \mathrm{H}^8$ by its image, which is the sublattice of the kernel which is divisible by k

$$\begin{aligned} E_2^0 &= \frac{\mathrm{Ker}(d_3 : \mathrm{H}^{\mathrm{even}} \to \mathrm{H}^{\mathrm{odd}})}{\Im(d_3 : \mathrm{H}^{\mathrm{odd}} \to \mathrm{H}^{\mathrm{even}})} = \frac{\mathrm{H}^8}{k\mathrm{H}^8} = \frac{\mathbb{Z}}{k\mathbb{Z}} = \mathbb{Z}_k \\ E_2^1 &= \frac{\mathrm{Ker}(d_3 : \mathrm{H}^{\mathrm{odd}} \to \mathrm{H}^{\mathrm{even}})}{\Im(d_3 : \mathrm{H}^{\mathrm{even}} \to \mathrm{H}^{\mathrm{odd}})} = \frac{\mathrm{H}^3}{k\mathrm{H}^3} = \frac{\mathbb{Z}}{k\mathbb{Z}} = \mathbb{Z}_k. \end{aligned} \tag{7.11}$$

The D-branes that have survived up to this point are Poincaré dual to the third and eighth cohomology classes, identifying them as D0 and D5-branes wrapping a point and a 5-cycle in SU(3) respectively. The nontrivial image of d_3 is the result of two MMS instantons, a D2-brane which violates D0 charge and a baryonic D7-brane which violates D5 charge. The branes excluded because they are not in the kernel of d_3 are the D3-brane wrapping the S^3 with D1 insertions and the D8-brane wrapping all of $SU(3)$ with k D6-brane insertions. The D6 insertions, for example, wrap the cycle N which is Poincaré dual to x^3 and also extend along one time direction.

Had we been considering the direct product of spheres $S^3 \times S^5$, Eq. (7.11) would have already been the twisted K-theory. But SU(3) is somewhat different. The 5-manifold N is not $spin^c$, and so a brane wrapping this 5-manifold may suffer from a Freed-Witten anomaly. The nontrivial cohomology of the non-$spin^c$ 5-cycle N is

$$\mathrm{H}^0(N) = \mathbb{Z}, \qquad \mathrm{H}^3(N) = \mathbb{Z}_2, \qquad \mathrm{H}^5(N) = \mathbb{Z}. \tag{7.12}$$

As N is not $spin^c$, W_3 is the nontrivial class in $\mathrm{H}^3(N)$. The pushforward of this class onto the cohomology of $SU(3)$ is $\mathbb{Z}_2$-torsion, but there are no torsion classes in the cohomology of $SU(3)$, which as we noted implies that Sq^3 is zero. Therefore W_3 is in the kernel of the pushforward that defines Sq^3 and so does not affect d_3. However if $W_3 + H$ is nonzero then the brane is nonetheless anomalous and so does not carry a K-theory charge. Therefore it must fail to be in the kernel of a higher AHSS differential. As the only other odd spacing between cohomology classes is 5, anomalous branes must not be in the kernel of d_5.

We still have not yet commented on when $W_3 + H$ is nonzero. In Ref. [7] the authors found that

$$W_3 + H = 1 + k \in \mathbb{Z}_2 = \mathrm{H}^3(N) \tag{7.13}$$

and so the D5 is anomalous if and only if k is even. Therefore d_5 is 2-torsion, and is nontrivial when k is even, which identifies it as

$$d_5 x^3 = \begin{cases} \frac{k}{2} x^8 & \text{if } k \text{ is even} \\ 0 & \text{if } k \text{ is odd} \end{cases}, \qquad d_5 x^8 = 0. \tag{7.14}$$

While nontrivial actions of d_3 describe Dp-branes on which D$(p-2)$-branes end, those of d_5 describe Dp-branes on which D$(p-4)$-branes end. In particular, when k is even, a baryonic D5-brane wrapping N has a FW anomaly which is canceled by $k/2$ D1-brane insertions which each end at a point on each timeslice. Similarly, a D4-brane which sweeps out the 5-cycle absorbs $k/2$ D0-branes, and so D0-brane charge is conserved modulo $k/2$ and not modulo k as one would have suspected using only (7.11).

One may now calculate the twisted K-theory of $SU(3)$ with k units of H flux. When k is odd, d_5 is trivial and so as a set the K-groups are isomorphic to E_2

$$\mathrm{K}_0^{H=k} \cong \mathrm{K}^0_{H=k} = \mathbb{Z}_k, \qquad \mathrm{K}_1^{H=k} \cong \mathrm{K}^1_{H=k} = \mathbb{Z}_k. \tag{7.15}$$

On the other hand when k is even one needs to take the quotient of the kernel of d_5 in E_2 by its image. d_5 kills every class in E_2^0 and so the kernel includes all of $E_2^0 = \mathbb{Z}_k$. On the other hand it only kills the odd classes in E_2^1, and so its odd kernel is $2E_2^1 = \mathbb{Z}_{k/2}$. The image does not contain any nonzero classes in E_2^1, and contains, besides zero, only the single class

$$\frac{k}{2} \in \mathbb{Z}_k = E_2^0 \tag{7.16}$$

in E_2^0, therefore one needs to quotient only by the $\mathbb{Z}_2$ subgroup of E_2^0 generated by the element

$k/2$. In all we have then found [7]

$$\begin{aligned}
\mathrm{K}_0^{H=k} \cong \mathrm{K}^0_{H=k} &= \frac{\mathrm{Ker}(d_5 : E_2^0 \to E_2^1)}{\Im(d_5 : E_2^1 \to E_2^0)} = \frac{\mathbb{Z}_k}{\mathbb{Z}_2} = \mathbb{Z}_{k/2} \\
\mathrm{K}_1^{H=k} \cong \mathrm{K}^1_{H=k} &= \frac{\mathrm{Ker}(d_5 : E_2^1 \to E_2^0)}{\Im(d_5 : E_2^0 \to E_2^1)} = \frac{\mathbb{Z}_{k/2}}{0} = \mathbb{Z}_{k/2}.
\end{aligned} \tag{7.17}$$

In particular there are $k/2$ topologically distinct even and also odd brane charges when k is even, whereas there are k even and k odd distinct charges when k is odd.

8 Problems

While the K-theory classifications of RR fields and D-branes has been reasonably successful, it suffers from several fundamental problems.

8.1 K-theory vs Homology Revisited

One of the easiest objections to make is that D-branes can wrap any homology cycle, and so D-branes can be classified by homology. D-branes wrapping homology cycles which are not in the kernels of some of the AHSS differentials will not carry K-theory charges, and so they will necessarily have anomalies which are canceled by the insertions of lower dimensional D-branes. These configurations, as we have frequently reiterated, are the baryons introduced by Witten in Ref. [13]. They are not inconsistent if the inserted branes are not themselves anomalous [26] and if the other end of the inserted branes can go to infinity or to a horizon or boundary.

However for some practical purposes one may wish not to include baryonic branes, as they often have infinite energies. Upon Kaluza-Klein reduction they tend to correspond to particles which are confined, for example to 't Hooft-Polyakov monopoles in an $N=2$ gauge theory whose supersymmetry is softly broken to $N=1$. Such monopoles may be confined by some number of vortices depending on the monopole charges [48], and each of these vortices is the dimensional reduction of one of the inserted branes. An example in which the pattern of the confinement in $N=1$ gauge theories can be read from relative homology has appeared in Ref. [49] and this discussion is generalized in Ref. [50]. Thus the homology classification of branes captures all brane charges, while the K-theory classification only captures the charges of nonbaryonic branes. Therefore the choice of classification depends on the physical question being asked.

Not only are the branes which are not closed under the AHSS differentials legitimate states, but the brane charges in the image are in a sense conserved. When a brane that carries a nontrivial homology charge but a trivial K-theory charge is destroyed by an AHSS instanton, the instanton leaves behind a RR field strength. This RR field strength in turn can be used to reconstruct the original brane charge. Thus, while the D-brane charge itself is not conserved, a combination of the brane charge and the flux is conserved. Combining this with some information from the integrals of the RR connections one reproduces the fact that improved field strengths are precisely gauge invariant and so there are conserved charges valued in de Rham cohomology. Thus, for some questions of conserved charges, the answer is not twisted K-theory but rather de Rham cohomology. One arrives at twisted K-theory when one drops all information about RR fluxes and tries to only classify the D-branes.

Similarly one may argue that RR field strengths should be classified by cohomology. Improved fields strengths are gauge invariant and are classified by twisted $d+H\wedge$ de Rham cohomology. However one may also argue that unimproved field strengths should be classified by integral cohomology. Some integral cohomology classes do not lift to twisted K-theory classes, but the corresponding field strengths are not inconsistent, they merely correspond to configurations with RR charge. Thus, if one is willing to consider configurations with RR charges, then unimproved RR field strengths may assume any value in integral cohomology and not just those which lift to K-theory classes.

Again one may ask whether RR field strengths which are exact under the AHSS differentials should be identified with zero. Such RR field strengths carry a K-theory charge which is equal to zero, as they can be shifted away by a large gauge transformation. These large gauge transformations also add an integral cohomology class to an RR gauge connection. Thus, just as the decay of K-theoretically trivial D-branes leads to an integral RR field strength, the decay of a K-theoretically trivial RR field strength leads to an integral RR gauge connection, for example

$$dC_p \longrightarrow dC_p - H \cup \Phi, \qquad C_{p-2} \longrightarrow C_{p-2} + \Phi \tag{8.1}$$

as in Eq. (3.9). While in the bulk such a shift, which is merely an integral shift of a Wilson loop, is unobservable, on the worldvolume of a D-brane it has a physical effect. The RR gauge connections couple to the worldvolume field strengths of D-branes, for example as a theta angle when the flux is codimension four. An integral shift of the theta angle leaves the gauge theory invariant, but may shift the charges assigned to some of the objects in a given solution. Thus in some applications one may wish to distinguish between the different integral classes of connections and so between the different cohomology classes of RR field strengths which are identical as K-theory classes.

Summarizing, D-branes are classified by both K-theory and by homology, and fluxes are classified by both K-theory and cohomology. However for certain physical applications one classification scheme may be preferable to the other. For example, as we saw in Sec. 7, twisted K-theory reproduces the spectrum of boundary states in some CFTs. We will see in Subsec. 9.2 that homology classes of D-branes classify the ranks of gauge groups in Klebanov-Strassler gauge theories whereas twisted K-theory classes classify the universality classes of the gauge theories.

8.2 The Star Problem

The RR field strengths of the democratic formulation of the type II supergravity are not independent. Instead they double count the degrees of freedom. This redundancy is killed by the star condition

$$G = \star G \tag{8.2}$$

which relates the improved p-form field strength to the improved $(10-p)$-form field strength. The operator $\star$ is the Hodge star, which is a continuous function of the metric. In particular, the ratio between two dual components of G varies continuously with the metric and so is in general irrational. One may think that this poses no problem, as G is a differential form which is not closed and so cannot be quantized in the usual way.

To see that this is a problem, consider the special case in which the NS 3-form H vanishes

$$H = 0, \qquad G_{p+1} = dC_p. \tag{8.3}$$

The $(p+1)$-forms dC_p are quantized, and so in this case G_{p+1} is quantized. This quantization is required for the D$(p-1)$-brane partition function to be well-defined, and in the K-theory perspective it is required because the dC's are Chern characters, which are always rational. However the ratio of two components depends on the star and so is generically irrational, which is a contradiction.

Physically this problem is solved by choosing a polarization, as is done in a similar context in Ref. [51]. That is, one must choose one independent set of half of the dC's, called a polarization, and quantize them, this is the largest set that can be simultaneously quantized. Then in the partition function one only sums over this half. Finally, one must check that the partition function would be the same given any other choice of polarization. For example, in QED on a compact spacetime one must choose between quantizing the electric and the magnetic field strengths, or perhaps some combination on various cycles, but it is impossible to quantize all of the fluxes on all of the cycles for a generic metric.

While physically this appears to be the correct way to make sense of the theory, it is difficult to interpret in the K-theory context. If one interprets the dC's as Chern characters, then only half of the Chern characters are well-defined. There appear to be three proposals for when such a choice can make sense.

First, one may need to replace K-theory with some sort of quantized version of K-theory. The K-groups $\mathrm{K}^0(M)$ and $\mathrm{K}^1(M)$ are isomorphic to the set of maps $[M \longrightarrow X]$ from M to a space X called the classifying space of K^0 or K^1

$$\mathrm{K}^0(M) = [M \longrightarrow BU(\infty)], \qquad \mathrm{K}^1(M) = [M \longrightarrow U(\infty)]. \tag{8.4}$$

Chern characters are pullbacks of cohomology classes corresponding to the homotopy classes of the classifying space X using the map (8.4) from spacetime to the classifying space of K-theory.

This is consistent with the fact that $BU(\infty)$ has nontrivial homotopy at all of the even degrees and $U(\infty)$ has nontrivial homotopy at all of the odd degrees

$$\pi_{2k}(BU(\infty)) = \pi_{2k+1}(U(\infty)) = \mathbb{Z}, \qquad \pi_{2k+1}(BU(\infty)) = \pi_k(U(\infty)) = 0 \tag{8.5}$$

and correspondingly K^0 is characterized by even-degree Chern characters and K^1 is characterized by odd degree Chern characters.

We want to choose a polarization, which means that only some of the Chern characters and so only some of the pullbacks may be well-defined. Only some of these pullbacks will be well-defined if, for example, one replaces these maps by sections of bundles in which the classifying space is nontrivially fibered over the spacetime. As no global section exists for a nontrivial fibration, one can only determine local sections over a submanifold of the base. The choice of submanifold may give a choice of polarization. However unfortunately while nontrivial $U(\infty)$ bundles over 10-dimensional manifolds always exist, $BU(\infty)$ bundles are not characterized by a 10-dimensional class, like ch_5 for a $U(\infty)$ bundle, and so it is not clear what fibration to use. Thus it is easier to quantize K^1 than K^0.

A second way out is to consider a manifold in which there is a natural polarization which is the cohomology of a submanifold, and then to only consider the K-theory of the submanifold. In fact, we have already done this when we applied K-theory to classify conserved D-brane charges. In this case our spacetime was $\mathbb{R} \times M^9$. When one direction is $\mathbb{R}$ it need not be quantized, but if we only insist that a K-class is defined on M^9 we may even choose to replace $\mathbb{R}$ with S^1 so that spacetime is compact. In either case, we only quantized the fluxes on the compact M^9, which were classified by the twisted K-theory of M^9. This is a polarization of the cohomology and of the K-theory by the Künneth theorem

$$\begin{aligned}
\mathrm{H}^p(M^9 \times S^1) &= \mathrm{H}^p(M^9) \otimes \mathrm{H}^0(S^1) \oplus \mathrm{H}^{p-1}(M^9) \otimes \mathrm{H}^1(S^1) \\
&= \mathrm{H}^p(M^9) \otimes \mathbb{Z} \oplus \mathrm{H}^{p-1}(M^9) \otimes \mathbb{Z} = \mathrm{H}^p(M^9) \oplus \mathrm{H}^{p-1}(M^9) \\
\mathrm{K}^p(M^9 \times S^1) &= \mathrm{K}^p(M^9) \otimes \mathrm{K}^0(S^1) \oplus \mathrm{K}^{p-1}(M^9) \otimes \mathrm{K}^1(S^1) \\
&= \mathrm{K}^p(M^9) \otimes \mathbb{Z} \oplus \mathrm{K}^{p-1}(M^9) \otimes \mathbb{Z} = \mathrm{K}^p(M^9) \oplus \mathrm{K}^{p-1}(M^9)
\end{aligned} \tag{8.6}$$

which demonstrates that the cohomology (K-theory) of $M^9 \times S^1$ is just two copies of the cohomology (K-theory) of M^9, one in which the forms have a leg around the circle and one in which they do not.

If the circle direction is considered to be time then the half of the classes with a leg on a circle are the electric fluxes and the other half are the magnetic fluxes. This analysis works equally well with the circle replaced by $\mathbb{R}$ as it was above, but then there is no Dirac quantization argument for the quantization of electric fluxes. In either case, however, it is consistent to quantize only the magnetic fluxes as we have done in the classification of conserved RR charges. Thus, when spacetime is of the form $M^9 \times \mathbb{R}^1$ or $M^9 \times S^1$ the star problem has a natural solution which allows one to continue to classify brane charges and magnetic RR field strengths as elements of the K-theory of M^9. It is still unknown whether the topology of our universe is of such a form, if it began at a fixed time in a big bang then it is not.

A third resolution to the star problem is to eliminate half of the Chern characters by replacing the classifying space of K-theory by something with less cohomology, in particular, which does not include the cohomology classes of the Chern characters to be eliminated. While the first resolution only worked for K^1, this resolution only works for K^0. The problem is that K^1 is classified by 5-classes $ch_{5/2}$ which are interrelated by the Hodge duality, and so one cannot simply eliminate the cohomology class in the classifying space $U(\infty)$ of K^1 whose pullback is $ch_{5/2}$, as one would lose all of the 5-classes. Physically, such a solution would either annihilate all of the components of IIB supergravity's self-dual 5-form G_5 or it would eliminate none of them, but it would not eliminate the one-half required for a polarization.

For example, we may choose a polarization of type IIA supergravity in which G_0, G_2 and G_4 are quantized and independent and G_6, G_8 and G_{10} are found by solving the star condition (8.2). Recall that G_{2p} is the Chern character ch_p which is the pullback of the cohomology class of the generator

$$\pi_{2p}(BU(\infty)) = \mathbb{Z}. \tag{8.7}$$

Therefore we need to replace $BU(\infty)$, the classifying space of K^0, with a new space whose sixth, eighth and tenth homotopy groups are trivial.

Rather than working with $BU(\infty)$, recall the fact that, by definition, $BU(\infty)$ is the classifying space not only of K^0 but also of $U(\infty)$ bundles. The homotopy groups of the classifying space BP of P-principle bundles are related to the homotopy groups of the fiber P by the theorem

$$\pi_k(BP) = \pi_{k-1}(P) \tag{8.8}$$

therefore the classes of $\pi_{2p}(BU(\infty)) = \mathbb{Z}$ correspond to the classes $\pi_{2p-1}(U(\infty)) = \mathbb{Z}$. Thus to eliminate the sixth, eighth and tenth homotopy groups of $BU(\infty)$ it suffices to remove the fifth, seventh and ninth homotopy groups of $U(\infty)$. This obviously drastically changes the Lie group $U(\infty)$, and the resulting Lie group is not uniquely defined as, for example, the higher homotopy groups are invisible to the 10-dimensional physics and so can apparently be chosen at will.

One particularly simple choice of truncation of $U(\infty)$ is LE_8, the centrally extended loop group of E_8, this is affine E_8 at a level which we will identify with G_0. The identification of the level with G_0 reproduces several Freed-Witten anomalies as topological obstructions to the existence of this bundle [38]. Considering massless IIA, the Romans mass G_0 is equal to zero and so the LE_8 fiber is just the Cartesian product of the free loop group of E_8 with $U(1)$. The only nontrivial homotopy groups, up to dimension 14, of this fiber are

$$\pi_1(LE_8) = \pi_2(LE_8) = \pi_3(LE_8) = \mathbb{Z}. \tag{8.9}$$

At first this may seem too big. We wanted the homotopy groups to be subgroups of those of $U(\infty)$, but $\pi_2(U(\infty)) = 0$ and so we have also added a class. This new class will be characterized by a characteristic 3-class which in the mathematics literature is called the Dixmier-Douady class. In Ref. [40] it was argued that this new class be identified with the H flux. Evidence came from the fact that, again, topological obstructions to the existence of the bundle reproduce known examples of Freed-Witten anomalies with this interpretation. Another suggestion that this new 3-class is the H field comes from the fact that an LE_8 bundle over a ten-manifold M carries the same data as an E_8 bundle over a circle bundle over the 10-manifold, that is to say, over the M-theory spacetime. An E_8 bundle is characterized by a four class. If one considers the E_8 bundle of [37, 6] then this four class is the M-theory 4-form and its dimensional reductions give the IIA 4-form RR field strength and 3-form NS field strength. In [40] it was argued that the characteristic 3-class of the LE_8 bundle is precisely the dimensional reduction of this M-theory 4-form, and so is indeed the NS 3-form.

While the replacement of the classifying space of K^0 with LE_8 may appear far-fetched, it simultaneously achieves two objectives. First, it naturally chooses a polarization and so solves the star problem in any background. This is an improvement over the second solution which required that every spatial slice of spacetime have the same topology. Secondly it naturally integrates the NS 3-form flux into the bundle framework, as it must be as IIA string theory admits 9-11 flips which interchange the H flux and the RR 4-form field strength dC_3. As an added bonus, if one correctly identifies a gravitational correction to the relation between dC_3 and the Pontrjagin class of the M-theory E_8 bundle, it produces the correct topology of the E_8 gauge bundles of the end of the world, as found from gravitational anomaly cancellation in Ref. [52].

8.3 S-duality Covariance

We have seen that the star problem does not appear in several variations of the K-theory classification. However all mild variations of the K-theory classification suffer from a more fundamental problem. All schemes treat the NSNS H field as part of the background data in which one seeks to classify RR fields or charges. Indeed, the H flux in many ways resembles the topological data more than it resembles the RR fields. For example, the topology of spacetime is determined by the metric, which is also a NSNS field and it is mixed with the H field, for example, in the Seiberg-Witten map [7] and also in T-duality [54] as we will review in Subsec. 9.1. By contrast, RR fields mix among themselves under T-duality.

While one can argue that morally the H field is or is not a part of the topological background data, in type IIB string theory the H flux and the RR field strength dC_2 are related by S-duality and this S-duality leaves the topology of the spacetime invariant. Therefore, in a given background, dC_2 and H must be classified identically. In Subsec. 9.2 we will argue that this means that one can, for example, fix dC_2 as part of the background data and then classify the S-duals of the D-branes by twisted K-theory with twist equal to dC_2. However none of these S-duals can be the correct classification, as every one of them gives a different consistency condition for the fields dC_2 and H.

The inconsistency of the K-theory classification and S-duality was demonstrated concretely in Ref. [6]. Consider the twisted K-theory classification of RR field strengths. An RR field strength lifts to K-theory if it is annihilated by d_3, for example the 3-form field strength dC_2 lifts if it satisfies

$$0 = d_3(dC_2) = (Sq^3 + H\cup)dC_2 = dC_2 \cup dC_2 + H \cup dC_2. \tag{8.10}$$

Here we have used the fact that Sq^3 squares three-classes. If dC_2 does not satisfy Eq. (8.10), it is still a consistent field configuration but the configuration will carry a D3-brane charge which is Poincaré dual to $d_3(dC_2)$, in other words

$$Q_{D3} = \mathrm{PD}(dC_2 \cup dC_2 + H \cup dC_2) \tag{8.11}$$

where PD is the Poincaré dual.

D3-brane charge is believed to be S-duality invariant. We will assume that even torsion D3-brane charges, like those in the image of Sq^3, are S-duality invariant. The authors of Ref. [6] then performed the S-duality transformation

$$dC_2 \longrightarrow dC_2, \qquad H \longrightarrow H + dC_2 \tag{8.12}$$

under which the D3-charge (8.11) becomes

$$Q_{D3} = \mathrm{PD}(dC_2 \cup dC_2 + H \cup dC_2 + dC_2 \cup dC_2) = \mathrm{PD}(H \cup dC_2) \tag{8.13}$$

where we have used the fact that $dC_2 \cup dC_2$ is $\mathbb{Z}_2$ torsion, as is the cup product of any odd degree class with itself. Notice that the charges (8.11) and (8.13) in general do not agree, unless one imposes that $dC_2 \cup dC_2 = 0$ which would correspond to imposing that W_3 and H must vanish independently on any D-brane worldvolume, a conjecture which has not been supported by the analysis of worldsheet anomalies in [14]. Therefore the K-theory classification, which rests upon (8.11), appears to be inconsistent with S-duality.

We can see this inconsistency in the language of MMS instantons by noting that NS5-branes may also form, sweep out a cycle and decay. In type IIB S-duality guarantees that NS5-branes will also be subject to a Freed-Witten anomaly

$$Q_{D3} = W_3 + dC_2 \tag{8.14}$$

which will be canceled by the insertion of Q_{D3} D3-branes. However in the derivation of the K-theory classification from the AHSS we have only considered D-brane MMS instantons. If one also includes NS5-brane MMS instantons one needs to further quotient the group of D3-brane charges by those charges which may be created or destroyed by NS5-brane MMS instantons.

Given that the twisted K-theory classification of RR fields and S-duality appear to conflict, there are several directions to proceed. Perhaps the easiest is to claim that S-duality simply is not understood for torsion fields and so we should not worry about it. A second approach is to abandon the K-theory classification and replace K-theory with a different generalized cohomology theory. This avenue has been explored in Refs. [55, 56, 57, 58, 59, 60] in which the authors instead use elliptic cohomology.

However it may be that even this is not radical enough. When we consider both the H flux and the RR fields to be independent fields, the equations of motion that describe them, even in the supergravity limit, are nonlinear. This is in contrast to the case in which the H flux is fixed in which case the equations for the RR field are linear and so the solutions form a linear space and so naturally have the structure of an abelian group. The solutions of a nonlinear set of equations do not obviously form a group. In the case of type IIB string theory on the conifold for example, one finds that pairs (H, dC_2) are classified, when they are both nonzero, by the semigroup of natural numbers which are the greatest common divisors of the three-classes, although perhaps if one motivates a choice of sign for these elements one may conclude that pairs are classified by the integers, which do form an abelian group. In general no argument has been presented that the combined set of NSNS and RR fields, subjected to the equations of motion for all of the fields and quotiented by all of the gauge symmetries, should have any additive structure.

The E_8 bundle framework discussed in the previous subsection provides a natural approach to finding topological classification schemes which naturally include both NSNS and RR fields and, unlike generalized cohomology theories do not impose that there exist a group structure. As we will see in Subsec. 9.1, the E_8 formalism already knows about T-duality, the dual circle is in the fiber itself. Thus it may be applied to type IIB at least in the case of backgrounds which are T-dual to type IIA backgrounds, although the bundle one arrives at in type IIB depends not just on the background, as in IIA, but also on the choice of circle to be T-dualized.

While no one knows what mathematical object classifies all NSNS and RR fields that solve all equations of motion quotiented by gauge invariances, the S-duality covariant generalization of Eq. (8.11) was guessed in Ref. [6] and proven in Ref. [26] using the Freed-Witten anomaly. The D3-brane charge is in general given by

$$Q_{D3} = \mathrm{PD}(dC_2 \cup dC_2 + H \cup dC_2 + H \cup H + P) \tag{8.15}$$

where P is independent of dC_2 and H and is S-duality invariant. Therefore if one finds P for any value of (dC_2, H) then one can use Eq. (8.15) to find the D3-brane charge for every value of (dC_2, H). Setting the D3-charge to zero one finds the S-duality covariant twisted K-theory condition on the 3-classes. Partial results on the value of the 6-class P were presented in Ref. [61]. Analogous conditions for the other RR field strengths are still lacking.

9 Applications

Despite its shortcomings, the K-theory classification has had several applications in different areas of string theory. For example it has been used to identify inconsistencies in string models that satisfy the usual tadpole cancellation conditions but in which certain probes suffer from global anomalies in their worldvolume gauge theories [62]. In these lectures we will mention two applications of the K-theory classification scheme. In Subsec. 9.1 we will describe how an isomorphism of twisted K-groups known as Takai duality led to a conjectured solution to the age old problem of finding a formula for the topology of a T-dual manifold given the original topology and H flux. Then in Subsec. 9.2 we will describe how an S-dual of the twisted K-theory classification classifies universality classes of cascading gauge theories. We will see that the homology classes of branes which lift to the same K-theory class correspond to different steps in the same cascade.

9.1 T-duality

When the H flux is topologically trivial, in other words when it represents the zero cohomology class, one can globally define the NS B field. If the compactification 10-manifold is topologically the product of a circle with a 9-manifold then, given the metric and B field, one can use the Buscher rules [63] to calculate the metric and B field of the compactification manifold which is T-dual with respect to the aforementioned circle. Using this prescription the topologies of the original compactification and the dual compactification are always identical.

It has been known for more than a decade [64] that if instead the compactification 10-manifold is a nontrivial circle fibration over a 9-manifold then in general the topology of the dual manifold is different from that of the original manifold. Over the years many examples of topology changing T-dualities have been found, see for example Ref. [65]. However the dual topology was often found through a somewhat laborious technique involving the exact metric and sometimes consistency conditions which may admit more than one unique solution. In the case of Calabi-Yau compactifications, for example, the metric is often unavailable and so no method of computing the topology of the T-dual manifold was known.

The twisted K-theory classification led to a conjectured formula for the topology and also the H flux of a T-dual manifold which, computationally, is far easier than early approaches even in the examples in which those approaches are feasible. Recall that magnetic fluxes on a timeslice are classified by K^1 in IIB, while in IIA they are classified by K^0. Similarly in IIB D-brane charges on a timeslice and also D-brane trajectories in spacetime are classified by K^0 while they are classified by K^1 in IIA. Type IIA and IIB compactifications are T-dual, and T-duality exchanges branes in IIA with branes in IIB and fluxes in IIA with fluxes in IIB. Therefore one may hope that T-duality exchanges K^0 and K^1.

More concretely, if one begins with a compactification on M with NS 3-form flux H and the T-dual manifold is $\hat{M}$ with NS 3-form flux $\hat{H}$ then one expects

$$\mathrm{K}^0_H(M) = \mathrm{K}^1_{\hat{H}}(\hat{M}), \qquad \mathrm{K}^1_H(M) = \mathrm{K}^0_{\hat{H}}(\hat{M}). \tag{9.1}$$

Thus the question of finding the T-dual compactification data $(\hat{M}, \hat{H})$ is reduced to the problem of finding a solution to Eq. (9.1).

In the language of algebraic K-theory a solution to Eq. (9.1), in the case in which M is a circle bundle over X, was provided nearly 20 years ago in Ref. [66]. Recall that the topology of the total space M of a circle bundle $M \to X$ is completely classified by the topology of X and a 2-class

$$c_1(M) \in \mathrm{H}^2(X; \mathbb{Z}) \tag{9.2}$$

called the Chern class of the circle bundle. Recasting the results of Ref. [66] in the language of topology, one finds that if M is a circle bundle over X with Chern class $c_1(M)$ then Eq. (9.1) is solved by an $\hat{M}$ which is a circle bundle over the same base X but with Chern class $c_1(\hat{M})$ and NS 3-form flux $\hat{H}$ satisfying

$$c_1(M) = \int_{\hat{S}^1} \hat{H}, \qquad c_1(\hat{M}) = \int_{S^1} H. \tag{9.3}$$

Here S^1 and $\hat{S}^1$ are the circle fibers of M and $\hat{M}$ respectively.

We have used the language of differential forms to write the relation between the NS fluxes and Chern characters as an integral, but in terms of integral cohomology the map from $\hat{H}$ to $c_1(M)$ and H to $c_1(\hat{M})$ is the pushforward π_* using the projection map π of the respective circle bundle. To completely specify the dual H flux one must also specify that T-duality leaves invariant any H flux which is in the image of the pullback π^* with respect to π. Intuitively, this means that H flux which is completely supported on X is unchanged by T-duality, instead T-duality merely exchanges the component of the H flux which has two legs along X and one along with fiber with a Chern class which has those same two legs along X.

In Ref. [54] the authors conjectured that the solution $(\hat{M}, \hat{H})$ of Eq. (9.3) is the T-dual compactification to (M, H). They showed that this conjecture reproduces all of the topology changing T-dualities that they knew of in the literature and locally reproduces the Buscher rules.

While the formula (9.3) is a success of the K-theory classification, it also follows from the competing E_8 classification. In this case the solution is manifestly unique, whereas in the K-theory case it is not known whether there are in some cases distinct solutions to Eq. (9.1). Recall that in the E_8 scheme one interprets the NS H flux as the degree 3 characteristic class of an LE_8 fibration. Using a construction reviewed in Ref. [38], the data of a G bundle over a circle bundle over X is equivalent to the data of an LG bundle over X, where LG is the trivially centrally extended based loop group of G. In what follows we will need to use the theorem

$$\pi_p(LG) = \begin{cases} \pi_p(G) \oplus \pi_{p+1}(G) & \text{if } p \geq 2 \\ \pi_p(G) \oplus \pi_{p+1}(G) \oplus \mathbb{Z} & \text{if } p = 1 \end{cases} \tag{9.4}$$

which allows one to construct the homotopy groups of LG from those of G.

Consider type IIA on the 10-manifold M^{10} which is a circle bundle over X^9. Then the data of the LE_8 bundle over M^{10} is captured precisely by the date of an LLE_8 bundle over X^9. Using Eqs. (8.9) and (9.4) one can now find the homotopy groups of LLE_8

$$\pi_1(LLE_8) = \mathbb{Z}^3, \qquad \pi_2(LLE_8) = \mathbb{Z}^2, \qquad \pi_3(LLE_8) = \mathbb{Z}. \tag{9.5}$$

The fact that the fundamental group of LLE_8 is dimension three means that the fiber contains three circles. The first two are the M-theory and type IIA circles which have been compactified. The third descends from $\pi_3(E_8) = \mathbb{Z}$ in the E_8 fibration over the M-theory spacetime.

In Ref. [39] it was conjectured that this third circle is the T-dual type IIB circle, and that the first two form the F-theory torus. This third circle is the dimensional reduction of the class $\pi_2(LE_8) = \mathbb{Z}$, whose characteristic class was the NS 3-form H. Therefore the characteristic class of the third circle, which is the Chern class $c_1(\hat{M})$ of the IIB circle bundle, is just the dimensional reduction of H on the IIA circle

$$c_1(\hat{M}) = \int_{S^1_{IIA}} H \tag{9.6}$$

in accordance with the conjecture (9.3).

9.2 The Klebanov-Strassler Cascade

In the previous subsection we saw that the K-theory classification led to the solution of an old problem. In this subsection we will describe a more mild success of the K-theory classification, we will argue that K-theory, twisted this time by the RR 3-form field strength dC_2, classifies universality classes of a class of cascading $SU(N+M)\times SU(N)$ gauge theories with a fixed step size M. In this section we will be using the S-dual of the usual K-theory classification in which one classifies F-strings, D3-branes and NS5-branes. The action of S-duality on higher branes and D-instantons is still a matter of debate in the literature and we will fortunately not need these branes in this example.

We have seen that homology groups are much larger than K-groups. In particular multiple homology classes correspond to the same K-class. We will now see that each of the homology classes corresponding to a single K-class describes one particular step in the cascade, and gives the ranks of the worldvolume gauge bundles at that step. In the S-duality covariant K-theory one needs to classify not only D-branes but also their duals. Now we are already using the S-dual of the ordinary K-theory, so in our case the S-duality covariant twisted K-theory is augmented by adding D-strings and D5-branes. The number of D5-branes will be equal to the step size M, and so the S-duality covariant twisted K-theory will also classify M and so will classify all of universality classes of all of the theories of this type.

In the mathematics literature the word *conifold* is used to describe any space which is a continuous manifold everywhere except for isolated ordinary regular double point singularities, which are locally the $2n-2$ real-dimensional solutions of the complex equation

$$\sum_{i=1}^{n} z_i^2 = 0 \tag{9.7}$$

in $\mathbb{C}^n$. We will instead refer to *the conifold*, which in the physics literature has come to mean the 6-manifold which is the solution of Eq. (9.7) in $\mathbb{C}^4$. Topologically the conifold is a cone over the product $S^2\times S^3$.

String theory on the conifold crossed with $\mathbb{R}^{3,1}$ was first considered in Ref. [67]. They found that a stack of N D3-branes at the tip of the cone are described by a particular $N=1$ supersymmetric gauge theory with gauge group $U(N)\times U(N)$ with charged chiral multiplets and a given superpotential. They were interested in the IR fixed point of the theory, in which the $U(1)$ factors decouple leaving an $SU(N)\times SU(N)$ conformal theory.

Later in Ref. [68] it was discovered that one can engineer the gauge group $SU(N+M)\times SU(N)$ by adding M D5-branes wrapped around the 2-cycle in the $S^2\times S^3$. The normal bundle to this S^2 is nontrivial and as a result the statistics of some of the worldvolume fields change [69]. As in [1], this nontrivial normal bundle couples D5-branes to the lower dimensional RR field C_4 so that the D5-brane carries one half of a Dirac unit of D3-brane charge. This gravitational charge also couples to the fundamental string worldsheet, and so it is often considered to be a part of the B field.

The half-integrality of the D3 charge does not violate Dirac's quantization argument, which ordinarily implies that D3-brane charge quantization is required for the well-definedness of the partition function of another D3-brane whose trajectory is deformed over an S^5 linking the original D3. Dirac's argument fails because the linking D3 would need to pass through the D5, at which point there would be a Hanany-Witten transition [70] creating a D1-brane. The contribution of this D1-brane to the partition function renders it well-defined. In general Dirac's quantization condition does not apply straightforwardly to branes that are dissolved in other branes.

In Ref. [71] the authors discovered that in a particular geometry, corresponding to the baryonic branch of the gauge theory, the gauge group $SU(N+M)\times SU(N)$ becomes strongly coupled in the infrared and admits a dual weakly coupled description with gauge group $SU(N)\times SU(N-M)$. This duality is similar to Seiberg duality, which would occur if the $SU(N)$ symmetry were not gauged. In their geometry the $SU(N)\times SU(N-M)$ gauge theory is also at its baryonic root, and so at weak coupling there is an effective gauge group $SU(N-M)\times SU(N-2M)$. This process continues until N is less than M. They named this series of Seiberg dualities the cascade.

The one-half unit of D3-brane charge will not be essential in what follows. The important feature of the M D5-branes will be that, by Gauss' Law, if S^3 is a 3-sphere linking all of the

D5's then

$$\int_{S^3} dC_2 = M. \tag{9.8}$$

This means that the S-dual twisted K-theory, which we recall classifies fundamental strings, D3-branes and NS5-branes, is twisted by M units.

This argument is a bit too fast. These lectures have been about smooth manifolds, and now we are considering a compactification with a singularity. We have not defined K-theory on singular spaces and in fact inequivalent definitions exist. In this example we are saved by the fact that the singularity can be eliminated in two different ways, each of which has little effect on the geometry far from the singularity. First, it may be blown up, which topologically means that the singularity is replaced by an S^2 which is homotopic to the S^2 in the base, intuitively only the S^3 collapses and the space is nonsingular. The second possibility is that the singularity may be deformed. Algebraically this may be done by replacing the defining equation (9.7) with

$$\sum_{i=1}^{n} z_i^2 = \xi \tag{9.9}$$

for some constant deformation parameter ξ. In this case the singularity is replaced by a 3-sphere which is homotopic to the 3-sphere on the base and again the space becomes nonsingular. Locality imposes that the physics far away from the singularity does not know whether or in which way the singularity was eliminated, and so when considering distant physics, like the K-theory of the base, one can often consider whichever smoothing is more convenient. We want a third cohomology class on which there is dC_2 flux and so we will consider the deformed conifold.

On the deformed conifold the S^2 is contractible. This means that the D5-branes may shrink to points, but they can never completely decay as they carry D3-brane charge. The fact that we have chosen the deformed conifold cannot affect the physics far away, and so in particular (9.8) still implies that the 3-form RR field strength is topologically nontrivial, and is the element M in the third cohomology group with compact support of the deformed conifold, which is $\mathbb{Z}$.

The cohomology with compact support of the deformed conifold consists of only two nontrivial classes

$$\mathrm{H}^3 = \mathbb{Z}, \qquad \mathrm{H}^6 = \mathbb{Z} \tag{9.10}$$

which are Poincaré dual to a brane wrapping the S^3 and to pointlike branes respectively. The N D3-branes are charged under the later, as are the M half D3-branes. We are interested in the S-dual of the K-theory classification, which we have stressed classifies only F1's, D3's and NS5's. This means that we will only try to classify the D3-branes and not the half D3-branes, which are really D5-branes. Thus cohomology classifies the number N of D3-branes, which is the same number N that appears in the rank of the gauge group $SU(N+M)\times SU(N)$.

The cohomology group $\mathrm{H}^3 = \mathbb{Z}$ classifies a different charge, that of branes wrapping the 3-sphere. However this 3-sphere has nontrivial dC_2 flux (9.8) which couples to the D3-brane worldvolume via, for example, the term

$$\mathcal{L}_{D3} \supset C_2 \wedge B. \tag{9.11}$$

This implies that C_2 flux carries fundamental string charge, and so dC_2 is the endpoint of a fundamental string. In this case the presence of M units of dC_2 implies that the D3-brane is a baryon on which M fundamental strings end, in line with the S-dual Freed-Witten anomaly. In the K-theory classification one does not assign a charge to baryonic configurations, and so these branes will not carry K-theory charges.

One can now calculate the twisted K-theory of the deformed conifold by applying the AHSS with twist dC_2 to the cohomology (9.10). The conifold itself is both *spin* and *spin*^{c} and the S^3 is also *spin*^{c} and so the Sq^3 term is equal to zero. Also there is no spacing between cohomology groups of more than 3, and so one only need consider the differential

$$d_3 = dC_2\cup : x^3 \mapsto Mx^6 \tag{9.12}$$

where x^3 and x^6 are the generators of H^3 and H^6 respectively. The twisted K-theory is then the quotient of kernel of d_3, which is $\mathrm{H}^6 = \mathbb{Z}$, by its image, which is $M\mathrm{H}^6 = M\mathbb{Z}$. This yields

$$\mathrm{K}^0_{dC_2} = \frac{\mathrm{Ker}(d_3 : \mathrm{H}^{\mathrm{even}} \longrightarrow \mathrm{H}^{\mathrm{odd}})}{\Im(d_3 : \mathrm{H}^{\mathrm{odd}} \longrightarrow \mathrm{H}^{\mathrm{even}})} = \frac{\mathbb{Z}}{M\mathbb{Z}} = \mathbb{Z}_M, \qquad \mathrm{K}^1_{dC_2} = 0 \tag{9.13}$$

where $K^1_{dC_2}$ vanishes because no elements of the odd cohomology are in the kernel of d_3.

The group $K^0_{dC_2}$ has M elements, of which the Jth is the lift of all of the cohomology classes corresponding to numbers N of D3-branes which are equal to J modulo M. In other words, the element J of K-theory corresponds to all of the gauge theories with gauge groups

$$SU(J+(K+1)M) \times SU(J+KM). \tag{9.14}$$

The gauge groups (9.14) are the set of gauge groups in a given cascade. Thus $K^0_{dC_2}$ parametrizes the possible cascades with step size M, or equivalent the endpoints of the cascade, which are the universality classes of the gauge theory when one restricts attention to the baryonic root vacua.

Acknowledgements

Jarah would like to thank the Modave Summer school for inviting him to give these lectures in an environment with many flammable materials. The work of J.E. is partially supported by IISN - Belgium (convention 4.4505.86), by the "Interuniversity Attraction Poles Programme – Belgian Science Policy" and by the European Commission programme MRTN-CT-2004-005104, in which he is associated to V. U. Brussel.

References

[1] R. Minasian and G. Moore, *K-theory and Ramond-Ramond charge*, J. High Energy Phys. **11** (1997) 002, [arXiv:hep-th/9710230].

[2] E. Witten, *D-Branes and K-Theory*, JHEP **12** (1998) 019, [arXiv:hep-th/9810188].

[3] A. Sen, *Tachyon Condensation on the Brane Anti-Brane System*, JHEP **9808** (1998) 012, [arXiv:hep-th/9805170].

[4] A. Kapustin, *D-Branes in a Topologically Nontrivial B-Field*, [arXiv:hep-th/9909089].

[5] G. Moore and E. Witten, *Self-Duality, Ramond-Ramond Fields, and K-Theory*, [arXiv:hep-th/9912279].

[6] E. Diaconescu, G. Moore and E. Witten, *E_8 Gauge Theory, and a Derivation of K-Theory from M-Theory*, Adv. Theor. Math. Phys. **6** (2003) 1031, [arXiv:hep-th/0005090].

[7] J. Maldacena, G. Moore and N. Seiberg, *D-Brane Instantons and K-Theory Charges*, JHEP **11** (2001) 062, [arXiv:hep-th/0108100].

[8] P. Bouwknegt and V. Mathai, *D-branes, B fields and twisted K theory*, JHEP **0003** (2000) 007, [arXiv:hep-th/0002023].

[9] P. K. Townsend, *p-Brane Democracy*, [arXiv:hep-th/9507048].

[10] M. B. Green, C. M. Hull and P. K. Townsend, *D-Brane Wess-Zumino Actions, T-Duality and the Cosmological Constant*, [arXiv:hep-th/9604119].

[11] E. Bergshoeff, R. Kallosh, T. Ortin, D. Roest and A. Van Proeyen, *New Formulations of D=10 Supersymmetry and D8-O8 Domain Walls*, [arXiv:hep-th/0103233].

[12] G. W. Moore and G. Segal, *D-Branes and K-Theory in 2D Topological Field Theory*, [arXiv:hep-th/0609042].

[13] E. Witten, *Baryons and Branes in Anti de Sitter Space*, [arXiv:hep-th/9805112].

[14] D. S. Freed and E. Witten, *Anomalies in String Theory with D-Branes*, Asian J. Math. **3** (1999) 819, [arXiv:hep-th/9907189].

[15] M. J. Douglas and S. H. Shenker, *Dynamics of SU(N) Supersymmetric Gauge Theory*, [arXiv:hep-th/9503163].

[16] C. Bohr, B. Hanke, and D. Kotschick, *Cycles, submanifolds, and structures on normal bundles*, Manuscripta Math. **108** (2002) 483, [arXiv:math.GT/0011178].

[17] R. M. G. Reis and R. J. Szabo, *Geometric K-Homology of Flat D-Branes*, [arXiv:hep-th/0507043].

[18] J. Evslin and H. Sati, *Can D-Branes Wrap Nonrepresentable Cycles?* [arXiv:hep-th/0607045].

[19] C. Bachas, M. Douglas and C. Schweigert, *Flux Stabilization of D-Branes*, [arXiv:hep-th/0003037].

[20] W. Taylor, *D-Branes in B Fields*, [arXiv:hep-th/0004141].

[21] M. Green, J. A. Harvey and G. Moore, *I-Brane Inflow and Anomalous Couplings on D-Branes*, [arXiv:hep-th/9605033].

[22] A. Adams, J. Polchinski and E. Silverstein, *Don't Panic! Closed String Tachyons in ALE Spacetimes*, [arXiv:hep-th/0108075].

[23] M. Fabinger and P. Hořava *Casimir Effect between World-Branes in Heterotic M-Theory*, Nucl Phys. **B580**, 243, (2000), [arXiv:hep-th/0002073].

[24] P. Hořava, *Type IIA D-Branes, K-Theory, and Matrix Theory*, Adv. Theor. Math. Phys. **2** (1999) 1373, [arXiv:hep-th/9812135].

[25] J. Evslin, *Twisted K-Theory as a BRST Cohomology*, [arXiv:hep-th/0605049].

[26] P. Bouwknegt, J. Evslin, B. Jurco, V. Mathai and H. Sati, *Flux Compactifications on Projective Spaces and the S-Duality Puzzle*, Adv. Theor. Math. Phys. **10** (2006) 345, [arXiv:hep-th/0501110].

[27] R. Thom, *Quelques propriétés globales des variétés différentiables*, Comment. Math. Helv. **28** (1954) 17.

[28] D. Sullivan, *René Thom's work on geometric homology and bordism*, Bull. Amer. Math. Soc. (N.S.) **41** (2004) 341.

[29] R. Myers, *Dielectric-Branes*, JHEP **9912**, 22 (1999), [arXiv:hep-th/9910053].

[30] M. Atiyah and G. Segal, *Twisted K-theory and Cohomology*, [arXiv:math.KT/0510674].

[31] J. Evslin and U. Varadarajan, *K-Theory and S-Duality: Starting over from Square 3*, J. High Energy Phys. **03** (2003) 026, [arXiv:hep-th/0112084].

[32] G. R. Cavalcanti, *Formality in Generalized Kahler Geometry*, [arXiv:math.DG/0603596].

[33] G. R. Cavalcanti, *New Aspects of the dd^c Lemma*, [arXiv:math.DG/0501406].

[34] L. Martucci and P. Smyth, *Supersymmetric D-branes and Calibrations on General $N = 1$ Backgrounds*, [arXiv:hep-th/0507099].

[35] L. Martucci, *D-Branes on Generalized $N = 1$ Backgrounds: Superpotentials and D-terms*, [arXiv:hep-th/0602129].

[36] M. Gualtieri, *Generalized Complex Geometry*, [arXiv:math.DG/0401221].

[37] E. Witten, *On Flux Quantization in M-Theory and the Effective Action*, [arXiv:hep-th/9609122].

[38] J. Evslin, *From E_8 to F via T*, [arXiv:hep-th/0311235].

[39] J. Evslin, *Twisted K-Theory from Monodromies*, [arXiv:hep-th/0302081].

[40] A. Adams and J. Evslin, *The Loop Group of E_8 and K-Theory from $11d$*, JHEP **02**, 29 (2003), [arXiv:hep-th/0203218].

[41] J. A. Harvey, *Topology of the Gauge Group in Noncommutative Gauge Theory*, [arXiv:hep-th/0105242].

[42] C. Maccaferri, *Chan-Paton Factors and Higgsing from Vacuum String Field Theory*, [arXiv:hep-th/0506213].

[43] D. S. Freed, *Dirac Charge Quantization and Generalized Differential Cohomology*, [arXiv:hep-th/0011220].

[44] P. S. Aspinwall and R. Y. Donagi, *The Heterotic String, the Tangent Bundle, and Derived Categories*, [arXiv:hep-th/9806094].

[45] P. S. Aspinwall, *D-Branes on Calabi-Yau Manifolds*, [arXiv:hep-th/0403166].

[46] S. Fredenhagen and V. Schomerus, *Branes on Group Manifolds, Gluon Condensates and Twisted K-Theory*, [arXiv:hep-th/0012164].

[47] A. Yu. Alekseev and V. Schomerus, *D-branes in the WZW Model*, [arXiv:hep-th/9812193].

[48] R. Auzzi, S. Bolognesi and J. Evslin, *Monopoles Can be Confined by 0, 1 or 2 Vortices*, [arXiv:hep-th/0411074].

[49] E. Witten, *Branes and the Dynamics of QCD*, [arXiv:hep-th/9706109].

[50] S. Bolognesi and J. Evslin, *Stable vs Unstable Vortices in SQCD*, [arXiv:hep-th/0506174].

[51] E. Witten, *Duality Relations Among Topological Effects in String Theory*, [arXiv:hep-th/9912086].

[52] P. Hořava and E. Witten, *Heterotic and Type I String Dynamics from Eleven Dimensions*, Nucl. Phys. **B460**, 506, (1996), [arXiv:hep-th/9510209].

[53] N. Seiberg and E. Witten, *String Theory and Noncommutative Geometry*, [arXiv:hep-th/9908142].

[54] P. Bouwknegt, J. Evslin and V. Mathai, *T-Duality: Topology Change from H-Flux*, [arXiv:hep-th/0306062].

[55] I. Kriz and H. Sati, *M Theory, Type IIA Superstrings, and Elliptic Cohomology*, [arXiv:hep-th/0404013].

[56] I. Kriz and H. Sati, *Type IIB String Theory, S-Duality, and Generalized Cohomology*, [arXiv:hep-th/0410293].

[57] I. Kriz and H. Sati, *Type II String Theory and Modularity*, [arXiv:hep-th/0501060].

[58] H. Sati, *Flux Quantization and the M-Theoretic Characters*, [arXiv:hep-th/0507106].

[59] I. Kriz, *Toward a derivation of E-theory from F-theory*, [arXiv:hep-th/0511011].

[60] H. Sati, *The Elliptic Curves in Gauge Theory, String Theory, and Cohomology*, [arXiv:hep-th/0511087].

[61] E. Diaconescu, D. Freed, and G. Moore, *The M-Theory 3-form and E_8 Gauge Theory*, [arXiv:hep-th/0312069].

[62] A. M. Uranga, *D-Brane Probes, RR Tadpole Cancellation and K-Theory Charge*, [arXiv:hep-th/0011048].

[63] T. H. Buscher, *Path Integral Derivation of Quantum Duality in Nonlinear Sigma Models*, Phys. Lett. **B201** (1988) 466.

[64] E. Alvarez, L. Alvarez-Gaume, J. L. F. Barbon and Y. Lozano, *Some Global Aspects of Duality in String Theory*, [arXiv:hep-th/9309039].

[65] S. Kachru, M. B. Schulz, P. K. Tripathy and S. P. Trivedi, *New Supersymmetric String Compactifications*, [arXiv:hep-th/0211182].

[66] I. Raeburn and J. Rosenberg, *Crossed products of continuous-trace C^*-algebras by smooth actions*, Trans. Amer. Math. Soc. **305** (1988) 1-45.

[67] I. R. Klebanov and E. Witten, *Superconformal Field Theory on Threebranes at a Calabi-Yau Singularity*, [arXiv:hep-th/9807080].

[68] S. S. Gubser and I. R. Klebanov, *Baryons and Domain Walls in an N=1 Superconformal Gauge Theory*, [arXiv:hep-th/9808075].

[69] J. M. Maldacena and C. Nunez, *Towards the Large* N *Limit of Pure* $N = 1$ *Super Yang-Mills*, [arXiv:hep-th/0008001].

[70] A. Hanany and E. Witten, *Type IIB Superstrings, BPS Monopoles and Three-Dimensional Gauge Dynamics*, [arXiv:hep-th/9611230].

[71] I. R. Klebanov and M. J. Strassler, *Supergravity and a Confining Gauge Theory: Duality Cascades and* χ*SB-Resolution of Naked Singularities*, [arXiv:hep-th/0007191].

An introduction to the mechanics of black holes

Geoffrey Compère[1]

Physique Théorique et Mathématique,
Université Libre de Bruxelles and
International Solvay Institutes
Campus Plaine C.P. 231, B-1050 Bruxelles, Belgium

Abstract. These notes provide a self-contained introduction to the derivation of the zero, first and second laws of black hole mechanics. The prerequisite conservation laws in gauge and gravity theories are also briefly discussed. An explicit derivation of the first law in general relativity is performed in appendix.

[1]Research Fellow of the National Fund for Scientific Research (Belgium). email: gcompere@ulb.ac.be

Preliminary remark. These notes (except the third chapter) are mainly based on previous reviews on thermodynamics of black holes [1, 2, 3, 4].

A black hole usually refers to a part of spacetime from which no future directed timelike or null line can escape to arbitrarily large distance into the outer asymptotic region. A white hole or white fountain is the time reversed concept which is believed not to be physically relevant, and will not be treated.

More precisely, if we denote by $\mathfrak{I}^+$ the future asymptotic region of a spacetime $(\mathcal{M}, g_{\mu\nu})$, e.g. null infinity for asymptotically flat spacetimes and timelike infinity for asymptotically anti-de Sitter spacetimes, the black hole region $\mathcal{B}$ is defined as

$$\mathcal{B} \equiv \mathcal{M} - I^-(\mathfrak{I}^+), \tag{0.1}$$

where I^- denotes the chronological past. The region $I^-(\mathfrak{I}^+)$ is what is usually referred to as the *domain of outer communication*, it is the set of points for which it is possible to construct a future directed timelike line to arbitrary large distance in the outer region.

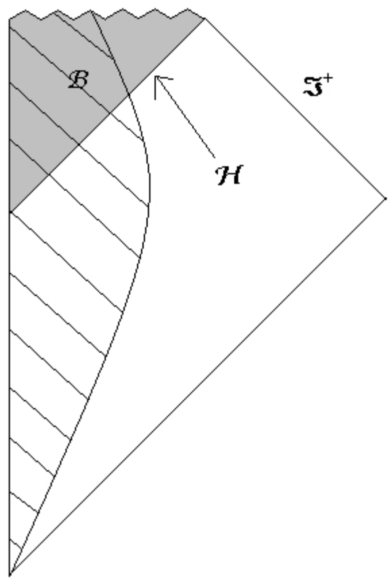


Figure 7.6: Penrose diagram of an asymptotically flat spacetime with spherically symmetric collapsing star. Each point is a $n-2$-dimensional sphere. Light rays propagate along 45° diagonals. The star region is hatched and the black hole region is indicated in grey.

The *event horizon* $\mathcal{H}$ of a black hole is then the boundary of $\mathcal{B}$. Let us denote $J^-(U)$ the causal past of a set of points $U \subset \mathcal{M}$ and $\bar{J}^-(U)$ the topological closure of J^-. We have $I^-(U) \subset J^-(U)$. The (future) event horizon of $\mathcal{M}$ can then equivalently be defined as

$$\mathcal{H} \equiv \bar{J}^-(\mathfrak{I}^+) - J^-(\mathfrak{I}^+), \tag{0.2}$$

i.e. the boundary of the closure of the causal past of $\mathfrak{I}^+$. See Fig. 7.6 for an example. The event horizon is a concept defined with respect to the entire causal structure of $\mathcal{M}$.

The event horizons are null hypersurfaces with peculiar properties. We shall develop their properties in section 1, what will allow us to sketch the proof of the area theorem [5]. The area theorem also called the second law of black hole mechanics because of its similarity with classical thermodynamics [6] is concerned with the dynamical evolution of sections of the event horizon at successive times.

In section 2, we introduce the notion of Killing horizon. This concept is adapted for black holes in equilibrium in stationary spacetimes. We will show how the zero law of mechanics, the consistency of a specific quantity defined on the horizon, directly comes out of the definitions.

The tools necessary to handle with the conservations laws in gravity theories are briefly introduced in section 3. In section 4, we derive the first law for two infinitesimally close equilibrium black holes [7].

Remark that the zero and first law of black hole mechanics may also be generalized to black holes in non-stationary spacetimes. This was done in the framework of "isolated horizons" very recently [8, 9]. However, in this introduction, we limit ourselves to the original notion of Killing horizon.

Notation In what follows, $\partial_\mu f = f_{,\mu}$ is the partial derivative, while $D_\mu f = f_{;\mu}$ denotes the covariant derivative.

1 Event horizons

1.1 Null hypersurfaces

Let $S(x^\mu)$ be a smooth function and consider the $n-1$ dimensional null hypersurface $S(x) = 0$, which we denote by $\mathcal{H}$. This surface will be the black hole horizon in the subsequent sections. It is a null hypersurface, i.e. such that its normal $\xi^\mu \sim g^{\mu\nu}\partial_\nu S$ is null,

$$\xi^\mu \xi_\mu \stackrel{\mathcal{H}}{=} 0. \tag{1.1}$$

The vectors η^μ tangent to $\mathcal{H}$ obey $\eta_\mu \xi^\mu|_{\mathcal{H}} = 0$ by definition. Since $\mathcal{H}$ is null, ξ^μ itself is a tangent vector, i.e.

$$\xi^\mu = \frac{dx^\mu(t)}{dt} \tag{1.2}$$

for some null curve $x^\mu(t)$ inside $\mathcal{H}$. One can then prove that $x^\mu(t)$ are null geodesics[2]

$$\xi^\nu \xi^\mu{}_{;\nu} \stackrel{\mathcal{H}}{=} \kappa \xi^\mu, \tag{1.4}$$

where κ measure the extent to which the parameterization is not affine. If we denote by l the normal to $\mathcal{H}$ which corresponds to an affine parameterization $l^\nu l^\mu{}_{;\nu} = 0$ and $\xi = f(x)\, l$ for some function $f(x)$, then $\kappa = \xi^\mu \partial_\mu \ln |f|$.

According to the Frobenius' theorem, a vector field v is hypersurface orthogonal if and only if it satisfies $v_{[\mu}\partial_\nu v_{\rho]} = 0$, see e.g. [10]. Therefore, the vector ξ satisfies the irrotationality condition

$$\xi_{[\mu}\partial_\nu \xi_{\rho]} \stackrel{\mathcal{H}}{=} 0. \tag{1.5}$$

A congruence is a family of curves such that precisely one curve of the family passes through each point. In particular, any smooth vector field define a congruence. Indeed, a vector field define at each point a direction which can be uniquely "integrated" along a curve starting from an arbitrary point.

Since $S(x)$ is also defined outside $\mathcal{H}$, the normal ξ defines a congruence but which is a null congruence only when restricted to $\mathcal{H}$. In order to study this congruence outside $\mathcal{H}$, it is useful to define a transverse null vector n^μ cutting off the congruence with

$$n^\mu n_\mu = 0, \qquad n_\mu \xi^\mu = -1. \tag{1.6}$$

The normalization -1 is chosen so that if we consider ξ to be tangent to an outgoing radial null geodesic, then n is tangent to an ingoing one, see Fig. (7.7). The normalization conditions (1.6) (imposed everywhere, $(n^2)_{;\nu} = 0 = (n \cdot \xi)_{;\nu}$) do not fix uniquely n. Let us choose one such n arbitrarily. The extent to which the family of hypersurfaces $S(x) = \mathit{const}$ are not null is given by

$$\varsigma \equiv \frac{1}{2}(\xi^2)_{;\mu} n^\mu \neq 0. \tag{1.7}$$

[2]Proof: Let $\xi_\mu = \tilde{f} S_{,\mu}$. We have

$$\begin{aligned} \xi^\nu \xi_{\mu;\nu} &= \xi^\nu \partial_\nu \tilde{f} S_{,\mu} + \tilde{f}\xi^\nu S_{,\mu;\nu} \\ &= \xi^\nu \partial_\nu \ln \tilde{f} \xi_\mu + \tilde{f}\xi^\nu S_{;\nu;\mu} \\ &= \xi^\nu \partial_\nu \ln \tilde{f} \xi_\mu + \tilde{f}\xi^\nu (\tilde{f}^{-1}\xi_\nu)_{;\mu} \\ &= \xi^\nu \partial_\nu \ln \tilde{f} \xi_\mu + \frac{1}{2}(\xi^2)_{,\mu} - \partial_\mu \ln \tilde{f} \xi^2. \end{aligned} \tag{1.3}$$

Now, as ξ is null on the horizon, any tangent vector η to $\mathcal{H}$ satisfy $(\xi^2)_{;\mu}\eta^\mu = 0$. Therefore, $(\xi^2)_{;\mu} \sim \xi_\mu$ and the right-hand side of (1.3) is proportional to ξ_μ on the horizon.

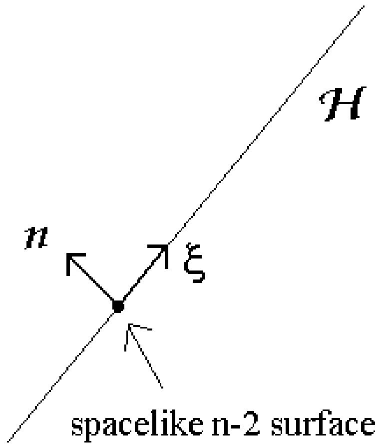


Figure 7.7: The null vector n is defined with respect to ξ.

The vectors η orthogonal to both ξ and n,

$$\eta^\mu \xi_\mu = 0 = \eta^\mu n_\mu, \tag{1.8}$$

span a $n-2$ dimensional spacelike subspace of $\mathcal{H}$. The metric can be written as

$$g_{\mu\nu} = -\xi_\mu n_\nu - \xi_\nu n_\mu + \gamma_{\mu\nu} \tag{1.9}$$

where $\gamma_{\mu\nu} = \gamma_{(\mu\nu)}$ is a positive definite metric with $\gamma_{\mu\nu}\xi^\mu = 0 = \gamma_{\mu\nu}n^\mu$. The tensor $\gamma^\mu_{\ \nu} = g^{\mu\alpha}\gamma_{\alpha\nu}$ provides a projector onto the $n-2$ spacelike tangent space to $\mathcal{H}$.

For future convenience, we also consider the hypersurface orthogonal null congruence l^μ with affine parameter τ that is proportional to ξ^μ on $\mathcal{H}$[3],

$$l^\mu l_\mu = 0, \qquad l^\nu l^\mu_{\ ;\nu} = 0, \qquad l^\mu \overset{\mathcal{H}}{\sim} \xi^\mu. \tag{1.10}$$

The vector field l extends ξ outside the horizon while keeping the null property.

1.2 The Raychaudhuri equation

In this section, we shall closely follow the reference [2]. We introduce part of the material needed to prove the area law.

Firstly, let us decompose the tensor $D_\mu\xi_\nu$ into the tensorial products of ξ, n and spacelike vectors η tangent to $\mathcal{H}$,[4]

$$D_\mu\xi_\nu \overset{\mathcal{H}}{=} v_{\mu\nu} - \xi_\nu(\kappa n_\mu + \gamma^\alpha_{\ \mu} n^\beta D_\alpha\xi_\beta) - \xi_\mu n^\alpha D_\alpha \xi_\nu, \tag{1.12}$$

where the orthogonal projection $v_{\mu\nu} = \gamma^\alpha_{\ \mu}\gamma^\beta_{\ \nu}D_\alpha\xi_\beta$ can itself be decomposed in symmetric and antisymmetric parts

$$v_{\mu\nu} = \theta_{\mu\nu} + \omega_{\mu\nu}, \qquad \theta_{[\mu\nu]} = 0, \qquad \omega_{(\mu\nu)} = 0. \tag{1.13}$$

The Frobenius irrotationality condition (1.5) is equivalent to $\omega_{\mu\nu}|_{\mathcal{H}} = 0$[5]. The tensor $\theta_{\mu\nu}$ is interpreted as the expansion rate tensor of the congruence while its trace $\theta = \theta_\mu^{\ \mu}$ is the

[3] We shall reserve the notation ξ^μ for vectors coinciding with l^μ on the horizon but which are not null outside the horizon.

[4] Proof: Let us first decompose $D_\mu\xi_\nu$ as

$$D_\mu\xi_\nu = v_{\mu\nu} + n_\mu(C_1 n_\nu + C_2\xi_\nu + C_3\eta_\nu) + \tilde\eta_\mu\xi_\nu + \hat\eta_\mu n_\nu - \xi_\mu\alpha_\nu, \tag{1.11}$$

where $v_{\mu\nu} = \gamma^\alpha_{\ \mu}\gamma^\beta_{\ \nu}v_{\alpha\beta}$ and η^μ, $\tilde\eta^\mu$, $\hat\eta^\mu$ are spacelike tangents to $\mathcal{H}$. Contracting with ξ^μ and using (1.4), we find $C_1 = 0 = C_3$, $C_2 = -\kappa$. Contracting with $\gamma^\mu_{\ \alpha}n^\nu$, we find $\tilde\eta_\mu = -\gamma^\alpha_{\ \mu}n^\beta D_\alpha\xi_\beta$. Contracting with $\gamma^\mu_{\ \alpha}\xi^\nu$, we find finally $\hat\eta_\mu = -1/2\gamma^\alpha_{\ \mu}D_\alpha(\xi^2) = 0$ thanks to (1.1).

[5] Proof: We have

$$\xi_{[\mu}\partial_\nu\xi_{\rho]} = \xi_{[\mu}D_\nu\xi_{\rho]} = \xi_{[\mu}v_{\nu\rho]} = \xi_{[\mu}\omega_{\nu\rho]}. \tag{1.14}$$

divergence of the congruence. Any smooth $n-2$ dimensional area element evolves according to

$$\frac{d}{dt}(d\mathcal{A}) = \theta \, d\mathcal{A}. \tag{1.15}$$

The shear rate is the trace free part of the strain rate tensor,

$$\sigma_{\mu\nu} = \theta_{\mu\nu} - \frac{1}{n-2}\theta\gamma_{\mu\nu}. \tag{1.16}$$

Defining the scalar $\sigma^2 = (n-2)\sigma_{\mu\nu}\sigma^{\mu\nu}$, one has

$$\xi_{\mu;\nu}\xi^{\nu;\mu} \stackrel{\mathcal{H}}{=} \frac{1}{n-2}(\theta^2+\sigma^2) + \kappa^2 + \varsigma^2, \tag{1.17}$$

where ς was defined in (1.7). Note also that the divergence of the vector field has three contributions,

$$\xi^\mu{}_{;\mu} \stackrel{\mathcal{H}}{=} \theta + \kappa - \varsigma. \tag{1.18}$$

Now, the contraction of the Ricci identity

$$v^\alpha{}_{;\mu;\nu} - v^\alpha{}_{;\nu;\mu} = -R^\alpha{}_{\lambda\mu\nu}v^\lambda, \tag{1.19}$$

implies the following identity

$$(v^\nu{}_{;\nu})_{;\mu}v^\mu = (v^\nu v^\mu{}_{;\nu})_{;\mu} - v^{\nu;\mu}v_{\mu;\nu} - R_{\mu\nu}v^\mu v^\nu, \tag{1.20}$$

valid for any vector field v. The formulae (1.17)-(1.18) have their equivalent for l as

$$l_{\mu;\nu}l^{\nu;\mu} = \frac{1}{n-2}(\theta_{(0)}^2 + \sigma_{(0)}^2), \qquad l^\mu{}_{;\mu} = \theta_{(0)}, \tag{1.21}$$

where the right hand side are expressed in terms of the expansion rate $\theta_{(0)} = \theta\frac{dt}{d\tau}$ and shear rate $\sigma_{(0)} = \sigma\frac{dt}{d\tau}$ with respect to the affine parameter τ. The identity (1.20) becomes

$$\frac{d\theta_{(0)}}{d\tau} = \dot{\theta}_{(0)} \stackrel{\mathcal{H}}{=} -\frac{1}{n-2}(\theta_{(0)}^2 + \sigma_{(0)}^2) - R_{\mu\nu}l^\mu l^\nu, \tag{1.22}$$

where the dot indicate a derivation along the generator. It is the final form of the Raychaudhuri equation for hypersurface orthogonal null geodesic congruences in n dimensions.

1.3 Properties of event horizons

As we have already mentioned, the main characteristic of event horizons is that they are null hypersurfaces. In the early seventies, Penrose and Hawking further investigated the generic properties of past boundaries, as event horizons. We shall only enumerate these properties below and refer the reader to the references [11, 3] for explicit proofs. These properties are crucial in order to prove the area theorem.

(i) *Achronicity property.* No two points of the horizon can be connected by a timelike curve.

(ii) The null geodesic generators of $\mathcal{H}$ may have past end-points in the sense that the continuation of the geodesic further into the past is no longer in $\mathcal{H}$.

(iii) The generators of $\mathcal{H}$ have no future end-points, i.e. no generator may leave the horizon.

The second property hold for example for collapsing stars where the past continuation of all generators leave the horizon at the time the horizon was formed. As a consequence of properties 2 and 3, null geodesics may enter $\mathcal{H}$ but not leave it.

1.4 The area theorem

The area theorem was initially demonstrated by Hawking [5]. We shall follow closely the reviews by Carter [2] and Townsend [3]. The theorem reads as follows.

Theorem 1 *Area law.* If

(i) Einstein's equations hold with a matter stress-tensor satisfying the null energy condition, $T_{\mu\nu}k^\mu k^\nu \geq 0$, for all null k^μ,

(ii) The spacetime is "strongly asymptotically predictable"

then the surface area $\mathcal{A}$ of the event horizon can never decrease with time.

The theorem was stated originally in 4 dimensions but it is actually valid in any dimension $n \geq 3$.

In order to understand the second requirement, let us remind some definitions. The future domain of dependence $D^+(\Sigma)$ of an hypersurface Σ is the set of points p in the manifold for which every causal curve through p that has no past end-point intersects Σ. The significance of $D^+(\Sigma)$ is that the behavior of solutions of hyperbolic PDE's *outside* $D^+(\Sigma)$ is not determined by initial data on Σ. If no causal curves have past end-points, then the behavior of solutions inside $D^+(\Sigma)$ is entirely determined in terms of data on Σ. The past domain of dependence $D^-(\Sigma)$ is defined similarly.

A Cauchy surface is a spacelike hypersurface which every non-spacelike curve intersects exactly once. It has as domain of dependence $D^+(\Sigma) \cup D^-(\Sigma)$ the manifold itself. If an open set $\mathcal{N}$ admits a Cauchy surface then the Cauchy problem for any PDE with initial data on $\mathcal{N}$ is well-defined. This is also equivalent to say that $\mathcal{N}$ is globally hyperbolic.

The requirement (ii) means that it should exist a globally hyperbolic submanifold of spacetime containing both the exterior spacetime *and* the horizon. It is equivalent to say there exists a family of Cauchy hypersurfaces $\Sigma(\tau)$, such that $\Sigma(\tau')$ is inside the domain of dependence of $\Sigma(\tau)$ if $\tau' > \tau$.

Now, the boundary of the black hole is the past event horizon $\mathcal{H}$. It is a null hypersurface with generator l^μ (that is proportional to ξ on $\mathcal{H}$). We can choose to parameterize the Cauchy surfaces $\Sigma(\tau)$ using the affine parameter τ of the null geodesic generator l.

The *area of the horizon* $\mathcal{A}(\tau)$ is then the area of the intersection of $\Sigma(\tau)$ with $\mathcal{H}$. We have to prove that $\mathcal{A}(\tau') > \mathcal{A}(\tau)$ if $\tau' > \tau$.

Sketch of the proof:

The Raychaudhuri equation for the null generator l reads as (1.22). Therefore, wherever the energy condition $R_{\mu\nu}l^\mu l^\nu \geq 0$ hold, the null generator will evolve subject to the inequality

$$\frac{d\theta_{(0)}}{d\tau} \leq -\frac{1}{n-2}\,\theta_{(0)}^2, \tag{1.23}$$

except on possible singular points as caustics. It follows that if $\theta_{(0)}$ becomes negative at any point p on the horizon (i.e. if there is a convergence) then the null generator can continue in the horizon for at most a finite affine distance before reaching a point p at which $\theta_{(0)} \to -\infty$, i.e. a point of infinite convergence representing a caustic beyond which the generators intersects.

Now, from the third property of event horizons above, the generators cannot leave the horizon. Therefore at least two generators cross at p inside $\mathcal{H}$ and, following Hawking and Ellis (Prop 4.5.12 of [11]), they may be deformed to a timelike curve, see figure 7.8. This is however impossible because of the achronicity property of event horizons. Therefore, in order to avoid the contradiction, the point p cannot exist and $\theta_{(0)}$ cannot be negative.

Since (at points where the horizon is not smooth) new null generators may begin but old ones cannot end, equation (1.15) implies that the total area $\mathcal{A}(\tau)$ cannot decrease with increasing τ,

$$\frac{d}{d\tau}\mathcal{A} \geq \oint \theta_{(0)}\, d\mathcal{A} \geq 0. \tag{1.24}$$

This completes the proof.

In particular, if two black holes with area $\mathcal{A}_1$ and $\mathcal{A}_2$ merge then the area $\mathcal{A}_3$ of the combined black hole have to satisfy

$$\mathcal{A}_3 > \mathcal{A}_1 + \mathcal{A}_2. \tag{1.25}$$

The area $A(\tau)$ do not change if $\theta = 0$ on the entire horizon $\mathcal{H}$. The black hole is then stationary.

Note that this derivation implicitly assume regularity properties of the horizon (as piecewise C^2) which may not be true for generic black holes. Recently these gaps in the derivation have been totally filled in [12, 4].

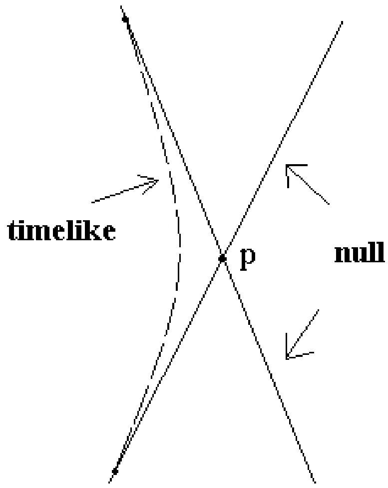


Figure 7.8: If two null generators of $\mathcal{H}$ cross, they may be deformed to a timelike curve.

2 Equilibrium states

2.1 Killing horizons

In any stationary and asymptotically flat spacetime with a black hole, the event horizon is a Killing horizon [11]. This theorem firstly proven by Hawking is called the rigidity theorem. It provides an essential link between event horizons and Killing horizons.[6]

A Killing horizon is a null hypersurface whose normal ξ is a Killing vector

$$\mathcal{L}_\xi g_{\mu\nu} = \xi_{\mu;\nu} + \xi_{\nu;\mu} = 0. \tag{2.1}$$

This additional property will allow us to explore many characteristics of black holes.

The parameter κ which we call now the surface gravity of $\mathcal{H}$ is defined in (1.4). In asymptotically flat spacetimes, the normalization of κ is fixed by requiring $\xi^2 \to -1$ at infinity (similarly, we impose $\xi^2 \to -\frac{r^2}{l^2}$ in asymptotically anti-de Sitter spacetimes).

For Killing horizons, the expansion rate $\theta_{\mu\nu} = \gamma_{(\mu}^{\ \alpha}\gamma_{\nu)}^{\ \beta} D_\alpha \xi_\beta = 0$, so $\theta = \sigma = 0$. Moreover, from (1.18) and (2.1), we deduce $\varsigma = \kappa$. Equation (1.17) then provides an alternative definition for the surface gravity,

$$\kappa^2 = -\frac{1}{2}\xi_{\mu;\nu}\xi^{\mu;\nu}|_{\mathcal{H}}. \tag{2.2}$$

Contracting (1.4) with the transverse null vector n, one has also

$$\kappa = \xi_{\mu;\nu}\xi^\mu n^\nu|_{\mathcal{H}} = \frac{1}{2}(\xi^2)_{,\mu} n^\mu|_{\mathcal{H}}. \tag{2.3}$$

The Raychaudhuri equation (1.22) also states in this case that

$$R_{\mu\nu}\xi^\mu\xi^\nu \stackrel{\mathcal{H}}{=} 0, \tag{2.4}$$

because l is proportional to ξ on the horizon.

From the decomposition (1.12), the irrotationality condition (1.5) and the Killing property $\xi_{[\mu;\nu]} = \xi_{\mu;\nu}$, one can write

$$\xi_{\mu;\nu} \stackrel{\mathcal{H}}{=} \xi_\mu q_\nu - \xi_\nu q_\mu, \tag{2.5}$$

where the covector q_μ can be fixed uniquely by the normalization $q_\mu n^\mu = 0$. Using (2.3), one can further decompose the last equation in terms of $(n, \xi, \{\eta\})$ as

$$\xi_{\mu;\nu} \stackrel{\mathcal{H}}{=} -\kappa(\xi_\mu n_\nu - \xi_\nu n_\mu) + \xi_\mu \hat{\eta}_\nu - \hat{\eta}_\mu \xi_\nu, \tag{2.6}$$

[6]The theorem further assumes the geometry is analytic around the horizon. Actually, there exist a counter-example to the rigidity theorem as stated in Hawking and Ellis [11] but under additional assumptions such as global hyperbolicity and simple connectedness of the spacetime, the result is totally valid [13].

where $\hat{\eta}$ satisfy $\hat{\eta} \cdot \xi = 0 = \hat{\eta} \cdot n$. In particular, it shows that for any spacelike tangent vectors η, $\tilde{\eta}$ to $\mathcal{H}$, one has $\xi_{\mu;\nu}\eta^\mu\tilde{\eta}^\nu \stackrel{\mathcal{H}}{=} 0$.

2.2 Zero law

We are now in position to prove that the surface gravity κ is constant on the horizon under generic conditions. More precisely,

Theorem 2 *Zero law.* [7] If

(i) The spacetime (M, g) admits a Killing vector ξ which is the generator of a Killing horizon $\mathcal{H}$,

(ii) Einstein's equations hold with matter satisfying the dominant energy condition, i.e. $T_{\mu\nu}l^\nu$ is a non-spacelike vector for all $l^\mu l_\mu \leq 0$,

then the surface gravity κ of the Killing horizon is constant over $\mathcal{H}$.

Using the aforementioned properties of null hypersurfaces and Killing horizons, together with

$$\xi_{\nu;\mu;\rho} = R_{\mu\nu\rho}{}^{\tau}\xi_\tau, \tag{2.7}$$

which is valid for Killing vectors, one can obtain (see [2] for a proof)

$$\dot{\kappa} = \kappa_{,\mu}\xi^\mu \stackrel{\mathcal{H}}{=} 0, \tag{2.8}$$

$$\kappa_{,\mu}\eta^\mu \stackrel{\mathcal{H}}{=} -R_{\mu\nu}\xi^\mu\eta^\nu, \tag{2.9}$$

for all spacelike tangent vectors η. Now, from the dominant energy condition, $R_{\mu\nu}\xi^\mu$ is not spacelike. However, the Raychaudhuri equation implies (2.4). Therefore, $R_{\mu\nu}\xi^\mu$ must be zero or proportional to ξ_ν and $R_{\mu\nu}\xi^\mu\eta^\nu = 0$.

This theorem has an extension when gravity is coupled to electromagnetism. If the Killing vector field ξ is also a symmetry of the electromagnetic field up to a gauge transformation, $\mathcal{L}_\xi A_\mu + \partial_\mu \epsilon = 0$, one can also prove that the electric potential

$$\Phi = -(A_\mu\xi^\mu + \epsilon)|_{\mathcal{H}} \tag{2.10}$$

is constant on the horizon.

3 Conservation laws

"*Anybody who looks for a magic formula for "local gravitational energy-momentum" is looking for the right answer to the wrong question. Unhappily, enormous time and effort were devoted in the past to trying to "answer this question" before investigators realized the futility of the enterprise*"

Misner, Thorne and Wheeler [14]

According to Misner, Thorne and Wheeler, the Principle of Equivalence forbids the existence of a localized energy-momentum stress-tensor for gravity. No experiment can be designed to measure a notion of local energy of the gravitational field because in a locally inertial frame the effect of gravity is locally suppressed. However, it is meaningful to ask what is energy content of *region* or of the *totality* of a spacetime. For the first issue, we refer the reader to the literature on recent quasi-local methods [15, 16], see also [17] for the link with pseudo-tensors. Here, we shall deal with the second issue, the best studied and oldest topic, namely, the definition of global conservation laws for gravity (and for general gauge theories).

It exists an overabundant literature over conservations laws. Some methods are prominent but none impose itself as the best one, each of them having overlapping advantages, drawbacks and scope of application. The following presentation will therefore reflect only a biased and narrow view on the topic.

3.1 The generalized Noether theorem

Let us begin the discussion by recalling the first Noether theorem.

Theorem 3 *First Noether Theorem* Any equivalence class of continuous global symmetries of a lagrangian $L\, d^n x$ is in one-to-one correspondence with an equivalence class of conserved currents J^μ, $\partial_\mu J^\mu = 0$.

Here, two global symmetries are equivalent if they differ by a gauge transformation and by a symmetry generated by a parameter vanishing on-shell. Two currents J^μ, J'^μ are equivalent[7] if they differ by a trivial current,

$$J^\mu \sim J'^\mu + \partial_\nu k^{[\mu\nu]} + t^\mu(\frac{\delta L}{\delta \phi}), \qquad t^\mu \approx 0, \tag{3.1}$$

where t^μ depends on the equations of motion (i.e. vanishes on-shell). This theorem is essential in order to define the energy in classical mechanics or in field theories. For example, the total energy of the field associated with a time translation $(\partial_t)^\mu$ on a spacelike Cauchy surface Σ is defined as $E = \int_\Sigma J_\mu n^\mu$ where n^μ is the unit normal to Σ and $J_\mu = T_{\mu\nu}(\partial_t)^\nu$ where $T_{\mu\nu}$ is the conserved stress-tensor of the field ($\partial_\mu T^{\mu\nu} = 0$).

In diffeomorphic invariant theories, the infinitesimal coordinate transformations generated by a vector ξ are pure gauge transformations. The first Noether theorem implies that *all currents J_ξ associated to infinitesimal diffeomorphisms are trivial*[8].

The main lesson of Theorem 3 is that for gauge theories, one should *not* look at conserved currents. In order to generalize the Noether theorem, it is convenient to introduce the notation for $n-p$ forms,

$$(d^{n-p}x)_{\mu_1\dots\mu_{n-p}} = \frac{1}{p!(n-p)!}\epsilon_{\mu_1\dots\mu_n} dx^{\mu_{n-p+1}} \wedge \cdots \wedge dx^{\mu_n}. \tag{3.2}$$

The Noether current can be reexpressed as a $n-1$ form as $J = J^\mu (d^{n-1}x)_\mu$ which is closed, i.e. $dJ = 0$.

Now, we shall see that the conservation laws for gauge theories are *lower degree conservation laws*, involving conserved $n-2$ forms, $k = k^{[\mu\nu]}(d^{n-2}x)_{\mu\nu}$, i.e. such that $dk = 0$ or $\partial_\nu k^{[\mu\nu]} = 0$. Indeed, in the nineties, the following theorem was proved [18, 19], see also [20, 21] for related work and [22] for introductions to local cohomology,

Theorem 4 *Generalized Noether Theorem* Any parameter of a gauge transformation vanishing on-shell such that the parameter itself is non zero on-shell is in one-to-one correspondence with $n-2$-forms k that are conserved on-shell, $dk \approx 0$ (up to trivial $n-2$-forms[9] and up to the addition of the divergence of a $n-3$-form).

Essentially, the theorem amounts first to identify the class of gauge transformations vanishing on-shell with non-trivial gauge parameters with the cohomology group $H^n_2(\delta|d)$ and the class of non-trivial conserved $n-2$ forms with $H^{n-2}_0(d|\delta)$. The proof of the theorem then reduces to find an isomorphism between these cohomology classes.

As a first example, in electromagnetism (that may be coupled to gravity), the trivial gauge transformations $\delta A_\mu = \partial_\mu c = 0$ are generated by constants $c \in \mathbb{R}$. The associated $n-2$-form is simply

$$k_c[A, g] = c\ (d^{n-2}x)_{\mu\nu}\sqrt{-g}F^{\mu\nu} \tag{3.3}$$

which is indeed closed (outside sources that are not considered here) when the equations of motion hold. The closeness of k indicate that the electric charge

$$\mathcal{Q}_E = \oint_S k_{c=1}, \tag{3.4}$$

[7] In $n = 1$ dimension, two currents differing by a constant $J \in \mathbb{R}$ are also considered as equivalent.

[8] For example, in general relativity, one has $\delta L = \delta g_{\mu\nu}\frac{\delta L}{\delta g_{\mu\nu}} + \partial_\mu \Theta^\mu(g, \delta g)$ for some $\Theta(g, \delta g)$. For a diffeomorphism, one has $\delta g_{\mu\nu} = D_\mu \xi_\nu + D_\nu \xi_\mu$, $\mathcal{L}_\xi L = \partial_\mu(\xi^\mu L)$ and $\delta g_{\mu\nu}\frac{\delta L}{\delta g_{\mu\nu}} = -2\sqrt{-g}D_\mu \xi_\nu = -2\partial_\mu(\sqrt{-g}G^{\mu\nu}\xi_\nu)$ where $G_{\mu\nu}$ is the Einstein tensor. What is usually called the Noether current is $J^\mu = \Theta^\mu(g, \mathcal{L}_\xi g) - \xi^\mu L$. However, it is trivial because (as implied by Noether theorem) it exists a $k^{\mu\nu} = k^{[\mu\nu]}$ such that $J^\mu = -2\sqrt{-g}G^{\mu\nu}\xi_\nu + \partial_\nu k^{[\mu\nu]}$.

[9] Trivial $n-2$-forms include superpotentials that vanish on-shell and topological superpotentials that are closed off-shell, see [22] for a discussion of topological conservation laws.

i.e. the integral of k over a closed surface S at constant time, does not depend on time and may be freely deformed in vacuum regions[10].

In generally covariant theories, the trivial diffeomorphisms $\delta g_{\mu\nu} = \mathcal{L}_\xi g_{\mu\nu}$ are generated by the Killing vectors[11]

$$\mathcal{L}_\xi g_{\mu\nu} = \xi_{\mu;\nu} + \xi_{\nu;\mu} = 0. \tag{3.5}$$

However, here comes the problem: there is *no* solution to the Killing equation for *arbitrary* fields. Therefore, *none* vector ξ can be associated to a *generically* conserved $n-2$ form. The hope to associate a conserved quantity

$$\mathcal{Q}_\xi[g] = \oint_S K_\xi[g], \qquad dK_\xi \approx 0, \qquad K_\xi \not\approx 0, \tag{3.6}$$

for a given vector ξ to all solutions g of general relativity is definitively annihilated.

What is the way out? In fact, there are many ways out with different methods and results. In these notes, we will weaken the requirements enormously by selecting special surfaces S, vectors ξ and classes of metrics g and by allowing the conserved quantities

$$\mathcal{Q}_\xi[g,\bar{g}] = \oint_S K_\xi[g,\bar{g}], \qquad dK_\xi \approx 0, \qquad K_\xi \not\approx 0, \tag{3.7}$$

to depend on some background solution $\bar{g}$. If we normalize the charge of the background (typically Minkowski or anti-de Sitter spacetime) to zero, $\mathcal{Q}_\xi[g,\bar{g}]$ will provide a well-defined quantity associated to g and ξ. Let us now explain how the construction works for some particular cases in Einstein gravity.

3.2 The energy in general relativity

Let us only derive the conservation laws obtained originally by Arnowitt, Deser and Misner [23, 24] and Abbott and Deser [25]. For simplicity, only Einstein's gravity is discussed but other gauge theories admits similar structures.

Let us first linearize the Einstein-Hilbert lagrangian $L^{EH}[g]$ with $g = \bar{g} + h$ around a solution $\bar{g}$. It can be shown that the linearized lagrangian $L^{free}[h]$ is gauge invariant under

$$\delta h_{\mu\nu} = \mathcal{L}_\xi \bar{g}_{\mu\nu}, \tag{3.8}$$

where ξ is an arbitrary vector. Now, if the background $\bar{g}_{\mu\nu}$ admits exact Killing vectors, the generalized Noether theorem says that it exists $n-2$ forms $k_\xi[h,\bar{g}]$ which are conserved when h satisfies the linearized equations of motion. In fact, the Killing vectors enumerate all non-trivial solutions to $\mathcal{L}_\xi \bar{g}_{\mu\nu} \approx 0$ in that case and the $n-2$ forms $k_\xi[h,\bar{g}]$ are thus the only non-trivial conserved forms [26].

For Einstein gravity, the $n-2$ form associated to a Killing vector ξ of $\bar{g}$ is well-known to be [27, 28, 29, 30]

$$k_\xi[h,\bar{g}] = -\delta_h K_\xi[g] - i_\xi \Theta[h,\bar{g}] \tag{3.9}$$

where $i_\xi = \xi^\mu \frac{\partial}{\partial dx^\mu}$ is the inner product and the Komar term and the Θ term are given by

$$K_\xi[g] = \frac{\sqrt{-g}}{16\pi G}(D^\mu \xi^\nu - D^\nu \xi^\mu)(d^{n-2}x)_{\mu\nu}, \tag{3.10}$$

$$\Theta[h,g] = \frac{\sqrt{-g}}{16\pi G}(g^{\mu\alpha} D^\beta h_{\alpha\beta} - g^{\alpha\beta} D^\mu h_{\alpha\beta})(d^{n-1}x)_\mu. \tag{3.11}$$

Here, the variation δ_h acts on g and the result of the variation is evaluated on $\bar{g}$.

Now, the point is that this result may be lifted to the full interacting theory (at least) in two different ways :

[10] Explicitly, let S be some surface $r = const$, $t = const$. The r component of $dk = 0$ is $\partial_t k^{tr} + \partial_A k^{Ar} = 0$ where $A = \theta, \phi$ are the angular coordinates. The time derivative of $\mathcal{Q}$ is then given by $\oint_S \partial_t k^{tr} (d^{n-2}x)_{tr} = -\oint_S \partial_A k^{Ar} d\mathcal{A} = 0$ by Stokes theorem. Similarly, the t component of $dk = 0$ is $\partial_r k^{rt} + \partial_A k^{At} = 0$ and $\partial_r \mathcal{Q} = 0$ too.

[11] Trivial diffeomorphisms must also satisfy $\mathcal{L}_\xi \phi^i = 0$ if other fields ϕ^i are present.

(i) For suitable classes of spacetimes $(\mathcal{M}, g)$ with *boundary conditions* in an asymptotic region such that *the linearized theory applies* around some symmetric background $\bar{g}$ in the asymptotic region.

(ii) For classes of solutions $(\mathcal{M}, g)$ with a set of exact Killing vectors.

To illustrate the first case, let take as an example asymptotically flat spacetimes at spatial infinity $r \to \infty$. This class of spacetimes is constrained by the condition $g_{\mu\nu} - \eta_{\mu\nu} = O(1/r)$, where η is the Minkowski metric. We then consider the linearized field $h_{\mu\nu} = g_{\mu\nu} - \eta_{\mu\nu}$ around the background Minkowski metric. Under some appropriate additional boundary conditions, it can be shown that the linearized theory applies at infinity: since Minkowski spacetime admits Killing vectors $\bar{\xi}$, $k_{\bar{\xi}}[h, \bar{g}]$[12] are $n-2$-form conserved in the asymptotic region, i.e. *asymptotically conserved*[13]. The translations, rotations and boosts of Minkowski spacetime are thus associated to energy-momentum and angular momentum. These are the familiar ADM expressions. The conserved quantities in anti-de Sitter spacetime can also be constructed that way.

In the second case, one applies the linearized theory around a family of solution $g_{\mu\nu}$ which have ξ as an exact Killing vector. This allows one to compute the charge difference between $g_{\mu\nu}$ and an infinitely close metric $g_{\mu\nu} + \delta g_{\mu\nu}$. As $dk_\xi[\delta g, g] = 0$ in the whole spacetime, the charge difference $\oint_S k_\xi[\delta g, g]$ does not depend one the choice of integration surface,

$$\oint_S k_\xi[\delta g, g] = \oint_{S'} k_\xi[\delta g, g], \tag{3.12}$$

where S, S' are any $n-2$ surfaces, usually chosen to be $t = \mathit{const}$, $r = \mathit{const}$ in spherical coordinates. The total charge associated to ξ of a solution can then be defined by

$$Q_\xi[g, \bar{g}] = \oint_S \int_{\bar{g}}^{g} k_\xi[\delta g'; g'], \tag{3.13}$$

where $\bar{g}_{\mu\nu}$ is a background solution with charge normalized to zero and g' is the integration variable. The outer integral is performed along a path of solutions. This definition is only meaningful if the charge does not depend on the path, which amount to what is called the *integrability condition*

$$\oint_S (\delta_1 k_\xi[\delta_2 g; g] - \delta_2 k_\xi[\delta_1 g; g]) = 0. \tag{3.14}$$

As a conclusion, we have sketched how one obtains the promised definition of charge (3.7) in the two aforementioned cases. In the first (asymptotic) case, $K_\xi[g, \bar{g}] = k_\xi[h, \bar{g}]$ where $h = g - \bar{g}$ is the linearized field at infinity. In the second (exact) case, $K_\xi[g, \bar{g}] = \int_{\bar{g}}^{g} k_\xi[\delta g'; g']$, where one integrates the linearized form k_ξ along a path of solutions.

Finally note that all results presented here in covariant language have their analogue in Hamiltonian form [31, 29, 32].

4 Quasi-equilibrium states

In 3+1 dimensions, stationary axisymmetric black holes are entirely characterized by their mass and their angular momentum. This fact is part of the *uniqueness theorems*, see [33] for a review. In n dimensions, the situation is more complicated. First, the black hole may rotate in different perpendicular planes. In 3+1 dimensions, the rotation group $SO(3)$ has only one Casimir invariant, but in n dimensions, it has $D \equiv \lfloor (n-1)/2 \rfloor$ Casimirs. Therefore, one expects that, in general, a black hole will have D conserved angular momenta. This is what happens in the higher dimensional Kerr and Reissner-Nordstrøm black holes [34]. Remark that the generalization of rotating black holes to anti-de Sitter backgrounds was done only very recently [35]. So far so good; this is not a big deal with respect to uniqueness.

[12] Note for completeness that the theory of asymptotically conserved $n-2$ forms also allows for an extended notion of symmetry, asymptotic symmetries, where $\mathcal{L}_\xi \bar{g}$ tends to zero only asymptotically.

[13] In fact, boundary conditions are chosen such that the charges are finite, conserved and form a representation of the Poincaré algebra.

The worrying (but interesting) point is that higher dimensions allow for more exotic horizon topologies than the sphere. For example, *black ring* solutions where found [36] recently in 5 dimensions with horizon topology $S^1 \times S^2$. The initial idea of the uniqueness theorems that stationary axisymmetric black holes are entirely characterized by a few number of charges at infinity is thus not valid in higher dimensions. In what follows, we shall derive the first law of black hole mechanics without using uniqueness results.

From now, we restrict ourselves to stationary and axisymmetric black holes with Killing horizon, having ∂_t and ∂_{φ^a} , $1 \leq a \leq D$ as Killing vectors. We allow for arbitrary horizon topology, only assuming the horizon is connected. The Killing generator of the horizon is then a combination of the Killing vectors,

$$\xi = \frac{\partial}{\partial t} + \Omega^a \frac{\partial}{\partial \varphi^a}, \tag{4.1}$$

where Ω^a are called the angular velocities at the horizon.

4.1 The first law for Einstein gravity

We are now set up to present the terms of the first law:

Theorem 5 *First law.* Let $(\mathcal{M}, g)$ and $(\mathcal{M} + \delta\mathcal{M}, g + \delta g)$ be two slightly different stationary black hole solutions of Einstein's equations with Killing horizon. The difference of energy $\mathcal{E}$, angular momenta $\mathcal{J}_a$ and area $\mathcal{A}$ of the black hole are related by

$$\delta\mathcal{E} = \Omega^a\, \delta\mathcal{J}_a + \frac{\kappa}{8\pi}\delta\mathcal{A}, \tag{4.2}$$

where Ω^a are the angular velocities at the horizon and κ is the surface gravity.

This *equilibrium state version* of the first law of black hole mechanics is essentially a balance sheet of energy between two stationary black holes[14]. We shall prove that it comes directly from the equality of the charge related to ξ at the horizon spacelike section H and at infinity, as (3.12),

$$\oint_{S^\infty} k_\xi[\delta g_{\mu\nu}; g_{\mu\nu}] = \oint_H k_\xi[\delta g_{\mu\nu}; g_{\mu\nu}]. \tag{4.3}$$

The energy and angular momenta of the black hole are defined as[15]

$$\delta\mathcal{E} = \oint_{S^\infty} k_{\partial_t}[\delta g_{\mu\nu}; g_{\mu\nu}], \qquad \delta\mathcal{J}^a = -\oint_{S^\infty} k_{\partial_{\varphi^a}}[\delta g_{\mu\nu}; g_{\mu\nu}] \tag{4.4}$$

Therefore, the left-hand side of (4.3) is by definition given by

$$\oint_{S^\infty} k_\xi[\delta g_{\mu\nu}; g_{\mu\nu}] = \delta\mathcal{E} - \Omega\delta\mathcal{J} \tag{4.5}$$

Using (3.9), we may rewrite the right-hand side of (4.3) as

$$\oint_H k_\xi[\delta g_{\mu\nu}; g_{\mu\nu}] = -\delta\oint_H K_\xi[g] + \oint_H K_{\delta\xi}[g] - \oint_H \xi\cdot\Theta[\delta g; g], \tag{4.6}$$

where the variation of ξ that cancels between the two first terms on the right-hand side is put for later convenience.

On the horizon, the integration measure for $n-2$-forms is given by

$$\sqrt{-g}(d^{n-2}x)_{\mu\nu} = \frac{1}{2}(\xi_\mu n_\nu - n_\mu\xi_\nu)d\mathcal{A}, \tag{4.7}$$

[14]It also exists a physical process version, where an infinitesimal amount of matter is send through the horizon from infinity.

[15]The relative sign difference between the definitions of $\mathcal{E}$ and $\mathcal{J}^a$ trace its origin to the Lorentz signature of the metric [28].

where $d\mathcal{A}$ is the angular measure on H. Using the properties of Killing horizons, the Komar integral on the horizon becomes

$$\oint_H K_\xi[g] = -\frac{\kappa \mathcal{A}}{8\pi G}, \tag{4.8}$$

where $\mathcal{A}$ is the area of the horizon. Now, it turns out that

$$\oint_H K_{\delta\xi}[g] - \oint_H \xi \cdot \Theta[\delta g; g] \quad = \quad -\delta\kappa \frac{\mathcal{A}}{8\pi G}. \tag{4.9}$$

The computation which is straightforward but lengthly is done explicitly in Appendix A [16] *without* assuming specific invariance properties under the variation as done in the original derivation [7] and subsequent derivations thereof [1, 28].

The right-hand side of (4.3) is finally given by

$$\oint_H k_\xi[\delta g_{\mu\nu}; g_{\mu\nu}] = \frac{\kappa}{8\pi G}\delta\mathcal{A}, \tag{4.10}$$

as it should and the first law is proven. We can see in this derivation that the first law is a *geometrical* law in the sense that it relates the geometry of Killing horizons to the geometric measure of energy and angular momenta.

Remark finally that the derivation was done in arbitrary dimensions, without hypotheses on the topology of the horizon and for arbitrary stationary variations. The first law also applies in particular for extremal black holes by taking $\kappa = 0$.

4.2 Extension to electromagnetism

It is straightforward to extend the present considerations to the coupled Einstein-Maxwell system. The original derivation was given in [7], see also [3, 37] for alternative derivations. We assume that no magnetic monopole is present, so that the potential A is regular everywhere.

According to the generalized Noether theorem, we first need to extend the notion of symmetry. The gauge transformations of the fields $(g_{\mu\nu}, A_\mu)$ that are zero when the equations of motion are satisfied are given by

$$\mathcal{L}_\xi g_{\mu\nu} \quad = \quad 0, \tag{4.11}$$

$$\mathcal{L}_\xi A_\mu + \partial_\mu \epsilon \quad = \quad 0. \tag{4.12}$$

These equations are the generalized Killing equations. We consider only stationary and axisymmetric black holes with Killing horizon which have as a solution to these equations $(\xi, \epsilon) = (\partial_t, 0)$, $(\xi, \epsilon) = (\partial_{\varphi^a}, 0)$ for $a = 1...D$ and $(\xi, \epsilon) = (0, \epsilon \in \mathbb{R})$. The conserved quantities associated to these symmetry parameters are respectively the energy, the angular momenta and the electric charge. The existence of the electric charge independently from the other charges suggests that it will show up in the first law.

The conserved superpotential associated to a symmetry parameter (ξ, ϵ) can be shown to be

$$k^{tot}_{\xi,\epsilon}[\delta g, \delta A; g, A] = k^{grav}_\xi - \delta K^{em}_{\xi,\epsilon} + K^{em}_{\delta\xi,\delta\epsilon} - \xi \cdot \Theta^{em} \tag{4.13}$$

where $k^{grav}_\xi[\delta g, g]$ is the gravitational contribution (3.9)[17],

$$K^{em}_{\xi,\epsilon}[g, A] = \frac{\sqrt{-g}}{16\pi G}[F^{\mu\nu}(\xi^\alpha A_\alpha + \epsilon)](d^{n-2}x)_{\mu\nu} \tag{4.14}$$

and

$$\Theta^{em}[g, A; \delta A] = \frac{\sqrt{-g}}{16\pi G} F^{\alpha\mu}\delta A_\alpha (d^{n-1}x)_\mu. \tag{4.15}$$

Let us now look at the fundamental equality (4.3) of the first law where we choose ξ as the Killing generator and $\epsilon = 0$. The superpotential $k = k^{tot}$ should contain the electromagnetic contributions as well.

[16] I thanks G. Barnich for his suggestion of this computation.

[17] We use "geometrized units" for the electromagnetic field; the lagrangian is $L = \frac{\sqrt{-g}}{16\pi G}(R - \frac{1}{4}F^2)$.

On the one hand, the energy and angular momenta are still defined by (4.4) where k is given by (4.13). For usual potentials, the electromagnetic contributions vanish at infinity and only the gravitational contributions are important[18]. Equation (4.5) still hold. On the other hand, at the horizon, the electromagnetic field is not negligible and we have

$$\oint_H k^{tot}_{\xi,0} = \frac{\kappa}{8\pi}\delta\mathcal{A} - \delta\oint_H K^{em}_{\xi,\epsilon} + \oint_H K^{em}_{\delta\xi,0} - \oint_H \xi\cdot\Theta^{em}. \tag{4.16}$$

Now, remember that the zero law extended to electromagnetism said that $\Phi = -(\xi^\alpha A_\alpha)|_H$ is constant on the horizon[19]. Therefore, we have directly

$$\oint_H K^{em}_{\xi,0} = -\Phi\mathcal{Q}. \tag{4.17}$$

In order to work out the two remaining terms on the right hand side of (4.16), let us rewrite equation (4.12) as

$$\mathcal{L}_\xi A_\mu = \xi^\nu F_{\nu\mu} + (\xi^\alpha A_\alpha)_{,\mu} = 0. \tag{4.18}$$

Since Φ is constant on the horizon, this equation says that the electric field $E^\mu \equiv F^{\mu\nu}\xi_\nu$ with respect to ξ satisfy $\eta_\mu E^\mu \overset{\mathcal{H}}{=} 0$ for all tangent vectors η^μ, therefore E^μ is proportional to ξ^μ or more precisely,

$$F^{\mu\nu}\xi_\nu \overset{\mathcal{H}}{=} (F^{\nu\rho}\xi_\nu n_\rho)\xi^\mu. \tag{4.19}$$

Using the last relation with (4.7), it is easy to show that

$$\oint_H K^{em}_{\delta\xi,0} - \oint_H \xi\cdot\Theta^{em} \quad = \quad -\delta\Phi\mathcal{Q} - \oint_H \frac{d\mathcal{A}}{4\pi G} n_\mu F^{\mu\nu}\delta A_\nu \xi^2. \tag{4.20}$$

The last term vanishes because of (1.1). Finally, we obtain the first law valid when electromagnetic fields are present,

$$\delta\mathcal{E} - \Omega_a\delta\mathcal{J}^a = \frac{\kappa}{8\pi G}\delta\mathcal{A} + \Phi\delta\mathcal{Q}. \tag{4.21}$$

Note finally that the first applies for any theory of gravity, from arbitrary diffeomorphic invariant Lagrangians [28], to black holes in non-conventional background geometries [38, 39] or to black objects in string theory or supergravity [40, 41].

In addition, Hawking [42] discovered that $T \equiv \frac{\kappa\hbar}{2\pi}$ is the temperature of the quantum radiation emitted by the black hole. Moreover, as tells us the second law, $S \equiv \frac{\mathcal{A}}{4\hbar G}$ is a quantity that can classically only increase with time. It suggests [43] to associate an entropy S to the black hole. This striking occurrence of thermodynamics emerging out of the structure of gravity theories is a fundamental issue to be further explained in still elusive quantum gravity theories.

4.3 Extension to any theory of gravity

Let us finally briefly review the proposal of Iyer and Wald [28, 44] for the extension of the first law to an arbitrary diffeomorphic invariant Lagrangian. Let

$$L[\phi] = L[g_{\mu\nu}, R_{\mu\nu\rho\sigma;(\alpha)}], \tag{4.22}$$

be the lagrangian where ϕ denote all the fields of the theory and (α) denotes an arbitrary number of symmetrized derivatives.

According to the authors, the charge difference between the solutions ϕ and $\phi+\delta\phi$ corresponding to the vector ξ can be written as (4.6) [28, 44] where the Komar term $\oint K_\xi$ is given by

$$\oint (d^{n-2}x)_{\mu\nu}\left(\sqrt{-g}\xi^\mu W^\nu[\phi] + Y^{\mu\nu}[\mathcal{L}_\xi\phi,\phi] - \frac{\delta L}{\delta R_{\mu\nu\alpha\beta}}\xi_{\alpha;\beta} - (\mu\leftrightarrow\nu)\right) \tag{4.23}$$

[18] Note however that if the gauge potential tends to a constant at infinity, the quantities (4.4) will contain a contribution from the electric charge. We choose for convenience a potential vanishing at infinity. The contributions from the electromagnetic fields will then comes only from the surface integral over the horizon.

[19] We previously assumed that $\mathcal{L}_\xi A_\mu = 0$. So $\epsilon = 0$ here.

and where $\Theta[\delta\phi;\phi]$, $W^\mu[\phi]$ and $Y^{\mu\nu}[\mathcal{L}_\xi\phi,\phi]$ have a generic form which we shall not need here. The energy and angular momentum are defined as previously by (4.4) where the $n-2$ form k_ξ is defined by (4.6).

Therefore, equations (4.3)-(4.5)-(4.6) are still valid with appropriate $K_\xi[\phi]$ and $\Theta[\delta\phi;\phi]$. Now, we assume that the surface gravity κ is not vanishing and that the horizon generators are geodesically complete to the past. Then, it exists [45] a special $n-2$ spacelike surface on the horizon, the bifurcation surface, where the Killing vector ξ vanish. For Killing vectors ξ, i.e. such that $\mathcal{L}_\xi\phi^i = 0$ and for solutions of the equations of motion, k_ξ is a closed form, (3.12) hold and we may evaluate the right-hand side of (4.6) on the bifurcation surface. There, assuming regularity conditions, only the third term in the Komar expression (4.23) contributes [28, 46] and the right-hand side of (4.6) becomes

$$\frac{\kappa}{2\pi}\delta\mathcal{S}, \tag{4.24}$$

where the "higher order in the curvature" entropy is defined by

$$\mathcal{S} = -8\pi\oint_H (d^{n-2}x)_{\mu\nu}\frac{\delta L}{\delta R_{\mu\nu\alpha\beta}}\xi_\alpha n_\beta. \tag{4.25}$$

The first law (4.2) therefore holds with appropriate definitions of energy, angular momentum and entropy.

For the Einstein-Hilbert lagrangian, we have

$$\frac{\delta L^{EH}}{\delta R_{\mu\nu\alpha\beta}} = \frac{\sqrt{-g}}{32\pi G}(g^{\mu\alpha}g^{\nu\beta} - g^{\mu\beta}g^{\nu\alpha}), \tag{4.26}$$

and the entropy (4.25) reduces to the familiar expression $\mathcal{A}/4G$.

Acknowledgments

I am grateful to G. Barnich, V. Wens, M. Leston and S. Detournay for their reading of the manuscript and for the resulting fruitful discussions. Thanks also to all participants and organizers of the Modave school for this nice time. This work is supported in part by a "Pôle d'Attraction Interuniversitaire" (Belgium), by IISN-Belgium, convention 4.4505.86, by the National Fund for Scientific Research (FNRS Belgium), by Proyectos FONDECYT 1970151 and 7960001 (Chile) and by the European Commission program MRTN-CT-2004-005104, in which the author is associated to V.U. Brussel.

Appendix

A Completion of the proof of the first law

Let us prove the relation (4.9). We consider any stationary variation of the fields $\delta g_{\mu\nu}$, $\delta\xi^\mu$, i.e. such that

$$\mathcal{L}_\xi\delta g_{\mu\nu} + \mathcal{L}_{\delta\xi}g_{\mu\nu} = 0. \tag{A1}$$

The variation is chosen to commute with the total derivative, i.e. the coordinates are left unchanged $\delta x^\mu = 0$.

Using the decomposition (4.7), the left-hand side of equation (4.9) can written explicitly as

$$\begin{aligned}\oint_H K_{\delta\xi}[g] - \oint_H \xi\cdot\Theta[\delta g; g] \;&=\; \oint_H \frac{d\mathcal{A}}{16\pi G}\Big(-\delta\xi^{\mu;\nu}(\xi_\mu n_\nu - \xi_\nu n_\mu) \\ &\quad +\xi^\mu(\delta g_{\mu\nu}{}^{;\nu} - g^{\alpha\beta}\delta g_{\alpha\beta;\mu})\Big).\end{aligned} \tag{A2}$$

We have to relate this expression to the variation of the surface gravity κ. This is merely an exercise of differential geometry.

Since the horizon $S(x) = 0$ stay at the same location in x^μ, the covariant vector normal to the horizon $\xi_\mu = f\partial_\mu S$, where f is a κ-dependent normalization function, satisfies

$$\delta\xi_\mu \stackrel{\mathcal{H}}{=} \delta f \xi_\mu, \tag{A3}$$

where $\delta\xi_\mu \equiv \delta(g_{\mu\nu}\xi^\nu)$. From the variation of (1.1) and of the second normalization condition (1.6), one obtains

$$\delta\xi^\mu\xi_\mu \stackrel{\mathcal{H}}{=} 0, \qquad \delta n^\mu \xi_\mu \stackrel{\mathcal{H}}{=} \delta f, \tag{A4}$$

which shows that $\delta\xi^\mu$ has no component along n^μ and δn^μ has a component along n^μ which equals $-\delta f$[20].

Let us develop the variation of κ starting from the definition (2.3). One has

$$\begin{aligned}\delta\kappa &= \frac{1}{2}(\xi^\mu\xi_\mu)_{;\nu}\delta n^\nu + \frac{1}{2}(\delta\xi_\mu\xi^\mu + \xi_\mu\delta\xi^\mu)_{;\nu}n^\nu, &\text{(A6)}\\ &= \frac{1}{2}\delta\xi_{\mu;\nu}(\xi^\mu n^\nu + \xi^\nu n^\mu) + \xi^\mu{}_{;\nu}(\delta\xi_\mu n^\nu + \xi_\mu\delta n^\nu)\\ &\quad + \frac{1}{2}n^\nu(\xi_\mu\delta\xi^\mu)_{;\nu} - \frac{1}{2}n^\nu\mathcal{L}_\xi\delta\xi_\nu, &\text{(A7)}\end{aligned}$$

where all expressions are implicitly pulled-back on the horizon. The first term in (A7) is recognized as $-\frac{1}{2}\delta\xi_{\mu;\nu}g^{\mu\nu}$ after using (1.9), (A3) and (2.6). According to (A3)-(A4), the second term can be written as

$$\xi^\mu{}_{;\nu}(\delta\xi_\mu n^\nu + \xi_\mu\delta n^\nu) = \xi_{\mu;\nu}\xi^\mu\delta\eta^\nu, \tag{A8}$$

for some $\delta\eta^\nu$ tangent to $\mathcal{H}$. This term vanishes thanks to (2.6). The third term can be written as

$$\frac{1}{2}n^\nu(\xi_\mu\delta\xi^\mu)_{;\nu} = -\frac{1}{2}n^\nu\mathcal{L}_{\delta\xi}\xi_\nu + n_\nu\xi_\mu\delta\xi^{\mu;\nu}. \tag{A9}$$

Now, the Lie derivative of $\delta\xi_\mu$ along ξ can be expressed as

$$\mathcal{L}_\xi\delta\xi_\mu = -\mathcal{L}_{\delta\xi}\xi_\mu, \tag{A10}$$

by using the Killing equation (2.1) and its variation (A1). The fourth term can then be written as

$$-\frac{1}{2}n^\nu\mathcal{L}_\xi\delta\xi_\nu = \frac{1}{2}n^\nu\mathcal{L}_{\delta\xi}\xi_\nu. \tag{A11}$$

Adding all the terms, the variation of the surface gravity becomes

$$\begin{aligned}\delta\kappa &= -\frac{1}{2}(\delta\xi_\mu)^{;\mu} + \delta\xi^{\mu;\nu}\xi_\mu n_\nu,\\ &= -\frac{1}{2}\delta g_{\mu\nu}{}^{;\mu}\xi^\nu - \frac{1}{2}\delta\xi^\mu{}_{;\mu} + \frac{1}{2}\delta\xi^{\mu;\nu}(\xi_\mu n_\nu - \xi_\nu n_\mu) + \frac{1}{2}\delta\xi^{\mu;\nu}(\xi_\mu n_\nu + \xi_\nu n_\mu),\\ &= -\frac{1}{2}\delta g_{\mu\nu}{}^{;\mu}\xi^\nu - \delta\xi^\mu{}_{;\mu} + \frac{1}{2}\delta\xi^{\mu;\nu}(\xi_\mu n_\nu - \xi_\nu n_\mu) + \frac{1}{2}\delta\xi^{\mu;\nu}\gamma_{\mu\nu}. &\text{(A12)}\end{aligned}$$

The last line is a consequence of (1.9). Contracting (A1) with $g^{\mu\nu}$ we also have

$$\delta\xi^\mu{}_{;\mu} = -\frac{1}{2}\xi^\mu g^{\alpha\beta}\delta g_{\alpha\beta;\mu}. \tag{A13}$$

Finally, the last term in (A12) reduces to $\frac{1}{2}\delta t^\alpha{}_{|\alpha}$ where $|\alpha$ denotes the covariant derivative with respect to the $n-2$ metric $\gamma_{\mu\nu}$ and $\delta t^\mu = \gamma^\mu{}_\nu\delta\xi^\nu$ is the pull-back of $\delta\xi^\mu$ on $\mathcal{H}$. Indeed, one has

$$\begin{aligned}\frac{1}{2}\delta\xi^{\mu;\nu}\gamma_{\mu\nu} &= \frac{1}{2}\delta t^{\mu;\nu}\gamma_{\mu\nu}, &\text{(A14)}\\ &= \frac{1}{2}(\delta t^\mu{}_{,\nu}\gamma_\mu{}^\nu + \Gamma_{\mu;\nu\alpha}\gamma^{\mu\nu}\delta t^\alpha), &\text{(A15)}\\ &= \frac{1}{2}\delta t^\mu{}_{|\mu}, &\text{(A16)}\end{aligned}$$

[20]Note also the following property that is useful in order to prove the first law in the way of [28]. Using (2.5) and (A3), we have

$$\delta\xi_{\mu,\nu} - \delta\xi_{\nu,\mu} \stackrel{\mathcal{H}}{=} \xi_\mu(\delta f q_\nu + \delta q_\nu) - \xi_\nu(\delta f q_\mu + \delta q_\mu). \tag{A5}$$

It implies in particular that the expression $\delta\xi_{[\mu;\nu]}$ has no tangential-tangential component, $\delta\xi_{[\mu;\nu]}\eta^\mu\tilde{\eta}^\nu \stackrel{\mathcal{H}}{=} 0$, $\forall\eta,\tilde{\eta}$ orthogonal to $\mathcal{H}$.

where $|\mu$ denotes the covariant derivative with respect to the $n-2$ metric $\gamma_{\mu\nu}$. The first line uses (2.1)-(A4) and $\gamma_{\mu\nu}\xi^\nu = 0 = \gamma_{\mu\nu}n^\nu$. The last line uses the decomposition (1.9) and $\delta t^\alpha n_\alpha = 0 = \delta t^\alpha \xi_\alpha$. We have finally the result

$$\delta\kappa = -\frac{1}{2}\delta g_{\mu\nu}{}^{;\mu}\xi^\nu + \frac{1}{2}\xi^\mu g^{\alpha\beta}\delta g_{\alpha\beta;\mu} + \frac{1}{2}\delta\xi^{\mu;\nu}(\xi_\mu n_\nu - \xi_\nu n_\mu) + \frac{1}{2}(\gamma^\mu{}_\nu \delta\xi^\nu)_{|\mu}. \tag{A17}$$

Expression (A2) is therefore equals to

$$\oint_H K_{\delta\xi}[g] - \oint_H \xi\cdot\Theta[\delta g; g] \quad = \quad -\oint_H \frac{d\mathcal{A}}{8\pi G}\delta\kappa, \tag{A18}$$

and the result (4.9) follows because $\delta\kappa$ is constant on the horizon.

Remark that in classical derivations [7, 1], it is assumed that the Killing vectors ∂_t and ∂_φ have the same components before and after the variation,

$$\delta(\partial_t)^\mu = \delta(\partial_\varphi)^\mu = 0.$$

One then has $\delta\xi^\mu = \delta\Omega(\partial_\varphi)^\mu$ and the variation of κ reduces to the well-known expression

$$\delta\kappa = -\frac{1}{2}\delta g_{\mu\nu}{}^{;\mu}\xi^\nu + \delta\Omega(\partial_\varphi)^{\mu;\nu}\xi_\mu n_\nu.$$

References

[1] B. Carter, "Black hole equilibrium states," *in Black holes, eds. C. DeWitt and B. S. DeWitt, Gordon and Breach* (1973).

[2] B. Carter, "Mathematical foundations of the theory of relativistic stellar and black hole congurations," *Gravitation in Astrophysics, ed. B. Carter and J. B. Hartle, Cargese Summer School 1986, Plenum Press, New York and London 1987.*

[3] P. K. Townsend, "Black holes," gr-qc/9707012.

[4] R. M. Wald, "The thermodynamics of black holes," *Living Rev. Rel.* **4** (2001) 6, gr-qc/9912119.

[5] S. W. Hawking, "Black holes in general relativity," *Commun. Math. Phys.* **25** (1972) 152–166.

[6] J. D. Bekenstein, "Black holes and the second law," *Nuovo Cim. Lett.* **4** (1972) 737–740.

[7] J. M. Bardeen, B. Carter, and S. W. Hawking, "The four laws of black hole mechanics," *Commun. Math. Phys.* **31** (1973) 161–170.

[8] A. Ashtekar *et al.*, "Isolated horizons and their applications," *Phys. Rev. Lett.* **85** (2000) 3564–3567, gr-qc/0006006.

[9] A. Ashtekar and B. Krishnan, "Isolated and dynamical horizons and their applications," *Living Rev. Rel.* **7** (2004) 10, gr-qc/0407042.

[10] R. Wald, *General Relativity.* The University of Chicago Press, 1984.

[11] S. Hawking and G. Ellis, *The large scale structure of space-time.* Cambridge University Press, 1973.

[12] P. T. Chrusciel, E. Delay, G. J. Galloway, and R. Howard, "The area theorem," *Annales Henri Poincare* **2** (2001) 109–178, gr-qc/0001003.

[13] P. T. Chrusciel, "On rigidity of analytic black holes," *Commun. Math. Phys.* **189** (1997) 1–7, gr-qc/9610011.

[14] C. Misner, K. Thorne, and J. Wheeler, *Gravitation.* W.H. Freeman, New York, 1973.

[15] J. D. Brown and J. York, James W., "The microcanonical functional integral. 1. the gravitational field," *Phys. Rev.* **D47** (1993) 1420–1431, gr-qc/9209014.

[16] J. D. Brown and J. York, James W., "Quasilocal energy and conserved charges derived from the gravitational action," *Phys. Rev.* **D47** (1993) 1407–1419.

[17] C.-C. Chang, J. M. Nester, and C.-M. Chen, "Pseudotensors and quasilocal gravitational energy- momentum," *Phys. Rev. Lett.* **83** (1999) 1897–1901, gr-qc/9809040.

[18] G. Barnich, F. Brandt, and M. Henneaux, "Local BRST cohomology in the antifield formalism. I. General theorems," *Commun. Math. Phys.* **174** (1995) 57–92, hep-th/9405109.

[19] G. Barnich, F. Brandt, and M. Henneaux, "Local brst cohomology in einstein yang-mills theory," *Nucl. Phys.* **B455** (1995) 357–408, hep-th/9505173.

[20] R. Bryant and P. Griffiths, "Characteristic cohomology of differential systems I: General theory," *J. Am. Math. Soc.* **8** (1995) 507.

[21] I. Anderson, "The variational bicomplex," tech. rep., Formal Geometry and Mathematical Physics, Department of Mathematics, Utah State University, 1989.

[22] C. G. Torre, "Local cohomology in field theory with applications to the Einstein equations," hep-th/9706092. Lectures given at 2nd Mexican School on Gravitation and Mathematical Physics, Tlaxcala, Mexico, 1-7 Dec 1996.

[23] R. Arnowitt, S. Deser, and C. Misner, *Gravitation, an Introduction to Current Research*, ch. 7. The Dynamics of General Relativity, pp. 227–265. Wiley, New York, 1962.

[24] R. Arnowitt, S. Deser, and C. W. Misner, "The dynamics of general relativity," gr-qc/0405109.

[25] L. F. Abbott and S. Deser, "Stability of gravity with a cosmological constant," *Nucl. Phys.* **B195** (1982) 76.

[26] G. Barnich, S. Leclercq, and P. Spindel, "Classification of surface charges for a spin 2 field on a curved background solution," gr-qc/0404006.

[27] L. F. Abbott and S. Deser, "Stability of gravity with a cosmological constant," *Nucl. Phys.* **B195** (1982) 76.

[28] V. Iyer and R. M. Wald, "Some properties of Noether charge and a proposal for dynamical black hole entropy," *Phys. Rev.* **D50** (1994) 846–864, gr-qc/9403028.

[29] I. M. Anderson and C. G. Torre, "Asymptotic conservation laws in field theory," *Phys. Rev. Lett.* **77** (1996) 4109–4113, hep-th/9608008.

[30] R. M. Wald and A. Zoupas, "A general definition of conserved quantities in general relativity and other theories of gravity," *Phys. Rev.* **D61** (2000) 084027, gr-qc/9911095.

[31] T. Regge and C. Teitelboim, "Role of surface integrals in the Hamiltonian formulation of general relativity," *Ann. Phys.* **88** (1974) 286.

[32] G. Barnich and F. Brandt, "Covariant theory of asymptotic symmetries, conservation laws and central charges," *Nucl. Phys.* **B633** (2002) 3–82, hep-th/0111246.

[33] M. Heusler, "Stationary black holes: Uniqueness and beyond," *Living Rev. Rel.* **1** (1998) 6.

[34] R. C. Myers and M. J. Perry, "Black holes in higher dimensional space-times," *Ann. Phys.* **172** (1986) 304.

[35] G. W. Gibbons, H. Lu, D. N. Page, and C. N. Pope, "The general Kerr-de Sitter metrics in all dimensions," hep-th/0404008.

[36] R. Emparan and H. S. Reall, "A rotating black ring in five dimensions," *Phys. Rev. Lett.* **88** (2002) 101101, hep-th/0110260.

[37] S. Gao, "The first law of black hole mechanics in einstein-maxwell and einstein-yang-mills theories," *Phys. Rev.* **D68** (2003) 044016, `gr-qc/0304094`.

[38] G. W. Gibbons, M. J. Perry, and C. N. Pope, "The First Law of Thermodynamics for Kerr-Anti-de Sitter Black Holes," `hep-th/0408217`.

[39] G. Barnich and G. Compere, "Conserved charges and thermodynamics of the spinning goedel black hole," *Phys. Rev. Lett.* **95** (2005) 031302, `hep-th/0501102`.

[40] G. T. Horowitz and D. L. Welch, "Exact three-dimensional black holes in string theory," *Phys. Rev. Lett.* **71** (1993) 328–331, `hep-th/9302126`.

[41] K. Copsey and G. T. Horowitz, "The role of dipole charges in black hole thermodynamics," `hep-th/0505278`.

[42] S. W. Hawking, "Particle creation by black holes," *Commun. Math. Phys.* **43** (1975) 199–220.

[43] J. D. Bekenstein, "Generalized second law of thermodynamics in black hole physics," *Phys. Rev.* **D9** (1974) 3292–3300.

[44] V. Iyer and R. M. Wald, "A comparison of Noether charge and Euclidean methods for computing the entropy of stationary black holes," *Phys. Rev.* **D52** (1995) 4430–4439, `gr-qc/9503052`.

[45] I. Racz and R. M. Wald, "Extension of space-times with killing horizon," *Class. Quant. Grav.* **9** (1992) 2643–2656.

[46] T. Jacobson, G. Kang, and R. C. Myers, "On black hole entropy," *Phys. Rev.* **D49** (1994) 6587–6598, `gr-qc/9312023`.

Basics of Open String Field Theory

Carlo Maccaferri[1]

Theoretische Natuurkunde, Vrije Universiteit Brussel,
Physique Théorique et Mathématique, Université Libre de Bruxelles,
International Solvay Institutes
Pleinlaan 2, B-1050 Brussels, Belgium

Abstract. We give an elementary introduction to Open String Field Theory and the physics of Tachyon Condensation in the context of the Bosonic String. After a general overview on the subject we define the basic ingredients of OSFT and compute the 0 level Tachyon Potential. We define in a quite detailed way the 3 string vertex and we compute the matter Neumann coefficients in the zero momentum sector. We then outline the properties of the theory at the Tachyon Vacuum and we end with some speculations about the emergence of closed string degrees of freedom.

[1]email: carlo.maccaferri@ulb.ac.be

1 Introduction

String Theory is, at the time of writing, a very (if not the only) promising way to describe our universe in a consistent and unified theoretical framework. It provides a perturbative formulation of quantum gravity and it incorporates non abelian gauge theories. Our perturbative understanding is well established and leads to the formulation of the celebrated 5 (super)string models (Type IIA/B, $SO(32)$–Type I, $SO(32)/E_8 \times E_8$–Heterotic). These theories are (perturbatively) theories of open (type I) and closed (the others) supersymmetric strings in ten dimensional flat space time; such strings vibrate generating (as harmonics) an infinite tower of particles, some of them are massless with appropriate polarization tensors. The effective field theory of such massless fields is a supergravity theory in ten dimension (coupled to Super Yang Mills in the Type I and Heterotic cases).

1.1 D–branes

In these supergravity theories there are black–hole like solutions which are extended in space. These solutions have a definite tension (mass per unit volume) and are charged by the massless p–forms of the corresponding string spectrum / supergravity multiplet. One of the main results of the last decade is the recognition that such supergravity solitons admit a microscopic string theory description: they are Dirichlet branes (D–branes). They can be described as hypersurfaces in space time on which open strings are constrained to end (in this sense $SO(32)$–type I is a theory of (unoriented) open strings ending on 16 space–time filling D9–branes). However these objects are not just boundary conditions, they are genuine dynamical objects that can move in spacetime; moreover they are physical sources for closed strings. This can be understood in the following way. Imagine to have 2 parallel D–branes and consider an open string connecting the two D–branes, then consider the one loop partition function of this string; graphically this corresponds to a cylinder connecting the two D–branes which, in turn, can be interpreted as an exchange of a closed string between two sources.

This example shows that open and closed strings are deeply intertwined and cannot be studied separately: a theory of open strings generates closed string poles at one loop and, on the other hand, closed strings are sourced by the D–branes on which open strings live on.

The discovery of D–branes has been a key element to understand that the five distinct string theories just mentioned above (plus a still not defined theory, dubbed M–Theory, whose low energy limit is eleven dimensional supergravity) are related to each others by suitable duality transformations. This web of dualities points towards the existence of a single theory to be formulated with a unique and complete set of variables, which reduces to the known superstring theories on particular points of its moduli space. This still hypothetical theory should give a non–perturbative definition of quantum gravity and, as such, should be background independent: space time itself should arise dynamically as a coherent state from nothing.

1.2 Tachyons

In this scenario the relevance of tachyons is fundamental. From a point particle point of view tachyons are particles that propagate faster than light, violating causality. Equivalently they are relativistic particles with negative $mass^2$. This is however a fake understanding of tachyons and, from a first quantized point of view, a tachyon is just an inconsistency of a theory. On the other hand, in field theory, we know that the concept of mass (of a scalar) arises from the quadratic term of the scalar potential around a stationary point. If the quadratic term is positive then, by quantizing the theory, we get a massive particle but, if the quadratic term is negative, we get a tachyon that simply indicates that we are quantizing the theory on an unstable vacuum, perturbation theory breaks down and some phase transition takes place, driving the theory to a new stable vacuum, with a different perturbative spectrum.

There are many unstable vacua in string theory, signaled by corresponding tachyons on the string spectrum. The simplest example to think about is just the 26 dimensional closed bosonic string in flat space. This theory is not supersymmetric, does not contain fermions (it is in fact not very realistic...) and has a low lying state that is a tachyon. Although this is the simplest tachyon that one encounters in the first study of string theory, it is also the most mysterious one: indeed it signals the instability of the 26 dimensional bosonic spacetime itself and it is not clear at all if some decaying process can bring it to a new stable spacetime (maybe the 10 dimensional supersymmetric spacetimes?).

There is another very simple tachyon, the open string one. A theory of open strings is however a theory of D–branes, as open strings are just excitations of them. If an open string theory contains a tachyon, this can only mean that the corresponding D–brane system is unstable. In particular all bosonic D–branes are not charged and they all contain a tachyon in their spectra. In this sense the 26 dimensional open string tachyon is just the signal of the instability of the space filling D25–brane who does not have any charge protecting it. These examples might look academical as they are in the realm of the bosonic string, there are however other open string tachyons in the supersymmetric string, the bosonic case is just a simpler example of the same kind of phenomenon.

Type–IIA/B theories contain stable D–branes of even/odd dimensions, these branes are stable because they carry RR–charge and source the corresponding RR–massless fields of the closed string sector, they are BPS states and break half of the supersymmetry of the bulk space time. From an open string point of view they don't contain tachyons in their spectra because the NS tachyon has been swept away by the GSO projections needed to keep modular invariance at one loop. Since these branes are charged they possess an orientation given by the corresponding RR–p form which is a volume form for the brane's worldvolume. A D–brane of opposite orientation is just a D–brane with opposite RR–charge, an anti–D–brane. Now, if a D–brane and an anti–D–brane are placed parallely at a distance less then the fundamental string length, there is a tachyon corresponding to the lowest state of open strings stretched between the two branes, that arises because such open strings undergo the opposite GSO projection. This tachyon is just the signal that a D–brane/anti–D–brane system is unstable as it does not posses a global RR–charge.

There are also single D–branes which are unstable, these are the branes of wrong dimensionality (odd for Type–IIA, even for Type–IIB), the non–BPS D–branes. The instability is due to the fact that these branes are not charged as there are no RR–p forms in the closed string sector that can couple to them. And there is again the corresponding tachyon in the open string sector coming from the opposite GSO–projection w.r.t. the stable case.

Where these instabilities drive the theory? Is there a stable vacuum to decay to? The answer (at least for the case of open string tachyons) is yes: an unstable system of branes decays to a vacuum where it ceases to exist and its mass is converted in closed string radiation that can propagate in the bulk. This phenomenon is known as Tachyon Condensation.

Why the study of tachyon condensation is important? It is so because the decay of unstable objects is a physical process that interpolates between two different vacua (the unstable one and the stable one). In other words, tachyon condensation is a natural path to explore the (open)–string landscape and, hence, to address in an explicit physical example the study of the elusive concept of background independence.

Needless to say that open string tachyon condensation is just (one of) the starting point(s) for the project of a background–independent formulation of string theory. The mysterious closed string tachyon (who represents the instabilities of space–time itself) still waits for a convincing interpretation. Nevertheless it appears that the physics of open tachyon condensation is still rich enough to get insights into non–perturbative string theory. This is so because of the profound (an not yet fully understood) relation between open and closed strings.

1.3 Open/Closed Duality

One of the latest biggest achievements of string theory is the AdS/CFT correspondence which states (in its strongest formulation) that quantum Type–IIB closed string theory on $AdS_5 \times S^5$ with N units of RR 5–form flux is dual to $\mathcal{N} = 4$ $U(N) - SYM$ theory which lives on the projective 4–dimensional boundary of AdS. This correspondence basically states that a (perturbative) quantum theory of gravity on a given asymptotic background is fully captured by a Yang–Mills theory which is, in turn, the low energy limit of open string theory on N D3–branes in flat space: in other words the open strings dynamics on D–branes in flat space gives us a quantum theory of gravity in a space time which is the result of the back–reaction of the branes on the original flat geometry. This is not the stating that the full closed string Hilbert space (with all the possible changes in the closed string background) is captured by a particular D–branes configuration, but it means that a complete quantum formulation of open string theory on such D–brane configuration gives a consistent and unitary quantum theory of gravity on a given asymptotic spacetime. A change in the D–brane system produces a different back reaction, hence a different spacetime. It is maybe too much optimistic to think that all closed strings background can be obtained in this way, but this is certainly an interesting way of thinking at background independence.

1.4 String Field Theories

These examples show that, even if we only know the perturbative expansion around some particular background, there are quite convincing physical reasons to believe that we can understand how string theory backgrounds are dynamically connected. However we have to face the problem that the perturbative formulation of string theory is explicitly non background independent. This in fact is mostly a consequence of the first quantized formulation. History teaches us that the most complete theory of particles has been achieved by passing from first quantization to second quantization, that is from Quantum Mechanics to Quantum Field Theory. It is only in the framework of QFT that one can have control of the vacua of a theory and how such vacua are dynamically connected via nonperturbative effects (tunneling, dynamical symmetry breaking, confinement, etc...). In the theory of particles we see that the right language to describe physics is to promote every particle with a corresponding space–time field and then proceeding with quantization. What about strings? Even in first quantization we see that the quantum fluctuations of a single string give rise to an infinite set of particles: some of them are massless, some of them may be tachyonic and infinite of them are massive. Passing from first quantization to second quantization leads to a QFT with an infinite number of space time fields: the task seems impossible both from a conceptual (infinity means no knowledge in physics) and from a computational (infinite interactions for a given physical process) point of view. However string theory is not just a theory of infinite interacting particles, there is order inside. What marks the difference with respect to particles is the conformal symmetry of the worldsheet theory: this symmetry gives us a consistent and unique interacting scheme (at every order in perturbation theory) starting with non interacting strings. This is like having a rule that gives us (unambiguously) vertices of Feynmann diagrams from the free propagators! In this sense first quantized string theory contains informations about the full non perturbative theory, they are just hidden inside.

A vacuum of string theory is identified once the string propagates in such a way that the corresponding worldsheet theory is conformal, in other words a vacuum of string theory is a two dimensional conformal field theory. What about the string spectrum (the infinite on–shell particles obtained from the vibration of the string)? They are perturbations of the vacuum, hence they correspond to (infinitesimal) deformations of the underlying conformal field theory. However these are not generic deformations but are such as to preserve conformal symmetry, they are marginal deformations. In this language the string's landscape has an intriguing description: it is the space of two–dimensional field theories. Some points in this space are conformal field theories and correspond to exact string backgrounds (vacua). Around each vacuum there are marginal directions which deform the CFT while maintaining conformal invariance, this infinitesimal deformations are the string excitations around that particular vacuum. Some of these infinitesimal deformations can be exponentiated to a finite one, giving a one parameter family of CFT's/strings vacua. There can be also vacuum points which cannot be connected through (time independent) marginal deformations but that are the result of an RG–flow to some IR fixed point. In this language non–perturbative string theory can be identified with the dynamics of two dimensional field theories. It is evident that such an understanding is equivalent to a second quantized formulation of string theory: a String Field Theory.

In a String Field Theory framework, the basic degrees of freedom are all the possible deformations (string fields) of a given reference conformal field theory one starts with. Such theories admit classical solutions which are in one–to–one correspondence to exact backgrounds of string theory which, in principle, can be completely disconnected from the starting background. They also have, being "field" theories, an off–shell extension of the corresponding first quantized theory: hence they can properly describe non perturbative transitions between different vacua. There are formulations of closed and open string field theories.

While closed string field theory has a complicated non polynomial form that has proven to be resistent to any kind of analytic treatment, Open String Field Theory has a remarkable simple structure which is of Chern–Simons form. The theory is simple enough to do numerical studies on the structure of its vacua. It is fair to say that a complete formulation (where explicit computations can be performed) exists up to now only for the bosonic open string and for the NS sector of the open superstring. In both cases a study of tachyon condensation has been proved possible and a non trivial tachyon potential has been seen to emerge from the level truncated action of (Super) Open String Field Theory.

Very recently the first non trivial solution representing the open string Tachyon vacuum has been derived by Martin Schnabl, [17]. This puts the Open Bosonic String Field Theory in a privileged status as it opens the way for future analytic studies on the non perturbative dynamics of String Theory.

These lectures are meant to be a (very) basic introduction to OSFT, the interested reader is strongly suggested to improve her/his understanding on the already existing reviews [2, 3, 4, 5, 6, 7], where a complete set of references can be found.

2 Open String Field Theory: an outline

Open String Field Theory, [1], is a second quantized formulation of the open bosonic string. Its fundamental degrees of freedom are the open string fields, namely all kinds of vertex operators (primary and not primary) that can be inserted at the boundary of a given bulk CFT, which represents a (once and for all) fixed closed string background, for example flat space–time.

The explicit action of OSFT is derived starting from the perturbative vacuum representing a given (exactly solvable) Boundary Conformal Field Theory. In most application this $BCFT$ is the D25–brane's one, with Neumann boundary conditions on all the (non–interacting) space–time directions.

We will begin being rather formal, concentrating on the abstract properties of the various objects that define the string field theory action. We will then give precise and computable definitions.

2.1 Open Bosonic Strings

To start with, we recall some general feature of the first quantized bosonic string on flat space.

One starts with the σ-model which describes the embedding of the coordinates X^μ from the worldsheet Σ

$$S = -\frac{1}{4\pi\alpha'}\int_\Sigma \sqrt{-\gamma}\gamma^{ab}\,\partial_a X^\mu\,\partial_b X_\mu \tag{2.1}$$

Gauge fixing the metric to $\gamma_{ab} = e^\phi\delta_{ab}$ and adding the corresponding ghost/antighost system in order to stay on the gauge slice, leaves us with the combined matter plus ghost CFT

$$S = -\frac{1}{4\pi\alpha'}\int_\Sigma d^2z\,\partial X^\mu\bar\partial X_\mu + \frac{1}{\pi}\int_\Sigma d^2z\ \ (b\bar\partial c + \bar b\partial\bar c) \tag{2.2}$$

The BRST charge of the above gauge fixing is given by

$$Q = \frac{1}{2\pi i}\oint dz\,(cT_{matter} + :bc\partial c:) \tag{2.3}$$

where T_{matter} is the stress tensor of the X^μ CFT (D–free bosons).

It is well known that

$$Q^2 = 0 \quad \Leftrightarrow \quad D = 26 \tag{2.4}$$

since in this case the total central charge is vanishing and the worldsheet theory is non–anomalous.

The set of physical states is given by vertex operators of ghost number 1 in the cohomology of Q.

$$Q|V_{phys}\rangle = 0, \quad |V_{phys}\rangle \neq Q|W\rangle \tag{2.5}$$

The previous relation is intended to be applied on insertion of operators on the boundary of the Riemann surface Σ (this is because we are dealing with *open* strings).

Via the state–operator correspondence (which always holds in a radial quantized CFT) we can associate to any operator a corresponding state in the set of boundary deformations of our CFT.

$$|\phi\rangle = \phi(z=0)|0\rangle \tag{2.6}$$

where $|0\rangle$ is the SL(2,R) invariant vacuum (of ghost number zero).

2.2 Linearized action

We want a space-time action whose variation yields the physical condition coming from the BRST quantization of the worldsheet BCFT.

The job is readily done by the following

$$S_{kin}[\psi] = \langle \psi, Q_{BRST}\psi\rangle \tag{2.7}$$

In the above formula ψ is a classical string field: a generic vertex operator of ghost number 1; Q_{BRST} is the first quantized BRST operator and the inner product $\langle\cdot,\cdot\rangle$ is the *bpz* inner product, relative to the $BCFT_0$ in consideration (the D25–brane); namely

$$\begin{aligned} \langle \phi, \psi\rangle &= \langle I\circ\phi(0)\;\psi(0)\rangle_{BCFT_0} \\ I(z) &= -\frac{1}{z} \end{aligned} \tag{2.8}$$

Note that

$$\langle A\rangle \neq 0 \quad \Rightarrow \quad gh(A) = +3 \tag{2.9}$$

This condition asks for the classical string field ψ to be at ghost number 1.

By varying the kinetic action we get the linearized equation of motion

$$Q_{BRST}|\psi\rangle = 0 \tag{2.10}$$

which is the usual on–shell condition for vertex operator of ghost number 1. The action possesses a (reducible) gauge invariance

$$\delta\psi = Q_{BRST}|\Lambda\rangle \tag{2.11}$$

for a generic string field Λ of ghost number 0, this gauge symmetry is reducible because we can have string fields of any negative ghost number, hence we have to mod out the previous gauge transformation by Q_{BRST}–closed string fields of ghost number zero, and so on. This is achieved using BV quantization (this is very well explained in the review by Thorn).

We have then seen that the non trivial solution of the linearized equation of motion are in one to one correspondence with the usual open string spectrum.

To see how this works practically, let's plug in the action the tachyon vertex operator

$$|\psi\rangle = \int dk t(k) e^{ikx} c_1|0\rangle \tag{2.12}$$

Due to the fact that

$$\{Q_{BRST}, b_0\} = L_0 \tag{2.13}$$

and

$$b_0|\psi\rangle = 0 \tag{2.14}$$

We have that

$$\langle\psi|Q_{BRST}|\psi\rangle = \langle\psi|c_0 L_0|\psi\rangle \tag{2.15}$$

Leading to

$$\langle\psi, Q_{BRST}\psi\rangle = \int dkt(-k)(k^2-1)t(k) = \int dxt(x)(\Box - 1)t(x) \tag{2.16}$$

This is nothing but the free space time action for the open string tachyon. In the same way one can get the free actions for the photon and for the other massive modes of the open string.

2.3 The interacting action

The kinetic action describes the small fluctuations of the perturbative vacuum that are identified by the (infinitesimal) boundary marginal deformations of $BCFT_0$. The simplest covariant way to introduce interactions is to add a cubic term to the action

$$S[\psi] = -\frac{1}{g_o^2}\left(\frac{1}{2}\langle\psi, Q_{BRST}\psi\rangle + \frac{1}{3}\langle\psi, \psi * \psi\rangle\right) \tag{2.17}$$

Note that we have normalized the action with the open string coupling constant g_o.

The cubic term is constructed using the operation $*$ which is an associative non–commutative product in the Hilbert space of string fields.

$$(\psi_1 * \psi_2) * \psi_3 = \psi_1 * (\psi_2 * \psi_3) \tag{2.18}$$

The Q_{BRST} operator is a derivation of the $*$–algebra

$$Q_{BRST}(\psi_1 * \psi_2) = (Q_{BRST}\psi_1) * \psi_2 + (-1)^{|\psi_1|}\psi_1 * (Q_{BRST}\psi_2), \tag{2.19}$$

where $|\psi_1|$ is the grassmannality of the string field, the ghost number in the case of the bosonic string.

We also ask for ciclicity

$$\langle A, B * C\rangle = (-1)^{|A|(|B|+|C|)}\langle B, C * A\rangle \tag{2.20}$$

Using the fact that

$$\langle Q_{BRST}(...)\rangle_{BCFT_0} = 0 \tag{2.21}$$

one can easily prove that the above action is invariant under the following gauge transformation

$$\delta\psi = Q_{BRST}|\Lambda\rangle + [\psi, \Lambda]_* \tag{2.22}$$

This infinitesimal gauge transformation can be extended to a finite one

$$\psi' = \mathrm{e}^{\Lambda}\left(Q_{BRST} + \psi\right)\mathrm{e}^{-\Lambda}, \tag{2.23}$$

where the exponentials are in the $*$–product sense. The addition of just a cubic coupling to make the action interacting can seem a bit arbitrary and maybe simple too much. This is however the only consistent choice one can make as it gives a unique and complete covering of the moduli space of Riemann surfaces with boundary, [12]. In other words, any worldsheet of an arbitrary number of external legs and loops can be uniquely recovered by an appropriate Feynmann diagram build with the cubic vertex and the propagator.

The equation of motion are obtained by varying the action with respect to ψ and reads

$$Q_{BRST}|\psi\rangle + |\psi\rangle * |\psi\rangle = 0 \tag{2.24}$$

Given a solution ψ_0 of the equation of motion one can shift the the string field in the following way

$$\psi = \psi_0 + \phi \tag{2.25}$$

Then the action can be rewritten as

$$S[\psi] = S[\psi_0] + -\frac{1}{g_o^2}\left(\frac{1}{2}\langle\phi, Q_{\psi_0}\phi\rangle + \frac{1}{3}\langle\phi, \phi * \phi\rangle\right) \tag{2.26}$$

where the new kinetic operator Q_{ψ_0} is defined

$$Q_{\psi_0}\phi = Q_{BRST}\phi + \{\psi_0, \phi\}_* \tag{2.27}$$

The quantity $S[\psi_0]$ is the action evaluated at the classical solution, if this solution is static (it has no kinetic energy), then this quantity corresponds to the static energy of ψ_0, in particular

$$-\frac{S[\psi_0]}{V^{(26)}} = \tau_{\psi_0} \tag{2.28}$$

where τ_{ψ_0} is the *tension* of ψ_0, the space–averaged energy mod space.

3 The $*$–product

The key element that makes String Field Theory an interacting theory is the promotion of the string field Hilbert space to a non commutative algebra. As already said in the previous chapter this is achieved by introducing a multiplication rule between string fields, the $*$ product. It is time now to explore its definition and its properties. We will first give an heuristic definition based only on the embedding coordinates in the target space $X^\mu(\sigma)$ (the *matter* sector). The matter string field can be understood as a functional of the string embedding coordinates (Schrodinger representation)

$$|\psi\rangle \quad \Rightarrow \quad \psi[X^\mu(\sigma)] = \langle X^\mu(\sigma)|\psi\rangle \tag{3.1}$$

$$\tag{3.2}$$

the states $|X^\mu(\sigma)\rangle$ are the open string position eigenstates

$$\hat{X}^\mu(\sigma)\,|X^\mu(\sigma)\rangle = X^\mu(\sigma)\,|X^\mu(\sigma)\rangle \tag{3.3}$$

$$\langle X^\mu(\sigma)|Y^\nu(\sigma')\rangle = \eta^{\mu\nu}\delta[X(\sigma) - Y(\sigma')] \tag{3.4}$$

The worldsheet parameter σ spans the whole open string and it lies in the interval $[0, \pi]$. For the definition of the $*$ product it is necessary to split the string into its left and right part, so we define

$$\hat{l}^\mu(\sigma) = \hat{X}^\mu(\sigma) \qquad 0 \le \sigma < \frac{\pi}{2} \tag{3.5}$$

$$\hat{r}^\mu(\sigma) = \hat{X}^\mu(\pi - \sigma) \qquad \frac{\pi}{2} < \sigma \le \pi \tag{3.6}$$

The midpoint $\sigma = \frac{\pi}{2}$ cannot be left/right decomposed so we will treat it as a separate coordinate (even if it is a part of a continuum).

$$\hat{x}_m^\mu = \hat{X}^\mu\left(\frac{\pi}{2}\right) \tag{3.7}$$

given these definitions the string field can be expressed as a functional of the midpoint and left/right degrees of freedom

$$\psi[X^\mu(\sigma)] = \psi[x_m^\mu;\, l^\mu(\sigma), r^\mu(\sigma)] \tag{3.8}$$

The bpz inner product can be expressed as a functional integration with respect *all* degrees of freedom

$$\begin{aligned}
\langle\psi|\phi\rangle &= \int \mathcal{D}X(\sigma)\,\langle\psi|X(\sigma)\rangle\,\langle X(\sigma)|\phi\rangle \\
&= \int \mathcal{D}X(\sigma)\,\psi[X(\pi-\sigma)]\,\phi[\sigma] \\
&= \int dx_m \mathcal{D}l(\sigma)\mathcal{D}r(\sigma)\,\psi[x_m;\, l(\sigma), r(\sigma)]\,\phi[x_m, r(\sigma), l(\sigma)]
\end{aligned} \tag{3.9}$$

Note that this operation consists in gluing two strings with opposite left/right orientation. Since all the degrees of freedom are integrated, one is left with just a pure number. This is reminiscent of the trace of the product of two infinite matrices.

The star product between two string fields is another string field defined in the following way

$$(\psi * \phi)[x_m;\, l(\sigma), r(\sigma)] = \int \mathcal{D}y(\sigma)\psi[x_m;\, l(\sigma), y(\sigma)]\phi[x_m;\, y(\sigma), r(\sigma)] \tag{3.10}$$

This operation consists in identifying the left half of the first string with the right half of the second string, integrating the overlapping degrees of freedom as to reproduce a third string. This is analogous to the multiplication of two infinite matrices.

4 Conformal definition of the cubic vertex

The representation we gave of the star product is very intuitive, but it is not very convenient to do practical computation. Here we want to give a precise definition of the cubic vertex of OSFT based on conformal field theory. Let's consider the most general coupling between 3 string fields

$$\langle \psi, \phi * \chi \rangle = \tag{4.1}$$

The definitions of the next section imply that we have to compute

$$= \int \mathcal{D}x\mathcal{D}y\mathcal{D}z; \psi[z, y]\phi[y, x]\chi[x, z] \tag{4.2}$$

This can be interpreted as a correlation function on the 2 dimensional world sheet quantum field theory which is a *conformal* field theory. Using the operator/state correspondence we can define our string fields as generic vertex operators inserted at the origin of the complex upper half plane

$$|A\rangle = A(z = 0)|0\rangle \tag{4.3}$$

Let us consider three unit semidisks in the upper half z_a $(a = 1, 2, 3)$ plane. Each one represents the string freely propagating in semicircles from the origin (world-sheet time $\tau = -\infty$) to the unit circle $|z_a| = 1$ $(\tau = 0)$, where the interaction is supposed to take place. We map each unit semidisk to a 120° wedge of the complex plane via the following conformal maps:

$$f_a(z_a) = \alpha^{2-a} f(z_a)\,,\ a = 1, 2, 3 \tag{4.4}$$

where

$$f(z) = \Big(\frac{1 + iz}{1 - iz}\Big)^{\frac{2}{3}} \tag{4.5}$$

Here $\alpha = e^{\frac{2\pi i}{3}}$ is one of the three third roots of unity. In this way the three semidisks are mapped to nonoverlapping (except at the interaction points $z_a = i$) regions in such a way as to fill up a unit disk centered at the origin. The curvature is zero everywhere except at the center of the disk, which represents the common midpoint of the three strings in interaction.

The interaction vertex is defined by a correlation function on the disk in the following way

$$\langle \psi, \phi * \chi \rangle = \langle f_1 \circ \psi(0)\, f_2 \circ \phi(0)\, f_3 \circ \chi(0)\rangle \tag{4.6}$$

4.1 The level 0 tachyon potential

Let's see how this work for the zero momentum tachyon, given by

$$|T\rangle = T\, c(0)|0\rangle \tag{4.7}$$

We have then to compute

$$\langle T, T * T\rangle = T^3\langle f_1 \circ c(0)\, f_2 \circ c(0)\, f_3 \circ c(0)\rangle \tag{4.8}$$

Since $c(z)$ is a primary field of weight -1 we have

$$f \circ c(z) = \frac{1}{f'(z)}\, c(f(z)) \tag{4.9}$$

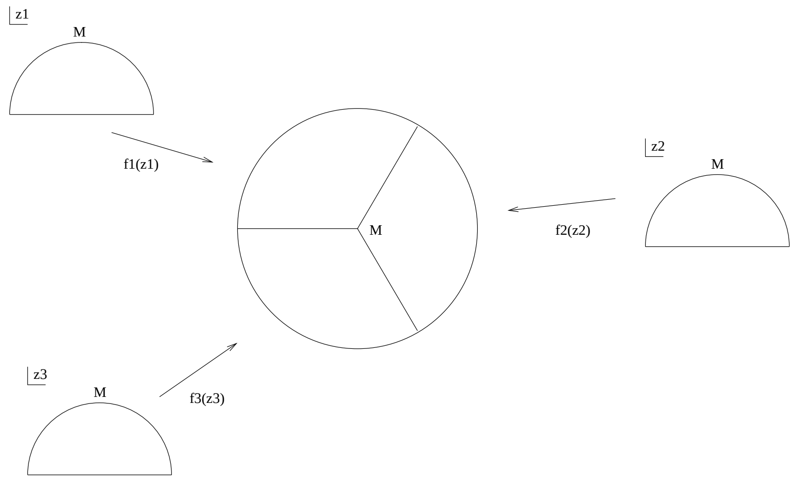


Figure 8.9: *The conformal maps from the three unit semidisks to the three-wedges unit disk*

In particular, using the explicit definitions and the well known

$$\langle c(z_1)\, c(z_2)\, c(z_3)\rangle = (z_1 - z_2)(z_2 - z_3)(z_3 - z_1) \tag{4.10}$$

we get explicitly

$$\begin{aligned} &\langle f_1 \circ c(0)\, f_2 \circ c(0)\, f_3 \circ c(0)\rangle \\ = \frac{1}{f_1'(0) f_2'(0) f'3(0)} &\langle c(f_1(0))\, c(f_2(0))\, c(f_3(0))\rangle \;=\; \sqrt{3}\,\frac{81}{64} \end{aligned} \tag{4.11}$$

Collecting the result of the kinetic term computation, we can extract the *tachyon potential*, that is (minus) the action for the zero momentum tachyon vertex operator

$$V(T) = -S(T) = \frac{1}{g^2}\left(\frac{1}{2}\langle T, QT\rangle + \frac{1}{3}\langle T, T * T\rangle\right) = M_{D25}(2\pi^2)\left(-\frac{1}{2}T^2 + \frac{1}{3}\sqrt{3}\,\frac{81}{64}\,T^3\right) \tag{4.12}$$

where we have used the relation between the D25–brane tension and the open string coupling constant ($\alpha' = 1$)

$$M_{D25} = \frac{1}{2\pi^2} \tag{4.13}$$

The critical points of this potential are at $T = 0$ (*perturbative vacuum*) and at $T = T^* = \frac{64\sqrt{3}}{81}$ (*non–perturbative vacuum*). At the non–perturbative vacuum we get

$$V(T^*) = -0.684 M_{D25} \tag{4.14}$$

so we already see, although we have truncated our string field to just the tachyon vertex operator (zero level), the energy density of the new vacuum accounts for 68% of the D25 D–brane tension!

5 Three strings vertex and matter Neumann coefficients

A very convenient representation for explicit computations with many vertex operators is via the definition of the 3–strings vertex. This object lives on three copies of the string Hilbert space and defines the $*$ product in the following way

$${}_3\langle\psi * \phi| =_{123} \langle V_3|\, |\psi\rangle_1 |\phi\rangle_2 \tag{5.1}$$

The three strings vertex of Open String Field Theory is given by

$$|V_3\rangle = \int d^{26}p_{(1)} d^{26}p_{(2)} d^{26}p_{(3)} \delta^{26}(p_{(1)} + p_{(2)} + p_{(3)}) \exp(-E)\, |0, p\rangle_{123} \tag{5.2}$$

where

$$E = \sum_{a,b=1}^{3} \left(\frac{1}{2} \sum_{m,n\geq 1} \eta_{\mu\nu} a_m^{(a)\mu\dagger} V_{mn}^{ab} a_n^{(b)\nu\dagger} + \sum_{n\geq 1} \eta_{\mu\nu} p_{(a)}^{\mu} V_{0n}^{ab} a_n^{(b)\nu\dagger} + \frac{1}{2} \eta_{\mu\nu} p_{(a)}^{\mu} V_{00}^{ab} p_{(b)}^{\nu} \right) \qquad (5.3)$$

Summation over the Lorentz indices $\mu, \nu = 0, \ldots, 25$ is understood and η denotes the flat Lorentz metric. The operators $a_m^{(a)\mu}, a_m^{(a)\mu\dagger}$ denote the non–zero modes matter oscillators of the a–th string, which satisfy

$$[a_m^{(a)\mu}, a_n^{(b)\nu\dagger}] = \eta^{\mu\nu} \delta_{mn} \delta^{ab}, \qquad m, n \geq 1 \qquad (5.4)$$

$p_{(r)}$ is the momentum of the a–th string and $|0,p\rangle_{123} \equiv |p_{(1)}\rangle \otimes |p_{(2)}\rangle \otimes |p_{(3)}\rangle$ is the tensor product of the Fock vacuum states relative to the three strings. $|p_{(a)}\rangle$ is annihilated by the annihilation operators $a_m^{(a)\mu}$ and it is eigenstate of the momentum operator $\hat{p}_{(a)}^{\mu}$ with eigenvalue $p_{(a)}^{\mu}$. The normalization is

$$\langle p_{(a)} | p'_{(b)} \rangle = \delta_{ab} \delta^{26}(p + p')$$

The symbols $V_{nm}^{ab}, V_{0m}^{ab}, V_{00}^{ab}$ are called the Neumann coefficients .

An important ingredient in the following are the *bpz* transformation properties of the oscillators

$$bpz(a_n^{(a)\mu}) = (-1)^{n+1} a_{-n}^{(a)\mu}$$

Our purpose here is to discuss the definition and the properties of the three strings vertex by exploiting as far as possible the conformal definition given in the previuos section. To start with, we consider the string propagator at two generic points of this disk. The Neumann coefficients V_{NM}^{ab} are nothing but the Fourier modes of the propagator with respect to the original coordinates z_a. For the sake of simplicity we will just review the non–zero modes (zero momentum sector).

From the definition of the vertex it is clear that we have

$$V_{nm}^{ab} = \langle V_{123} | a_{-n}^{(a)} a_{-m}^{(b)} | 0 \rangle_{123} \qquad (5.5)$$

On the other hand, from the conformal definition of the vertex

$$\langle V_{123} | a_{-n}^{(a)} a_{-m}^{(b)} | 0 \rangle_{123} = \langle f_a \circ a_n^{\dagger} \, f_b \circ a_m^{\dagger} \rangle \qquad (5.6)$$

This allows as to directly compute the Neumann coefficients, using

$$a_n^{\dagger} = \frac{1}{\sqrt{n}} \oint dz \, \frac{1}{z^n} \, i\partial X(z) \qquad (5.7)$$

Indeed we have

$$\begin{aligned} V_{mn}^{ab} &= \frac{1}{\sqrt{nm}} \oint dz \oint dw \frac{1}{z^n w^m} \langle f_a \circ i\partial X(z) \, f_b \circ i\partial X(w) \rangle & (5.8) \\ &= -\frac{1}{\sqrt{nm}} \oint \frac{dz}{2\pi i} \oint \frac{dw}{2\pi i} \frac{1}{z^n} \frac{1}{w^m} f'_a(z) \frac{1}{(f_a(z) - f_b(w))^2} f'_b(w) & (5.9) \end{aligned}$$

where we have used the fact that ∂X is a weight 1 primary

$$f \circ \partial X(z) = f'(z) \partial X(f(z)) \qquad (5.10)$$

and the two–point function (propagator)

$$\langle \partial X(z) \, \partial X(w) \rangle = \frac{1}{(z-w)^2} \qquad (5.11)$$

It is easy to check that

$$\begin{aligned} V_{mn}^{ab} &= V_{nm}^{ba} \\ V_{mn}^{ab} &= (-1)^{n+m} V_{mn}^{ba} \\ V_{mn}^{ab} &= V_{mn}^{a+1,b+1} \end{aligned} \qquad (5.12)$$

In the last equation the upper indices are defined mod 3.

Let us consider the decomposition

$$V^{ab}_{mn} = \frac{1}{3}\Big(C_{nm} + \bar{\alpha}^{a-b} U_{nm} + \alpha^{a-b}\bar{U}_{nm}\Big) \tag{5.13}$$

After some algebra one gets

$$\begin{aligned} C_{nm} &= \frac{-1}{\sqrt{nm}} \oint \frac{dz}{2\pi i} \oint \frac{dw}{2\pi i} \frac{1}{z^n}\frac{1}{w^m}\Big(\frac{1}{(1+zw)^2} + \frac{1}{(z-w)^2}\Big) \\ U_{nm} &= \frac{-1}{3\sqrt{nm}} \oint \frac{dz}{2\pi i} \oint \frac{dw}{2\pi i} \frac{1}{z^n}\frac{1}{w^m}\Big(\frac{f^2(w)}{f^2(z)} + 2\frac{f(z)}{f(w)}\Big)\Big(\frac{1}{(1+zw)^2} + \frac{1}{(z-w)^2}\Big) \\ \bar{U}_{nm} &= \frac{-1}{3\sqrt{nm}} \oint \frac{dz}{2\pi i} \oint \frac{dw}{2\pi i} \frac{1}{z^n}\frac{1}{w^m}\Big(\frac{f^2(z)}{f^2(w)} + 2\frac{f(w)}{f(z)}\Big)\Big(\frac{1}{(1+zw)^2} + \frac{1}{(z-w)^2}\Big) \end{aligned} \tag{5.14}$$

The integrals can be directly computed in terms of the Taylor coefficients of f. The result is

$$C_{nm} = (-1)^n \delta_{nm} \tag{5.15}$$

$$\begin{aligned} U_{nm} &= \frac{1}{3\sqrt{nm}} \sum_{l=1}^{m} l \Big[(-1)^n B_{n-l}B_{m-l} + 2b_{n-l}b_{m-l}(-1)^m \\ &\quad -(-1)^{n+l}B_{n+l}B_{m-l} - 2b_{n+l}b_{m-l}(-1)^{m+l}\Big] \end{aligned} \tag{5.16}$$

$$\bar{U}_{nm} = (-1)^{n+m} U_{nm} \tag{5.17}$$

where we have set

$$\begin{aligned} f(z) &= \sum_{k=0}^{\infty} b_k z^k \\ f^2(z) &= \sum_{k=0}^{\infty} B_k z^k, \qquad \text{i.e.} \qquad B_k = \sum_{p=0}^{k} b_p b_{k-p} \end{aligned} \tag{5.18}$$

Eqs.(5.15, 5.16, 5.17) are obtained by expanding the relevant integrands in powers of z, w and correspond to the pole contributions around the origin. We notice that the above integrands have poles also outside the origin, but these poles either are not in the vicinity of the origin of the z and w plane, or, like the poles at $z = w$, simply give vanishing contributions. By changing $z \to -z$ and $w \to -w$, it is easy to show that

$$(-1)^n U_{nm}(-1)^m = \bar{U}_{nm}, \qquad \text{or} \qquad CU = \bar{U}C, \qquad C_{nm} = (-1)^n \delta_{nm} \tag{5.19}$$

In the second part of this equation we have introduced a matrix notation which we will use throughout. One can use this representation for (5.16, 5.17) to make computer calculations. For instance it is easy to show that the equations

$$\sum_{k=1}^{\infty} U_{nk} U_{km} = \delta_{nm}, \qquad \sum_{k=1}^{\infty} \bar{U}_{nk} \bar{U}_{km} = \delta_{nm} \tag{5.20}$$

are satisfied to any desired order of approximation. Each identity follows from the other by using (5.19). Using (5.20), together with the decomposition (5.13), it is easy to establish the commutativity relation (written in matrix notation)

$$[CV^{ab}, CV^{a'b'}] = 0 \tag{5.21}$$

for any a, b, a', b'. This relation is fundamental for the next developments.
It is common to define

$$\begin{aligned} X &= CV^{11} \\ X_+ &= CV^{12} \\ X_- &= CV^{21} \end{aligned} \tag{5.22}$$

Using (5.20), together with the decomposition (5.13), it is easy to establish the following linear and non linear relations (written in matrix notation).

$$\begin{aligned} X + X_+ + X_- &= 1 \\ X^2 + X_+^2 + X_-^2 &= 1 \\ X_+^3 + X_-^3 &= 2X^3 - 3X^2 + 1 \\ X_+ X_- &= X^2 - X \\ [X, X_\pm] &= 0 \\ [X_+, X_-] &= 0 \end{aligned} \tag{5.23}$$

These very important properties encode the associativity of the matter star product.

6 String Field Theory at the Tachyon Vacuum

6.1 Sen's Conjectures

Open String Field Theory is formulated around the D25-brane vacuum, exhibiting an instability due to the presence of the open string tachyon. As reviewed in the introduction, such an instability is understood as the instability of the D25–brane itself. Indeed bosonic D–branes (as well as D–branes anti D–branes pairs and non–BPS D–branes in superstring theory) do not possess any charge which can prevent them from decaying. To see if there is a stable point in the tachyon potential is a task that can be taken over by looking at the space time effective action of string field theory. Explicit numerical computations can be performed if the string spectrum is truncated up to a certain level, so as to have a finite number of spacetime fields. Level truncation is an approximation scheme by which one can recover a more and more precise effective field theory from the exact but somehow formal string field theory action. It consists in expanding the string field up to a certain level (the eigenvalue of the N operator) and in explicitly computing the action using the prescriptions of the previous section to compute $*$–products and *bpz*–inner products. A truncated string field takes the form

$$|\Psi\rangle = (\phi(x) + A_\mu(x) a_1^{\mu\dagger} + \ldots) c_1 |0\rangle \tag{6.1}$$

Plugging this expression in the action one ends up with a local action for the component fields up to a certain level

$$\mathcal{S}(\Psi) = \int d^{26}x F(\varphi_i, \partial\varphi_i, ...) \tag{6.2}$$

This action is a purely spacetime action and one can extract form it an effective tachyon potential.

Here we do not attempt at all to give a review of the level truncation computations from which the tachyon potential has been obtained, we just quote that a strong evidence that a local minimum exists has been achieved (see [2] for a pedagogical review of the level truncation technique). By truncating the action at a finite level one ends up with an effective tachyon potential, that have the qualitative form showed in figure

Lot of computations has been done to assure that the following statements about the tachyon vacuum are true

- The energy difference between the perturbative vacuum and the tachyon vacuum exactly matches with the D25–brane energy, hence it represents a configuration with no D–branes at all,
- The cohomology around the tachyon vacuum is trivial (at least at ghost number one), indicating that there are no physical perturbative open string states around it
- Lower dimensional D–branes can be obtained as tachyonic lumps in which, along the transverse directions, the tachyon reaches its minimum in the potential at $\pm\infty$ and it is vanishing at the origin, the energy of such lump solutions matches with the lower dimensional D–branes energy.

Such statements are known as Sen's conjectures and, thanks to the great amount of evidence reached, they are now accepted as the fundamental properties of Tachyon Condensation.

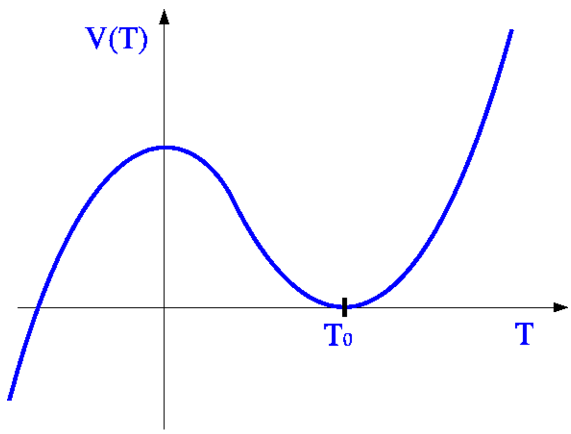


Figure 8.10: *The tachyon potential of open string field theory: the local maximum represents the unstable D25–brane, while the local minimum is the tachyon vacuum*

6.2 OSFT with no open strings

The second Sen's conjecture states that at the tachyon vacuum the theory we get is rather strange. In particular it states that there are no physical excitation. This is a spectacular phenomenon that has no counterpart in usual local field theories. This is well understandable from a physical point of view. We are dealing with a theory of open strings, and open strings states are what we call the physical excitations. Now, if we are sitting in a vacuum with no D-branes at all we cannot expect to have a physical open string spectrum, as open strings cannot attach anywhere. How can the tachyon vacuum BRST charge be trivial when the one we are starting with is not? The answer can be found in non trivial nature of the classical solution describing the tachyon vacuum. We remind that given a classical solution of OSFT, the induced BRST charge / kinetic term is given by

$$Q_{\psi_0}\phi = Q_{old}\phi + \psi_0 * \phi + \phi * \psi_0 \tag{6.3}$$

Let's now introduce the identity string field, $|I\rangle$. By definition it is the identity of the star algebra.

$$|I\rangle * |\phi\rangle = |\phi\rangle * |I\rangle = |\phi\rangle \tag{6.4}$$

One can now show that if there exists a string field A of ghost number -1 such that

$$I = Q_{\psi_0} A, \tag{6.5}$$

then the cohomology of Q_{ψ_0} is empty at any ghost number. The proof of this very important statement is embarrassingly simple

$$0 = Q_{\psi_0}\phi = A*Q_{\psi_0}\phi = -Q_{\psi_0}(A*\phi)+(Q_{\psi_0}A)*\phi = -Q_{\psi_0}(A*\phi)+I*\phi = -Q_{\psi_0}(A*\phi)+\phi \tag{6.6}$$

that is

$$Q_{\psi_0}\phi = 0 \quad \Leftrightarrow \quad \phi = Q_{\psi_0}(A * \phi) \tag{6.7}$$

This string field A has been recently found by Ellwood and Schnabl, [16], and is based on the first non trivial classical solution representing the tachyon vacuum given by Schnabl [17].

6.3 Vacuum String Field Theory

The remarkable properties of the tachyon vacuum suggest that Open String Field Theory should take its simplest form around it. As we have seen in section 2, when we expand OSFT around a classical solution the action is reproduced up to a shift in the kinetic operator, (2.27). So the only thing we need to write down OSFT at the tachyon vacuum is the new kinetic operator which, in turn, is known if the classical solution representing the tachyon vacuum is known. Alternatively one can use Sen's conjectures to guess the form of the kinetic operator. In [14] a conjecture was put forward under the name of Vacuum String Field Theory. In this model the BRST operator is taken to be pure ghost: this is a particular implementation of the universality of the tachyon vacuum. In particular the proposed kinetic operator takes the form of a c-midpoint insertion, [13]

$$\mathcal{Q} = \frac{1}{2i}\left(c(i) - c(-i)\right) \tag{6.8}$$

We recall that this operator has trivial cohomology due to the relation

$$\{\mathcal{Q}, b_0\} = 1 \quad \Rightarrow \quad \mathcal{Q}\psi = 0 \rightarrow \psi = \mathcal{Q}(b_0\psi) \tag{6.9}$$

Since both the star product and the kinetic operator are matter–ghost factorized, it is natural to search for solutions of the equation of motion which are matter/ghost factorized too. In particular, starting from the VSFT equation of motion

$$\mathcal{Q}\psi + \psi * \psi = 0 \tag{6.10}$$

and making the factorization ansatz

$$\psi = \psi_m \otimes \psi_{gh} \tag{6.11}$$

one ends up with the following equations, in the ghost and matter sector

$$\mathcal{Q}\psi_{gh} + \psi_{gh} *_{gh} \psi_{gh} = 0 \tag{6.12}$$

$$\psi_m *_m \psi_m = \psi_m \tag{6.13}$$

The equation in the matter sector defines idempotents (projectors) of the matter star algebra. One very interesting topic is the classification of projectors of the star algebra, as it corresponds to a classification of D–branes. From this analysis it emerges that a single D–brane can be described by a rank 1 projector and a set of N D–branes by a rank N projector (that is the sum of N orthogonal rank 1 projectors).

Rank 1 projectors can be easily defined starting from split left/right open string functionals

$$\psi[X(\sigma)] = \psi[l(\sigma), r(\sigma)] = \psi_L[l(\sigma)]\psi_R[r(\sigma)] \tag{6.14}$$

When we compute the $*$–square we get

$$(\psi_L[l(\sigma)]\psi_R[r(\sigma)]) * (\psi_L[l(\sigma)]\psi_R[r(\sigma)]) = K\ \psi_L[l(\sigma)]\psi_R[r(\sigma)] \tag{6.15}$$

where (formally)

$$K = \int \mathcal{D}y(\sigma)\, \psi_R(\sigma)\psi_L(\pi - \sigma), \qquad \sigma \in \left[\frac{\pi}{2}, \pi\right] \tag{6.16}$$

So, if ψ_L and ψ_R are chosen in such a way that $K \neq 0$ a rank 1 projector is simply given by $\frac{1}{K}\psi$. One very simple choice is to choose twist invariant states (L$\leftrightarrow$R), that is

$$\psi_L[l(\sigma)] = \psi_R[r(\pi - \sigma)] \tag{6.17}$$

with the normalization condition on ψ_R

$$\int \mathcal{D}y(\sigma)\, \psi_R[y(\sigma)]\, \psi_R[y(\sigma)] = 1 \tag{6.18}$$

It is clear that such a state projects into a 1 dimensional subspace of the half string Hilbert space, spanned by all possible functions $r(\sigma)$. It is also clear how to construct other projectors that are orthogonal wrt it, just choose another half string functional $\phi_R[r(\sigma)]$ such that

$$\int \mathcal{D}y(\sigma)\, \psi_R[y(\sigma)]\, \phi_R[y(\sigma)] = 0 \tag{6.19}$$

This means that for every function $f(\sigma)$ I can construct the "characteristic" half string functional ψ_f

$$\psi_f[y(\sigma)] = \delta[f(\sigma) - y(\sigma)] \tag{6.20}$$

Each characteristic functional defines a projector and characteristic functionals relative to different functions are of course orthogonal.

This quite abstract construction of star algebra projectors can be made very precise using the conformal or the operatorial definition of the star product, see the reviews in the references.

As a last remark we would like to point out that star algebra projector can be used not just as solutions of VSFT (which is in some sense too much singular and needs some not yet understood regularization) but they are also relevant for solving the original OSFT equation of motion, [18].

More on this subject can be found on the contribution to these proceedings by N. Bouatta.

6.4 Closed Strings in OSFT

So far we have been dealing with open strings and D-branes. We know that this is not the end of the story as unitarity forces us to include a closed string sector. One of the biggest results in the eighties in OSFT was that one loop non planar SFT diagrams contained on shell closed string poles. This lead to the believing that closed strings can be a derived concept, the fundamental degrees of freedom being the open string fields. Are closed strings elementary excitations? or they are maybe automatically generated when one is dealing with the full Quantum Open String Field Theory. The reader might recognize that the answer to this question can provide a microscopic derivation of the gauge/gravity correspondence. At the moment there are not satisfactory proof of this, nonetheless some progresses have been made in this direction.

OSFT is a gauge theory and, as in any gauge theory, a prominent role is played by gauge invariant operators.

It is simple to show that to any *on shell* closed string state we can associate a gauge invariant operator by inserting the closed string vertex operator at the *midpoint* of the identity string field. In particular let's consider the following second quantized operator/ functional

$$\mathcal{O}_V[\psi] = \langle I|V(\pi/2)|\psi\rangle \tag{6.21}$$

It is easy to show that this operator is gauge invariant if and only if

$$\{Q, V(\pi/2)\} = \{\bar{Q}, V(\pi/2)\} = 0 \tag{6.22}$$

that is $V(\sigma)$ is an on-shell closed string vertex operator. One can show that the Feynamnn rules derived by adding such gauge invariant operators to the OSFT action gives a complete covering of Riemann surfaces with open plus closed string punctures and at least one boundary. This physically means that with the use of just open string variables one can compute the scattering of closed strings off a brane using open string propagators and cubic open string vertices. As in AdS/CFT closed strings are viewed as non dynamical sources which are needed by gauge invariance.

The question is now: can we get *purely* closed string amplitudes from OSFT? The naive answer to this question is no, we cannot: any OSFT diagram will always contain at least one boundary, where there are the boundary conditions of the D–brane we are working on. But now we know that at the tachyon vacuum there are no D-branes, so what happens to the boundary? One can show, [13], that due to the suppression of open string modes at the tachyon vacuum, the integral over the Schwinger parameters (defining the lengths of propagators) get localized to zero length. This in turn implies that any boundary collapses to zero size as well, becoming a closed string puncture. This very heuristic picture can be made more precise: The scattering of n on-shell closed string states off the Tachyon Vacuum is identical to the scattering of n+1 closed strings with no D-branes, the extra closed string state being given by the open string boundary which shrinks to zero size. What is left to be seen is how much of closed string physics one can achieve in this way. In particular what is not clear at all is if one can derive a genuine closed string field theory starting from open strings, or if one is just stuck to closed string S–matrices. Needless to say, understanding this issue would be a breakthrough in Quantum Gravity.

Acknowledgments

I thank the organizer for the opportunity of lecturing in an exciting environment, where lot of nice discussions have come out. C.M. is supported in part by the Belgian Federal Science Policy

Office through the Interuniversity Attraction Pole P5/27, in part by the European Commission FP6 RTN programme MRTN-CT-2004-005104 and in part by the "FWO-Vlaanderen" through project G.0428.06.

References

[1] E.Witten, *Noncommutative Geometry and String Field Theory*, Nucl.Phys. **B268** (1986) 253.

[2] K.Ohmori, *A review of tachyon condensation in open string field theories*, [hep-th/0102085].

[3] C. B. Thorn, *"String Field Theory,"* Phys. Rept. **175** (1989) 1.

[4] W. Taylor and B. Zwiebach, *"D-branes, tachyons, and string field theory,"* arXiv:hep-th/0311017.

[5] A. Sen, *"Tachyon dynamics in open string theory,"* Int. J. Mod. Phys. A **20** (2005) 5513 [arXiv:hep-th/0410103].

[6] I.Ya.Aref'eva, D.M.Belov, A.A.Giryavets, A.S.Koshelev, P.B.Medvedev, *Noncommutative field theories and (super)string field theories*, [hep-th/0111208].

[7] L. Bonora, C. Maccaferri, D. Mamone and M. Salizzoni, *Topics in string field theory,* [hep-th/0304270].

[8] A.Leclair, M.E.Peskin, C.R.Preitschopf, *String Field Theory on the Conformal Plane. (I) Kinematical Principles*, Nucl.Phys. **B317** (1989) 411.

[9] A.Leclair, M.E.Peskin, C.R.Preitschopf, *String Field Theory on the Conformal Plane. (II) Generalized Gluing*, Nucl.Phys. **B317** (1989) 464.

[10] D.J.Gross and A.Jevicki, *Operator Formulation of Interacting String Field Theory*, Nucl.Phys. **B283** (1987) 1.

[11] D.J.Gross and A.Jevicki, *Operator Formulation of Interacting String Field Theory, 2*, Nucl.Phys. **B287** (1987) 225.

[12] B. Zwiebach, *"A Proof That Witten's Open String Theory Gives A Single Cover Of Moduli Space,"* Commun. Math. Phys. **142** (1991) 193.

[13] D.Gaiotto, L.Rastelli, A.Sen and B.Zwiebach, *Ghost Structure and Closed Strings in Vacuum String Field Theory*, [hep-th/0111129].

[14] L.Rastelli, A.Sen and B.Zwiebach, *String field theory around the tachyon vacuum*, Adv. Theor. Math. Phys. **5** (2002) 353 [hep-th/0012251].

[15] L. Rastelli, A. Sen and B. Zwiebach, *"Vacuum string field theory,"* arXiv:hep-th/0106010.

[16] I. Ellwood and M. Schnabl, *"Proof of vanishing cohomology at the tachyon vacuum,"* arXiv:hep-th/0606142.

[17] M. Schnabl, *"Analytic solution for tachyon condensation in open string field theory,"* arXiv:hep-th/0511286.

[18] L. Rastelli and B. Zwiebach, *"Solving open string field theory with special projectors,"* arXiv:hep-th/0606131.

String field theory around the tachyon vacuum

Nazim Bouatta [1]

Physique Théorique et Mathématique,
Université Libre de Bruxelles and
International Solvay Institutes
Campus Plaine C.P. 231, B-1050 Bruxelles, Belgium

Abstract. This is a brief review of vacuum string field theory (VSFT), an approach to open string field theory around the stable vacuum of the tachyon. We discuss the sliver state explaining its role as projector in the space of half-string basis. We review the construction of D-brane solutions in vacuum string field theory. We show that in the sliver basis the star product correspond to a matrix product.

[1] Chercheur FRIA. email: nbouatta@ulb.ac.be

1 Introduction and summary

In a series of papers [1, 2, 3, 4, 5] Rastelli, Sen and Zwiebach (RSZ) proposed an analytic approach to tachyon condensation by introducing a new formulation to open string field theory − vacuum string field theory (VSFT). This theory uses the open string tachyon vacuum to formulate the dynamics. Among all possible open string backgrounds the tachyon vacuum is particularly natural given its physically expected uniqueness as the endpoint of all processes of tachyon condensation. As opposed to the conventional OSFT, where the kinetic operator is the BRST operator Q_B, in VSFT the kinetic operator $\mathcal{Q}$ is non-dynamical and is built solely out of worldsheet ghost fields.

VSFT is structurally much simpler than conventional OSFT. Indeed, it is possible to construct analytically classical solutions representing arbitrary D-p branes, with correct ratios of tensions, thereby providing a non-trivial check on the correctness of the RSZ proposal. The key ansatz that makes this analysis possible is that the string field solution representing a D-brane factorizes into a ghost part Ψ_g and a matter part Ψ_m, with Ψ_g the same for all D-branes, and Ψ_m different for different D-branes [2].

The matter part of the string field satisfies a very simple equation: a projector equation (Ψ_m squares to itself under $*$-multiplication). Two points of view have been useful for solving this equation. In the geometric method [4], the $*$-product is defined by the gluing of Riemann surfaces. In the algebraic method [6], one relies on the operator representation of the $*$-product using flat space oscillator modes. In both approaches a key role is played by the sliver state (a rank one projector), a solution of the matter string field equations which can alternatively be viewed geometrically as the surface state associated with a specific one-punctured disk, or algebraically as a squeezed state, *i.e.* the exponential of an oscillator bilinear acting on the vacuum. Solutions representing various (multiple) D-branes in vacuum string field theory are obtained as (superpositions of) various deformations of the sliver state (higher rank projectors).

The algebraic approach, while more tied to the choice of a flat-space background, provides very explicit expressions for the string fields. More crucially, in the algebraic approach we can take direct advantage of the insight that D-brane solutions are projectors onto the half-string state space. The intuitive left/right splitting picture provided by the functional representation can be turned into a completely algebraic procedure to obtain multiple D-brane solutions.

2 Vacuum string field theory

In principle the analysis of the tachyon vacuum is straightforward. One must: (i) find the classical solution Φ_0 representing the vacuum with no D-brane, (ii) expand the string field action setting $\Phi = \Phi_0 + \Psi$, where Ψ is the fluctuation field, and (iii) analyze the spectrum of Ψ using the resulting kinetic term. Nevertheless, in practice this has not been simple to carry out, first and foremost because until recentely there was no known closed form expression for Φ_0 (in a very important work schnabl was able to found an analytic expression for Φ_0 [7, 8], see also [9]).

In VSFT we analyze the problem from a different angle. Instead of trying to construct the classical solution Φ_0, expand the SFT action around Φ_0, and attempt field redefinitions to bring the kinetic term to a simple form, we shall make an inspired guess about the form of the SFT action expanded around the tachyon vacuum. Then we will try to check that this action satisfies the various consistency requirements. Since the string field theory action is cubic in the string field, shifting the string field by a classical solution does not change the cubic interaction term. Thus we only need to guess the quadratic term of the shifted action. The requirement that the new action is obtained from the original one by a shift of the string field by a classical solution of the equations of motion puts constraints on the possible choices of the quadratic term, − these basically correspond to the requirement that the shifted action also has an infinite parameter gauge invariance like the original SFT action. Besides these constraints the action must satisfy the following additional requirements:

(i) The kinetic operator must have vanishing cohomology. This would imply absence of physical open string states around the tachyon vacuum.

(ii) The kinetic operator must be universal, namely, it must be possible to express the kinetic operator without reference to the particular configuration that we start with.

(iii) The action must have classical solutions representing the original D-brane configuration, as well as lump solutions representing D-branes of all lower dimensions.

We begin with the cubic string field theory action :

$$S(\Phi) = -\frac{1}{g_o^2}\left[\frac{1}{2}\langle\,\Phi\,,\,Q_B\,\Phi\rangle + \frac{1}{3}\langle\,\Phi\,,\,\Phi * \Phi\rangle\right], \tag{2.1}$$

where g_o is the open string coupling constant and Φ is the open string field, conventionally taken to be Grassmann odd and of ghost number one for the classical action. In addition, Q_B is the BRST operator, $\langle\cdot,\cdot\rangle$ is a bilinear inner product based on BPZ conjugation and $*$ denotes star-multiplication of string fields. The consistency of this classical action is guaranteed by the following identities involving the BRST operator

$$\begin{aligned} &Q_B^2 = 0, \\ &Q_B(A * B) = (Q_B A) * B + (-1)^A A * (Q_B B)\,, \\ &\langle\, Q_B A, B\,\rangle = -(-)^A \langle A, Q_B B\rangle\,, \end{aligned} \tag{2.2}$$

and identities involving the inner product and the star operation

$$\begin{aligned} &\langle A, B\rangle = (-)^{AB}\langle B, A\rangle \\ &\langle\, A\,, B * C\,\rangle = \langle A * B\,,\,C\,\rangle \\ &A * (B * C) = (A * B) * C\,. \end{aligned} \tag{2.3}$$

In the sign factors, the exponents $A, B, \cdots$ denote the Grassmanality of the state, and should be read as $(-)^A \equiv (-)^{\epsilon(A)}$ where $\epsilon(A) = 0\,(\mathrm{mod}\ 2)$ for A Grassmann even, and $\epsilon(A) = 1\,(\mathrm{mod}\ 2)$ for A Grassmann odd. We also have:

$$\begin{aligned} &\epsilon(A * B) = \epsilon(A) + \epsilon(B) \\ &\mathrm{gh}(A * B) = \mathrm{gh}(A) + \mathrm{gh}(B)\,, \end{aligned} \tag{2.4}$$

where gh denotes ghost number, and we take the ghost number of the SL(2,R) vacuum to be zero. Equations (2.2) and (2.3) guarantee that the above action is invariant under the gauge transformations:

$$\delta\Phi = Q_B\Lambda + \Phi * \Lambda - \Lambda * \Phi\,, \tag{2.5}$$

for any Grassmann-even ghost-number zero state Λ.

Let Φ_0 be the string field configuration describing the tachyon vacuum, a solution of the classical field equations following from the action in (2.1):

$$Q_B\Phi_0 + \Phi_0 * \Phi_0 = 0\,. \tag{2.6}$$

If $\widetilde{\Phi} = \Phi - \Phi_0$ denotes the shifted open string field, then the cubic string field theory action expanded around the tachyon vacuum has the form:

$$S(\Phi_0 + \widetilde{\Phi}) = S(\Phi_0)\; - \frac{1}{g_o^2}\left[\frac{1}{2}\langle\,\widetilde{\Phi}\,,\,Q\,\widetilde{\Phi}\rangle + \frac{1}{3}\langle\,\widetilde{\Phi}\,,\,\widetilde{\Phi} * \widetilde{\Phi}\rangle\right]. \tag{2.7}$$

Here $S(\Phi_0)$ is a constant, which according to the energetics part of the tachyon conjectures equals the mass M of the D-brane when the D-brane extends over a space-time of finite volume. Indeed, the potential energy $V(\Phi_0) = -S(\Phi_0)$ associated to this string field configuration should equal minus the mass of the brane. The kinetic operator Q is given in terms of Q_B and Φ_0 as:

$$Q\widetilde{\Phi} = Q_B\widetilde{\Phi} + \Phi_0 * \widetilde{\Phi} + \widetilde{\Phi} * \Phi_0\,. \tag{2.8}$$

More generally, on arbitrary string fields one would define

$$QA = Q_B A + \Phi_0 * A - (-1)^A A * \Phi_0\,. \tag{2.9}$$

The consistency of the action (2.7) is guaranteed from the consistency of the one in (2.1). Since neither the inner product nor the star multiplication have changed, the identities in (2.3) still hold. One can readily check that the identities in (2.2) hold when Q_B is replaced by Q. Just as (2.1) is invariant under the gauge trasformations (2.5), the action in (2.7) is invariant under $\delta\widetilde{\Phi} = Q\Lambda + \widetilde{\Phi} * \Lambda - \Lambda * \widetilde{\Phi}$ for any Grassmann-even ghost-number zero state Λ.

For the analysis around this final vacuum it suffices to study the action

$$S_0(\widetilde{\Phi}) \equiv -\frac{1}{g_o^2}\left[\frac{1}{2}\langle\,\widetilde{\Phi}\,,\,Q\,\widetilde{\Phi}\rangle + \frac{1}{3}\langle\,\widetilde{\Phi}\,,\,\widetilde{\Phi} * \widetilde{\Phi}\rangle\right]. \tag{2.10}$$

Even if we knew Φ_0 explicitly and constructed $S_0(\widetilde{\Phi})$ using eq.(2.10), this may not be the most convenient form of the action. Typically a nontrivial field redefinition is necessary to bring the shifted SFT action to the a simpler form representing the new background.

In proposing a simple form of the tachyon action, one have in mind field redefinitions of the action in (2.10) that leave the cubic term invariant but simplify the operator Q in (2.9) by transforming it into a simpler operator $\mathcal{Q}$. To this end one consider field redefinitions of the type

$$\widetilde{\Phi} = e^{K}\,\Psi\,, \tag{2.11}$$

where K is a ghost number zero Grassmann even operator. In addition, we require

$$\begin{aligned} &K(A * B) = (KA) * B + A * (KB)\,,\\ &\langle\, KA, B\,\rangle = -\langle A, KB\rangle\,. \end{aligned} \tag{2.12}$$

These properties guarantee that the form of the cubic term is unchanged and that after the field redefinition the action takes the form

$$\mathcal{S}(\Psi) \equiv -\frac{1}{g_0^2}\left[\frac{1}{2}\langle\,\Psi\,,\,\mathcal{Q}\,\Psi\rangle + \frac{1}{3}\langle\,\Psi\,,\,\Psi * \Psi\rangle\right], \tag{2.13}$$

where

$$\mathcal{Q} = e^{-K} Q e^{K}\,. \tag{2.14}$$

Again, gauge invariance only requires:

$$\begin{aligned} &\mathcal{Q}^2 = 0,\\ &\mathcal{Q}(A * B) = (\mathcal{Q}A) * B + (-1)^A A * (\mathcal{Q}B)\,,\\ &\langle\, \mathcal{Q}A, B\,\rangle = -(-)^A\langle A, \mathcal{Q}B\rangle\,. \end{aligned} \tag{2.15}$$

These identities hold by virtue of (2.12) and (2.14). We will proceed here postulating a $\mathcal{Q}$ that satisfies these identities as well as other conditions ($\mathcal{Q}$ must be universal with a vanishing cohomology and with a ghost number one).

One can satisfy the previous requirements by by letting $\mathcal{Q}$ be constructed purely from ghost operators. In particular the ghost number one operators

$$\mathcal{Q} \equiv \frac{1}{2i}(c(i) - c(-i)), \tag{2.16}$$

The operators $\mathcal{Q}$ have a vanishing cohomology since for e the operator b_0 satisfies $\{\mathcal{Q}, b_0\} = 1$. It then follows that whenever $\mathcal{Q}\psi = 0$, we have $\psi = \{\mathcal{Q}, b_0\}\psi = \mathcal{Q}(b_0\psi)$, showing that ψ is $\mathcal{Q}$ trivial. Finally, since they are built from ghost oscillators, all $\mathcal{Q}$'s are manifestly universal.

3 Construction of the D-brane solution

3.1 Matter-ghost factorization

In VSFT one can give a direct construction of the classical solutions representing various D-branes and verify that the ratios of their tensions agree with the known answer. RSZ used an ansatz where the solution Ψ representing a D-brane has a factorized form $\Psi_m \otimes \Psi_g$, with Ψ_m and Ψ_g being string fields built solely out of matter and ghost operators respectively. Such factorized form is clearly compatible with the structure of the relevant string field equation since the kinetic operator $\mathcal{Q}$ does not mix matter and ghost sectors, and moreover, as is familiar, the star product also factors into the matter and ghost sectors. More explicitly, given string fields

$A = A_m \otimes A_g$ and $B = B_m \otimes B_g$, we have $A * B = (A_m *^m B_m) \otimes (A_g *^g B_g)$, where $*^m$ and $*^g$ denote multiplication rules in the matter and ghost sectors respectively. While the matter factor Ψ_m is clearly different for the various D-branes, one assume that the ghost factor Ψ_g is common to all the D-branes.

The equations of motion of string field theory:

$$\mathcal{Q}\Psi = -\Psi * \Psi \,, \tag{3.1}$$

must have a space-time independent solution describing the D25-brane, and also lump solutions of all codimensions describing lower dimensional D-branes. We shall look for solutions of the form:

$$\Psi = \Psi_m \otimes \Psi_g \,, \tag{3.2}$$

where Ψ_g denotes a state obtained by acting with the ghost oscillators on the SL(2,R) invariant vacuum of the ghost CFT, and Ψ_m is a state obtained by acting with matter oscillators on the SL(2,R) invariant vacuum of the matter CFT. Let us denote by $*^g$ and $*^m$ the star product in the ghost and matter sector respectively. Eq.(3.1) then factorizes as

$$\mathcal{Q}\Psi_g = -\Psi_g *^g \Psi_g \,, \tag{3.3}$$

and

$$\Psi_m = \Psi_m *^m \Psi_m \,. \tag{3.4}$$

Such a factorization is possible since $\mathcal{Q}$ is made purely of ghost operators.

In looking for the solutions describing D-branes of various dimensions we shall assume that Ψ_g remains the same for all solutions, whereas Ψ_m is different for different D-branes. Given two static solutions of this kind, described by Ψ_m and Ψ'_m, the ratio of the energy associated with these two solutions is obtained by taking the ratio of the actions associated with the two solutions. For a string field configuration satisfying the equation of motion (2.7), the action (2.1) is given by

$$\mathcal{S}|_\Psi = -\frac{1}{6g_0^2}\,\langle\,\Psi\,,\,\mathcal{Q}\,\Psi\,\rangle\,. \tag{3.5}$$

Thus with the ansatz (3.2) the action takes the form:

$$\mathcal{S}|_\Psi = -\frac{1}{6g_0^2}\,\langle\,\Psi_g\,|\,\mathcal{Q}\,\Psi_g\,\rangle_g\,\langle\Psi_m|\Psi_m\rangle_m \equiv K\langle\Psi_m|\Psi_m\rangle_m\,, \tag{3.6}$$

where $\langle\,|\,\rangle_g$ and $\langle\,|\,\rangle_m$ denote BPZ inner products in ghost and matter sectors respectively. $K = -(6g_0^2)^{-1}\langle\Psi_g|\mathcal{Q}\Psi_g\rangle_g$ is a constant factor calculated from the ghost sector which remains the same for different solutions. Thus we see that the ratio of the action associated with the two solutions is

$$\frac{\mathcal{S}|_{\Psi'}}{\mathcal{S}|_\Psi} = \frac{\langle\Psi'_m|\Psi'_m\rangle_m}{\langle\Psi_m|\Psi_m\rangle_m}\,. \tag{3.7}$$

The ghost part drops out of this calculation.

3.2 The oscillator representation of the sliver

In this section we discuss the oscillator representation of the sliver. Here we wish to review the construction of the matter part of the sliver in the oscillator representation and consider the basic algebraic properties that guarantee that the multiplication of two slivers gives a sliver. In fact we follow the discussion of Kostelecky and Potting [6] who gave the first algebraic construction of a state that would star multiply to itself in the matter sector.

In order to star multiply two states $|A\rangle$ and $|B\rangle$ we must calculate

$$(|A\rangle * |B\rangle)_3 = {}_1\langle A|_2\langle B|V_3\rangle_{123}\,, \tag{3.8}$$

where $|\,\rangle_r$ denotes a state in the r-th string Hilbert space, and $|\,\rangle_{123}$ denotes a state in the product of the Hilbert space of three strings. The key ingredient here is the three-string vertex $|V_3\rangle_{123}$. While the vertex has nontrivial momentum dependence, if the states A and B are at zero momentum, the star product gives a zero momentum state that can be calculated using

$$|V_3\rangle_{123} = \exp\Bigl(-\frac{1}{2}\sum_{r,s} a^{(r)\dagger}\cdot V^{rs}\cdot a^{(s)\dagger}\Bigr)|0\rangle_{123}\,, \tag{3.9}$$

and the rule $\langle 0|0\rangle = 1$. Here the V^{rs}, with $r, s = 1, 2, 3$, are infinite matrices V^{rs}_{mn} ($m, n = 1, \cdots \infty$) satisfying the cyclicity condition $V^{rs} = V^{r+1,s+1}$ and the symmetry condition $(V^{rs})^T = V^{sr}$. These properties imply that out of the nine matrices, three: V^{11}, V^{12} and V^{21}, can be used to obtain all others. $a_m^{(r)\mu\dagger}$ ($0 \le \mu \le 25$) denote oscillators in the r-th string Hilbert space. For simplicity, the Lorentz and the oscillator indices, and the Minkowski matrix $\eta_{\mu\nu}$ used to contract the Lorentz indices, have all been suppressed in eq.(3.9). We shall follow this convention throughout the paper.

One now introduces

$$M^{rs} \equiv CV^{rs}\,, \quad \text{with} \quad C_{mn} = (-1)^m \delta_{mn}, \quad m, n \ge 1\,. \tag{3.10}$$

These matrices can be shown to satisfy the following properties:

$$\begin{aligned} &CV^{rs} = V^{sr}C\,, \quad (V^{rs})^T = V^{sr}\,, \\ &(M^{rs})^T = M^{rs}\,, \quad CM^{rs}C = M^{sr}\,, \quad [M^{rs}, M^{r's'}] = 0\,. \end{aligned} \tag{3.11}$$

In particular note that all the M matrices commute with each other. Defining $X \equiv M^{11}$, the three relevant matrices are X, M^{12} and M^{21}. Explicit formulae exist that allow their explicit computation [10, 11, 12].

They can be shown in general grounds to satisfy the following useful relations:

$$\begin{aligned} &X + M^{12} + M^{21} = 1\,, \\ &M^{12}M^{21} = X^2 - X\,, \\ &(M^{12})^2 + (M^{21})^2 = 1 - X^2\,, \\ &(M^{12})^3 + (M^{21})^3 = 2X^3 - 3X^2 + 1 = (1 - X)^2(1 + 2X)\,, \end{aligned} \tag{3.12}$$

which gives

$$(M^{12} - M^{21})^2 = (1 - X)(1 + 3X)\,. \tag{3.13}$$

The state in the matter Hilbert space that multiplies to itself turns out to take the form

$$|\Psi\rangle = \mathcal{N}^{26} \exp\Big(-\frac{1}{2}\, a^\dagger \cdot S \cdot a^\dagger\Big)|0\rangle\,, \quad \mathcal{N} = \{\det(1 - X)\det(1 + T)\}^{1/2}\,, \quad S = CT \tag{3.14}$$

where the matrix T satisfies $CTC = T$ and the equation

$$XT^2 - (1 + X)T + X = 0\,, \tag{3.15}$$

which gives

$$T = (2X)^{-1}(1 + X - \sqrt{(1 + 3X)(1 - X)}). \tag{3.16}$$

.

We conclude this subsection by constructing a pair of projectors that will be very useful in explicit computations of star products. We define the matrices

$$\begin{aligned} L &= \frac{1}{(1 + T)(1 - X)}\Big[M^{12}(1 - TX) + T(M^{21})^2\Big]\,, \\ R &= \frac{1}{(1 + T)(1 - X)}\Big[M^{21}(1 - TX) + T(M^{12})^2\Big]\,. \end{aligned} \tag{3.17}$$

One readily verifies that they satisfy the following properties:

$$L^T = L\,, \quad R^T = R\,, \quad CLC = R\,, \tag{3.18}$$

and more importantly

$$\begin{aligned} L + R &= 1\,, \\ L - R &= \frac{M^{12} - M^{21}}{\sqrt{(1 - X)(1 + 3X)}}\,. \end{aligned} \tag{3.19}$$

From eq.(3.13) we see that the square of the second right hand side is the unit matrix. Thus $(L-R)^2 = 1$, and this together with the squared version of the first equation gives

$$LR = 0\,. \tag{3.20}$$

This equation is also easily verified directly. Multiplying the first equation in (3.19) by L and alternatively by R we get

$$LL = L\,, \quad RR = R\,. \tag{3.21}$$

This shows that L and R are projection operators into orthogonal subspaces, and the C exchanges these two subspaces. We will see later on that L and R project into the oscillators of the right and the left half of the string respectively.

4 Coherent states, higher rank projectors and multiple D-branes

4.1 Coherent states

As seen earlier, the matter part of the sliver state is given by

$$|\Xi\rangle = \mathcal{N}^{26}\exp\Big(-\frac{1}{2}a^\dagger\cdot S\cdot a^\dagger\Big)|0\rangle\,. \tag{4.1}$$

Coherent states are defined by letting exponentials of the creation operator act on the vacuum. Treating the sliver as the vacuum we introduce coherent like states of the form

$$|\Xi_\beta\rangle = \exp\Big(\sum_{n=1}^{\infty}(-)^{n+1}\beta_{\mu n}a_n^{\mu\dagger}\Big)|\Xi\rangle = \exp(-a^\dagger\cdot C\beta)|\Xi\rangle\,. \tag{4.2}$$

As built, the states satisfy a simple BPZ conjugation property:

$$\langle\Xi_\beta| = \langle\Xi|\exp\Big(\sum_{n=1}^{\infty}\beta_{n\mu}a_n^{\mu}\Big) = \langle\Xi|\exp(\beta\cdot a)\,. \tag{4.3}$$

One can show that two coherent states $*$ multiply in the following way:

$$|\Xi_{\beta_1}\rangle * |\Xi_{\beta_2}\rangle = \exp\Big(-\mathcal{C}(\beta_1,\beta_2)\Big)|\Xi_{\rho_1\beta_1+\rho_2\beta_2}\rangle \tag{4.4}$$

where

$$\rho_1 \equiv L\,, \quad \rho_2 \equiv R. \tag{4.5}$$

This is a useful relation that allows one to compute $*$-products of slivers acted by oscillators by simple differentiation. In particular, using the previous equation we get

$$\begin{aligned}(a_{m_1}^{\mu_1\dagger}\cdots a_{m_k}^{\mu_k\dagger}|\Xi\rangle) * (a_{n_1}^{\nu_1\dagger}\cdots a_{n_l}^{\nu_l\dagger}|\Xi\rangle) \quad &= \quad (-1)^{\sum_{i=1}^{k}(m_i+1)+\sum_{j=1}^{l}(n_j+1)} \\ &\Big(\frac{\partial}{\partial\beta_{1m_1\mu_1}}\cdots\frac{\partial}{\partial\beta_{1m_k\mu_k}}\frac{\partial}{\partial\beta_{2n_1\nu_1}}\cdots\frac{\partial}{\partial\beta_{2n_l\nu_l}}\big(|\Xi_{\beta_1}\rangle * |\Xi_{\beta_2}\rangle\big)\Big)_{\beta_1=\beta_2=0}\,.\end{aligned} \tag{4.6}$$

Since $\rho_1+\rho_2 = 1$, for $\beta_1 = \beta_2$ the $*$ product of two coherent states reduces to

$$|\Xi_\beta\rangle * |\Xi_\beta\rangle = \exp\Big(-\mathcal{C}(\beta,\beta)\Big)|\Xi_\beta\rangle\,. \tag{4.7}$$

It follows from (4.7) that by adjusting the normalization of the Ξ_β state

$$P_\beta \equiv \exp\Big(\mathcal{C}(\beta,\beta)\Big)|\Xi_\beta\rangle\,, \tag{4.8}$$

we obtain projectors

$$P_\beta * P_\beta = P_\beta\,. \tag{4.9}$$

By generalizing the following procedure one can describe a multiple D-brane configuration.

4.2 The Comma Vertex

We introduce new oscillators s which annihilate the sliver by performing a Bogoliubov transformation from the original oscillators a

$$s = \frac{1}{\sqrt{1-T^2}}(a + CTa^\dagger), \quad s^\dagger = \frac{1}{\sqrt{1-T^2}}(a^\dagger + CTa), \tag{4.10}$$

$$s|\Xi\rangle = 0. \tag{4.11}$$

The 3-string vertex in this new basis (up to normalization) turns out to be [6, 13]

$$|V_3\rangle = \exp\left(-\sum_{r=1}^{3} s_r^\dagger CLs_{r+1}^\dagger\right)|\Xi\rangle_{123} = \exp\left(-\frac{1}{2}\sum_{r,s=1}^{3} s_r^\dagger \widehat{V}_3^{rs} s_s^\dagger\right)|\Xi\rangle_{123}, \tag{4.12}$$

where $\widehat{V}_3^{rs}$ is given by

$$C\widehat{V}_3 = (1 - CV_3T)^{-1}(CV_3 - T) = \begin{pmatrix} 0 & L & R \\ R & 0 & L \\ L & R & 0 \end{pmatrix}. \tag{4.13}$$

The identity string field in this new basis are found to be

$$|I\rangle = e^{-\frac{1}{2}s^\dagger Cs^\dagger}|\Xi\rangle, \quad . \tag{4.14}$$

Using the projections L and R, we can define the splitting of s into two parts

$$s = \sum_\alpha s_{L\alpha}e_\alpha + s_{R\alpha}f_\alpha, \tag{4.15}$$

where e_α and f_α are the orthonormal basis of the eigenspace of L and R:

$$e_\alpha \cdot e_\beta = f_\alpha \cdot f_\beta = \delta_{\alpha\beta}, \quad e_\alpha \cdot f_\beta = 0, \quad f_\alpha = -Ce_\alpha, \tag{4.16}$$

$$Le_\alpha = e_\alpha, \quad Re_\alpha = 0, \quad Lf_\alpha = 0, \quad Rf_\alpha = f_\alpha, \tag{4.17}$$

$$L = \sum_\alpha e_\alpha e_\alpha^T, \quad R = \sum_\alpha f_\alpha f_\alpha^T. \tag{4.18}$$

Since s_L and s_R commute,

$$[s_{L\alpha}, s_{L\beta}^\dagger] = [s_{R\alpha}, s_{R\beta}^\dagger] = \delta_{\alpha\beta}, \quad [s_{L\alpha}, s_{R\beta}^\dagger] = 0, \tag{4.19}$$

the Hilbert space factorizes into the Fock spaces of s_L and s_R

$$\mathcal{A}_{str} = \mathcal{H}_L \otimes \mathcal{H}_R. \tag{4.20}$$

In terms of $s_{L,R}$, the 3-string vertex takes a simple form ("comma" form)

$$|V_3\rangle = \exp\left(s_{1R}^\dagger s_{2L}^\dagger + s_{2R}^\dagger s_{3L}^\dagger + s_{3R}^\dagger s_{1L}^\dagger\right)|\Xi\rangle_{123}, \tag{4.21}$$

where

$$s_L^\dagger s_R^\dagger = \sum_\alpha s_{L\alpha}^\dagger s_{R\alpha}^\dagger. \tag{4.22}$$

The identity string field become

$$|I\rangle = e^{s_R^\dagger s_L^\dagger}|\Xi\rangle, \quad . \tag{4.23}$$

4.3 Matrix representation of the $*$ product

First, we construct the string field oscillators $A_\alpha, A_\alpha^\dagger$ [14] which satisfy the canonical commutation relations

$$[A_\alpha, A_\beta^\dagger]_\star = \delta_{\alpha\beta}|I\rangle. \tag{4.24}$$

To find the explicit expression of the string field oscillators, it is convenient to introduce the coherent states

$$\begin{aligned} |z_L, z_R\rangle &= \exp(z_L s_L^\dagger + z_R s_R^\dagger)|\Xi\rangle, && (4.25)\\ I(z_L, z_R) &= \exp(z_L s_L^\dagger + z_R s_R^\dagger)|I\rangle. && (4.26)\end{aligned}$$

The multiplication rules of these coherent states are found to be

$$\begin{aligned} |z_L, z_R\rangle \star |w_L, w_R\rangle &= e^{z_R w_L}|z_L, w_R\rangle, && (4.27)\\ I(z_L, z_R) \star |w_L, w_R\rangle &= e^{z_R w_L}|z_L + w_L, w_R\rangle, && (4.28)\\ |z_L, z_R\rangle \star I(w_L, w_R) &= e^{z_R w_L}|z_L, z_R + w_R\rangle, && (4.29)\\ I(z_L, z_R) \star I(w_L, w_R) &= e^{z_R w_L} I(z_L + w_L, z_R + w_R). && (4.30)\end{aligned}$$

Using (4.30), we can check that the string fields

$$A_\alpha = s_{R\alpha}^\dagger|I\rangle, \quad A_\alpha^\dagger = s_{L\alpha}^\dagger|I\rangle, \tag{4.31}$$

satisfy the relation (4.24). Since the sliver state is annihilated by the string field oscillators

$$A_\alpha \star |\Xi\rangle = |\Xi\rangle \star A_\alpha^\dagger = 0, \tag{4.32}$$

the sliver can be identified with a projection onto the Fock vacuum $|0\rangle\rangle$ of the string field oscillators:

$$|\Xi\rangle = |0\rangle\rangle\langle\langle 0|. \tag{4.33}$$

This identification is justified by the fact that the sliver is naturally factorized into the vacuum of s_L and s_R:

$$|\Xi\rangle = |\Xi_L\rangle \otimes |\Xi_R\rangle \tag{4.34}$$

where

$$s_{L\alpha}|\Xi_L\rangle = 0, \quad s_{R\alpha}|\Xi_R\rangle = 0. \tag{4.35}$$

In this subsection we construct a mapping between open string fields and matrices. Since the identity string field has the form $|I\rangle = e^{s_L^\dagger s_R^\dagger}|\Xi\rangle$, the string field oscillators (4.31) can be rewritten as

$$\begin{aligned} A_\alpha &= s_{L\alpha}|I\rangle = s_{R\alpha}^\dagger|I\rangle, && (4.36)\\ A_\alpha^\dagger &= s_{L\alpha}^\dagger|I\rangle = s_{R\alpha}|I\rangle. && (4.37)\end{aligned}$$

Using (4.30) recursively, we can show the following relations:

$$\begin{aligned} A_{\alpha_1} \star \cdots \star A_{\alpha_k} &= s_{L\alpha_1}\cdots s_{L\alpha_k}|I\rangle = s_{R\alpha_1}^\dagger \cdots s_{R\alpha_k}^\dagger|I\rangle, && (4.38)\\ A_{\alpha_1}^\dagger \star \cdots \star A_{\alpha_k}^\dagger &= s_{L\alpha_1}^\dagger \cdots s_{L\alpha_k}^\dagger|I\rangle = s_{R\alpha_1}\cdots s_{R\alpha_k}|I\rangle, && (4.39)\\ [A_\alpha, A_\beta^\dagger]_\star &= [s_{L\alpha}, s_{L\beta}^\dagger]|I\rangle = [s_{R\alpha}, s_{R\beta}^\dagger]|I\rangle. && (4.40)\end{aligned}$$

The star exponential of string field oscillators turns out to be

$$e_\star^{z_L A^\dagger} = e^{z_L s_L^\dagger}|I\rangle, \quad e_\star^{z_R A} = e^{z_R s_R^\dagger}|I\rangle. \tag{4.41}$$

From (4.28), (4.29) and (4.41), we can see that the coherent states of $s_{L,R}$ and A_α are related as

$$|z_L, z_R\rangle = e^{z_L s_L^\dagger}|I\rangle \star |\Xi\rangle \star e^{z_R s_R^\dagger}|I\rangle = e^{z_L A^\dagger}|0\rangle\rangle\langle\langle 0|e^{z_R A} = |z_L\rangle\rangle\langle\langle z_R|. \tag{4.42}$$

Note that this relation is consistent with the trace

$$\mathrm{Tr}\Big(|z_L, z_R\rangle\Big) = \langle I|z_L, z_R\rangle = e^{z_L z_R}. \tag{4.43}$$

By expanding (4.42) in terms of $z_{L,R}$, we find that the occupation number state of $s_{L,R}$ corresponds to the matrix element between the number states of string field oscillators:

$$|n,m\rangle = \prod_{\alpha,\beta} \frac{s_{L\alpha}^{\dagger n_\alpha} s_{R\beta}^{\dagger m_\beta}}{\sqrt{n_\alpha! m_\beta!}} |\Xi\rangle = \prod_\alpha \frac{A_\alpha^{\dagger n_\alpha}}{\sqrt{n_\alpha!}} |0\rangle\rangle\langle\langle 0| \prod_\beta \frac{A_\alpha^{m_\beta}}{\sqrt{m_\beta!}} = |n\rangle\rangle\langle\langle m|. \tag{4.44}$$

We can see that the number states behave exactly as matrix elements:

$$|n\rangle\rangle\langle\langle m| \star |k\rangle\rangle\langle\langle l| = \delta_{m,k} |n\rangle\rangle\langle\langle l|, \tag{4.45}$$

$$\mathrm{Tr}\Big(|n\rangle\rangle\langle\langle m|\Big) = \delta_{n,m}. \tag{4.46}$$

Now we can construct a mapping between string fields and matrices as follows. Since every state can be expanded in terms of the number basis of $s_{L,R}$, every string field can be written as a matrix:

$$|\Psi\rangle = \sum_{n,m} \Psi_{n,m} |n,m\rangle = \sum_{n,m} \Psi_{n,m} |n\rangle\rangle\langle\langle m|. \tag{4.47}$$

Especially, the identity string field corresponds to the identity matrix

$$|I\rangle = e^{s_L^\dagger s_R^\dagger} |\Xi\rangle = \sum_n |n,n\rangle = \sum_n |n\rangle\rangle\langle\langle n|. \tag{4.48}$$

Note that the 3-string vertex can also be written in terms of the number basis

$$|V_3\rangle_{123} = \sum_{l,m,n} |l\rangle\rangle\langle\langle m|_1 \otimes |m\rangle\rangle\langle\langle n|_2 \otimes |n\rangle\rangle\langle\langle l|_3. \tag{4.49}$$

We can see that this vertex represents the matrix multiplication.

Acknowledgments

N.B. would like to thank G. Barnich and C. Maccaferri for discussions.
N.B. is supported in part by a "Pole d'Attraction Interuniversitaire" (Belgium), by IISN-Belgium, convention 4.4505.86, by Proyectos FONDECYT 1970151 and 7960001 (Chile) and by the European Commission program MRTN-CT-2004-005104, in which this author is associated to V.U. Brussels.

References

[1] L. Rastelli, A. Sen and B. Zwiebach, "String field theory around the tachyon vacuum," Adv. Theor. Math. Phys. **5** (2002) 353 [arXiv:hep-th/0012251].

[2] L. Rastelli, A. Sen and B. Zwiebach, "Classical solutions in string field theory around the tachyon vacuum," Adv. Theor. Math. Phys. **5** (2002) 393 [arXiv:hep-th/0102112].

[3] L. Rastelli, A. Sen and B. Zwiebach, "Half strings, projectors, and multiple D-branes in vacuum string field theory," JHEP **0111** (2001) 035 [arXiv:hep-th/0105058].

[4] L. Rastelli, A. Sen and B. Zwiebach, "Boundary CFT construction of D-branes in vacuum string field theory," JHEP **0111** (2001) 045 [arXiv:hep-th/0105168].

[5] L. Rastelli, A. Sen and B. Zwiebach, "Vacuum string field theory," arXiv:hep-th/0106010.

[6] V. A. Kostelecky and R. Potting, "Analytical construction of a nonperturbative vacuum for the open bosonic string," Phys. Rev. D **63** (2001) 046007 [arXiv:hep-th/0008252].

[7] M. Schnabl, "Analytic solution for tachyon condensation in open string field theory," arXiv:hep-th/0511286.

[8] I. Ellwood and M. Schnabl, "Proof of vanishing cohomology at the tachyon vacuum," arXiv:hep-th/0606142.

[9] L. Rastelli and B. Zwiebach, "Solving open string field theory with special projectors," arXiv:hep-th/0606131.

[10] D. J. Gross and A. Jevicki, "Operator Formulation Of Interacting String Field Theory," Nucl. Phys. B **283** (1987) 1.

[11] D. J. Gross and A. Jevicki, "Operator Formulation Of Interacting String Field Theory. 2," Nucl. Phys. B **287** (1987) 225.

[12] E. Cremmer, A. Schwimmer and C. B. Thorn, "The vertex function in Witten's formulation of string field theory," Phys. Lett. B **179** (1986) 57.

[13] K. Furuuchi and K. Okuyama, "Comma vertex and string field algebra," JHEP **0109** (2001) 035 [arXiv:hep-th/0107101].

[14] T. Kawano and K. Okuyama, "Open string fields as matrices," JHEP **0106** (2001) 061 [arXiv:hep-th/0105129].

DISTA-UPO/07

Lectures on Hopf Algebras, Quantum Groups and Twists

Paolo Aschieri[1]

Centro Studi e Ricerche Enrico Fermi, Compendio Viminale, I-00184, Roma, Italy

Dipartimento di Scienze e Tecnologie Avanzate
Universitá del Piemonte Orientale, and INFN - Sezione di Torino
Via Bellini 25/G 15100 Alessandria, Italy

Abstract. Lead by examples we introduce the notions of Hopf algebra and quantum group. We study their geometry and in particular their Lie algebra (of left invariant vector fields). The examples of the quantum $Sl(2)$ Lie algebra and of the quantum (twisted) Poincaré Lie algebra $iso_\theta(3,1)$ are presented.

[1]E-mail: aschieri@mfn.unipmn.it

1 Introduction

Hopf algebras where initially considered more than Half a century ago. New important examples, named quantum groups, where studied in the 80's [1, 2, 4], they arose in the study of the quantum inverse scattering method in integrable systems [3]. Quantum groups can be seen as symmetry groups of noncommutative spaces, this is one reason they have been investigated in physics and mathematical physics (noncommutative spaces arise as quantization of commutative ones). The emergence of gauge theories on noncommutative spaces in open string theory in the presence of a NS 2-form background [7] has further motivated the study of noncommutative spaces and of their quantum group symmetry properties.

We here introduce the basic concepts of quantum group and of its Lie algebra of infinitesimal transformations. We pedagogically stress the connection with the classical (commutative) case and we treat two main examples, the quantum $su(2)$ Lie algebra and the quantum Poincaré Lie algebra.

Section 1 shows how commutative Hopf algebras emerge from groups. The quantum group $SL_q(2)$ is then presented and its corresponding universal enveloping algebra $U_q(sl(2))$ discussed. The relation between $SL_q(2)$ and $U_q(sl(2))$ is studied in Section 5. The quantum $sl(2)$ Lie algebra i.e. the algebra of infinitesimal transformations is then studied in Section 6. Similarly the geometry of Hopf algebras obtained from (abelian) twists is studied via the example of the Poincaré Lie algebra.

One aim of these lectures is to concisely introduce and relate all three aspects of quantum groups: 1) deformed algebra of functions [4], 2) deformed universal enveloping algebra [1, 2], and 3) quantum Lie algebra [8], that encodes the construction of the (bicovariant) differential calculus and geometry, most relavant for physical applications. A helpful review for the first and second aspects is [5], for quantum Lie algebra we refer to [6] and [9]. The (abelian) twist case, that is an interesting subclass, can be found in [10] and in [11].

In the appendix for reference we review some basic algebra notions and define Hopf algebras diagramatically.

2 Hopf algebras from groups

Let us begin with two examples motivating the notion of Hopf algebra. Let G be a finite group, and $A = Fun(G)$ be the set of functions from G to complex numbers $\mathbb{C}$. $A = Fun(G)$ is an algebra over $\mathbb{C}$ with the usual sum and product $(f+h)(g) = f(g)+h(g)$, $(f\cdot h) = f(g)h(g)$, $(\lambda f)(g) = \lambda f(g)$, for $f,h \in Fun(G)$, $g \in G$, $\lambda \in \mathbb{C}$. The unit of this algebra is I, defined by $I(g) = 1$, $\forall g \in G$. Using the group structure of G (multiplication map, existence of unit element and inverse element), we can introduce on $Fun(G)$ three other linear maps, the coproduct Δ, the counit ε, and the coinverse (or antipode) S:

$$\begin{aligned} \Delta(f)(g,g') &\equiv f(gg'), \quad \Delta : Fun(G) \to Fun(G)\otimes Fun(G) && (2.1)\\ \varepsilon(f) &\equiv f(1_G), \quad \varepsilon : Fun(G) \to \mathbb{C} && (2.2)\\ (Sf)(g) &\equiv f(g^{-1}), \quad S : Fun(G) \to Fun(G) && (2.3)\end{aligned}$$

where 1_G is the unit of G.

In general a coproduct can be expanded on $Fun(G)\otimes Fun(G)$ as:

$$\Delta(f) = \sum_i f_1^i \otimes f_2^i \equiv f_1 \otimes f_2, \tag{2.4}$$

where $f_1^i, f_2^i \in A = Fun(G)$ and $f_1 \otimes f_2$ is a shorthand notation we will often use in the sequel. Thus we have:

$$\Delta(f)(g,g') = (f_1\otimes f_2)(g,g') = f_1(g)f_2(g') = f(gg'). \tag{2.5}$$

It is not difficult to verify the following properties of the co-structures:

$$\begin{aligned} &(id\otimes\Delta)\Delta = (\Delta\otimes id)\Delta \quad \text{(coassociativity of } \Delta) && (2.6)\\ &(id\otimes\varepsilon)\Delta(a) = (\varepsilon\otimes id)\Delta(a) = a && (2.7)\\ &m(S\otimes id)\Delta(a) = m(id\otimes S)\Delta(a) = \varepsilon(a)I && (2.8)\end{aligned}$$

and

$$\Delta(ab) = \Delta(a)\Delta(b), \quad \Delta(I) = I \otimes I \tag{2.9}$$
$$\varepsilon(ab) = \varepsilon(a)\varepsilon(b), \quad \varepsilon(I) = 1 \tag{2.10}$$
$$S(ab) = S(b)S(a), \quad S(I) = I \tag{2.11}$$

where $a, b \in A = Fun(G)$ and m is the multiplication map $m(a \otimes b) \equiv ab$. The product in $\Delta(a)\Delta(b)$ is the product in $A \otimes A$: $(a \otimes b)(c \otimes d) = ab \otimes cd$.

For example the coassociativity property (2.6), $(id \otimes \Delta)\Delta(f) = (\Delta \otimes id)\Delta(f)$ reads $f_1 \otimes (f_2)_1 \otimes (f_2)_2 = (f_1)_1 \otimes (f_1)_2 \otimes f_2$, for all $f \in A$. This equality is easily seen to hold by applying it on the generic element (g, g', g'') of $G \times G \times G$, and then by using associativity of the product in G.

An algebra A (not necessarily commutative) endowed with the homomorphisms $\Delta : A \to A \otimes A$ and $\varepsilon : A \to \mathbb{C}$, and the linear and antimultiplicative map $S : A \to A$ satisfying the properties (2.6)-(2.11) is a *Hopf algebra*. Thus $Fun(G)$ is a Hopf algebra, it encodes the information on the group structure of G.

As a second example consider now the case where G is a group of matrices, a subgroup of GL given by matrices $T^a{}_b$ that satisfy some algebraic relation (for example orthogonality conditions). We then define $A = Fun(G)$ to be the algebra of polynomials in the matrix elements $T^a{}_b$ of the defining representation of G and in $\det T^{-1}$; i.e. the algebra is generated by the $T^a{}_b$ and $\det T^{-1}$.

Using the elements $T^a{}_b$ we can write an explicit formula for the expansion (2.4) or (2.5): indeed (2.1) becomes

$$\Delta(T^a{}_b)(g, g') = T^a{}_b(gg') = T^a{}_c(g)T^c{}_b(g'), \tag{2.12}$$

since T is a matrix representation of G. Therefore:

$$\Delta(T^a{}_b) = T^a{}_c \otimes T^c{}_b. \tag{2.13}$$

Moreover, using (2.2) and (2.3), one finds:

$$\varepsilon(T^a{}_b) = \delta^a_b \tag{2.14}$$
$$S(T^a{}_b) = (T^{-1})^a{}_b. \tag{2.15}$$

Thus the algebra $A = Fun(G)$ of polynomials in the elements $T^a{}_b$ and $\det T^{-1}$ is a Hopf algebra with co-structures defined by (2.13)-(2.15) and (2.9)-(2.11).

The two example presented concern commutative Hopf algebras. In the first example the information on the group G is equivalent to that on the Hopf algebra $A = Fun(G)$. We constructed A from G. In order to recover G from A notice that every element $g \in G$ can be seen as a map from A to $\mathbb{C}$ defined by $f \to f(g)$. This map is multiplicative because $fh(g) = f(g)h(g)$. The set G can be obtained from A as the set of all nonzero multiplicative linear maps from A to $\mathbb{C}$ (the set of characters of A).

Concerning the group structure of G, the product is recovered from the coproduct in A via (2.5), i.e. gg' is the new character that associates to any $f \in A$ the complex number $f_1(g)f_2(g')$. The unit of G is the character ε; the inverse g^{-1} is defined via the antipode of A.

In the second example, in order to recover the topology of G, we would need a C^*-algebra completion of the algebra $A = Fun(G)$ of polynomial functions. Up to these topological (C^*-algebra) aspects, we can say that the information concerning a matrix group G can be encoded in its commutative Hopf algebra $A = Fun(G)$.

In the spirit of noncommutative geometry we now consider noncommutative deformations $Fun_q(G)$ of the algebra $Fun(G)$. The space of points G does not exist anymore, by noncommutative or quantum space G_q is meant the noncommutative algebra $Fun_q(G)$. Since G is a group then $Fun(G)$ is a Hopf algebra; the noncommutative Hopf algebra obtained by deformation of $Fun(G)$ is then usually called *Quantum group*. The term quantum stems for the fact that the deformation is obtained by quantizing a Poisson (symplectic) structure of the algebra $Fun(G)$ [1].

Metti tutte le cose che non sono state possibili toccare, inclusa la derivata di Jackson.

Note on Noncommutative Geometry We have traded the finite set G for the finite dimensional commutative algebra $A = Fun(G)$ and the finite group G for the Hopf algebra $A = Fun(G)$. This equivalence holds also in the case in which G is a topological space (or group).

In general if X is a compact topological space, the algebra A of continuous functions from X to $\mathbb{C}$ has the norm, $||f|| = sup_{x\in X}|f(x)|$ where $|f(x)|$ is the modulus of $f(x)$. and a map $*$ (involution) given by defining f^* to be the complex conjugate of $f(x)$. The norm and the involution $*$ are compatible in the sense that $||ff^*|| = ||f||^2$, and we say that A is a C^*-algebra. The set X is recovered as before as the set of characters χ of A, (the topology on X is the weakest topology for which the functions $\hat{f}$ defined by $\hat{f}(\chi) = \chi(f)$ are continuous).

More importantly, by a theorem of Gel'fand every commutative C^*-algebra with a unit element can be considered the C^*-algebra of functions on a compact topological space. In short to specify a space X is equivalent to specify a C^*-algebra A, and viceversa.

The starting point of noncommutatve gemetry is to relax the commutativity of A, i.e. to consider noncommutative deformations of A. In the spirit of the Gel'fand theorem we consider these noncommutative C^*-algebras as describing a noncommutative deformation of the corresponding topological space X. The space of points X does not exist anymore, instead we have a noncomutative space described by the noncommutative C^*-algebra A.

If the space X is a group then the algebra A will be endowed with extra structure, that of a topological Hopf algebra.

3 Quantum groups. The example of $SL_q(2)$

Following [4] we consider quantum groups defined as the associative algebras A freely generated by non-commuting matrix entries $T^a{}_b$ satisfying the relation

$$R^{ab}{}_{ef}T^e{}_cT^f{}_d = T^b{}_fT^a{}_eR^{ef}{}_{cd} \tag{3.1}$$

and some other conditions depending on which classical group we are deforming (see later). The matrix R controls the non-commutativity of the $T^a{}_b$, and its elements depend continuously on a (in general complex) parameter q, or even a set of parameters. For $q \to 1$, the so-called "classical limit", we have

$$R^{ab}{}_{cd} \stackrel{q\to 1}{\longrightarrow} \delta^a_c\delta^b_d, \tag{3.2}$$

i.e. the matrix entries $T^a{}_b$ commute for $q = 1$, and one recovers the ordinary $Fun(G)$.

The associativity of A leads to a consistency condition on the R matrix, the quantum Yang–Baxter equation:

$$R^{a_1b_1}{}_{a_2b_2}R^{a_2c_1}{}_{a_3c_2}R^{b_2c_2}{}_{b_3c_3} = R^{b_1c_1}{}_{b_2c_2}R^{a_1c_2}{}_{a_2c_3}R^{a_2b_2}{}_{a_3b_3}. \tag{3.3}$$

For simplicity we rewrite the "RTT" equation (3.1) and the quantum Yang–Baxter equation as

$$R_{12}T_1T_2 = T_2T_1R_{12} \tag{3.4}$$

$$R_{12}R_{13}R_{23} = R_{23}R_{13}R_{12}, \tag{3.5}$$

where the subscripts 1, 2 and 3 refer to different couples of indices. Thus T_1 indicates the matrix $T^a{}_b$, T_1T_1 indicates $T^a{}_cT^c{}_b$, $R_{12}T_2$ indicates $R^{ab}{}_{cd}T^d{}_e$ and so on, repeated subscripts meaning matrix multiplication. The quantum Yang–Baxter equation (3.5) is a condition sufficient for the consistency of the RTT equation (3.4). Indeed the product of three distinct elements $T^a{}_b$, $T^c{}_d$ and $T^e{}_f$, indicated by $T_1T_2T_3$, can be reordered as $T_3T_2T_1$ via two differents paths:

$$T_1T_2T_3 \begin{array}{c} \nearrow \\ \\ \searrow \end{array} \begin{array}{c} T_1T_3T_2 \to T_3T_1T_2 \\ \\ T_2T_1T_3 \to T_2T_3T_1 \end{array} \begin{array}{c} \searrow \\ \\ \nearrow \end{array} T_3T_2T_1 \tag{3.6}$$

by repeated use of the RTT equation. The relation (3.5) ensures that the two paths lead to the same result.

The algebra A ("the quantum group") is a noncommutative Hopf algebra whose co-structures are the same of those defined for the commutative Hopf algebra $Fun(G)$ of the previous section, eqs. (2.13)-(2.15), (2.9)-(2.11).

Note Define $\hat{R}^{ab}{}_{cd} = R^{ba}{}_{cd}$. Then the quantum Yang-Baxter equation becomes the braid relation

$$\hat{R}_{23}\hat{R}_{12}\hat{R}_{23} = \hat{R}_{12}\hat{R}_{23}\hat{R}_{12} \ . \tag{3.7}$$

If $\hat{R}$ satisfies $\hat{R}^2 = id$ we have that $\hat{R}$ is a representation of the permutation group. In the more general case $\hat{R}$ is a representation of the braid group. The $\hat{R}$-matrx can be used to construct invariants of knots.

Let us give the example of the quantum group $SL_q(2) \equiv Fun_q(SL(2))$, the algebra freely generated by the elements α, β, γ and δ of the 2×2 matrix

$$T^a{}_b = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \tag{3.8}$$

satisfying the commutations

$$\begin{gathered} \alpha\beta = q\beta\alpha, \quad \alpha\gamma = q\gamma\alpha, \quad \beta\delta = q\delta\beta, \quad \gamma\delta = q\delta\gamma \\ \beta\gamma = \gamma\beta, \quad \alpha\delta - \delta\alpha = (q - q^{-1})\beta\gamma, \qquad q \in \mathbb{C} \end{gathered} \tag{3.9}$$

and

$$\det_q T \equiv \alpha\delta - q\beta\gamma = I. \tag{3.10}$$

The commutations (3.9) can be obtained from (3.1) via the R matrix

$$R^{ab}{}_{cd} = \begin{pmatrix} q & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & q - q^{-1} & 1 & 0 \\ 0 & 0 & 0 & q \end{pmatrix} \tag{3.11}$$

where the rows and columns are numbered in the order 11, 12, 21, 22.

It is easy to verify that the "quantum determinant" defined in (3.10) commutes with α, β, γ and δ, so that the requirement $\det_q T = I$ is consistent. The matrix inverse of $T^a{}_b$ is

$$(T^{-1})^a{}_b = (\det_q T)^{-1} \begin{pmatrix} \delta & -q^{-1}\beta \\ -q\gamma & \alpha \end{pmatrix} \ . \tag{3.12}$$

The coproduct, counit and coinverse of α, β, γ and δ are determined via formulas (2.13)-(2.15) to be:

$$\begin{gathered} \Delta(\alpha) = \alpha \otimes \alpha + \beta \otimes \gamma, \quad \Delta(\beta) = \alpha \otimes \beta + \beta \otimes \delta \\ \Delta(\gamma) = \gamma \otimes \alpha + \delta \otimes \gamma, \quad \Delta(\delta) = \gamma \otimes \beta + \delta \otimes \delta \end{gathered} \tag{3.13}$$

$$\varepsilon(\alpha) = \varepsilon(\delta) = 1, \quad \varepsilon(\beta) = \varepsilon(\gamma) = 0 \tag{3.14}$$

$$S(\alpha) = \delta, \quad S(\beta) = -q^{-1}\beta, \quad S(\gamma) = -q\gamma, \quad S(\delta) = \alpha. \tag{3.15}$$

Note The commutations (3.9) are compatible with the coproduct Δ, in the sense that $\Delta(\alpha\beta) = q\Delta(\beta\alpha)$ and so on. In general we must have

$$\Delta(R_{12}T_1T_2) = \Delta(T_2T_1R_{12}), \tag{3.16}$$

which is easily verified using $\Delta(R_{12}T_1T_2) = R_{12}\Delta(T_1)\Delta(T_2)$ and $\Delta(T_1) = T_1 \otimes T_1$. This is equivalent to proving that the matrix elements of the matrix product $T_1T'_1$, where T' is a matrix [satisfying (3.1)] whose elements *commute* with those of $T^a{}_b$, still obey the commutations (3.4).

Note $\Delta(\det_q T) = \det_q T \otimes \det_q T$ so that the coproduct property $\Delta(I) = I \otimes I$ is compatible with $\det_q T = I$.

Note The condition (3.10) can be relaxed. Then we have to include the central element $\zeta = (\det_q T)^{-1}$ in A, so as to be able to define the inverse of the q-matrix $T^a{}_b$ as in (3.12), and the coinverse of the element $T^a{}_b$ as in (2.15). The q-group is then $GL_q(2)$. The reader can deduce the co-structures on ζ: $\Delta(\zeta) = \zeta \otimes \zeta$, $\varepsilon(\zeta) = 1$, $S(\zeta) = \det_q T$.

4 Universal enveloping algebras and $U_q(sl(2))$

Another example of Hopf algebra is given by any ordinary Lie algebra g, or more precisely by the universal enveloping algebra $U(g)$ of a Lie algebra g, i.e. the algebra, with unit I, of polynomials in the generators χ_i modulo the commutation relations

$$[\chi_i, \chi_j] = C_{ij}{}^{k} \chi_k \ . \tag{4.1}$$

Here we define the co-structures on the generators as:

$$\Delta(\chi_i) = \chi_i \otimes I + I \otimes \chi_i \quad \Delta(I) = I \otimes I \tag{4.2}$$
$$\varepsilon(\chi_i) = 0 \qquad \varepsilon(I) = 1 \tag{4.3}$$
$$S(\chi_i) = -\chi_i \qquad S(I) = I \tag{4.4}$$

and extend them to all $U(g)$ by requiring Δ and ε to be linear and multiplicative, $\Delta(\chi\chi') = \Delta(\chi)\Delta(\chi')$, $\varepsilon(\chi\chi') = \varepsilon(\chi)\varepsilon(\chi')$, while S is linear and antimultiplicative. In order to show that the construction of the Hopf algebra $U(g)$ is well given, we have to check that the maps Δ, ε, S are well defined. We give the proof for the coproduct. Since $[\chi, \chi']$ is linear in the generators we have

$$\Delta[\chi, \chi'] = [\chi, \chi'] \otimes I + I \otimes [\chi, \chi'] \ , \tag{4.5}$$

on the other hand, using that Δ is multiplicative we have

$$\Delta[\chi, \chi'] = \Delta(\chi)\Delta(\chi') - \Delta(\chi')\Delta(\chi) \tag{4.6}$$

it is easy to see that these two expressions coincide.

The Hopf algebra $U(g)$ is noncommutative but it is cocommutative, i.e. for all $\zeta \in U(g)$, $\zeta_1 \otimes \zeta_2 = \zeta_2 \otimes \zeta_1$, where we used the notation $\Delta(\zeta) = \zeta_1 \otimes \zeta_2$. We have discussed deformations of commutative Hopf algebras, of the kind $A = Fun(G)$, and we will see that these are related to deformations of cocommutatative Hopf algebras of the kind $U(g)$ where g is the Lie algebra of G.

We here introduce the basic example of deformed universal enveloping algebra: $U_q(sl(2))$ [1, 2], which is a deformation of the usual enveloping algebra of $sl(2)$,

$$[X^+, X^-] = H \quad , \quad [H, X^\pm] = 2X^\pm \ . \tag{4.7}$$

The Hopf algebra $U_q(sl(2))$ is generated by the elements K_+, K_-, X_+ and X_- and the unit element I, that satisfy the relations

$$[X_+, X_-] = \frac{K_+^2 - K_-^{-2}}{q - q^{-1}} \ , \tag{4.8}$$
$$K_+ X_\pm K_- = q^{\pm 1} X_\pm \ , \tag{4.9}$$
$$K_+ K_- = K_- K_+ = I \ . \tag{4.10}$$

The parameter q that appears in the right hand side of the first two equations is a complex number. It can be checked that the algebra $U_q(sl(2))$ becomes a Hopf algebra by defining the following costructures

$$\Delta(X_\pm) = X^\pm \otimes K_+ + K_- \otimes X_\pm \ , \quad \Delta(K_\pm) = K_\pm \otimes K_\pm \tag{4.11}$$
$$\varepsilon(X_\pm) = 0 \qquad \varepsilon(K_\pm) = 1 \tag{4.12}$$
$$S(X_\pm) = -q^{\pm 1} X_\pm \qquad S(K_\pm) = K_\mp \tag{4.13}$$

If we define $K_+ = 1 + \frac{q}{2}\mathcal{H}$ then we see that in the limit $q \to 1$ we recover the undeformed $U(sl(2))$ Hopf algebra.

The Hopf algebra $U_q(sl(2))$ is not cocommutative, however the noncocommutativity is under control, as we now show. We set $q = e^h$, consider h a formal parameter and allow for power series in h. We are considering a topological completion of $U_q(sl(2))$, this is equivalently generated by the three generators $X_\pm$ and H, where we have $K_\pm = e^{\pm hH/2}$. In this case there exists an element $\mathcal{R}$ of $U_q(sl(2)) \widehat{\otimes} U_q(sl(2))$ (also the usual tensorproduct $\otimes$ has to be extended to allow for power series), called universal R-matrix that governs the noncocommutativity of the coproduct Δ,

$$\sigma\Delta(\zeta) = \mathcal{R}\Delta(\zeta)\mathcal{R}^{-1} \ , \tag{4.14}$$

where σ is the flip operation, $\sigma(\zeta \otimes \xi) = \xi \otimes \zeta$. The element $\mathcal{R}$ explicitly reads

$$\mathcal{R} = q^{\frac{H\otimes H}{2}} \sum_{n=0}^{\infty} \frac{(1-q^{-2})^n}{[n]!} (q^{H/2}X_+ \otimes q^{-H/2}X_-)^n q^{n(n-1)/2} \tag{4.15}$$

where $[n] \equiv \frac{q^n - q^{-n}}{q-q^{-1}}$, and $[n]! = [n][n-1]\ldots 1$.

The universal $\mathcal{R}$ matrix has further properties, that strucure $U_q(sl(2))$ to be a quasitriangular Hopf algebra. Among these properties we mention that $\mathcal{R}$ is invertible and that it satifies the Yang-Baxter equation

$$\mathcal{R}_{12}\mathcal{R}_{13}\mathcal{R}_{23} = \mathcal{R}_{23}\mathcal{R}_{13}\mathcal{R}_{12} \tag{4.16}$$

where we used the notation $\mathcal{R}_{12} = \mathcal{R} \otimes I$, $\mathcal{R}_{23} = I \otimes \mathcal{R}$ and $\mathcal{R}_{13} = \mathcal{R}^{\alpha} \otimes I \otimes \mathcal{R}_{\alpha}$, where $\mathcal{R} = \mathcal{R}^{\alpha} \otimes \mathcal{R}_{\alpha}$ (sum over α understood).

5 Duality

Consider a finite dimensional Hopf algebra A, the vector space A' dual to A is also a Hopf algebra with the following product, unit and costructures [we use the notation $\psi(a) = \langle \psi, a \rangle$ in order to stress the duality between A' and A]: $\forall \psi, \phi \in A'$, $\forall a, b \in A$

$$\langle \psi\phi, a \rangle = \langle \psi \otimes \phi, \Delta a \rangle \ , \quad \langle I, a \rangle = \varepsilon(a) \tag{5.1}$$

$$\langle \Delta(\psi), a \otimes b \rangle = \langle \psi, ab \rangle \ , \quad \varepsilon(\psi) = \langle \psi, I \rangle \tag{5.2}$$

$$\langle S(\psi), a \rangle = \langle \psi, S(a) \rangle \tag{5.3}$$

where $\langle \psi \otimes \phi \ , \ a \otimes b \rangle \equiv \langle \psi, a \rangle \, \langle \phi, b \rangle$. Obviously $(A')' = A$ and A and A' are dual Hopf algebras.

In the infinite dimensional case the definition of duality between Hopf algebras is more delicate because the coproduct on A' might not take values in the subspace $A' \otimes A'$ of $(A \otimes A)'$ and therefore is ill defined. We therefore use the notion of pairing: two Hopf algebras A and U are paired if there exists a bilinear map $\langle \ , \ \rangle \ : U \otimes A \to \mathbb{C}$ satisfying (5.1) and (5.2), (then (5.3) can be shown to follow as well).

The Hopf algebras $Fun(G)$ and $U(g)$ described in Section 2 and 3 are paired if g is the Lie algebra of G. Indeed we realize g as left invariant vectorfields on the group manifold. Then the pairing is defined by

$$\forall t \in g, \forall f \in Fun(G), \quad \langle t, f \rangle = t(f)|_{1_G} \ ,$$

where 1_G is the unit of G, and more in general is well defined by[2]

$$\forall\, tt'...t'' \in U(g), \forall f \in Fun(G), \quad \langle tt'...t'', f \rangle = t(t'...(t''(f)))|_{1_G} \ .$$

[2]In order to see that relations (5.1),(5.2) hold, we recall that t is left invariant if $TL_g(t|_{1_G}) = t|_g$, where TL_g is the tangent map induced by the left multiplication of the group on itself: $L_g g' = gg'$. We then have

$$t(f)|_g = \left(TL_g t|_{1_G}\right)(f) = t[f(g\tilde{g})]|_{\tilde{g}=1_G} = t[f_1(g)f_2(\tilde{g})]|_{\tilde{g}=1_G} = f_1(g)\, t(f_2)|_{1_G}$$

and therefore

$$\langle \tilde{t}\, t, f \rangle = \tilde{t}(t(f))|_{1_G} = \tilde{t}f_1|_{1_G} t f_2|_{1_G} = \langle \tilde{t} \otimes t, \Delta f \rangle \ ,$$

and

$$\langle t, fh \rangle = t(f)|_{1_G} h|_{1_G} + f|_{1_G} t(h)|_{1_G} = \langle \Delta(t), f \otimes h \rangle \ .$$

The duality between the Hopf algebras $Fun(Sl(2))$ and $U(sl(2)$ holds also in the deformed case, so that the quantum group $Sl_q(2)$ is dual to $U_q(sl(2))$. In order to show this duality we introduce a subalgebra (with generators $L^{\pm}$) of the algebra of linear maps from $Fun(Sl(2))$ to $\mathbb{C}$. We then see that this subalgebra has a natural Hopf algebra structure dual to $Fun(Sl(2))$. Finally we see in formula (5.22) that this subalgebra is just $U_q(sl(2))$. This duality is important because it allows to consider the elements of $U_q(sl(2))$ as (left invariant) differential operators on $Fun(Sl(2))$. This is the first step for the construction of a differential calculus on the quantum group $Fun(Sl(2))$.

The $L^{\pm}$ functionals

The linear functionals ${L^{\pm a}}_b$ are defined by their value on the elements ${T^a}_b$:

$$ {L^{\pm a}}_b({T^c}_d) = \langle {L^{\pm a}}_b, {T^c}_d \rangle = {R^{\pm ac}}_{bd}, \tag{5.4} $$

where

$$ (R^+)^{ac}{}_{bd} \equiv q^{-1/2} R^{ca}{}_{db} \tag{5.5} $$

$$ (R^-)^{ac}{}_{bd} \equiv q^{1/2} (R^{-1})^{ac}{}_{bd}, \tag{5.6} $$

The inverse matrix R^{-1} is defined by

$$ {R^{-1ab}}_{cd} {R^{cd}}_{ef} \equiv \delta^a_e \delta^b_f \equiv {R^{ab}}_{cd} {R^{-1cd}}_{ef}. \tag{5.7} $$

To extend the definition (5.4) to the whole algebra A, we set:

$$ {L^{\pm a}}_b(ab) = {L^{\pm a}}_g(a) {L^{\pm g}}_b(b), \quad \forall a, b \in A \tag{5.8} $$

so that, for example,

$$ {L^{\pm a}}_b({T^c}_d {T^e}_f) = {R^{\pm ac}}_{gd} {R^{\pm ge}}_{bf}. \tag{5.9} $$

In general, using the compact notation introduced in Section 2,

$$ L_1^{\pm}(T_2 T_3 ... T_n) = R_{12}^{\pm} R_{13}^{\pm} ... R_{1n}^{\pm}. \tag{5.10} $$

As it is esily seen from (5.9), the quantum Yang-Baxter equation (3.5) is a necessary and sufficient condition for the compatibility of (5.4) and (5.8) with the RTT relations: $L_1^{\pm}(R_{23} T_2 T_3 - T_3 T_2 R_{23}) = 0$.

Finally, the value of $L^{\pm}$ on the unit I is defined by

$$ {L^{\pm a}}_b(I) = \delta^a_b. \tag{5.11} $$

It is not difficult to find the commutations between ${L^{\pm a}}_b$ and ${L^{\pm c}}_d$:

$$ R_{12} L_2^{\pm} L_1^{\pm} = L_1^{\pm} L_2^{\pm} R_{12} \tag{5.12} $$

$$ R_{12} L_2^{+} L_1^{-} = L_1^{-} L_2^{+} R_{12}, \tag{5.13} $$

where the product $L_2^{\pm} L_1^{\pm}$ is by definition obtained by duality from the coproduct in A, $L_2^{\pm} L_1^{\pm}(a) \equiv (L_2^{\pm} \otimes L_1^{\pm})\Delta(a)$, $\forall a \in A$. For example consider

$$ R_{12}(L_2^+ L_1^+)(T_3) = R_{12}(L_2^+ \otimes L_1^+)\Delta(T_3) = R_{12}(L_2^+ \otimes L_1^+)(T_3 \otimes T_3) = q\, R_{12} R_{32} R_{31} $$

and

$$ L_1^+ L_2^+ (T_3) R_{12} = q\, R_{31} R_{32} R_{12} $$

so that the equation (5.12) is proven for L^+ by virtue of the quantum Yang–Baxter equation (3.3), where the indices have been renamed $2 \to 1, 3 \to 2, 1 \to 3$. Similarly, one proves the remaining "RLL" relations.

Note As mentioned in [4], L^+ is upper triangular, L^- is lower triangular (this is due to the upper and lower triangularity of R^+ and R^-, respectively). From (5.12) and (5.13) we have

$$ {L^{\pm a}}_a {L^{\pm b}}_b = {L^{\pm b}}_b {L^{\pm b}}_b \ ; \quad {L^{+a}}_a {L^{-b}}_b = {L^{-b}}_b {L^{+a}}_a = \varepsilon \ . \tag{5.14} $$

we also have

$$L^{\pm 1}{}_1 L^{\pm 2}{}_2 = \varepsilon \; . \tag{5.15}$$

The algebra of polynomials in the $L^{\pm}$ functionals becomes a Hopf algebra paired to $Sl_q(2)$ by defining the costructures via the duality (5.4):

$$\Delta(L^{\pm a}{}_b)(T^c{}_d \otimes T^e{}_f) \equiv L^{\pm a}{}_b(T^c{}_d T^e{}_f) = L^{\pm a}{}_g(T^c{}_d)L^{\pm g}{}_b(T^e{}_f) \tag{5.16}$$

$$\varepsilon(L^{\pm a}{}_b) \equiv L^{\pm a}{}_b(I) \tag{5.17}$$

$$S(L^{\pm a}{}_b)(T^c{}_d) \equiv L^{\pm a}{}_b(S(T^c{}_d)) \tag{5.18}$$

cf. [(5.2), (5.3)], so that

$$\Delta(L^{\pm a}{}_b) = L^{\pm a}{}_g \otimes L^{\pm g}{}_b \tag{5.19}$$

$$\varepsilon(L^{\pm a}{}_b) = \delta^a_b \tag{5.20}$$

$$S(L^{\pm a}{}_b) = L^{\pm a}{}_b \circ S \; . \tag{5.21}$$

This Hopf algebra is $U_q(sl(2))$ because it can be checked that relations (5.12), (5.13), (5.14), (5.15) fully characterize the $L^{\pm}$ functionals, so that the algebra of polynomials in the *symbols* $L^{\pm a}{}_b$ that satisfy the relations (5.12), (5.13), (5.14), (5.15) is isomorphic to the algebra generated by the $L^{\pm}$ *functionals* on $U_q(sl(2))$. An explicit relation between the $L^{\pm}$ matrices and the generators $X^{\pm}$ and $K^{\pm}$ of $U_q(sl(2))$ introduced in the previous section is obtained by comparing the "RLL" commutation relations with the $U_q(sl(2))$ Lie algebra relations, we obtain

$$L^+ = \begin{pmatrix} K_- & q^{-1/2}(q-q^{-1})X_+ \\ 0 & K_+ \end{pmatrix} \quad , \quad L^- = \begin{pmatrix} K_+ & 0 \\ q^{1/2}(q^{-1}-q)X_- & K_- \end{pmatrix} \; . \tag{5.22}$$

6 Quantum Lie algebra

We now turn our attention to the issue of determining the Lie algebra of the quantum group $Sl_q(2)$, or equivalently the quantum Lie algebra of the universal enveloping algebra $U_q(sl(2))$.

In the undeformed case the Lie algebra of a universal enveloping algebra U (for example $U(sl(2))$) is the unique linear subspace g of U of primitive elements, i.e. of elements χ that have coproduct:

$$\Delta(\chi) = \chi \otimes 1 + 1 \otimes \chi \; . \tag{6.1}$$

Of course g generates U and g is closed under the usual commutator bracket [,],

$$[u, v] = uu - vu \in g \quad \text{for all } u, v \in g \; . \tag{6.2}$$

The geometric meaning of the bracket $[u, v]$ is that it is the adjoint action of g on g,

$$[u, v] = ad_u\, v \tag{6.3}$$

$$ad_u\, v := u_1 v S(u_2) \tag{6.4}$$

where we have used the notation $\Delta(u) = u_1 \otimes u_2$, where a sum over u_1 and u_2 is understood. Recalling that $\Delta(u) = u \otimes 1 + 1 \otimes u$ and that $S(u) = -u$, from (6.4) we immediately obtain (6.3). In other words, the commutator $[u, v]$ is the Lie derivative of the left invariant vectorfield u on the left invariant vectorfield v. More in general the adjoint action of U on U is given by

$$ad_\xi\, \zeta = \xi_1 \zeta S(\xi_2) \; , \tag{6.5}$$

where we used the notation (sum understood) $\Delta(\xi) = \xi_1 \otimes \xi_2$.

In the deformed case the coproduct is no more cocommutative and we cannot identify the Lie algebra of a deformed universal enveloping algebra U_q with the primitive elements of U_q, they are too few to generate U_q. We then have to relax this requirement. There are three natural conditions that according to [8] the q-Lie algebra of a q-deformed universal enveloping algebra U_q has to satisfy, see [9, 12], and [13] p. 41. It has to be a linear subspace g_q of U_q such that

$$i) \qquad g_q \text{ generates } U_q \ , \tag{6.6}$$
$$ii) \qquad \Delta(g_q) \subset g_q \otimes 1 + U_q \otimes g_q \ , \tag{6.7}$$
$$iii) \qquad [g_q, g_q] \subset g_q \ . \tag{6.8}$$

Here now Δ is the coproduct of U_q and $[\ ,\ ]$ denotes the q-bracket

$$[u, v] = ad_u\, v = u_1 v S(u_2) \ . \tag{6.9}$$

where we have used the coproduct notation $\Delta(u) = u_1 \otimes u_2$. Property $iii)$ is the closure of g_q under the adjoint action. Property $ii)$ implies a minimal deformation of the Leibnitz rule.

From these conditions, that do not in general single out a unique subspace g_q, it follows that the bracket $[u, v]$ is quadratic in u and v, that it has a deformed antisymmetry property and that it satisfies a deformed Jacoby identity.

In the example $U_q = U_q(sl(2))$ we have that a quantum $sl(2)$ Lie algebra is spanned by the four linearly independent elements

$$\chi^{c_1}{}_{c_2} = \frac{1}{q - q^{-1}}[L^{+c_1}{}_b S(L^{-b}{}_{c_2}) - \delta^{c_1}_{c_2}\varepsilon] \ . \tag{6.10}$$

In the commutative limit $q \to 1$, we have $\chi^2{}_2 = -\chi^1{}_1 = H/2$, $\chi^1{}_2 = X_+$, $\chi^2{}_1 = X_-$, and we recover the usual $sl(2)$ Lie algebra.

The q-Lie algebra commutation relations explicitly are

$$\chi_1\chi_+ - \chi_+\chi_1 + (q^4 - q^2)\chi_+\chi_2 = q^3\chi_+$$

$$\chi_1\chi_- - \chi_-\chi_1 - (q^2 - 1)\chi_-\chi_-2 = -q\chi_-$$

$$\chi_1\chi_2 - \chi_2\chi_1 = 0$$

$$\chi_+\chi_- - \chi_-\chi_+ - (1 - q^2)\chi_1\chi_2 + (1 - q^2)\chi_2\chi_2 = q(\chi_1 - \chi_2)$$

$$\chi_2\chi_+ - q^2\chi_+\chi_2 = -q\chi_+$$

$$\chi_2\chi_- - q^{-2}\chi_-\chi_2 = q^{-1}\chi_-$$

where we used the composite index notation

$${}^{a_1}{}_{a_2} \to {}^i \ , \quad {}_{b_1}{}^{b_2} \to {}_j \ , \quad \text{and} \quad i, j = 1, +, -, 2 \ .$$

These q-lie algebra relations can be compactly written [9][3]

$$[\chi_i, \chi_j] = \chi_i\chi_j - \Lambda^{rs}{}_{ji}\chi_s\chi_r \ , \tag{6.11}$$

where $\Lambda^{a_2}{}_{a_1}{}^{d_2}{}_{d_1}{}^{c_1}{}_{c_2}{}^{b_1}{}_{b_2} = S(L^{+b_1}{}_{a_1})L^{-a_2}{}_{b_2}(T^{c_1}{}_{d_1}S(T^{d_2}{}_{c_2}))$. The q-Jacoby identities then read

$$[\chi_i, [\chi_j, \chi_r]] = [[\chi_i, \chi_j], \chi_r]] + \Lambda^{kl}{}_{ji}[\chi_l, [\chi_k, \chi_r]] \ . \tag{6.12}$$

[3]Relation to the conventions of [8, 9] (here underlined): $\chi_i = -S^{-1}\underline{\chi}_i$, $f_j{}^i = S^{-1}\underline{f}^i{}_j$.

7 Deformation by twist and quantum Poincaré Lie algebra

In this last section, led by the example the Pincaré Lie algebra, we rewiew a quite general method to deform the Hopf algebra $U(g)$, the universal enveloping algebra of a given Lie algebra g. It is based on a twist procedure. A twist element $\mathcal{F}$ is an invertible element in $Ug \otimes Ug$. A main property $\mathcal{F}$ has to satisfy is the cocycle condition

$$(\mathcal{F} \otimes 1)(\Delta \otimes id)\mathcal{F} = (1 \otimes \mathcal{F})(id \otimes \Delta)\mathcal{F} \ . \tag{7.1}$$

Consider for example the usual Poincaré Lie algebra $iso(3,1)$:

$$\begin{aligned}
[P_\mu, P_\nu] &= 0 \ , \\
[P_\rho, M_{\mu\nu}] &= i(\eta_{\rho\mu} P_\nu - \eta_{\rho\nu} P_\mu) \ , &(7.2)\\
[M_{\mu\nu}, M_{\rho\sigma}] &= -i(\eta_{\mu\rho} M_{\nu\sigma} - \eta_{\mu\sigma} M_{\nu\rho} - \eta_{\nu\rho} M_{\mu\sigma} + \eta_{\nu\sigma} M_{\mu\rho}) \ , &(7.3)
\end{aligned}$$

A twist element is given by

$$\mathcal{F} = e^{\frac{i}{2}\theta^{\mu\nu} P_\mu \otimes P_\nu} \ , \tag{7.4}$$

where $\theta^{\mu\nu}$ (despite the indices $\mu\nu$ notation) is a real antisymmetric matrix of dimensionful constants (the previous deformation parameter q was a consant too). We consider $\theta^{\mu\nu}$ fundamental physical constants, like the velocity of light c, or like $\hbar$. In this setting symmetries will leave $\theta^{\mu\nu}$, c and $\hbar$ invariant. The inverse of $\mathcal{F}$ is

$$\mathcal{F} = e^{\frac{-i}{2}\theta^{\mu\nu} P_\mu \otimes P_\nu} \ .$$

This twist satisfies the cocycle condition (7.1) because the Lie algebra of momenta is abelian.

The Poincaré Hopf algebra $U^{\mathcal{F}}(iso(3,1))$ is a deformation of $U(iso(3,1))$. As algebras $U^{\mathcal{F}}(iso(3,1)) = U(iso(3,1))$; but $U^{\mathcal{F}}(iso(3,1))$ has the new coproduct

$$\Delta^{\mathcal{F}}(\xi) = \mathcal{F}\Delta(\xi)\mathcal{F}^{-1} \ , \tag{7.5}$$

for all $\xi \in U(iso(3,1))$. In order to write the explicit expression for $\Delta^{\mathcal{F}}(P_\mu)$ and $\Delta^{\mathcal{F}}(M_{\mu\nu})$, we use the Hadamard formula

$$Ad_{e^X} Y = e^X \ Y \ e^{-X} = \sum_{n=0}^{\infty} \frac{1}{n!} \underbrace{[X,[X,...[}_{n} X,Y]] = \sum_{n=0}^{\infty} \frac{(adX)^n}{n!} \ Y$$

and the relation $[P \otimes P', M \otimes 1] = [P, M] \otimes P'$, and thus obtain [16], [15]

$$\begin{aligned}
\Delta^{\mathcal{F}}(P_\mu) &= P_\mu \otimes 1 + 1 \otimes P_\mu \ , \\
\Delta^{\mathcal{F}}(M_{\mu\nu}) &= M_{\mu\nu} \otimes 1 + 1 \otimes M_{\mu\nu} \\
&\quad - \frac{1}{2}\theta^{\alpha\beta}\left((\eta_{\alpha\mu} P_\nu - \eta_{\alpha\nu} P_\mu) \otimes P_\beta + P_\alpha \otimes (\eta_{\beta\mu} P_\nu - \eta_{\beta\nu} P_\mu)\right) .
\end{aligned} \tag{7.6}$$

We have constructed the Hopf algebra $U^{\mathcal{F}}(iso(3,1)$: it is the algebra generated by $M_{\mu\nu}$ and P_μ modulo the relations (7.2), and with coproduct (7.6) and counit and antipode that are as in the undeformed case:

$$\varepsilon(P_\mu) = \varepsilon(M_{\mu\nu}) = 0 \quad , \quad S(P_\mu) = -P_\mu \quad , \quad S(M_{\mu\nu}) = -M_{\mu\nu} \ . \tag{7.7}$$

This algebra is a symmetry algebra of the noncommutative spacetime $x^\mu x^\nu - x^\nu x^\mu = i\theta^{\mu\nu}$.

In general given a Lie algebra g, and a twist $\mathcal{F} \in U(g) \otimes U(g)$, formula (7.5) defines a new coproduct that is not cocommutative. We call $U(g)^{\mathcal{F}}$ the new Hopf algebra with coproduct $\Delta^{\mathcal{F}}$, counit $\varepsilon^{\mathcal{F}} = \varepsilon$ and antipode $S^{\mathcal{F}}$ that is a deformation of S [14][4]. By definition as algebra $U(g)^{\mathcal{F}}$ equals $U(g)$, only the costructures are deformed.

[4]Explicitly, if we write $\mathcal{F} = \mathrm{f}^{\alpha} \otimes \mathrm{f}_{\alpha}$ and define the element $\chi = \mathrm{f}^\alpha S(\mathrm{f}_\alpha)$ (that can be proven to be invertible) then for all elements $\xi \in U(g)$, $S^{\mathcal{F}}(\xi) = \chi S(\xi)\chi^{-1}$.

We now construct the quantum Poincaré Lie algbra $iso^{\mathcal{F}}(3,1)$. Following the previous section, the Poincaré Lie algebra $iso^{\mathcal{F}}(3,1)$ must be a linear subspace of $U^{\mathcal{F}}(iso(3,1))$ such that if $\{t_i\}_{i=1,...n}$ is a basis of $iso^{\mathcal{F}}(3,1)$, we have (sum understood on repeated indices)

$$\begin{array}{ll} i) & \{t_i\} \text{ generates } U^{\mathcal{F}}(iso(3,1)) \\ ii) & \Delta^{\mathcal{F}}(t_i) = t_i \otimes 1 + f_i{}^j \otimes t_j \\ iii) & [t_i, t_j]_{\mathcal{F}} = C_{ij}{}^k t_k \end{array}$$

where $C_{ij}{}^k$ are structure constants and $f_i{}^j \in U^{\mathcal{F}}(iso(3,1))$ $(i,j = 1,...n)$. In the last line the bracket $[\ ,\ ]_{\mathcal{F}}$ is the adjoint action:

$$[t, t']_{\mathcal{F}} := ad^{\mathcal{F}}_t\, t' = t_{1_{\mathcal{F}}} t' S(t_{2_{\mathcal{F}}}) \ , \tag{7.8}$$

where we used the coproduct notation $\Delta^{\mathcal{F}}(t) = t_{1_{\mathcal{F}}} \otimes t_{2_{\mathcal{F}}}$. The statement that the Lie algebra of $U^{\mathcal{F}}(iso(3,1))$ is the undeformed Poincaré Lie algebra (7.2) is not correct because conditions $ii)$ and $iii)$ are not met by the generators P_μ and $M_{\mu\nu}$. There is a canonical procedure in order to obtain the Lie algebra $iso^{\mathcal{F}}(3,1)$ of $U^{\mathcal{F}}(iso(3,1))$ [10, 11]. Consider the elements

$$\begin{aligned} P^{\mathcal{F}}_\mu &:= \bar{\mathrm{f}}^\alpha(P_\mu)\bar{\mathrm{f}}_\alpha = P_\mu \ , & (7.9) \\ M^{\mathcal{F}}_{\mu\nu} &:= \bar{\mathrm{f}}^\alpha(M_{\mu\nu})\bar{\mathrm{f}}_\alpha = M_{\mu\nu} - \frac{i}{2}\theta^{\rho\sigma}[P_\rho, M_{\mu\nu}]P_\sigma \\ &\qquad = M_{\mu\nu} + \frac{1}{2}\theta^{\rho\sigma}(\eta_{\mu\rho}P_\nu - \eta_{\nu\rho}P_\mu)P_\sigma & (7.10) \end{aligned}$$

Their coproduct is

$$\begin{aligned} \Delta^{\mathcal{F}}(P_\mu) &= P_\mu \otimes 1 + 1 \otimes P_\mu \ , \\ \Delta^{\mathcal{F}}(M^{\mathcal{F}}_{\mu\nu}) &= M^{\mathcal{F}}_{\mu\nu} \otimes 1 + 1 \otimes M^{\mathcal{F}}_{\mu\nu} + i\theta^{\alpha\beta} P_\alpha \otimes [P_\beta, M_{\mu\nu}] \ . \end{aligned} \tag{7.11}$$

The counit and antipode are

$$\varepsilon(P_\mu) = \varepsilon(M^{\mathcal{F}}_{\mu\nu}) = 0 \quad , \quad S(P_\mu) = -P_\mu \quad , \quad S(M^{\mathcal{F}}_{\mu\nu}) = -M^{\mathcal{F}}_{\mu\nu} - i\theta^{\rho\sigma}[P_\rho, M_{\mu\nu}]P_\sigma \ . \tag{7.12}$$

The elements $P^{\mathcal{F}}_\mu$ and $M^{\mathcal{F}}_{\mu\nu}$ are generators because they satisfy condition $i)$ (indeed $M_{\mu\nu} = M^{\mathcal{F}}_{\mu\nu} + \frac{i}{2}\theta^{\rho\sigma}[P_\rho, M^{\mathcal{F}}_{\mu\nu}]P_\sigma$). They are deformed *infinitesimal* generators because they satisfy the Leibniz rule $ii)$ and because they close under the Lie bracket $iii)$. Explicitly

$$\begin{aligned} [P_\mu, P_\nu]_{\mathcal{F}} &= 0 \ , \\ [P_\rho, M^{\mathcal{F}}_{\mu\nu}]_{\mathcal{F}} &= i(\eta_{\rho\mu}P_\nu - \eta_{\rho\nu}P_\mu) \ , \\ [M^{\mathcal{F}}_{\mu\nu}, M^{\mathcal{F}}_{\rho\sigma}]_{\mathcal{F}} &= -i(\eta_{\mu\rho}M^{\mathcal{F}}_{\nu\sigma} - \eta_{\mu\sigma}M^{\mathcal{F}}_{\nu\rho} - \eta_{\nu\rho}M^{\mathcal{F}}_{\mu\sigma} + \eta_{\nu\sigma}M^{\mathcal{F}}_{\mu\rho}) \ . \end{aligned} \tag{7.13}$$

We notice that the structure constants are the same as in the undeformed case, however the adjoint action $[M^{\mathcal{F}}_{\mu\nu}, M^{\mathcal{F}}_{\rho\sigma}]_{\mathcal{F}}$ is not the commutator anymore, it is a deformed commutator quadratic in the generators and antisymmetric:

$$\begin{aligned} [P_\mu, P_\nu]_{\mathcal{F}} &= [P_\mu, P_\nu] \ , \\ [P_\rho, M^{\mathcal{F}}_{\mu\nu}]_{\mathcal{F}} &= [P_\rho, M^{\mathcal{F}}_{\mu\nu}] \ , \\ [M^{\mathcal{F}}_{\mu\nu}, M^{\mathcal{F}}_{\rho\sigma}]_{\mathcal{F}} &= [M^{\mathcal{F}}_{\mu\nu}, M^{\mathcal{F}}_{\rho\sigma}] - i\theta^{\alpha\beta}[P_\alpha, M_{\rho\sigma}][P_\beta, M_{\mu\nu}] \ . \end{aligned} \tag{7.14}$$

From (7.13) we immediately obtain the Jacoby identities:

$$[t\,, [t', t'']_{\mathcal{F}}]_{\mathcal{F}} + [t'\,, [t'', t]_{\mathcal{F}}]_{\mathcal{F}} + [t''\,, [t, t']_{\mathcal{F}}]_{\mathcal{F}} = 0 \ , \tag{7.15}$$

for all $t, t', t'' \in iso^{\mathcal{F}}(3,1)$.

Acknowledgments

I would like to thank the organizers and the participants of the second Modave Summer School in Mathematical Physics for the fruitful and refreshing week, characterized by lectures, discussions, group study, all in a community atmosphere.

This work is partially supported by the European Community Marie Curie contracts MERG-CT-2004-006374 (European Reintegrtion Grant) and MRTN-CT-2004-005104 (Human Potential programme), and by the Italian MIUR contract PRIN-2005023102.

Appendix

A Algebras, Coalgebras and Hopf algebras

In the introduction we have motivated the notion of Hopf algebra. We here review some basic definitions in linear algebra and show how Hopf algebras merge algebras and coalgebras structures in a symmetric (specular) way [17]).

We recall that a module by definition is an abelian group. The group operation is denoted $+$ (additive notation). A vector space A over $\mathbb{C}$ (or $\mathbb{R}$) is a $\mathbb{C}$-module, i.e. there is an action $(\lambda, a) \to \lambda a$ of the group $(\mathbb{C} - \{0\}, \cdot)$ on the module A,

$$(\lambda'\lambda)a = \lambda(\lambda' a) \ , \tag{A1}$$

and this action is compatible with the addition in A and in $\mathbb{C}$, i.e. it is compatible with the module structure of A and of $\mathbb{C}$:

$$\lambda(a + a') = \lambda a + \lambda a' \quad , \quad (\lambda + \lambda')a = \lambda a + \lambda' a \ . \tag{A2}$$

An **algebra** A over $\mathbb{C}$ with unit I, is a vector space over $\mathbb{C}$ with a multiplication map, that we denote $\cdot$ or μ,

$$\mu \,:\, A \times A \to A \tag{A3}$$

that is $\mathbb{C}$-bilinear: $(\lambda a) \cdot (\lambda' b) = \lambda\lambda'(a \cdot b)$, that is associative and that for all a satisfies $a \cdot I = I \cdot a = a$.

These three properties can be stated diagrammatically. $\mathbb{C}$-bilinearity of the product $\mu : A \times A \to A$, is equivalently expressed as linearity of the map $\mu : A \otimes A \to A$. Associativity reads,

$$\begin{array}{ccc} A \otimes A \otimes A & \xrightarrow{\mu\otimes id} & A \otimes A \\ {\scriptstyle id\otimes\mu}\big\downarrow & & {\scriptstyle \mu}\big\downarrow \\ A \otimes A & \xrightarrow{\mu} & A \end{array}$$

Finally the existence of the unit I such that for all a we have $a \cdot I = I \cdot a = a$ is equivalent to the existence of a linear map

$$i : \mathbb{C} \to A \tag{A4}$$

such that

$$\begin{array}{ccc} \mathbb{C} \otimes A & \xrightarrow{i\otimes id} & A \otimes A \\ {\scriptstyle \simeq}\big\downarrow & & {\scriptstyle \mu}\big\downarrow \\ A & \xrightarrow{id} & A \end{array}$$

and

$$\begin{array}{ccc} A \otimes \mathbb{C} & \xrightarrow{id\otimes i} & A \otimes A \\ {\scriptstyle \simeq}\big\downarrow & & {\scriptstyle \mu}\big\downarrow \\ A & \xrightarrow{id} & A \end{array}$$

where $\simeq$ denotes the canonical isomorphism between $\mathbb{C} \otimes A$ and A. The unit I is then recovered as $i(1) = I$.

A **coalgebra** A over $\mathbb{C}$ is a vector space with a linear map $\Delta : A \to A \otimes A$ that is coassociative, $(id \otimes \Delta)\Delta = (\Delta \otimes id)\Delta$, and a linear map $\varepsilon : A \to \mathbb{C}$, called counit that satisfies $(id \otimes \varepsilon)\Delta(a) =$

$(\varepsilon \otimes id)\Delta(a) = a$. These properties can be expressed diagrammatically by reverting the arrows of the previous diagrams:

$$\begin{tikzcd} A \otimes A \otimes A & A \otimes A \arrow[l, "\Delta\otimes id"'] \\ A \otimes A \arrow[u, "id\otimes\Delta"] & A \arrow[l, "\Delta"] \arrow[u, "\Delta"'] \end{tikzcd}$$

$$\begin{tikzcd} \mathbb{C} \otimes A & A \otimes A \arrow[l, "\varepsilon\otimes id"'] \\ A \arrow[u, "\simeq"] & A \arrow[l, "id"] \arrow[u, "\Delta"'] \end{tikzcd}$$

and

$$\begin{tikzcd} \mathbb{C} \otimes A & A \otimes A \arrow[l, "id\otimes\varepsilon"'] \\ A \arrow[u, "\simeq"] & A \arrow[l, "id"] \arrow[u, "\Delta"'] \end{tikzcd}$$

We finally arrive at the

Definiton A **bialgebra** A over $\mathbb{C}$ is a vectorspace A with an algebra and a coalgebra structure that are compatible, i.e.
1) the coproduct Δ is an algebra map between the algebra A and the algebra $A \otimes A$, where the product in $A \otimes A$ is $(a \otimes b)(c \otimes d) = ac \otimes bd$,

$$\Delta(ab) = \Delta(a)\Delta(b) \quad , \quad \Delta(I) = I \otimes I \tag{A5}$$

2) The counit $\varepsilon : A \to \mathbb{C}$ is an algebra map

$$\varepsilon(ab) = \varepsilon(a)\varepsilon(b) \quad , \quad \varepsilon(I) = 1 \ . \tag{A6}$$

Definition A **Hopf algebra** is a bialgebra with a linear map $S : A \to A$, called antipode (or coinverse), such that

$$\mu(S \otimes id)\Delta(a) = \mu(id \otimes S)\Delta(a) = \varepsilon(a)I \ . \tag{A7}$$

It can be proven that the antipode S is unique and antimultiplicative

$$S(ab) = S(b)S(a) \ .$$

From the definition of bialgebra it follows that $\mu : A \otimes A \to A$ and $i : \mathbb{C} \to A$ are coalgebra maps, i.e., $\Delta \circ \mu = \mu \otimes \mu \circ \underline{\Delta}$, $\varepsilon \otimes \mu = \underline{\varepsilon}$ and $\Delta \circ i = i \otimes i \circ \Delta_{\mathbb{C}}$, $\varepsilon \circ i = \varepsilon_{\mathbb{C}}$, where the coproduct and counit in $A \otimes A$ are given by $\underline{\Delta}(a \otimes b) = a_1 \otimes b_1 \otimes a_2 \otimes b_2$ and $\underline{\varepsilon} = \varepsilon \otimes \varepsilon$, while the coproduct in $\mathbb{C}$ is the map $\Delta_{\mathbb{C}}$ that identifies $\mathbb{C}$ with $\mathbb{C} \otimes \mathbb{C}$ and the counit is $\varepsilon_{\mathbb{C}} = id$. Viceversa if A is an algebra and a coalgebra and μ and i are coalgebra maps then it follows that Δ and ε are algebra maps.

One can write diagrammatically equations (A5), (A6), (A7), and see that the Hopf algebra definition is invariant under inversion of arrows and exchange of structures with costructures, with the antipode going into itself. In this respect the algebra and the coalgebra structures in a Hopf algebra are specular.

References

[1] V. G. Drinfeld, “Hopf algebras and the quantum Yang-Baxter equation,” Sov. Math. Dokl. **32** (1985) 254 [Dokl. Akad. Nauk Ser. Fiz. **283** (1985) 1060]. J. Sov. Math. **41** (1988) 898 [Zap. Nauchn. Semin. **155** (1986) 18].

[2] M. Jimbo, “A q-difference analogue of U(g) and the Yang-Baxter equation” Lett. Math. Phys. **10** (1985) 63, “A Q Analog of U (Gl (N+1)), Hecke Algebra and the Yang-Baxter equation,” Lett. Math. Phys. **11** (1986) 247.

[3] L. D. Faddeev, E. K. Sklyanin and L. A. Takhtajan, “The Quantum Inverse Problem Method. 1,” Theor. Math. Phys. **40** (1980) 688 [Teor. Mat. Fiz. **40** (1979) 194].

[4] L. D. Faddeev, N. Y. Reshetikhin and L. A. Takhtajan, Quantization of Lie groups and Lie algebras,” Leningrad Math. J. **1**, 193 (1990) [Alg. Anal. **1**, 178 (1989)].

[5] S. Majid, “Quasitriangular Hopf Algebras and Yang-Baxter equations,” Int. J. Mod. Phys. A **5** (1990) 1.

[6] B. Zumino, “Introduction to the differential geometry of quantum groups,” UCB-PTH-91-62 *These notes covers part of a plenary talk given at 10th IAMP Conf., Leipzig, Germany, Jul 30 - Aug 9, 1991*

[7] N. Seiberg and E. Witten, JHEP **9909** (1999) 032 [arXiv:hep-th/9908142].

[8] S. L. Woronowicz, “Differential Calculus On Compact Matrix Pseudogroups (Quantum Groups),” Commun. Math. Phys. **122** (1989) 125.

[9] P. Aschieri and L. Castellani, *An Introduction to noncommutative differential geometry on quantum groups,* Int. J. Mod. Phys. A **8** (1993) 1667 hep-th/9207084.

[10] P. Aschieri, M. Dimitrijevic, F. Meyer and J. Wess, “Noncommutative geometry and gravity,” Class. Quant. Grav. **23** (2006) 1883, hep-th/0510059.

[11] P. Aschieri, “Noncommutative symmetries and gravity,” J. Phys. Conf. Ser. **53** (2006) 799 [arXiv:hep-th/0608172].

[12] P. Schupp, P. Watts and B. Zumino, *Bicovariant quantum algebras and quantum Lie algebras,* Commun. Math. Phys. **157** (1993) 305 hep-th/9210150.

[13] P. Aschieri, *On the geometry of inhomogeneous quantum groups,* math.qa/9805119, Scuola Normale Superiore di Pisa, Pubblicazioni Classe di Scienze, Collana Tesi.

[14] V. G. Drinfeld, “On constant quasiclassical solutions of the Yang-Baxter equations,” Soviet Math. Dokl. **28** (1983) 667-671, Quasi-Hopf Algebras, Lengingrad Math. J. **1** (1990) 1419 [Alg. Anal. **1N6** (1989) 114].

[15] J. Wess, *Deformed coordinate spaces: Derivatives*, Lecture given at the Balkan workshop BW2003, August 2003, Vrnjacka Banja, Serbia , hep-th/0408080.

[16] M. Chaichian, P. Kulish, K. Nishijima, and A. Tureanu, *On a Lorentz-invariant interpretation of noncommutative space-time and its implications on noncommutative QFT*, Phys. Lett. **B604**, 98 (2004), hep-th/0408069.

[17] M.E. Sweedler, *Hopf Algebras*, Benjamin, New Yourk (1969),
E. Abe, *Hopf Algebras*, Cambridge University Press, Cambridge (1980)
S. Majid, *Foundations of quantum group theory*, Cambridge: University Press (1995) 606 p.