Topology for Physicists

M. Zirnbauer
Institut für Theoretische Physik,
Universität zu Köln

SS 2011

Contents

1 De Rham Theory

1.1 Differential complexes

Definition. A direct sum of vector spaces $C = \bigoplus_{q\in\mathbb{Z}} C^q$ indexed by the integers (or a subset thereof) is called a *differential complex* if it is equipped with a linear mapping (actually, a collection of linear mappings),

$$\ldots \longrightarrow C^{q-1} \xrightarrow{d} C^q \xrightarrow{d} C^{q+1} \longrightarrow \ldots$$

such that $d^2 = 0$. The linear operator d is called the *differential operator* of the complex C.

One says that the elements of $Z^q(C) := (\ker d) \cap C^q$ are *closed*, while those of $B^q(C) := (\operatorname{im} d) \cap C^q$ are *exact*. [In certain contexts, the elements of $Z^q(C)$ are also referred to as q-*co-cycles*, those of $B^q(C)$ as q-*co-boundaries*.]

The *cohomology* of C is the direct sum of vector spaces $H(C) = \bigoplus_{q\in\mathbb{Z}} H^q(C)$ defined by

$$H^q(C) = Z^q(C)/B^q(C).$$

An *exact sequence* in this context is a differential complex (C, d) with vanishing cohomology, i.e., with the property that $Z^q(C) = B^q(C)$ for all q.

Reminder. A group action $X \times G \to X$, $(x, g) \mapsto xg$ on a topological space X induces an equivalence relation

$$x \sim x' \Leftrightarrow \exists\, g \in G:\; x' = xg$$

on X. Such a relation $\sim$ organizes the elements of X into equivalence classes, which we write as $[x] = [xg]$. The space of equivalence classes $[x]$ is denoted by X/G and referred to as the quotient of X by G. (It is a topological space by the so-called quotient topology.)

These general notions define what is meant by the quotient space $H^q(C) = Z^q(C)/B^q(C)$. Indeed, the vector space $Z^q(C)$ of closed elements is acted upon by the vector space $B^q(C)$ of exact elements (viewed as an abelian group), with the group action simply being the operation of vector addition. The equivalence class $[z] \in H^q(C)$ of a closed element $z \in Z^q(C)$ is called a cohomology class. By definition, $[z]$ is the set of all closed elements $z' \in Z^q(C)$ which differ from z by an exact element, i.e., $z' = z + b$ for some $b = d\beta \in B^q(C)$. Note that by the law of addition

$$[z] + [z'] := [z + z'],$$

the quotient of two vector spaces is still a vector space.

1.1.1 **Example:** $\Omega(\mathbb{R}^3)$

An example of a differential complex known to every physicist is the following. Let C^0, C^1, C^2, C^3 be the spaces of differentiable functions, vector fields, axial vector fields, and pseudoscalar functions, respectively, each with domain of definition $\mathbb{R}^3$. We set $C^q = 0$ for $q < 0$ and $q > 3$. Then $C = \bigoplus_{q\in\mathbb{Z}} C^q$ is a differential complex [and essentially the same thing as what is denoted conventionally by $\Omega(\mathbb{R}^3)$] with differential operator

$$0 \longrightarrow C^0 \xrightarrow{\text{grad}} C^1 \xrightarrow{\text{curl}} C^2 \xrightarrow{\text{div}} C^3 \longrightarrow 0 \tag{1.1}$$

because $\operatorname{curl} \circ \operatorname{grad} = 0$ and $\operatorname{div} \circ \operatorname{curl} = 0$. In this example one has the following cohomology:

1. $H^0 = \mathbb{R}$ because the only functions on $\mathbb{R}^3$ with vanishing gradient are the constant functions,

2. $H^1 = 0$ because every rotationless vector field in $\mathbb{R}^3$ is a gradient,

3. $H^2 = 0$ because every divergenceless axial vector field on $\mathbb{R}^3$ is a curl,

4. $H^3 = 0$ because every pseudoscalar function is the divergence of an axial vector field.

This cohomology gets more interesting when $\mathbb{R}^3$ is replaced by some domain $U \subset \mathbb{R}^3$. For example, if U consists of n connected components, then $H^0 = \mathbb{R}^n$ because we get to choose a constant value for our function on each connected component of U separately.

Let x, y, z be the coordinate functions associated with a Cartesian basis e_x, e_y, e_z of $\mathbb{R}^3$. If $U = \mathbb{R}^3 \setminus z$-axis, then $H^1 = \mathbb{R}$. In fact, every rotationless vector field on $\mathbb{R}^3 \setminus z$-axis is, modulo gradients, some multiple of

$$v = \frac{x e_y - y e_x}{x^2 + y^2} \,.$$

If we remove from $\mathbb{R}^3$ a single point, say the origin: $U = \mathbb{R}^3 \setminus \{o\}$, then the second cohomology becomes non-trivial, $H^2 = \mathbb{R}$, with generator

$$v = \frac{x e_x + y e_y + z e_z}{\sqrt{x^2 + y^2 + z^2}^{\,3}}$$

(viewed as an axial vector field). In other words, every divergenceless vector field on $\mathbb{R}^3 \setminus \{o\}$ is some multiple of v modulo curls.

Remark. If it worries you that you can't verify the claims made above, please be patient. It is the very purpose of the present chapter to develop the mathematical tools needed to compute this type of cohomology (which is called the de Rham cohomology).

1.1.2 **Example:** $\Omega_c(\mathbb{R}^3)$

Recycling the previous example, let us change the rules of the game slightly and require that all our differentiable functions and vector fields on $\mathbb{R}^3$ are *compactly supported*, i.e., vanish outside a finite and closed domain. The zeroth cohomology then becomes trivial, $H^0 = 0$. Indeed, there exists no compactly supported constant function on $\mathbb{R}^3$ other than the zero function.

The first and second cohomologies remain trivial ($H^1 = H^2 = 0$). However, the third cohomology $H^3 = Z^3/B^3$ now is non-trivial. In fact, $Z^3 \equiv C^3$ [since the last map of the differential complex (1.1) is the zero map] and B^3 with compact supports is strictly smaller than Z^3. To verify the last fact, note that the integral of a compactly supported function $f = \operatorname{div} v \in B^3$ vanishes:

$$\iiint f dx dy dz = \iint v \cdot d^2 n = 0 \,,$$

by Gauss' theorem. On the other hand, there certainly exist compactly supported C^∞ functions $f \in Z^3$ with non-zero integral.

Problem. Show that two compactly supported functions f and g differ by a divergence ($f - g = \operatorname{div} u$) if and only if they have the same integral, $\int_{\mathbb{R}^3} f\, dxdydz = \int_{\mathbb{R}^3} g\, dxdydz$. $\square$

By using the solution of the problem, one immediately sees that $H^3 = \mathbb{R}$.

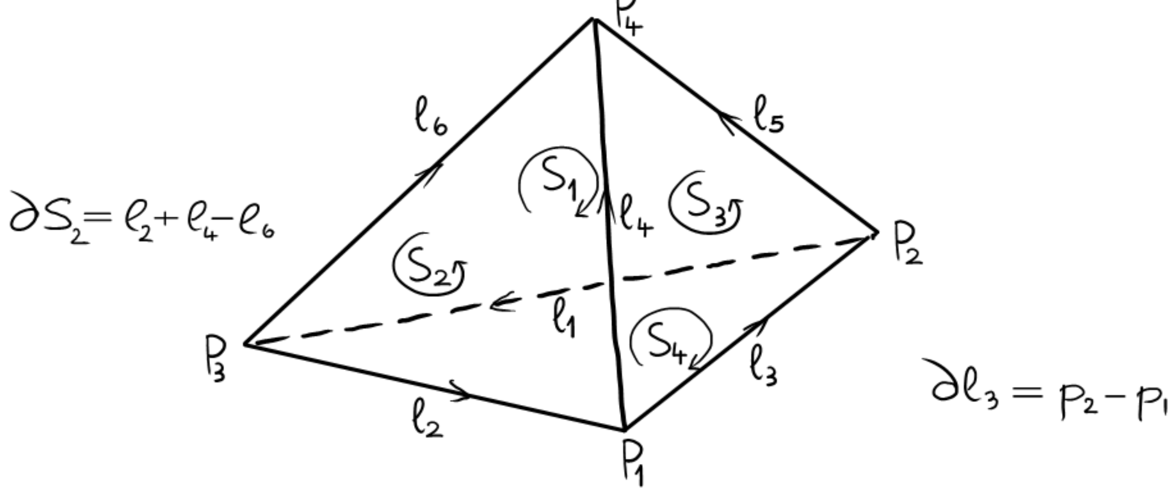
Remark. Let us make the following observation: in the case at hand, the cohomology H^q with compact supports is the same as the cohomology H^{3-q} without compact supports. This is no accident but reflects a principle known as *Poincaré duality.* We will meet the general statement of Poincaré duality later in the chapter.

1.1.3 Example: homology of a tetrahedron

The next class of example is of a combinatorial nature (and belongs to the realm of what's called homology). For simplicity let us consider the concrete situation of (the surface of) a tetrahedron. We associate with it a differential complex as follows.

A tetrahedron consists of four 0-cells (these are the vertices, or corners, or sites of the tetrahedron), six 1-cells (the edges, or links), and four 2-cells (the faces). Formal linear combinations of q-cells with real coefficients are called *q-chains.* They can be added and multiplied by scalars and thus form a vector space. The vector space of q-chains is denoted by C_q .

We now assign (in an arbitrary way) a sense of direction to each 1-cell and a sense of circulation to each 2-cell, see the figure below.



Then we have a boundary operator $\partial : \ C_q \to C_{q-1}$ which is defined in the following natural way. The boundary of a 0-chain always vanishes by decree. The boundary of a 1-cell is the 0-chain made from the end point with coefficient $+1$ and the starting point with coefficient -1. (This already defines $\partial : \ C_1 \to C_0$ by linear extension.) The boundary of a 2-cell S of the tetrahedron is the 1-chain $\partial S = \pm \ell_{i_1} \pm \ell_{i_2} \pm \ell_{i_3}$ made from the three 1-cells $\ell_{i_1}, \ell_{i_2}, \ell_{i_3}$ in its boundary, where the coefficient of ℓ_i in ∂S is $+1$ (-1) if the sense of direction of ℓ_i agrees (disagrees) with the sense of circulation of S. (Again, this already defines the boundary operator $\partial : \ C_2 \to C_1$.)

Problem. Show that this definition of boundary operator $\partial : \ C_q \to C_{q-1}$ satisfies $\partial^2 = 0$. $\square$

Thus our boundary operator ∂ has the property $\partial^2 = 0$ of a differential operator. There is, however, a slight difference: ∂ lowers the degree q whereas the definition above wants the differential operator to increase the degree. This can be repaired by letting $C^{-q} := C_q$ for $q = 0, 1, 2$ (and of course, $C^q = 0$ for $q < -2$ and $q > 0$). Alternatively, one may dualize the situation by defining $C^q := C_q^*$ (dual vector space) and taking the differential operator $d: C^q \to C^{q+1}$ to be the so-called *co-boundary operator*, i.e., the transpose of ∂.

Problem. Compute the (co)homology $H(C)$ of the differential complex $C = C^{-2} \oplus C^{-1} \oplus C^0$ of the tetrahedron with differential operator $d = \partial$.

1.1.4 Example: de Rham complex on $\mathbb{R}^n$

Let $x_1, x_2, \ldots, x_n$ be the standard linear coordinates of $\mathbb{R}^n$. By Ω we denote the exterior algebra (or Grassmann algebra) over $\mathbb{R}$ generated by the differentials $dx_1, dx_2, \ldots, dx_n$ with relations

$$dx_i\, dx_j = -dx_j\, dx_i\,.$$

(Note that $dx^i dx^i = 0$.) Ω is graded by $\Omega = \bigoplus_{q=0}^n \Omega^q$ where the vector space Ω^q has the basis

$$dx_{i_1}\, dx_{i_2} \cdots dx_{i_q} \quad (i_1 < i_2 < \ldots < i_q).$$

Note $\dim \Omega^q = \binom{n}{q}$ and $\dim \Omega = 2^n$. The C^∞ *differential forms* are the elements of

$$\Omega(\mathbb{R}^n) := \mathsf{C}^\infty(\mathbb{R}^n) \bigotimes_{\mathbb{R}} \Omega\,.$$

The algebra $\Omega(\mathbb{R}^n)$ inherits from Ω a grading $\Omega(\mathbb{R}^n) = \bigoplus_{q=0}^n \Omega^q(\mathbb{R}^n)$ by

$$\Omega^q(\mathbb{R}^n) := \mathsf{C}^\infty(\mathbb{R}^n) \bigotimes_{\mathbb{R}} \Omega^q.$$

The elements of $\Omega^q(\mathbb{R}^n)$ are called C^∞ differential forms of degree q, or q-forms for short. Thus a q-form is a sum of terms each of which is the product of a C^∞ function with an element of Ω^q.

There exists a differential operator called the *exterior derivative*,

$$d: \Omega^q(\mathbb{R}^n) \to \Omega^{q+1}(\mathbb{R}^n),$$

which is defined as follows. If $f \in \Omega^0(\mathbb{R}^n) \equiv \mathsf{C}^\infty(\mathbb{R}^n)$ then df is simply the differential:

$$df = \sum_i \frac{\partial f}{\partial x_i}\, dx_i\,.$$

If $\omega \in \Omega^q(\mathbb{R}^n)$ then

$$d\omega = d \sum \omega_{i_1 \ldots i_q}\, dx_{i_1} \cdots dx_{i_q} = \sum d\omega_{i_1 \ldots i_q}\, dx_{i_1} \cdots dx_{i_q}\,.$$

Problem. Show that d is an *anti-derivation*, i.e.,

$$d(\xi\, \eta) = (d\xi)\, \eta + (-1)^{\deg(\xi)} \xi\, d\eta.$$

Problem. Show that $d^2 = 0$. $\square$

The complex $\Omega(\mathbb{R}^n)$ together with the differential operator d is called the *de Rham complex* on $\mathbb{R}^n$. The kernel of d are the *closed forms*, the image of d the *exact forms*.

Definition. The *q-th de Rham cohomology* of $\mathbb{R}^n$ is the vector space

$$H^q_{\mathrm{dR}}(\mathbb{R}^n) = \{\text{closed } q\text{-forms on } \mathbb{R}^n\}/\{\text{exact } q\text{-forms on } \mathbb{R}^n\}.$$

Remark. All of the above makes sense for any open domain $U \subset \mathbb{R}^n$. Thus in the same vein we have $\Omega^q(U)$ and $H^q_{\mathrm{dR}}(U)$.

1.1.5 Snake lemma

Let A, B be a pair of differential complexes, with differential operators $d_A : A^q \to A^{q+1}$, $d_A^2 = 0$, and $d_B : B^q \to B^{q+1}$, $d_B^2 = 0$. A degree-preserving linear mapping $f : A \to B$ is called a *chain map* if it commutes with the differential operators of A and B:

$$d_B\, f = f\, d_A\,.$$

Fact. The map $f : A \to B$ descends to a map in cohomology:

$$f_* : H^q(A) \to H^q(B)\,.$$

Indeed, being a chain map f maps $\ker d_A$ into $\ker d_B$, and $\operatorname{im} d_A$ into $\operatorname{im} d_B$, so we can define the image under f of a cohomology class $a + d(A^{q-1}) \equiv [a] \in H^q(A)$ by

$$f_*([a]) := f(a) + d(B^{q-1}) \equiv [f(a)]. \quad \square$$

Example/Problem. Let $\phi : M \to N$ be a differentiable map between two manifolds. Then there is a map $\phi^* : \Omega(N) \to \Omega(M)$ called the *pullback* by ϕ. Look up the precise definition of pullback, and show that pullback is a chain map. $\square$

Next comes a quick reminder of something basic from linear algebra. We recall that a differential complex with vanishing cohomology is called an exact sequence. An exact sequence of three vector spaces A, B, C,

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0\,, \tag{1.2}$$

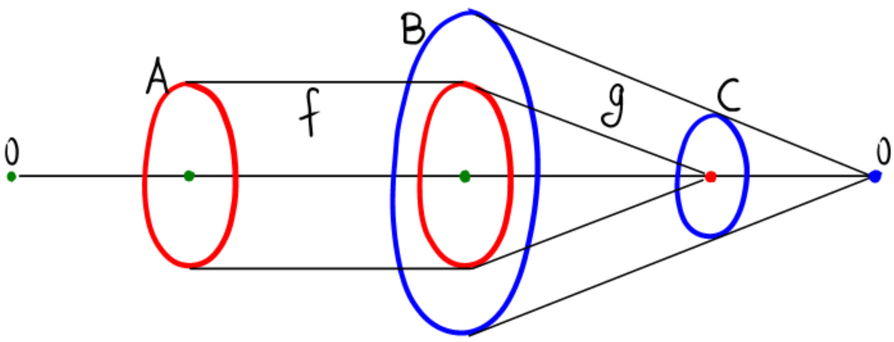
is called a *short exact sequence.* In this case the following properties are immediate:

$$\ker f = 0\,, \quad \operatorname{im} f = \ker g\,, \quad \operatorname{im} g = C\,.$$

Thus f is injective, g is surjective, and $g \circ f = 0$. It follows that the induced mapping

$$B/f(A) \to C\,, \quad b + \operatorname{im} f \mapsto g(b)\,,$$

is an isomorphism.



Next, consider a short exact sequence (1.2) of differential complexes A, B, C, with the additional property that the operators f and g are chain maps:

$$
\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
\ldots \arrow[r] & A^{q-1} \arrow[r, "d_A"] \arrow[d, "f"] & A^q \arrow[r, "d_A"] \arrow[d, "f"] & A^{q+1} \arrow[r] \arrow[d, "f"] & \ldots \\
\ldots \arrow[r] & B^{q-1} \arrow[r, "d_B"] \arrow[d, "g"] & B^q \arrow[r, "d_B"] \arrow[d, "g"] & B^{q+1} \arrow[r] \arrow[d, "g"] & \ldots \\
\ldots \arrow[r] & C^{q-1} \arrow[r, "d_C"] \arrow[d] & C^q \arrow[r, "d_C"] \arrow[d] & C^{q+1} \arrow[r] \arrow[d] & \ldots \\
 & 0 & 0 & 0 &
\end{tikzcd}
$$

Then, as we already know, the maps $f: A \to B$ and $g: B \to C$ descend to maps in cohomology:

$$f_* : H^q(A) \to H^q(B), \quad g_* : H^q(B) \to H^q(C).$$

What is less obvious is that one also has a canonical mapping

$$d_* : H^q(C) \to H^{q+1}(A).$$

This map d_* is defined as follows. [We now drop the subscript from d_A, d_B, and d_C. Which of these differential operators is meant will always be clear from the context.] Let $c \in C^q$. Then, since $g: B^q \to C^q$ is surjective there exists some $b \in B^q$ such that $g(b) = c$. If c is closed (i.e., $c \in \ker d$) then

$$g(db) = d\,g(b) = dc = 0\,,$$

and owing to $\ker g = \operatorname{im} f$ there exists some $a \in A^{q+1}$ such that $f(a) = db$. This element a is closed since f is injective and

$$f(da) = d\,f(a) = d^2 b = 0\,.$$

Thus for $[c] \in H^q(C)$ we tentatively make the assignment $[c] \stackrel{d_*}{\mapsto} [a] \in H^{q+1}(A)$.

To see that this defines an operator $d_* : H^q(C) \to H^{q+1}(A)$ as desired, we must check that the result $[a]$ does not depend on the choices made. Hence let c be replaced by another representative $c + d\gamma$ of $[c] \in H^q(C)$. Because g is surjective in degree $q-1$, there exists $\beta \in B^{q-1}$ such that

$$c + d\gamma = c + d\,g(\beta) = g(b + d\beta).$$

Thus b gets replaced by $b + d\beta$, which is a substitution that leaves db unchanged. Now, what happens if we replace b by $b + b_1$ with $\ker g \ni b_1 = f(a_1)$? In this case we get

$$d(b + b_1) = f(a) + d\,f(\alpha_1) = f(a + d\alpha_1)\,,$$

which again yields an unchanged result $[a + d\alpha_1] = [a]$ in cohomology. Thus our map $d_* : H^q(C) \to H^{q+1}(A)$ is indeed well-defined.

Problem ('snake lemma'). Prove the exactness of the long sequence

$$\ldots \xrightarrow{g_*} H^{q-1}(C) \xrightarrow{d_*} H^q(A) \xrightarrow{f_*} H^q(B) \xrightarrow{g_*} H^q(C) \xrightarrow{d_*} H^{q+1}(A) \xrightarrow{f_*} \ldots \quad \square$$

1.2 Mayer-Vietoris sequence

In Section 1.1.4 we introduced the de Rham complex of $\mathbb{R}^n$. Its differential operator, the exterior derivative d, has the important property of being independent of the chosen coordinate system. This property allows one to define the de Rham complex $\Omega(M)$ for any differentiable manifold M. To that end, one covers the manifold M by an atlas $\{U_\alpha\}$ of open subsets (or domains) U_α each of which is diffeomorphic to $\mathbb{R}^n$ (for $n = \dim M$). A differential form ω on M then is a collection of forms ω_U for U in the atlas of M such that $\omega_{U_\alpha}\big|_{U_\alpha \cap U_\beta} = \omega_{U_\beta}\big|_{U_\alpha \cap U_\beta}$ for every non-zero intersection $U_\alpha \cap U_\beta$. On each domain U_α one defines the exterior derivative as before (Section 1.1.4). By the coordinate-independence of the exterior derivative, these pieces of exterior derivative piece together to give a globally defined exterior derivative $d : \ \Omega^q(M) \to \Omega^{q+1}(M)$. The de Rham cohomology $H^q(M)$ is still the quotient of the closed q-forms on M by the exact q-forms.

The Mayer-Vietoris sequence is a powerful tool, which will let us understand a number of facts about $H^q(M)$. To introduce it, let $M = U \cup V$ with two open domains U and V, and let $U \cap V \neq \emptyset$ be their intersection. Defining the inclusion maps

$$i : \ U \to M, \quad j : \ V \to M, \quad k : \ U \cap V \to U, \quad l : \ U \cap V \to V,$$

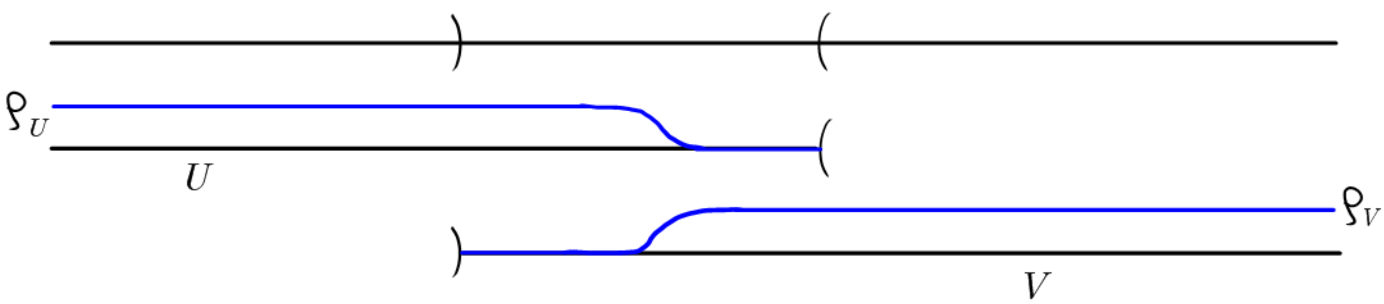
we claim that these give rise to a short exact sequence by the corresponding pullbacks:

$$0 \longrightarrow \Omega(M) \stackrel{(i^*, j^*)}{\longrightarrow} \Omega(U) \oplus \Omega(V) \stackrel{k^* - l^*}{\longrightarrow} \Omega(U \cap V) \longrightarrow 0\,. \tag{1.3}$$

Thus in the first non-trivial map of the sequence a differential form $\omega \in \Omega(M)$ is sent to its pair of restrictions (ω_U, ω_V) to U and V. In the second map one sends a pair $(\alpha, \beta) \in \Omega(U) \oplus \Omega(V)$ to the difference $\alpha - \beta$.

It is clear that (i^*, j^*) is injective, and its image equals the kernel of $k^* - l^*$. To prove the exactness of the sequence, it remains to show that $k^* - l^*$ is surjective. So, let $\omega \in \Omega(U \cap V)$. We need to find $\alpha \in \Omega(U)$ and $\beta \in \Omega(V)$ such that $\alpha - \beta$ agrees with ω on the intersection $U \cap V$. For this purpose let $\rho_U + \rho_V = 1$ be a partition of unity by smooth functions such that

$$\operatorname{supp} \rho_U \subset U\,, \quad \operatorname{supp} \rho_V \subset V.$$



Then $\alpha := \rho_V\,\omega$ is a form on U, $\beta := -\rho_U\,\omega$ is a form on V, and $\alpha - \beta = \rho_V\,\omega + \rho_U\,\omega = \omega$, as desired. Thus the map $k^* - l^*$ is indeed surjective.

Since the Mayer-Vietoris sequence (1.3) is exact, and the maps $(i^*,\, j^*)$ and $k^* - l^*$ (being pullbacks by inclusion maps) are chain maps, we get a long exact sequence (also called the Mayer-

Vietoris sequence)

$$\ldots \longrightarrow H^q(M) \xrightarrow{(i^*,j^*)} H^q(U) \oplus H^q(V) \xrightarrow{k^*-l^*} H^q(U \cap V) \xrightarrow{d_*} H^{q+1}(M) \longrightarrow \ldots$$

by the snake lemma of Section 1.1.5. Let us recall the definition of the operator d_* in the present context. Given a cohomology class $[\omega] \in H^q(U \cap V)$ we pick a representative ω and a pre-image $(\rho_V\,\omega, -\rho_U\,\omega)$ under the surjective map $k^* - l^*$. We then apply the exterior derivative, resulting in

$$d\,(\rho_V\,\omega, -\rho_U\,\omega) = (d\rho_V \wedge \omega, -d\rho_U \wedge \omega) \in H^{q+1}(U) \oplus H^{q+1}(V),$$

since $d\,\omega = 0$. Now owing to $\rho_U + \rho_V = 1$ we have $d\rho_U = -d\rho_V$, and therefore $d\rho_V \wedge \omega = -d\rho_U \wedge \omega$ makes sense as a form on $M = U \cup V$. One thus defines $d_* :\ H^q(U \cap V) \to H^{q+1}(M)$ by

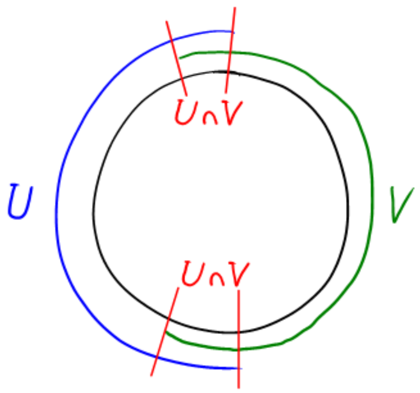
$$d_*[\omega] := [d\rho_V \wedge \omega] = [-d\rho_U \wedge \omega].$$

Note that the support of $d_*[\omega]$ is contained in $U \cap V$.

Problem. Why is $d_*[\omega]$ independent of the choice of partition of unity?

1.2.1 Example

As a simple application of the Mayer-Vietoris long exact sequence, we now use it to compute the de Rham cohomology of the circle S^1. To do so, we cover S^1 by two open domains U and V as shown in the next figure.



We then have the following long exact sequence:

$$\begin{tikzcd}
 & H^1(\mathrm{S}^1) \arrow[r] & H^1(U) \oplus H^1(V) \arrow[r] & H^1(U \cap V) \arrow[r] & 0 \\
0 \arrow[r] & H^0(\mathrm{S}^1) \arrow[r] & H^0(U) \oplus H^0(V) \arrow[r] & H^0(U \cap V) \arrow[ull, "d_*"']
\end{tikzcd}$$

We now recall that $\dim H^0(M)$ counts the number of connected components of M. Hence $H^0(U) = H^0(V) = \mathbb{R}$ and $H^0(U \cap V) = \mathbb{R}^2$. Moreover, we have $H^1(U) = H^1(V) = H^1(U \cap V) = 0$, since every 1-form in one dimension is exact. Thus our long exact sequence reads more explicitly like this:

$$\begin{tikzcd}
 & H^1(\mathrm{S}^1) \arrow[r] & 0 \arrow[r] & 0 \\
0 \arrow[r] & \mathbb{R} \arrow[r] & \mathbb{R} \oplus \mathbb{R} \arrow[r] & \mathbb{R} \oplus \mathbb{R} \arrow[ull, "d_*"']
\end{tikzcd}$$

We have that $d_* :\ H^0(U \cap V) \to H^1(\mathrm{S}^1)$ is surjective and that the image of the difference map $k^* - l^*$ is one-dimensional. By using $\mathrm{im}(k^* - l^*) = \ker d_*$ it follows that

$$H^1(\mathrm{S}^1) = \mathrm{im}\, d_*\big|_{H^0(U \cap V)} = H^0(U \cap V)/\ker d_* = \mathbb{R}^2/\mathbb{R} = \mathbb{R}\,.$$

1.2.2 Compact supports

The Mayer-Vietoris sequence has an analog in the setting of differential forms with compact supports. Turning to this case, we first observe that the story has to be run in a somewhat different way as the pullback of a compactly supported form need not be compactly supported in general (unless the maps used for pullback are proper). In fact, the good notion to use here turns out to be that of 'push forward'.

Let $M = U \cup V$ with U, V open as before, and consider the sequence of maps

$$0 \longleftarrow \Omega_c(M) \overset{i_*+j_*}{\longleftarrow} \Omega_c(U) \oplus \Omega_c(V) \overset{(k_*,-l_*)}{\longleftarrow} \Omega_c(U \cap V) \longleftarrow 0\,, \tag{1.4}$$

where the arrows now run from the right to the left and, for example, $i_* :\ \Omega_c(U) \to \Omega_c(M)$ is the mapping which extends a compactly supported form on U by zero to a form (still compactly supported) on M. The same goes for the other maps.

The short sequence (1.4) is still exact, and this time the exactness of the sequence is easy to see at every step. In particular, the last step is simpler than before. Indeed, for $\omega \in \Omega_c(M)$ consider the pair of forms $(\rho_U\, \omega, \rho_V\, \omega)$. Both of them are compactly supported, and their sum equals ω. We thus see that the map $\Omega_c(M) \overset{i_*+j_*}{\longleftarrow} \Omega_c(U) \oplus \Omega_c(V)$ is surjective.

The maps $i_* + j_*$ and $(k_*, -l_*)$ commute with the exterior derivative and hence are chain maps. Therefore, by the general principle of Section 1.1.5 we obtain another long exact sequence:

$$\ldots \longleftarrow H_c^{q+1}(U \cap V) \overset{d_*}{\longleftarrow} H_c^q(M) \overset{i_*+j_*}{\longleftarrow} H_c^q(U) \oplus H_c^q(V) \overset{(k_*,-l_*)}{\longleftarrow} H_c^q(U \cap V) \longleftarrow \ldots$$

which is called the Mayer-Vietoris sequence for compact supports. (Please be warned that, in order to avoid an overload of notation, we here refrain from inventing new symbols to denote the induced maps in cohomology.)

Let us again look briefly at how the operator d_* works in the present setting. We choose some partition of unity, $\rho_U + \rho_V = 1$, as before. For $\omega \in \Omega_c^q(M)$ we then form $(\rho_U\, \omega, \rho_V\, \omega) \in \Omega_c^q(U) \oplus \Omega_c^q(V)$ and take the exterior derivative,

$$d\,(\rho_U\, \omega, \rho_V\, \omega) = (d\rho_U \wedge \omega, d\rho_V \wedge \omega).$$

The operator $d_* :\ H_c^q(M) \to H_c^{q+1}(U \cap V)$ is then defined by

$$d_*[\omega] := [d\rho_U \wedge \omega] = [-d\rho_V \wedge \omega].$$

1.2.3 Example

As another simple application, let us recompute the de Rham cohomology of the circle, S^1, by using the long Mayer-Vietoris sequence with compact supports:

$$\begin{tikzcd}
0 & H_c^1(\mathrm{S}^1) \arrow[l] & H_c^1(U) \oplus H_c^1(V) \arrow[l] & H_c^1(U \cap V) \arrow[l] & \\
& H_c^0(\mathrm{S}^1) \arrow[urr, "d_*"] & H_c^0(U) \oplus H_c^0(V) \arrow[l] & H_c^0(U \cap V) \arrow[l] & 0\,. \arrow[l]
\end{tikzcd}$$

We have $H^0_c(U) = 0$ (recall that on an open U set there exist no compactly supported functions that have zero differential but aren't zero) and $H^1_c(U) = \mathbb{R}$ (because the total integral of a 1-form in one dimension is an obstruction to that 1-form being the differential of a compactly supported function). By filling in these cohomologies we obtain

$$\begin{tikzcd}
0 & H^1_c(\mathrm{S}^1) \arrow[l] & \mathbb{R}\oplus\mathbb{R} \arrow[l, "i_*+j_*"'] & \mathbb{R}\oplus\mathbb{R} \arrow[l, "(k_*,-l_*)"'] \\
 & H^0_c(\mathrm{S}^1) \arrow[urr, "d_*"] & 0 \arrow[l] & 0 \,. \arrow[l]
\end{tikzcd}$$

To compute $H^1_c(\mathrm{S}^1)$ we observe that the map $i_* + j_* :\ H^1_c(U) \oplus H^1_c(V) \to H^1_c(\mathrm{S}^1)$ is surjective. Thus

$$H^1_c(\mathrm{S}^1) = \mathrm{im}\,(i_* + j_*) = (H^1_c(U) \oplus H^1_c(V))/\ker\,(i_* + j_*).$$

Now $\ker\,(i_* + j_*)$ is one-dimensional, being generated by $([\alpha], [\beta]) \in H^1_c(U) \oplus H^1_c(V)$ subject to $\int_U \alpha + \int_V \beta = 0$. Hence

$$H^1_c(\mathrm{S}^1) = (\mathbb{R} \oplus \mathbb{R})/\mathbb{R} = \mathbb{R}\,.$$

Turning to $H^0_c(\mathrm{S}^1)$ we observe that $d_* :\ H^0_c(\mathrm{S}^1) \to H^1_c(U \cap V)$ is injective. Therefore

$$H^0_c(\mathrm{S}^1) \simeq \mathrm{im}\, d_* = \ker\,(k_*, -l_*)\,,$$

by the exactness of the sequence. The kernel of $(k_*, -l_*)$ has dimension one; it is generated by forms on $U \cap V$ whose total integral vanishes while the integral over either one of the two components of $U \cap V$ is non-zero. We thus conclude that $H^0_c(\mathrm{S}^1) = \mathbb{R}$.

In summary, it makes no difference for de Rham cohomology of S^1 whether we require compact supports or not:

$$H^0(\mathrm{S}^1) = H^0_c(\mathrm{S}^1) = \mathbb{R}\,, \quad H^1(\mathrm{S}^1) = H^1_c(\mathrm{S}^1) = \mathbb{R}\,.$$

This is no accident. In fact, one has $H^q_c(M) \equiv H^q(M)$ (for all q) whenever M is compact.

1.3 Poincaré duality

PD will be seen to be a fundamental result with many applications.

1.3.1 Integration of forms

It would not be appropriate here to give a tutorial in exterior calculus and integration of differential forms. [A good reference is the classical mechanics book of Arnold.] We will just look at a few cases to communicate to the unknowing reader that integration of differential forms is a very natural and easy process.

Let $E \in \Omega^1(\mathbb{R}^n)$ be a 1-form. Its integral along a (differentiable) curve γ: $[0,1] \to \mathbb{R}^n$ is

$$\int_\gamma E := \int_0^1 E_{\gamma(t)}(\gamma'(t))\, dt\,.$$

where $\gamma'(t) = \frac{d}{dt}\gamma(t)$. This definition is coordinate-independent and, in fact, invariant under reparametrization of the curve. If $x_1, x_2, \ldots, x_n$ are the standard coordinates of $\mathbb{R}^n$ (actually, any coordinate system will do for present purposes) and $E = \sum E_i \, dx_i$, then the coordinate expression for the line integral $\int_\gamma E$ is

$$\int_\gamma E = \sum_{i=1}^n \int_0^1 E_i(\gamma(t)) \, \gamma'(t)_i \, dt \, .$$

We turn to the case of a 2-form $B \in \Omega^2(\mathbb{R}^n)$. To integrate it, we need a parametrized surface, say $\sigma : \; [0,1]^2 \to \mathbb{R}^n$. The integral of B over σ is

$$\int_\sigma B := \int_0^1 \!\! \int_0^1 B_{\sigma(s,t)} \left(\frac{\partial}{\partial s} \sigma(s,t), \frac{\partial}{\partial t} \sigma(s,t) \right) \, ds \, dt \, .$$

In coordinates we have $B = \sum_{i<j} B_{ij} \, dx_i \, dx_j$ and the integral is expressed by

$$\int_\sigma B = \sum_{i,j=1}^n \int_0^1 \!\! \int_0^1 B_{ij}\big(\sigma(s,t)\big) \, \big(\partial_s \, \sigma(s,t)\big)_i \, \big(\partial_t \, \sigma(s,t)\big)_j \, ds \, dt \, ,$$

where the convention $B_{ij} = -B_{j\,i}$ is assumed.

It should be clear how this continues to higher degree.

1.3.2 Poincaré lemma

The Poincaré lemma says that the de Rham cohomology of $\mathbb{R}^n$ is trivial except in degree zero:

$$H^q(\mathbb{R}^n) = \begin{cases} \mathbb{R} & q = 0 \, , \\ 0 & \text{else} \, . \end{cases}$$

When the condition of compact support is imposed, the non-trivial cohomology moves to the top degree:

$$H^q_c(\mathbb{R}^n) = \begin{cases} \mathbb{R} & q = n \, , \\ 0 & \text{else} \, . \end{cases}$$

Let us indicate how the first statement is proved. For this we fix any reference point of $\mathbb{R}^n$, say the origin o, and for $q \geq 1$ define an operator $K : \; \Omega^q(\mathbb{R}^n) \to \Omega^{q-1}(\mathbb{R}^n)$, $\omega \mapsto K\omega$, by

$$(K\omega)_p(v_2, \ldots, v_q) = \int_0^1 \omega_{o+t(p-o)}(p - o, v_2, \ldots, v_q) \, t^{q-1} dt \, .$$

Problem. Show that $(dK + Kd)\omega = \omega$. $\square$

From the identity stated in the problem one immediately concludes that every closed form ($d\omega = 0$) of degree $q \geq 1$ in $\mathbb{R}^n$ is exact: $\omega = (dK + Kd)\omega = d(K\omega)$.

1.3.3 The statement of Poincaré duality

We begin with a few definitions. Let $\{U_\alpha\}$ be an atlas for an n-dimensional manifold M. One calls $\{U_\alpha\}$ a *good cover* of M if all non-empty finite intersections $U_{\alpha_1} \cap U_{\alpha_2} \cap \ldots \cap U_{\alpha_q}$ are diffeomorphic to $\mathbb{R}^n$.

Example. Recall that in Section 1.2.1 we used two open intervals U, V to cover the circle S^1. This cover is *not* good, as the intersection $U \cap V$ consists of two connected components and thus is diffeomorphic to *two* copies of $\mathbb{R}$, not just one. However, it is easy to produce a good cover by using three domains U, V, and W. Indeed, we can arrange for each of $U \cap V$, $V \cap W$ and $W \cap U$ to be connected, with empty intersection $U \cap V \cap W = \emptyset$. $\square$

Problem. What's the minimal cardinality of a good cover for the sphere S^2 ? $\square$

It is a fact that every manifold has a good cover. A manifold is said to be of *finite type* if it has a good cover of finite cardinality. (For example, every compact manifold is of finite type.)

A manifold M is called *orientable* if there exists a top-degree form $\omega \in \Omega^n(M)$ $(n = \dim M)$ which is everywhere non-zero.

Theorem (Poincaré duality). For every orientable n-dimensional manifold M of finite type one has an isomorphism

$$H^q(M) \simeq \left(H_c^{n-q}(M)\right)^*.$$

The proof of Poincaré duality will be discussed in a later section.

Here we continue with a reformulation. For this we recall from linear algebra that if a vector space V is equipped with a non-degenerate bilinear form $Q: V \otimes V \to \mathbb{R}$, then the linear mapping $\widetilde{Q}: V \to V^*$ by $v \mapsto Q(v, \cdot)$ is an isomorphism. The converse is also true. Similarly, the existence of a non-degenerate pairing $P: V \otimes W \to \mathbb{R}$ between two vector spaces V and W is equivalent to the existence of an isomorphism $\widetilde{P}: V \to W^*$, $v \mapsto P(v, \cdot)$.

By applying this general principle to our situation, we see that Poincaré duality amounts to the existence of a non-degenerate pairing $H^q(M) \otimes H_c^{n-q}(M) \to \mathbb{R}$. This pairing is given by integration:

$$\int : H^q(M) \otimes H_c^{n-q}(M) \to \mathbb{R}, \quad [\alpha] \otimes [\beta] \mapsto \int_M \alpha \wedge \beta .$$

Let us spend a few words verifying that this pairing is well-defined. First of all, since the second factor β is compactly supported by decree, the integral $\int_M \alpha \wedge \beta$ always converges, regardless of whether M is compact or not. Second, the integral does not depend on the choice of representative for either cohomology class. Indeed, if α_1 and α_2 are two representative of the same class $[\alpha]$, then $\alpha_1 = \alpha_2 + d\eta$ and

$$\int_M \alpha_1 \wedge \beta - \int_M \alpha_2 \wedge \beta = \int_M d\eta \wedge \beta = \int_M d(\eta \wedge \beta) = 0,$$

by $d\beta = 0$ and Stokes' theorem.

Remark. Stokes' theorem is the statement $\int_M d\omega = \int_{\partial M} \omega$. $\square$

Thus Poincaré duality can be restated as follows.

Theorem (Poincaré duality). For every orientable manifold M of finite type the pairing

$$H^q(M) \otimes H_c^{n-q}(M) \to \mathbb{R} \qquad (n = \dim M)$$

by integration is non-degenerate.

1.3.4 The Poincaré dual of a submanifold

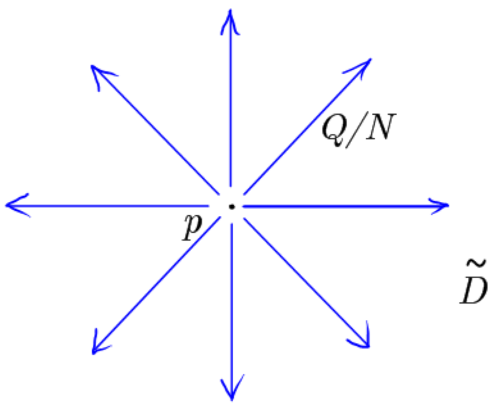
For a manifold M of dimension n, let $S \subset M$ be an oriented submanifold of dimension k. (Oriented means that there exists an everywhere non-vanishing k-form on S and one has fixed such a form.)

Also, let S be closed as a submanifold of M and denote by $\iota : S \to M$ the inclusion. If ω is any compactly supported k-form on M, then the integral $\int_S \iota^* \omega$ converges. What's more, integration along S descends to a linear functional on $[\omega] \in H_c^k(M)$ by Stokes' theorem. Thus we may regard S as defining an element of $\big(H_c^k(M)\big)^*$. By Poincaré duality $\big(H_c^k(M)\big)^* \simeq H^{n-k}(M)$ it follows that there exists a unique cohomology class, say $[\eta_S]$, in $H^{n-k}(M)$ such that

$$\int_S \iota^* \omega = \int_M \omega \wedge \eta_S$$

holds for every $[\omega] \in H_c^k(M)$. This form η_S (or rather its cohomology class $[\eta_S]$) is called the *closed* Poincaré dual of S. By the same token, one speaks of the closed Poincaré dual of a k-chain on M. (A k-chain on M is a linear combination of k-dimensional oriented submanifolds, actually k-cells, of M. In the present context all k-cells are required to be closed.)

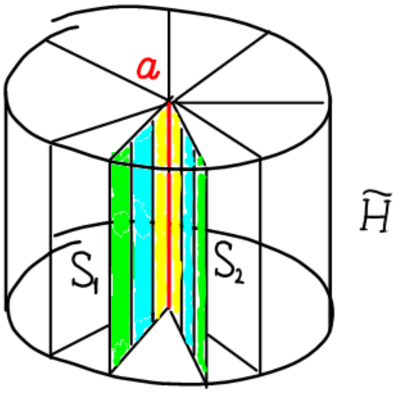
Example 1. Let $M = \mathbb{R}^3 \setminus \{p\}$ and consider the 1-chain $\widetilde{D} = (Q/N) \sum_{i=1}^N \gamma_i$ consisting of N rays γ_i each of which extends from the point p to infinity and carries 'electric flux' Q/N. Note that each such ray γ_i is closed as a submanifold of $\mathbb{R}^3 \setminus \{p\}$ (albeit not as a submanifold of $\mathbb{R}^3$).



Problem. Show that the Poincaré dual of the 1-chain $\widetilde{D}$ is the (cohomology class $[D] \in H^2(\mathbb{R}^3 \setminus \{p\})$ of the) closed 2-form $D = Q\,\tau_p$ where $\tau_p = \sin\theta_p \, d\theta_p \wedge d\phi_p$ is the solid-angle 2-form (expressed in spherical polar coordinates $\theta_p\,,\ \phi_p$) centered at p. $\square$

Remark. From the physics viewpoint, the closed 2-form $D = Q\,\tau_p$ should be interpreted as the electric excitation of a point charge Q at the position p. The benefit from Poincaré duality is that we may visualize this (perhaps somewhat abstract) electric-excitation 2-form D by the electric flux lines of the closed 1-chain $\widetilde{D}$. Of course, from the cohomological viewpoint it doesn't matter how we arrange the N rays; e.g., we might put them all on top of each other and consider a single ray from p to infinity. This flexibility stems from the fact that in cohomology one requires the equality $\int_{\widetilde{D}} \omega = \int_M \omega \wedge D$ to hold only for *closed* test forms ω (so the 'test' isn't very precise). On the other hand, in physics one might want such an equality to hold for *all* test forms ω. While that's too much to ask for, if we arrange the rays in a way guided by spherical symmetry then we do get a very good approximation $\int_{\widetilde{D}} \omega \approx \int_M \omega \wedge D$ by choosing N to be sufficiently large.

Example 2. Let $M = \mathbb{R}^3 \setminus a$ for some axis a, and consider the 2-chain $\widetilde{H} = (I/N)\sum_{i=1}^{N} S_i$ of a circular arrangement of N half planes S_i emanating from a and carrying 'magnetic voltage' I/N.



Problem. Show that the Poincaré dual of $\widetilde{H}$ is the closed 1-form $H = I\, d\vartheta_a$ where ϑ_a is the angular coordinate of a cylindrical coordinate system centered around a.

Remark. Again, in cohomology we could just use a single half-plane S and the Poincaré dual would still be $d\vartheta_a$. However, as physicists we prefer the circular arrangement (with a large number N of half planes) because then $\int_{\widetilde{H}} \omega = \int_M \omega \wedge H$ is not only true for closed test forms ω, but also holds approximately for non-closed ω. More concretely, our 2-chain $\widetilde{H}$ is a good approximation for the magnetic excitation 1-form H due to a stationary electric current I flowing along a. □

There exists a second notion of Poincaré dual, which is to be distinguished from the one above. Let now $\iota : S \to M$ be a compact submanifold (with $\dim S = k \leq n = \dim M$ as before). By the compactness of S the integral $\int_S \iota^*\omega$ makes sense for *any* k-form ω, no matter whether it has compact support or not. Integration along S again descends to a linear functional on $[\omega] \in H^k(M)$ by Stokes' theorem, and by Poincaré duality there is a unique cohomology class $[\eta'_S]$ in $H_c^{n-k}(M)$ such that

$$\int_S \iota^*\omega = \int_M \omega \wedge \eta'_S \quad \text{for all } [\omega] \in H^k(M)\,.$$

This cohomology class $[\eta'_S] \in H_c^{n-k}(M)$ is called the *compact* Poincaré dual of S.

Example 3. Let $\mathbb{R}^n$ be equipped with the Euclidean distance function $d(x,y) = |x-y|$ (that's just for our convenience), and consider some point $p \in \mathbb{R}^n$. The compact Poincaré dual of p is represented by a bump form ρ of mass one, say

$$\rho = f_{p,\epsilon}\, dx_1\, dx_2 \cdots dx_n\,, \qquad f_{p,\epsilon}(x) = \begin{cases} c_\epsilon\, \mathrm{e}^{-(\epsilon^2 - |x-p|^2)^{-1}} & |x-p| < \epsilon, \\ 0 & |x-p| \geq \epsilon, \end{cases}$$

where c_ϵ is a normalization constant ensuring that $\int \rho = 1$.

Remark. We have chosen a bump form which peaks at the point p. It should be emphasized that the whereabouts of the bump form don't matter at all in cohomology (as long as ρ is compactly supported and has mass one). Indeed, our Poincaré duality equation here reads $f(p) \equiv \int_p f = \int_{\mathbb{R}^n} f\,\rho$ and since $f \in H^0(\mathbb{R}^n)$ is a constant function, the equality holds if $\int \rho = 1$. Nevertheless, the choice of bump form localized at p is optimal in the sense that it achieves approximate equality even for the (non-cohomological) case of non-constant f. For example, we might regard ρ as the smooth charge density which results from 'smearing' a point charge at p.

Example 4. Let $M \subset \mathbb{R}^3$ be a spherical shell

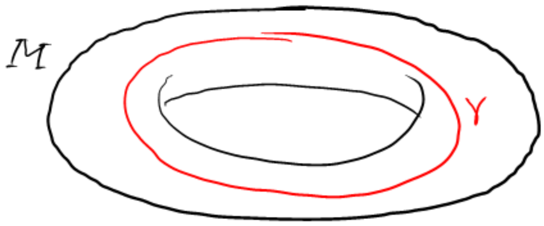
$$M = \{p \in \mathbb{R}^3 \mid R_1 < r(p) < R_2\}\,,$$

where r is the Euclidean distance from the origin, and take $S \subset M$ to be a sphere of radius R, $R_1 < R < R_2$. The compact Poincaré dual $[\eta'_S] \in H^1_c(M)$ then is the cohomology class of $\eta'_S = f(r)\,dr$ where $f(r)$ is any function with compact support in M and integral $\int_{R_1}^{R_2} f(r)\,dr = 1$. Speaking in physical terms, we may think of the radial electric field strength 1-form $E = V f(r) dr$ as the compact Poincaré dual of the 2-chain $\widetilde{E} := V \cdot S$ with electrical voltage V across S.

Example 5. Let $M \subset \mathbb{R}^3$ be a solid torus (or donut), described in cylindrical coordinates ρ, ϑ, z by

$$M := \left\{p \in \mathbb{R}^3 \mid (\rho(p) - R_1)^2 + z(p)^2 < R_2^2\right\},$$

and take $\gamma \in M$ to be the loop (or closed curve) which is the solution set of the equations $\rho = R_1$, $z = 0$. The compact Poincaré dual of γ is (the cohomology class of) a compactly supported closed 2-form $\eta'_\gamma = f(z, \rho)\,dz \wedge d\rho$ which integrates to unity along any cross section of the donut. We may think of $B \propto \eta'_S$ as the magnetic field strength 2-form due to an electric current circulating around the surface of the donut M.



1.3.5 Proof of Poincaré duality

The following lemma will be key to the proof of Poincaré duality.

Lemma ('Five Lemma'). Let there be two exact sequences and five linear maps $\alpha, \beta, \gamma, \delta, \varepsilon$ such that the following diagram commutes:

$$\begin{tikzcd}
\dots \arrow[r] & A \arrow[r, "f_1"] \arrow[d, "\alpha"'] & B \arrow[r, "f_2"] \arrow[d, "\beta"'] & C \arrow[r, "f_3"] \arrow[d, "\gamma"'] & D \arrow[r, "f_4"] \arrow[d, "\delta"'] & E \arrow[r] \arrow[d, "\varepsilon"'] & \dots \\
\dots \arrow[r] & A' \arrow[r, "f'_1"] & B' \arrow[r, "f'_2"] & C' \arrow[r, "f'_3"] & D' \arrow[r, "f'_4"] & E' \arrow[r] & \dots
\end{tikzcd}$$

Then if the maps $\alpha, \beta, \delta, \varepsilon$ are isomorphisms, so is the map γ.

Proof. By the commutativity of the diagram, the map β sends $\operatorname{im} f_1$ into $\operatorname{im} f'_1$, and the map γ sends $\operatorname{im} f_2$ into $\operatorname{im} f'_2$, and so on. Because β is an isomorphism it follows that $\gamma :\ \operatorname{im} f_2 \to \operatorname{im} f'_2$ is surjective, and because α is an isomorphism, the same goes for $\beta :\ \operatorname{im} f_1 \to \operatorname{im} f'_1$. Now the latter map must also be injective, or else $\beta :\ B \to B'$ would not be an isomorphism. Hence, $\operatorname{im} f_1$ is in bijection with $\operatorname{im} f'_1$ and we have

$$\ker f'_2 = \operatorname{im} f'_1 \simeq \operatorname{im} f_1 = \ker f_2\,,$$

by the exactness of both sequences. Owing to the rank-nullity theorem we obtain

$$\dim \operatorname{im} f_2' = \dim B' - \dim \ker f_2' = \dim B - \dim \ker f_2 = \dim \operatorname{im} f_2\,,$$

where $\dim B = \dim B'$ was used. Therefore, since $\gamma:\ \operatorname{im} f_2 \to \operatorname{im} f_2'$ is surjective, it must actually be bijective. This concludes the first part of the proof.

The second part of the proof begins with the observation that, again by the commutativity of the diagram, the map γ sends $\ker f_3$ into $\ker f_3'$. Thus γ pushes down to a map

$$\widetilde{\gamma}:\ C/\ker f_3 \to C'/\ker f_3'\,.$$

Because δ is an isomorphism it follows that $\widetilde{\gamma}$ is injective. We will now prove that $\widetilde{\gamma}$ is actually bijective, by showing that $C/\ker f_3$ and $C'/\ker f_3'$ have the same dimension.

By rank-nullity, $\dim\ C/\ker f_3 = \dim\ C'/\ker f_3'$ is equivalent to the statement that $\operatorname{im} f_3$ and $\operatorname{im} f_3'$ have the same dimension. The latter is true because $\operatorname{im} f_3 = \ker f_4$ and $\operatorname{im} f_3' = \ker f_4'$ by the exactness of the two sequences. Indeed, since δ and ε are isomorphisms, the kernels of the maps f_4 and f_4' have the same dimension. Thus $\widetilde{\gamma}$ is in fact bijective.

Altogether then, we have two isomorphisms

$$\gamma:\ \operatorname{im} f_2 \to \operatorname{im} f_2'\,, \qquad \widetilde{\gamma}:\ C/\operatorname{im} f_2 \to C'/\operatorname{im} f_2'\,,$$

where for the second one we used that $\operatorname{im} f_2 = \ker f_3$ and $\operatorname{im} f_2' = \ker f_3'$. Since C decomposes as $C \simeq \operatorname{im} f_2 \oplus (C/\operatorname{im} f_2)$ (and similar for C') it follows that $\gamma:\ C \to C'$ is an isomorphism. $\square$

For open sets U, V as before, we now pair the two Mayer-Vietoris sequences (with and without compact supports) to form the diagram

$$\begin{array}{ccccccc}
H^{q-1}(U\cap V) & \to & H^q(U\cup V) & \to & H^q(U)\oplus H^q(V) & \to & H^q(U\cap V) \\
\otimes & & \otimes & & \otimes & & \otimes \\
H_c^{n-q+1}(U\cap V) & \leftarrow & H_c^{n-q}(U\cup V) & \leftarrow & H_c^{n-q}(U)\oplus H_c^{n-q}(V) & \leftarrow & H_c^{n-q}(U\cap V) \\
\big\downarrow \int_{U\cap V} & & \big\downarrow \int_{U\cup V} & & \big\downarrow \int_U + \int_V & & \big\downarrow \int_{U\cap V} \\
\mathbb{R} & & \mathbb{R} & & \mathbb{R} & & \mathbb{R}
\end{array}$$

Lemma. The diagram above is (sign-)commutative.

Proof. Let us first show the commutativity of the middle square,

$$\begin{array}{ccc}
H^q(U\cup V) & \xrightarrow{(i^*,j^*)} & H^q(U)\oplus H^q(V) \\
\otimes & & \otimes \\
H_c^{n-q}(U\cup V) & \xleftarrow{i_*+j_*} & H_c^{n-q}(U)\oplus H_c^{n-q}(V) \\
\big\downarrow \int_{U\cup V} & & \big\downarrow \int_U + \int_V \\
\mathbb{R} & & \mathbb{R}\,.
\end{array}$$

Thus let $[\omega] \in H^q(U\cup V)$ and $([\alpha],[\beta]) \in H_c^{n-q}(U) \oplus H_c^{n-q}(V)$. Computing the pairing for the left row we have

$$[\omega]\otimes(i_*+j_*)([\alpha],[\beta]) \mapsto \int_{U\cup V} \omega\wedge(i_*\alpha + j_*\beta)\,,$$

and computing it for the right row

$$(i^*, j^*)([\omega]) \otimes ([\alpha], [\beta]) \mapsto \int_U i^*\omega \wedge \alpha + \int_V j^*\omega \wedge \beta\,.$$

The two pairings give the same value because $i^* i_*$ is the identity map and hence, e.g.,

$$\int_U i^*\omega \wedge \alpha = \int_U i^*(\omega \wedge i_*\alpha) = \int_{i(U)} \omega \wedge i_*\alpha = \int_{U\cup V} \omega \wedge i_*\alpha\,,$$

where the last equality holds because $i_*\alpha$ vanishes outside of U. The right square is very similar.

We turn to the left square:

$$\begin{array}{ccc} H^{q-1}(U\cap V) & \xrightarrow{d^*} & H^q(U\cup V) \\ \bigotimes & & \bigotimes \\ H_c^{n-q+1}(U\cap V) & \xleftarrow{d_*} & H_c^{n-q}(U\cup V) \\ \Big\downarrow \int_{U\cap V} & & \Big\downarrow \int_{U\cup V} \\ \mathbb{R} & & \mathbb{R}\,. \end{array}$$

Now let $[\xi] \in H^{q-1}(U\cap V)$ and $[\eta] \in H_c^{n-q}(U\cup V)$. We recall that $d^*[\xi] = [d\rho_V \wedge \xi] = [-d\rho_U \wedge \xi]$ (note the change of notation d^*) and $d_*[\eta] = [d\rho_U \wedge \eta] = [-d\rho_V \wedge \eta]$. The pairing on the left is

$$[\xi] \otimes d_*[\eta] \mapsto \int_{U\cap V} \xi \wedge d\rho_U \wedge \eta\,,$$

that on the right is

$$d^*[\xi] \otimes [\eta] \mapsto \int_{U\cup V} (-d\rho_U) \wedge \xi \wedge \eta = (-1)^q \int_{U\cup V} \xi \wedge d\rho_U \wedge \eta\,.$$

Because the support of $d\rho_U$ is localized in $U \cap V$ these agree but for a difference in sign. □

Remark. We can remove the sign difference, e.g., by redefining $d^* \to (-1)^q d^*$. Since the inversion of the sign of an operator changes neither its kernel nor its image, the Mayer-Vietoris sequence remains the same. □

Returning to our pair of Mayer-Vietoris sequences, we reverse the arrows in the second row to present the diagram in the following fashion:

$$\begin{array}{ccccccc} H^{q-1}(U\cap V) & \to & H^q(U\cup V) & \to & H^q(U)\oplus H^q(V) & \to & H^q(U\cap V) \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ H_c^{n-q+1}(U\cap V)^* & \to & H_c^{n-q}(U\cup V)^* & \to & H_c^{n-q}(U)^* \oplus H_c^{n-q}(V)^* & \to & H_c^{n-q}(U\cap V)^* \end{array}$$

The proof of Poincaré duality now proceeds by induction on the cardinality of a good cover. First of all, in the case of $M = \mathbb{R}^n$ Poincaré duality $H^q(\mathbb{R}^n) \simeq H_c^{n-q}(\mathbb{R}^n)^*$ holds as a result of the Poincaré lemma. Indeed, the only non-trivial cohomologies are $H^0(\mathbb{R}^n) = \mathbb{R} = H_c^n(\mathbb{R}^n)$ and the pairing $H^0(\mathbb{R}^n) \otimes H_c^n(\mathbb{R}^n) \to \mathbb{R}$ by $f \otimes \rho \mapsto \int f\rho$ is obviously non-degenerate.

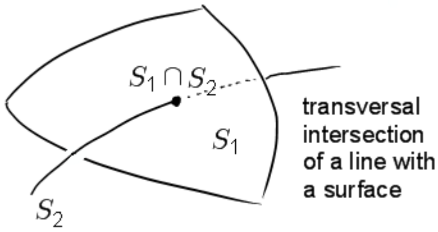
Next suppose that Poincaré duality holds for any manifold having a good cover with at most p open sets, and consider a manifold that has a good cover $\{U_0, \dots, U_p\}$ with $p+1$ open sets. Now $(U_0 \cup \dots \cup U_{p-1}) \cap U_p$ has a good cover with p open sets, namely $(U_0 \cap U_p, \dots, U_{p-1} \cap U_p)$. By hypothesis, Poincaré duality holds for U_p, for $U_0 \cup \dots \cup U_{p-1}$, and for $(U_0 \cup \dots \cup U_{p-1}) \cap U_p$. By applying the Five Lemma to the commutative diagram above, it then follows that Poincaré duality also holds for $U_0 \cup \dots \cup U_{q-1} \cup U_p$. This induction argument proves Poincaré duality for any orientable manifold that has a finite good cover.

1.3.6 Two properties of the Poincaré dual

Let us mention two geometric properties of Poincaré duality. For one thing, let $S_1 \subset M$ and $S_2 \subset M$ be two closed oriented submanifolds of dimension k_1 and k_2 respectively, and denote by $[\eta_{S_1}] \in H^{n-k_1}(M)$ and $[\eta_{S_2}] \in H^{n-k_2}(M)$ their Poincaré duals. The following statement gives a geometric interpretation of the operation of exterior multiplication (or wedge product).

Fact. If S_1 and S_2 intersect each other transversally, then

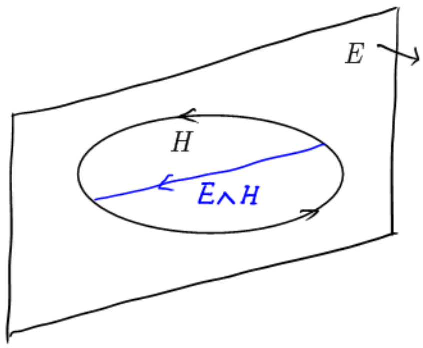
$$[\eta_{S_1 \cap S_2}] = [\eta_{S_1}] \wedge [\eta_{S_2}] \,.$$



Thus under Poincaré duality the (transversal) intersection of closed oriented submanifolds corresponds to the wedge product of forms.

Remark. The wedge product of two cohomology classes $[\alpha]$ and $[\beta]$ is defined to be $[\alpha] \wedge [\beta] := [\alpha \wedge \beta]$. This is a good definition as the wedge product of an exact and a closed form is always exact. Note also that $\dim(S_1 \cap S_2) = k_1 + k_2 - n$; indeed, there are n degrees of freedom and $(n - k_1) + (n - k_2)$ constraining equations. Thus $[\eta_{S_1}] \wedge [\eta_{S_2}] = [\eta_{S_1 \cap S_2}] \in H^{2n-k_1-k_2}(M)$ passes the test of counting dimensions.

Example. We illustrate the stated fact with an example from electrodynamics in $\mathbb{R}^3$. The energy current density of the electromagnetic field is the Poynting form $s = E \wedge H$ (traditionally called the Poynting vector), where E and H are the 1-forms of the electric field resp. magnetic excitation. In a static situation we may Poincaré-visualize the closed 1-forms E and H as closed 2-chains. The Poynting form $E \wedge H$ then is Poincaré dual to the closed 1-chain which is obtained by intersecting the 2-chains of E and H. Thus the lines of the electromagnetic energy current follow the lines of intersection of the surfaces of E and H. $\square$



The second property to be mentioned here is this.

Fact. Let $S \subset N$ be a closed oriented submanifold with Poincaré dual $[\eta_S]$. If $f^{-1}(S)$ denotes the pre-image of S under a mapping $f: M \to N$, then

$$[f^*\eta_S] = [\eta_{f^{-1}(S)}] \,,$$

i.e. under Poincaré duality the induced map on cohomology corresponds to the pre-image in geometry.

1.3.7 Künneth formula

Let us mention another useful property of de Rham cohomology which follows more or less directly from the Mayer-Vietoris sequence [for the proof see Bott & Tu].

Fact (Künneth formula). If M and N are manifolds of finite type, the de Rham cohomology of the direct product $M \times N$ is the tensor product of the de Rham cohomologies of the two factors:

$$H^r(M \times N) = \bigoplus_{p+q=r} H^p(M) \otimes H^q(N)\,.$$

The same formula holds in the case of the de Rham cohomology with compact supports.

1.3.8 Orientation line bundle

In this lecture course the theme of vector bundles will play a prominent role. (Actually, we have already been speaking about it, though not officially so). For present use with the introduction of twisted differential forms, we give the basic definitions right here.

Definition. Let $\pi : E \to M$ be a surjective map of manifolds whose fiber $\pi^{-1}(x) \equiv E_x$ is a real vector space for every $x \in M$. The map π is called a (smooth) *real vector bundle* of rank n if there exists an open cover $\{U_\alpha\}$ of M with fiber-preserving diffeomorphisms

$$\phi_\alpha : E\big|_{U_\alpha} = \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^n$$

such that $\phi_\alpha : E_x \to \mathbb{R}^n$ is a linear bijection for each $x \in U_\alpha$. The maps

$$\phi_\alpha \circ \phi_\beta^{-1} : (U_\alpha \cap U_\beta) \times \mathbb{R}^n \to (U_\alpha \cap U_\beta) \times \mathbb{R}^n, \quad (x, v) \mapsto (x, g_{\alpha,\beta}(x)v)\,,$$

are determined by so-called transition functions $g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(n, \mathbb{R})$. A vector bundle is called *flat* if there exists a trivialization $\{(U_\alpha, \phi_\alpha)\}$ such that all transition functions $g_{\alpha\beta}$ are constant. A *section* of the vector bundle $\pi : E \to M$ is a smooth map $s : M \to E$ with the property that $\pi \circ s$ is the identity map; thus $s(x) \in \pi^{-1}(x) = E_x$. The space of smooth sections of a vector bundle $E \to M$ is denoted by $\Gamma(E) \equiv \Gamma(M, E)$. $\square$

It is not possible in general to define an analog of the exterior derivative d on the sections of a vector bundle. (What's needed to differentiate sections is a so-called covariant derivative ∇.) However, if the vector bundle is flat, then d does make sense, as follows.

To define d on differential forms $\omega \in \Omega(M, E)$ with values in a flat vector bundle E, one fixes some basis $\{e^1, \ldots, e^n\}$ of $\mathbb{R}^n$. By using the trivialization maps $\phi_\alpha : E\big|_{U_\alpha} \to U_\alpha \times \mathbb{R}^n$ one introduces a basis of constant sections $e_\alpha^i = e_\alpha^i(x) = \phi_\alpha^{-1}(x, e^i)$ for every U_α. The exterior derivative $d\sigma$ of an E-valued differential form σ expressed on U_α as $\boldsymbol{\sigma}\big|_{U_\alpha} = \sum \sigma_i \otimes e_\alpha^i$ is then defined by

$$d\left(\sum_{i=1}^n \sigma_i \otimes e_\alpha^i\right) := \sum_{i=1}^n (d\sigma_i) \otimes e_\alpha^i\,.$$

We must check that this definition does not depend on the use of ϕ_α or ϕ_β on $U_\alpha \cap U_\beta$. Thus let $\sigma = \sum \sigma_i \otimes e_\alpha^i = \sum \tau_j \otimes e_\beta^j$. The coefficient functions are related by

$$\tau_j(x) = \sum c_{ij}\, \sigma_i(x) \qquad (x \in U_\alpha \cap U_\beta)\,,$$

where the coefficients c_{ij} are constants determined by

$$e_\alpha^i = \sum c_{ij}\, e_\beta^j\,.$$

(Since the transition functions of the flat vector bundle are constant by choice, so are the c_{ij}.) Now we can do our check:

$$
\begin{aligned}
d\left(\sum \tau_j \otimes e_\beta^j\right) &= \sum (d\tau_j) \otimes e_\beta^j = \sum_{i,j} (c_{ij}\, d\sigma_i) \otimes e_\beta^j \\
&= \sum_i (d\sigma_i) \otimes \sum_j c_{ij}\, e_\beta^j = \sum d\sigma_i \otimes e_\alpha^i = d\left(\sum \sigma_i \otimes e_\alpha^i\right).
\end{aligned}
$$

Thus we have $d|_{U_\alpha} = d|_{U_\beta}$ as required.

Problem. Show that the definition of the exterior derivative d is independent of the choice of (constant) trivialization $\{\phi_\alpha\}$. $\square$

By construction, the transition functions of a vector bundle satisfy the cocycle condition $g_{\alpha\beta}(x)\, g_{\beta\gamma}(x) = g_{\alpha\gamma}(x)$ on triple intersections $U_\alpha \cap U_\beta \cap U_\gamma \neq \emptyset$. Conversely, a cocycle $\{g_{\alpha\beta}\}$ with values $g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(n, \mathbb{R})$ determines a rank-n real vector bundle.

Let now $\{U_\alpha, \psi_\alpha\}$ be an atlas of coordinate maps $\psi_\alpha : U_\alpha \to \mathbb{R}^n$ for a manifold M and take

$$
x_\alpha^i : U_\alpha \xrightarrow{\psi_\alpha} \mathbb{R}^n \xrightarrow{x^i} \mathbb{R}
$$

to be the local coordinate functions given by standard coordinates $x^1, \ldots, x^n$ for $\mathbb{R}^n$. Then a top-dimensional form $\omega \in \Omega^n(M)$ is expressed by

$$
\omega = f_\alpha\, dx_\alpha^1 \wedge \cdots \wedge dx_\alpha^n = f_\beta\, dx_\beta^1 \wedge \cdots \wedge dx_\beta^n, \quad f_\beta = J_{\alpha\beta}\, f_\alpha, \quad J_{\alpha\beta} = \mathrm{Det}\left(\frac{\partial x_\alpha^i}{\partial x_\beta^j}\right),
$$

on any non-empty intersection $U_\alpha \cap U_\beta$. Note that by the multiplicativity of the determinant, the Jacobian $J_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathbb{R}$ satisfies the cocycle condition $J_{\alpha\beta} J_{\beta\gamma} = J_{\alpha\gamma}$ and so does the function $\mathrm{sign}(J_{\alpha\beta}) = J_{\alpha\beta}/|J_{\alpha\beta}|$. (The Jacobian $J_{\alpha\beta}$ never vanishes on $U_\alpha \cap U_\beta$.)

Definition. The *orientation line bundle* of a manifold M is the rank-1 real vector bundle $L \to M$ with transition functions $g_{\alpha\beta} = \mathrm{sign}(J_{\alpha\beta}) : U_\alpha \cap U_\beta \to \mathrm{GL}(1, \mathbb{R})$ (actually, O_1). $\square$

Thus an element of L is specified by a number $r \in \mathbb{R} \simeq \pi^{-1}(x)$ over a point $x \in M$. The number r depends on the coordinate chart $\phi_\alpha : U_\alpha \to \mathbb{R}$ used and changes sign when the orientation of the coordinate basis is reversed. Note that L is a flat vector bundle.

Fact. If a manifold M is orientable, then its orientation line bundle L is trivial, i.e. has a section with no zeroes.

Proof. Let M be orientable. Then by definition there exists a top form $\omega \in \Omega^n(M)$ whose local expressions $\omega|_{U_\alpha} = f_\alpha\, dx_\alpha^1 \wedge \cdots \wedge dx_\alpha^n$ have coefficients $f_\alpha : U_\alpha \to \mathbb{R}$ with no zeroes. Put $s_\alpha := f_\alpha/|f_\alpha|$. The transition rule $f_\beta = J_{\alpha\beta} f_\alpha$ implies the transition rule $s_\beta = \mathrm{sign}(J_{\alpha\beta}) s_\alpha$, which means that $s_\alpha : U_\alpha \to \mathbb{R}$ is the local expression of a globally defined section s of $L \to M$. Since ω has no zeroes, neither does s.

Problem. Prove the reverse implication: if the orientation line bundle $L \to M$ has a section with no zeroes, then M is orientable.

Example. The space of symmetric unitary matrices (say, of dimension $n \times n$) is of some prominence in theoretical physics. This space fails to be orientable for $n \geq 2$.

1.3.9 Twisted differential forms

Definition. A *twisted differential form* τ on M is a form with values in the orientation line bundle L of M. One writes $\tau \in \Omega(M, L)$. A twisted form of top degree is called a *density.* $\square$

The whole formalism of differential forms carries over to twisted forms. In particular, since the orientation line bundle is a flat vector bundle, one has a canonical exterior derivative $d : \Omega^q(M, L) \to \Omega^{q+1}(M, L)$ and hence a twisted de Rham cohomology; this still exists with and without compact supports, and is denoted by $H^\bullet_c(M, L)$ and $H^\bullet(M, L)$ respectively.

Twisted differential forms can be integrated. In the important case of a density, the integral (if it converges) exists for any manifold M, orientable or not. To give a few details, a density $\omega \in \Omega^n(M, L)$ by definition has the local expression

$$\omega\big|_{U_\cap U_\beta} = f_\alpha\, dx^1_\alpha \wedge \cdots \wedge dx^n_\alpha \otimes s_\alpha = f_\beta\, dx^1_\beta \wedge \cdots \wedge dx^n_\beta \otimes s_\beta\,,$$

where $f_\alpha = f_\beta/|J_{\alpha\beta}|$ and $s_\alpha = \operatorname{sign}(J_{\alpha\beta})\, s_\beta$. This may be reorganized as

$$\omega\big|_{U_\alpha \cap U_\beta} = f_\alpha\, |dx^1_\alpha \wedge \cdots \wedge dx^n_\alpha| = f_\beta\, |dx^1_\beta \wedge \cdots \wedge dx^n_\beta|\,,$$

where the transition rule is

$$|dx^1_\alpha \wedge \cdots \wedge dx^n_\alpha| = |J_{\alpha\beta}|\; |dx^1_\beta \wedge \cdots \wedge dx^n_\beta|\,.$$

The integral of a density ω is defined as the iterated Riemann integral

$$\int_M \omega := \sum_\alpha \int_{U_\alpha} \rho_\alpha f_\alpha dx^1_\alpha \cdots dx^n_\alpha\,,$$

where $\sum \rho_\alpha = 1$ is a partition of unity (subordinate to the cover $\{U_\alpha\}$). This definition does not require M to be oriented (or even orientable).

Examples. Important examples of twisted differential forms are furnished by the inhomogeneous Maxwell equations, namely (Gauss): $dD = \rho$ and (Ampére-Maxwell) $dH = j + \dot{D}$. The quantities appearing in these equations are the electric charge density $\rho \in \Omega^3(\mathbb{R}^3, L)$, the electric current density $j \in \Omega^2(\mathbb{R}^3, L)$, the electric excitation $D \in \Omega^2(\mathbb{R}^3, L)$ and the magnetic excitation $H \in \Omega^1(\mathbb{R}^3, L)$. Further examples are provided by the electromagnetic energy density $\frac{1}{2}(E \wedge D + B \wedge H) \in \Omega^3(\mathbb{R}^3, L)$ and energy current density $E \wedge H \in \Omega^2(\mathbb{R}^3, L)$.

Theorem (Poincaré duality). On an n-dimensional manifold M of finite type the pairings

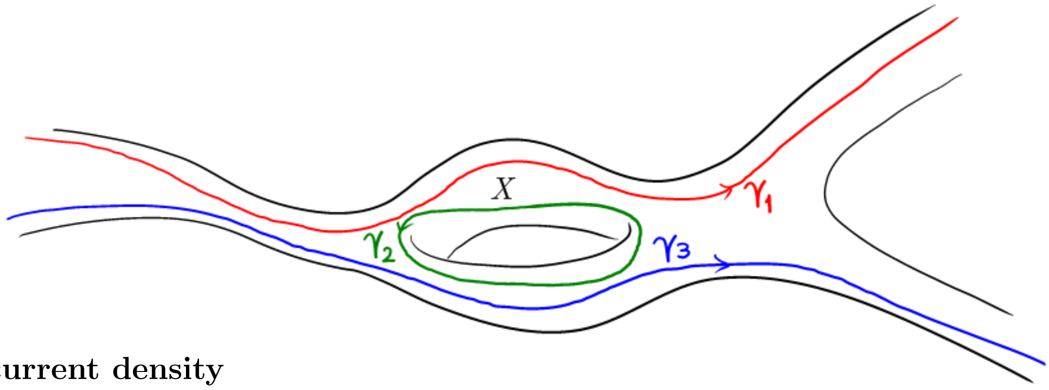
$$\int : \; H^q(M) \otimes H^{n-q}_c(M, L) \to \mathbb{R} \quad \text{and} \quad \int : \; H^q_c(M) \otimes H^{n-q}(M, L) \to \mathbb{R}$$

by integration are non-degenerate.

Remark. This is the optimal version of Poincaré duality, as it does not involve any orientation for M. By the notion of Poincaré dual of a submanifold, it leads to the correct way of drawing pictures of the electromagnetic field:

1.4 Application: d.c. electrical transport

We now embark on a substantial example illustrating all aspects of the theory developed so far: we will review some fundamental aspects of electrical transport theory, using the model of an electrical conductor as an n-dimensional manifold X (where $n = 1, 2, 3$ in reality) with open ends.



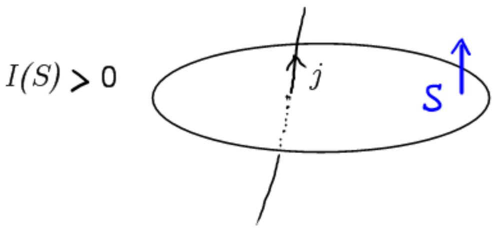
1.4.1 Charge and current density

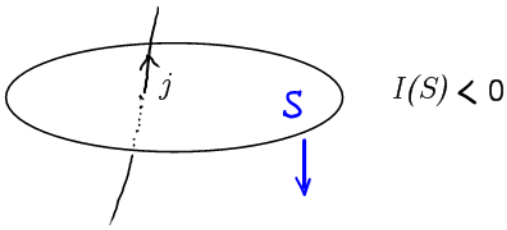
Finding the total amount of electric charge in a domain $U \subset X$ is a counting exercise that does not require U to be oriented or even orientable. Accordingly, in the continuum approximation one models the electric charge density ρ on X as a twisted n-form, $\rho \in \Omega^n(X, L)$. The electric charge $Q(U)$ in $U \subset X$ is computed from ρ by integration: $Q(U) = \int_U \rho$.

The electric current density, commonly denoted by j in physics, is the quantity that encodes the information about the flow of the electric charges. The proper mathematical model for it is a twisted $(n-1)$-form, $j \in \Omega^{n-1}(X, L)$. By integrating j over a $(n-1)$-dimensional submanifold S in X, one obtains the electric current through S:

$$I(S) := \int_S j\,.$$

$I(S)$ comes with a sign which depends on the choice of outer orientation of S (by which we mean a choice of direction of passing through the hypersurface S).





If S is a boundary, say $S = \partial U$, the law of conservation of electric charge says that $I(S) = -\frac{d}{dt}Q(U)$. The differential version of this law is $d\,j = -\dot{\rho}$.

In a stationary situation, where $\dot{\rho} = 0$, the electric current density $j \in \Omega^{n-1}(X, L)$ is closed: $d\,j = 0$. If we are not interested in the fine details of j but want only its period integrals, i.e. integrals over closed hypersurfaces, then there is no loss in sending j to its twisted de Rham cohomology class, $[j] \in H^{n-1}(X, L)$. The cohomology class $I := [j]$ is called the (total) *current.*

Example. For the conductor X shown above, the cohomology $H^2(X, L)$ is 3-dimensional. A basis is given by the Poincaré duals of the 1-cycles $\gamma_1, \gamma_2, \gamma_3$.

1.4.2 Current vector field

Let us now assume that X comes with a canonical volume density $\mathrm{dvol}_X \in \Omega^n(X, L)$. Then there exists an isomorphism

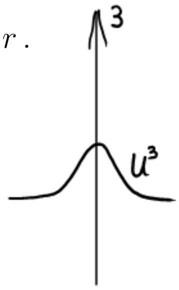
$$\Gamma(TX) \to \Omega^{n-1}(X, L)\,, \quad v \mapsto \iota(v)\,\mathrm{dvol}_X\,,$$

between vector fields, or sections of the tangent vector bundle $\Gamma(TX)$, and twisted $(n-1)$-forms. The operator $\iota(v)$ is the operator of contraction with the vector field v; for example,

$$\iota(v)\,B = \iota(v)\left(\sum_{i<j} B_{ij}\,dx^i \wedge dx^j\right) = \sum_{i<j}\left(v^i B_{ij}\,dx^j - v^j B_{ij}\,dx^i\right)$$

is the Lorentz force on a particle of (charge $q = -1$ and) velocity v in a magnetic field B. To give another example, let ρ be the charge density of a charged fluid with velocity field u. In that case the electric current density is the contraction of u with ρ. For $\rho = f\,|dx^1 \wedge dx^2 \wedge dx^3| = f\,dx^1 \wedge dx^2 \wedge dx^3 \otimes r$ where r denotes the section of the orientation line bundle which assigns to every point of $\mathbb{R}^3$ a right-handed Cartesian system $\{e_1, e_2, e_3\}$, this looks as follows:

$$\begin{aligned} j = \iota(u)\rho &= \iota(u)\left(f\,dx^1 \wedge dx^2 \wedge dx^3 \otimes r\right) \\ &= f\left(u^1 dx^2 \wedge dx^3 - u^2 dx^1 \wedge dx^3 + u^3 dx^1 \wedge dx^2\right) \otimes r\,. \end{aligned}$$



Problem. Show that $\omega = u^3 dx^1 \wedge dx^2 \otimes r \in \Omega^2(\mathbb{R}^3, L)$ with a bump function $u^3 = u^3(x_1, x_2)$ and integral (say, over the 12-plane) $\int \omega = 1$ is Poincaré dual to the 3-axis with orientation arrow pointing in the positive direction. □

In the present context, we may use the isomorphism $\Gamma(TX) \to \Omega^{n-1}(X, L)$ to think of the electric current density j in terms of the vector field v which yields j upon contraction with dvol_X :

$$\iota(v)\,\mathrm{dvol}_X = j\,.$$

v is called the vector field of the electric current, or current vector field for short.

Let us mention in passing that by the integral of the vector field v over an $(n-1)$-dimensional submanifold $S \subset X$ one means $\int_S v := \int_S \iota(v)\,\mathrm{dvol}_X$. The divergence of v is given by

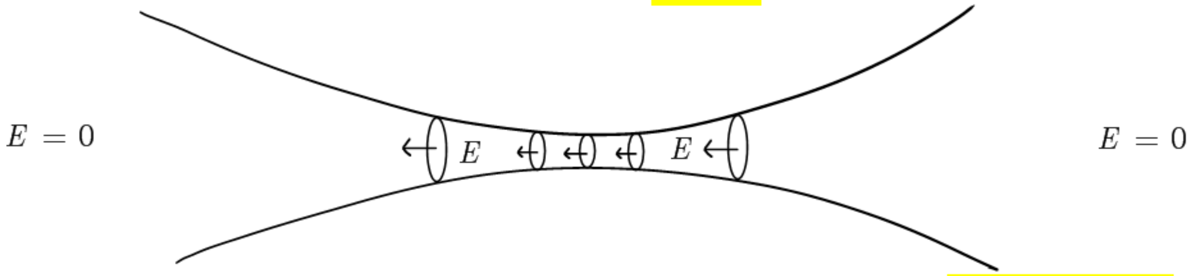
$$\mathrm{div}(v)\,\mathrm{dvol}_X = d\,\iota(v)\,\mathrm{dvol}_X\,.$$

Thus only a volume density (in particular, no metric tensor!) is needed in order to define div.

1.4.3 Voltage

The electric field strength is a 1-form, E, while the magnetic field strength is a 2-form, B. As part of Maxwell's theory the field strengths obey Faraday's law of induction: $d\,E = -\dot{B}$. Thus E is closed if $\dot{B} = 0$. Let us then consider sending E to its cohomology class, $E \mapsto [E]$. In a strictly static situation, it is a postulate of physics that the electric field has an electric potential: $E = -d\Phi$, so the de Rham cohomology class $[E] \in H^1(X)$ is always trivial in that case.

However, there exist two reasons why in a stationary situation $[E]$ may still become nontrivial. Firstly, it may happen that $\dot{B} = 0$ inside the conductor filling the region X, but $\dot{B} \neq 0$ somewhere outside. In that case E restricted to X is closed, but E need not be exact. Secondly, and more importantly, it is reasonable to assume (e.g., in the setting of mesoscopic physics) that E vanishes outside some bounded region of space. The proper notion to use for E then is the de Rham cohomology with compact supports, $[E] \in H^1_c(X)$. In the latter sense $[E]$ may be nonzero even in a truly static situation. $V := [E]$ is called the (static) voltage in physics.



From the force law for charged particles in an electromagnetic field, the electrical power (i.e., the rate of energy transfer from the field to the particles) is the integral

$$P = \int_X E \wedge j \; .$$

In a stationary situation where both E and j are closed, the electrical power descends to a pairing in cohomology:

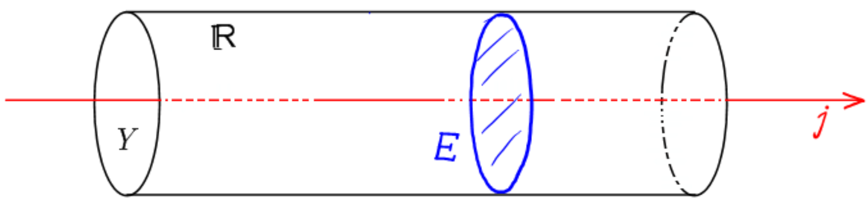
$$P: \; H^1_c(X) \otimes H^{n-1}(X, L) \to \mathbb{R} \, , \qquad ([E],[j]) \mapsto \int_X E \wedge j \, .$$

Poincaré duality says that this pairing is non-degenerate (for any X of finite type) or, in physics language: for every voltage $V = [E] \neq 0$ there is some current $I = [j] \neq 0$ so that the power $P(V, I)$ does not vanish, and the converse statement also holds.

Example. Let X be of the product form $X = \mathbb{R} \times Y$ with Y compact, closed, simply connected, and $\mathrm{dim} Y = n - 1$. Then by the Künneth formula,

$$H^1_c(X) = H^1_c(\mathbb{R}) \otimes H^0_c(Y) = \mathbb{R} \otimes \mathbb{R} = \mathbb{R} \, ,$$

and the voltage $V = [E] \in \mathbb{R}$ is given by the single number $\int E$ where the integral is along any path connecting the two ends $\{-\infty\} \times Y$ and $\{+\infty\} \times Y$ of X. One also has $H^{n-1}(X, L) = \mathbb{R}$, and the current $I = [j] \in \mathbb{R}$ is the number $I = \int j$, where the integral now is over the cross section $\{0\} \times Y$ or any $(n-1)$-cycle of X homologous to it. In this situation, the pairing between voltage and current by power is simply the product of the two numbers $\int E$ and $\int j$.



1.4.4 Conductance as a map in cohomology

Suppose that the electric charges of a physical system without external forces are at rest (so that $j = 0$). On imposing a driving force by means of an external electric field, one expects the system to respond with an electric current flow. For a sufficiently weak electric field the relation between E and j is linear in general, and one then calls the linear operator $\widehat{\sigma} : E \mapsto j$ the linear-response electrical conductivity. In the stationary limit of interest to us, one attaches to $\widehat{\sigma}$ the adjective 'd.c.' (standing for 'directed current' as opposed to 'alternating current', or 'a.c.').

Definition. The *linear-response electrical conductivity* is a linear mapping

$$\widehat{\sigma} : \Omega_c^1(X) \to \Omega^{n-1}(X, L) , \quad E \mapsto j = \widehat{\sigma}(E)$$

(depending, in general, on physical parameters such as gate voltages, magnetic fields, etc.). In the d.c. limit $\widehat{\sigma}$ has the following properties:

- $\widehat{\sigma}$ takes rotationless electric fields to divergenceless electric current densities, i.e., restricts to a linear mapping
$$\widehat{\sigma} : Z_c^1(X) \to Z^{n-1}(X, L) .$$
- The linear operator $\widehat{\sigma}$ possesses an integral kernel (with regularity properties not specified here). By using the one-to-one correspondence between vector fields and twisted $(n-1)$-forms by $v \leftrightarrow \iota(v)\,\mathrm{dvol}_X$, one may view this kernel as a bi-vector field and express $j = \widehat{\sigma}(E)$ in components with respect to some basis as
$$\widehat{\sigma}(E)^i(x) = \sum_j \int_X \sigma^{ij}(x, y) E_j(y)\,\mathrm{dvol}_X(y) .$$
- The components of the bi-vector field of $\widehat{\sigma}$ obey the Onsager relation
$$\sigma^{ij}(x, y; B) = \sigma^{j\,i}(y, x\,; -B) .$$
In words: changing the sign of the magnetic field strength B (and, more generally, changing the sign of all physical parameters which are odd w.r.t. time inversion) sends the bi-vector field of $\widehat{\sigma}$ to its transpose.

The situation at hand involves two differential complexes: the de Rham complex of compactly supported forms, $(\Omega_c(X), d)$, and the twisted de Rham complex $(\Omega(X, L), d)$. Recall that a linear mapping between differential complexes is called a chain map if it commutes with the differential operator d. The electrical conductivity is not a chain map but does share the following property.

Proposition. Under the postulates above, the d.c. linear-response electrical conductivity descends to a map $H_c^1(X) \to H^{n-1}(X, L)$ in cohomology.

Proof (sketch). Given that $\widehat{\sigma}$ takes closed electric fields to closed electric current densities by the first postulate, there is a well-defined induced map in cohomology if $\widehat{\sigma}(B_c^1(X)) \subset B^{n-1}(X, L)$. Thus, the statement to be proved is that if $E = -d\,\Phi$ with compactly supported Φ, then the twisted $(n-1)$-form $j = \widehat{\sigma}(E)$ is exact. Although this statement holds true in the general setting of a Riemannian manifold X with volume density dvol_X , we will assume X to be of Euclidean type and do the computation in Cartesian coordinates as follows.

For an arbitrary test form $\eta = \sum \eta_i \, dx^i \in \Omega^1_c(X)$ consider the integral

$$\int_X \widehat{\sigma}(-d\Phi) \wedge \eta = -\sum_i \int_X \left(\int_X \sum_j \sigma^{ij}(x,y) \frac{\partial}{\partial y^j} \Phi(y) \, \mathrm{dvol}_X(y) \right) \eta_i(x) \, \mathrm{dvol}_X(x) \, .$$

We partially integrate the inner integral, by using the property of compact support for Φ (and assuming sufficient regularity for the bi-vector field σ^{ij}):

$$-\int_X \sum_j \sigma^{ij}(x,y) \frac{\partial}{\partial y^j} \Phi(y) \, \mathrm{dvol}_X(y) = \int_X \Phi(y) \sum_j \frac{\partial}{\partial y^j} \sigma^{ij}(x,y) \, \mathrm{dvol}_X(y) \, .$$

Next we use the Onsager relation $\sigma^{ij}(x,y) = \tau^{j\,i}(y,x)$ where τ^{ij} denotes the conductivity tensor of the time-reversed system. By interchanging the order of integration we then obtain

$$\int_X \widehat{\sigma}(-d\Phi) \wedge \eta = \int_X \Phi(y) \sum_j \frac{\partial}{\partial y^j} \left(\int_X \sum_i \tau^{j\,i}(y,x) \, \eta_i(x) \, \mathrm{dvol}_X(x) \right) \mathrm{dvol}_X(y) \, .$$

The inner integral on the right-hand side can be written as

$$\int_X \sum_i \tau^{j\,i}(y,x) \, \eta_i(x) \, \mathrm{dvol}_X(x) = \widehat{\tau}(\eta) \, .$$

Thus we arrive at

$$\int_X \widehat{\sigma}(-d\,\Phi) \wedge \eta = \int_X \Phi \, d\,\widehat{\tau}(\eta) \, .$$

Finally, we take η to be closed (but otherwise arbitrary). Then $\widehat{\tau}(\eta)$ is closed by our first postulate and the integral on the right-hand side vanishes. By Poincaré duality, i.e. the non-degeneracy of the pairing $H^{n-1}(X,L) \otimes H^1_c(X) \to \mathbb{R}$ by integration, it follows that $\widehat{\sigma}(-d\Phi)$ must be zero in cohomology. Thus $\widehat{\sigma}(-d\Phi)$ is exact as claimed. $\square$

Problem. Assuming X to be Euclidean (for simplicity), show that the conductivity tensor satisfies

$$\sum_{i,j} \frac{\partial^2}{\partial x^i \partial y^j} \, \sigma^{ij}(x,y) = 0 \, . \, \square$$

We have demonstrated that the d.c. linear-response conductivity $\widehat{\sigma} : Z^1_c(X) \to Z^{n-1}(X,L)$ descends to a mapping in cohomology. This map, taking voltages $V = [E] \in H^1_c(X)$ to currents $I = [j] \in H^{n-1}(X,L)$, has a special name in physics.

Definition. The induced map,

$$G : \, H^1_c(X) \to H^{n-1}(X,L) \, ,$$

is called the d.c. linear-response electrical *conductance*. $\square$

By Poincaré duality, one can reformulate the conductance as

$$G : \, H^1_c(X) \xrightarrow{g} H^1_c(X)^* \simeq H^{n-1}(X,L) \, ,$$

where g, being a map between a vector space and its dual, has a canonical adjoint (or transpose), g^T. The Onsager relation restated at the cohomological level then says that

$$g^T\big|_B = g\big|_{-B} \, .$$

Thus in the absence of magnetic fields (and other parameters that break time-reversal symmetry) the conductance g is symmetric; in terms of the power $P([E],[j]) = \int E \wedge j$ this means that

$$P(V, I') = P(V, g(V')) = P(V', g(V)) = P(V', I)\,.$$

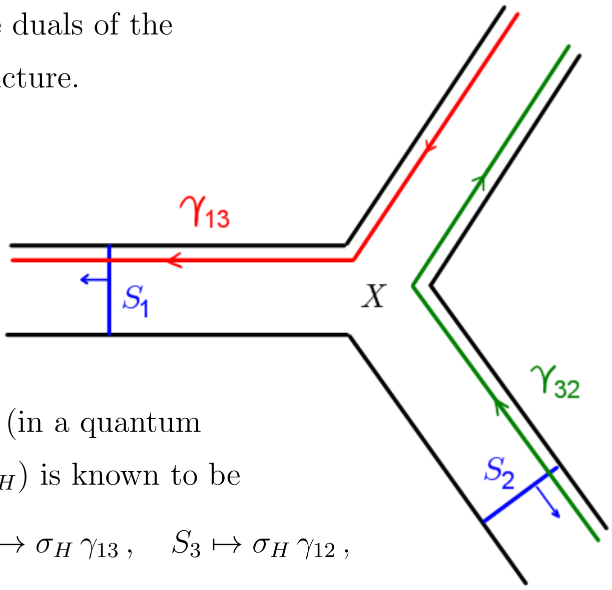
When a magnetic field (or other agents breaking time-reversal symmetry) are present, the conductance may have a skew-symmetric part. This part is called the non-dissipative (or Hall) part of the conductance; it does not contribute to the dissipated power $P(V', g(V))\big|_{V'=V}$.

Footnote. For a linear map $L:\ V \to W$, $V = W$, one has no way of telling in general whether L is symmetric or not (unless V is equipped with a non-degenerate quadratic form). However, for $W = V^*$ one can speak about symmetry or skew-symmetric without using any extra structure: one calls L symmetric if $L(v)(v') = L(v')(v)$ and skew if $L(v)(v') = -L(v')(v)$. An example of a symmetric map $L:\ V \to V^*$ is the tensor of the moments of inertia of a rigid body (with respect to some fixed point, say the center of mass) mapping angular velocities to angular momenta.

Example. In a quantum Hall (QH) insulator, i.e. a 2d electron gas exhibiting the quantum Hall effect, the symmetric part of the conductance vanishes while the skew-symmetric part, the Hall conductance, is quantized in (integer or fractional) units of the conductance quantum e^2/h. To illustrate how Poincaré duality helps to give a mathematical description of the situation, consider a quantum Hall insulator X with three leads. The cohomology of the voltage then is $H^1_c(X) = \mathbb{R}^2$, and the same goes for the cohomology $H^1(X, L)$ of the current.

A basis of $H^1_c(X)$ is supplied by the Poincaré duals of the cross sections S_1 and S_2 shown in the next picture.

A basis for $H^1(X, L)$ is provided by the (Poincaré duals of the) 1-cycles γ_{32} and γ_{13} .



The current response of the QH insulator (in a quantum Hall plateau regime with Hall conductivity σ_H) is known to be

$$S_1 \mapsto \sigma_H\, \gamma_{32}\,, \quad S_2 \mapsto \sigma_H\, \gamma_{13}\,, \quad S_3 \mapsto \sigma_H\, \gamma_{12}\,,$$

from experiments. Now the pairing by integration gives

$$P(S_1, \gamma_{32}) = 0 = P(S_2, \gamma_{13})\,, \quad P(S_1, \gamma_{13}) = 1\,, \quad P(S_2, \gamma_{32}) = -1\,,$$

which means that $\gamma_{13} \equiv S_1^*$ and $-\gamma_{32} = S_2^*$ are the basis elements of $H^1_c(X)^*$ which are dual to the basis S_1 and S_2 of $H^1_c(X)$. When expressed in such a basis, the current response takes the skew-symmetric form characteristic of non-dissipative transport:

$$g(S_1) = -\sigma_H\, S_2^*\,, \quad g(S_2) = +\sigma_H\, S_1^*\,.$$

Problem. By using your understanding of quantum Hall physics, make a similar analysis of the QH insulator with n leads.

2 Vector bundles and characteristic classes

By a *characteristic class* one means a cohomology class which is intrinsically associated with a vector bundle. Consider, for example, the two-sphere S^2. The cohomology classes $[B] \in H^2(\mathrm{S}^2) = \mathbb{R}$ are in one-to-one correspondence with total magnetic fluxes $\int_{\mathrm{S}^2} B$. A priori there exist no distinguished cohomology classes or magnetic fluxes in $H^2(\mathrm{S}^2)$. However, from the lecture course on Advanced QM we know that the *Dirac quantization condition* singles out those cohomology classes for which $\int_{\mathrm{S}^2} B \in \mathbb{Z}$ (in units of the flux quantum h/e). It turns out that the integrality of these classes derives from the existence of a vector bundle (whose sections have an interpretation as the wave functions of a charged particle in the field of a magnetic monopole) over S^2.

2.1 Euler class for rank 2

In this subsection we will meet the simplest example of a characteristic class: the Euler class of a real vector bundle of rank 2, which happens to be the same as the first Chern class of a complex vector bundle of rank 1. We will illustrate the Euler class by giving two examples: the Dirac quantization condition mentioned above, and the Berry line bundle of adiabatic quantum dynamics (popularly known by the phenomenon of "Berry phase").

We begin with some basic material about vector bundles.

2.1.1 Reduction of structure group

Suppose we are given a rank-n vector bundle $\pi : E \to M$ with trivialization $\{(U_\alpha, \phi_\alpha)\}$ and transition functions $g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(n, \mathbb{R})$. If $\{(U_\alpha, \tilde{\phi}_\alpha)\}$ is another trivialization, then there exist maps $\lambda_\alpha : U_\alpha \to \mathrm{GL}(n, \mathbb{R})$ such that $\phi_\alpha = \lambda_\alpha \tilde{\phi}_\alpha$. The structure functions $\tilde{g}_{\alpha\beta}$ for the new trivialization are $\tilde{g}_{\alpha\beta} = \lambda_\alpha^{-1} g_{\alpha\beta} \lambda_\beta$. Indeed,

$$g_{\alpha\beta} = \phi_\alpha \phi_\beta^{-1} = \lambda_\alpha \tilde{\phi}_\alpha \tilde{\phi}_\beta^{-1} \lambda_\beta^{-1} = \lambda_\alpha \tilde{g}_{\alpha\beta} \lambda_\beta^{-1} .$$

Since the transition functions $g_{\alpha\beta}$ satisfy the cocycle condition $g_{\alpha\beta} g_{\beta\gamma} = g_{\alpha\gamma}$ (on $U_\alpha \cap U_\beta \cap U_\gamma$) so do the new transition functions $\tilde{g}_{\alpha\beta}$. Two cocycles related by $g_{\alpha\beta} = \lambda_\alpha \tilde{g}_{\alpha\beta} \lambda_\beta^{-1}$ are called *equivalent.*

Fact. Two vector bundles are isomorphic if and only if their cocycles relative to some open cover are equivalent. □

Definition. Given a vector bundle E with cocycle $\{g_{\alpha\beta}\}$, if it is possible to find an equivalent cocycle with values in a subgroup $H \subset \mathrm{GL}(n, \mathbb{R})$, one says that the structure group of E may be *reduced* to H. A vector bundle is called *orientable* if its structure group may be reduced to the group $\mathrm{GL}^+(n, \mathbb{R})$ of linear transformations of $\mathbb{R}^n$ with positive determinant. If E is orientable, a trivialization $\{(U_\alpha, \phi_\alpha)\}$ of E is called *oriented* if all transition functions $g_{\alpha\beta}$ have positive determinant. Two oriented trivializations $\{(U_\alpha, \phi_\alpha)\}$ and $\{(V_\beta, \psi_\beta)\}$ are *equivalent* if for every $x \in U_\alpha \cap V_\beta$ the linear mapping $(\phi_\alpha \psi_\beta^{-1})(x) : \mathbb{R}^n \to \mathbb{R}^n$ has positive determinant. This equivalence relation divides the oriented trivializations into two classes, each of which is called an *orientation class* (or orientation for short) of E. □

Remark. Orientability of E as a vector bundle is not the same as orientability of E as a manifold. However, for a tangent bundle $E = TM$, orientability of E as a vector bundle is equivalent to orientability of M as a manifold. □

A *Riemannian structure* on a vector bundle E is a symmetric bilinear form

$$\langle \cdot , \cdot \rangle_x : \; E_x \times E_x \to \mathbb{R} \, ,$$

which is positive definite and depends smoothly on the position x. Such a structure exists for any rank-n vector bundle E; this follows from the local factorization $\pi^{-1}(U_\alpha) \simeq U_\alpha \times \mathbb{R}^n$ by a partition of unity argument and the fact that $\mathbb{R}^n$ can be given a Euclidean structure.

Example. The tangent bundle $E = TM$ of a Riemannian manifold carries a canonical Riemannian structure given by the metric tensor of M.

Problem. Show that the structure group $\mathrm{Gl}(n, \mathbb{R})$ of a rank-n real vector bundle E can always be reduced to $\mathrm{O}(n)$, and if E is orientable, that it can be reduced to $\mathrm{SO}(n)$.

2.1.2 Euler class

Let $\pi : E \to M$ be an oriented real vector bundle of rank 2. ('Oriented' here means that one of the two orientation classes of the orientable vector bundle E has been singled out.) We fix a Riemannian structure on E and choose an oriented trivialization $\{(U_\alpha, \phi_\alpha)\}$ by orthonormal frames $\phi_\alpha : \; \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^2$. The transition functions then are maps

$$g_{\alpha\beta} : \; U_\alpha \cap U_\beta \to \mathrm{SO}(2) \, .$$

Let $d\theta$ be the standard angular 1-form for $\mathrm{SO}(2)$. By pullback we get a closed 1-form $\eta_{\alpha\beta} := g^*_{\alpha\beta}(d\theta)$ on each intersection $U_\alpha \cap U_\beta$. Due to the cocycle condition $g_{\alpha\beta} = g_{\alpha\gamma}\, g_{\gamma\beta}$ these 1-forms satisfy $\eta_{\alpha\beta} = \eta_{\alpha\gamma} + \eta_{\gamma\beta}$ on any triple intersection $U_\alpha \cap U_\beta \cap U_\gamma$. In particular, $\eta_{\alpha\beta} = -\eta_{\beta\alpha}$.

Let now $\sum \rho_\gamma = 1$ be a partition of unity subordinate to the cover $\{U_\gamma\}$ and consider

$$\xi_\alpha = \frac{1}{2\pi} \sum_\gamma \rho_\gamma \, \eta_{\gamma\alpha} \, ,$$

which is a 1-form on U_α. On the intersection $U_\alpha \cap U_\beta$ we get

$$2\pi(\xi_\beta - \xi_\alpha) = \sum_\gamma \rho_\gamma (\eta_{\gamma\beta} - \eta_{\gamma\alpha}) = \sum_\gamma \rho_\gamma (\eta_{\alpha\gamma} + \eta_{\gamma\beta}) = \eta_{\alpha\beta} \sum \rho_\gamma \, ,$$

and hence

$$\xi_\beta - \xi_\alpha = \frac{\eta_{\alpha\beta}}{2\pi} \, .$$

Since $\eta_{\alpha\beta}$ is closed, we have $d\xi_\alpha = d\xi_\beta$ on $U_\alpha \cap U_\beta$. Therefore, the locally defined 2-forms

$$e_\alpha := d\xi_\alpha = \frac{1}{2\pi} \sum_\gamma d\rho_\gamma \wedge \eta_{\gamma\alpha} \, ,$$

piece together to a globally defined 2-form $e \in Z^2(M)$. Note that e is a differential form on the base space M, but its construction requires the existence of the total space E.

Definition. The cohomology class $[e] \in H^2(M)$ of the closed 2-form e is called the *Euler class* of the oriented rank-2 real vector bundle $\pi :\ E \to M$.

Remark 1. Starting from a complex line bundle, i.e., a vector bundle with fiber $\pi^{-1}(x) \simeq \mathbb{C}$, one gets an orientable rank-2 real vector bundle by using the isomorphism $\mathbb{C} \simeq \mathbb{R}^2$ and forgetting the complex structure of $\mathbb{R}^2$. In such a situation, the cohomology class $[e]$ is also known as the (first) Chern class of the complex line bundle.

Remark 2. In the case of a trivial vector bundle E, the Euler class $[e]$ vanishes. Indeed, for $E \simeq M \times \mathrm{R}^2$ one may take all transition functions to be unity, so that $\eta_{\alpha\beta} = 0$ for all overlapping domains $U_\alpha \cap U_\beta$. Thus the Euler class is a measure of the twisting of the vector bundle.

2.1.3 Example: Euler class of $T^*\mathrm{S}^2$

Consider $T^*\mathrm{S}^2$, the cotangent bundle of the two-sphere. $T^*\mathrm{S}^2$ is orientable (because S^2 is), and we take it to be oriented by the right-hand rule, which is to say that a right-handed system is formed by the orientation of S^2 in conjunction with the normal pointing outward.

Let S^2 be covered by two open subsets $U_n = \mathrm{S}^2 \setminus \{s\}$ and $U_s = \mathrm{S}^2 \setminus \{n\}$ which are obtained by removing the south resp. north pole (not a good cover). Starting from spherical polar coordinates θ and ϕ, it is convenient to introduce a complex coordinate function z for U_n by $z = \tan(\theta/2)\,\mathrm{e}^{\mathrm{i}\phi}$. In this coordinate the north pole is at $z = 0$ and the south pole at $z = \infty$. A complex coordinate for U_s is $w = -\cot(\theta/2)\,\mathrm{e}^{-\mathrm{i}\phi} = -z^{-1}$; this is defined at the south pole ($w = 0$) and singular at the north pole ($w = \infty$).

To construct the Euler form e associated with $T^*\mathrm{S}^2$, we need a Riemannian structure. For this we regard S^2 as a Riemannian manifold with geometry given by the Fubini-Study metric, whose local coordinate expression is

$$d\theta^2 + \sin^2\theta\, d\phi^2 = \frac{4dz\, d\bar z}{(1+|z|^2)^2} = \frac{4dw\, d\bar w}{(1+|w|^2)^2}\,.$$

Let us then define orthonormal basis forms ϑ^z, $\vartheta^{\bar z}$ on U_n and ϑ^w, $\vartheta^{\bar w}$ on U_w by

$$\vartheta^z = \frac{2dz}{1+|z|^2}\,, \quad \vartheta^{\bar z} = \frac{2d\bar z}{1+|z|^2}\,, \quad \vartheta^w = \frac{2dw}{1+|w|^2}\,, \quad \vartheta^{\bar w} = \frac{2d\bar w}{1+|w|^2}\,.$$

Viewing $T^*\mathrm{S}^2$ as a complex line bundle, the transition functions now follow from

$$g_{ns}\, \vartheta^w = \vartheta^z = \frac{2d\,(-1/w)}{1+|w|^{-2}} = \frac{\bar w}{w}\, \vartheta^w = \mathrm{e}^{2\mathrm{i}\phi}\, \vartheta^w.$$

To convert to the real setting we evaluate g_{ns} on the real orthonormal frame $(\vartheta^w + \vartheta^{\bar w})/\sqrt{2}$ and $(\vartheta^w - \vartheta^{\bar w})/\sqrt{2}\mathrm{i}$. The result of this computation is

$$g_{ns} = \begin{pmatrix} \cos(2\phi) & \sin(2\phi) \\ -\sin(2\phi) & \cos(2\phi) \end{pmatrix}.$$

As a check, note that g_{ns} is defined on $U_n \cap U_s$ and that $\mathrm{Det}(g_{ns}) = 1 > 0$. By the prescription of the previous section we now get $\eta_{ns} = 2d\phi = -\eta_{sn}$ and hence the Euler form

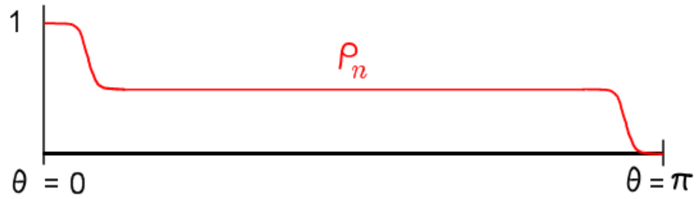
$$e = d\rho_n \wedge \frac{\eta_{ns}}{2\pi} = d\rho_n \wedge \frac{d\phi}{\pi} = -d\rho_s \wedge \frac{d\phi}{\pi} = d\rho_s \wedge \frac{\eta_{sn}}{2\pi}\,.$$

From this we see that the Euler class of the cotangent bundle $T^*\mathrm{S}^2$ is

$$[e] = -(2\pi)^{-1}[\sin\theta \, d\theta \wedge d\phi] \,.$$

In particular, we have $\int_{\mathrm{S}^2} e = -2$.

There exists a beautiful visualization of the Euler class by Poincaré duality. For this purpose, let the partition of unity be chosen in the particular way shown in the next figure.



Our Euler form e then consists of two bumps, each of integral -1. These may be visualized as the craters of two volcanoes, one each at the two (north and south) poles. From the extreme of this perspective, we may view the Euler class of $T^*\mathrm{S}^2$ as the Poincaré dual of the 0-chain consisting of the two points at the poles, each of weight -1. (For $E = T\mathrm{S}^2$ the weights would be $+1$.)

We take this opportunity to communicate without proof the following fact. The *Euler number* (or Euler characteristic) of an orientable manifold M is defined as the alternating sum

$$\chi(M) = \sum_{q=0}^{\dim M} (-1)^q \dim H^q(M) \,.$$

Since we have defined the Euler class only for the case of rank 2, the student of these notes can appreciate the following statement only for a 2-dimensional manifold M, although it holds actually for a manifold M of any dimension.

Theorem. For an orientable compact manifold M the Euler characteristic equals the integral of the Euler class $[e]$ associated with the tangent bundle $TM \to M$:

$$\chi(M) = \int_M e \,.$$

Problem. Verify this statement for the cases of $M = \mathrm{S}^2$ and $M = \mathrm{T}^2$. $\square$

Let us also mention that for the Euler class associated with a tangent bundle, there exists an alternative construction from geometric data (not the main subject of this lecture course). Indeed, let M be a Riemannian manifold with curvature tensor R, which is a 2-form with values in $\mathfrak{so}(TM)$. Since the elements of the Lie algebra $\mathfrak{so}(T_x M)$ are skew-symmetric, there exists a natural notion of Pfaffian of R, and one has $[e] = [\mathrm{Pf}(R)] / \sqrt{2\pi}^{\dim M}$. With this expression for the Euler class, the theorem for the case of $\dim M = 2$ is called the Gauss-Bonnet theorem (which is perhaps the simplest example of an index theorem). In the general case of any dimension, one speaks of the Gauss-Bonnet-Chern theorem.

2.1.4 Global angular form

Returning to the setting (cf. Section 2.1.2) of an oriented rank-2 real vector bundle $\pi : E \to M$, let E^0 denote the complement of the zero section $s^0 \in \Gamma(E)$, $s^0(x) = 0$. (Thus in E^0 the zero vector is missing from each fiber E_x. Note that in the case of a non-trivial vector bundle E, the bundle $E^0 \to M$ has no globally defined section, but this will be of no concern for what follows.)

Since E is oriented and admits a Riemannian structure, we can choose an oriented orthonormal trivialization $\{(U_\alpha, \phi_\alpha)\}$. Fixing a standard basis $\{e_1, e_2\}$ of the Euclidean plane $\mathbb{R}^2$, we get for each subset U_α an oriented orthonormal frame $\{e^1_\alpha, e^2_\alpha\}$ by $e^i_\alpha(x) := \phi_\alpha^{-1}(x, e_i)$. Such a frame defines polar coordinates r_α and θ_α on $E^0\big|_{U_\alpha}$ in the usual way. (To get a local coordinate system for $E^0\big|_{U_\alpha}$ you must add coordinates $x^1_\alpha, \ldots, x^m_\alpha$ for $U_\alpha \subset M$.) On $U_\alpha \cap U_\beta$ the radial coordinates r_α and r_β coincide, but the angular coordinates θ_α and θ_β differ by a rotation. In fact,

$$d\theta_\alpha - d\theta_\beta = \pi^* \eta_{\alpha\beta} \, .$$

Now, recalling the relation $\eta_{\alpha\beta} = 2\pi(\xi_\beta - \xi_\alpha)$ we obtain

$$\frac{d\theta_\alpha}{2\pi} + \pi^* \xi_\alpha = \frac{d\theta_\beta}{2\pi} + \pi^* \xi_\beta \, .$$

Hence these 1-forms, which are defined locally on $E^0\big|_{U_\alpha} \simeq U_\alpha \times (\mathbb{R}^2 \setminus \{0\})$, piece together to give a globally defined 1-form on E^0.

Definition. The 1-form $\psi \in \Omega^1(E^0)$ with local expression

$$\psi_\alpha = \frac{d\theta_\alpha}{2\pi} + \pi^* \xi_\alpha$$

on $E^0\big|_{U_\alpha}$ is called the *global angular form.*

Remark. From the local definition one sees that the global angular form has exterior derivative

$$d\psi = \pi^* e \, .$$

Thus, although the Euler form e fails (in general) to be exact as a form on M, it does become exact when pulled back to E^0, and the global angular form ψ is a potential for it.

2.2 Geometric structure from principal bundles

Going beyond issues of topology, we will now point out two things: (i) the global angular form ψ on E^0 determines a so-called *covariant derivative* (or *connection*) ∇ on E and, (ii) the Euler form e may be viewed as the *curvature* of that covariant derivative. We will first demonstrate this by a straightforward computation in local coordinates. Afterwards, we will provide some framework and perspective by describing the relevant constructions in differential geometry.

2.2.1 Covariant derivative and curvature

We begin with a few preparations (retaining the setting of Section 2.1.4). On overlapping domains $U_\alpha \cap U_\beta$ we have two equivalent expressions for a section $s \in \Gamma(E)$:

$$s(x) = \sum_{i=1}^{2} \sigma_i(x)\, e^i_\alpha(x) = \sum_{j=1}^{2} \tau_j(x)\, e^j_\beta(x) \, .$$

By definition, the transition function $g_{\alpha\beta} = U_\alpha \cap U_\beta \to \mathrm{SO}_2$ is the mapping

$$g_{\alpha\beta}(x) \;:\; \sum \tau_j(x)\, e_j \xrightarrow{\phi_\beta^{-1}(x)} \sum \tau_j(x)\, e_\beta^j(x) = s(x) = \sum \sigma_i(x)\, e_\alpha^i(x) \xrightarrow{\phi_\alpha(x)} \sum \sigma_i(x)\, e_i\,.$$

Fixing α, β we write $g(x) \equiv g_{\alpha\beta}(x)$ for short and read off the relations

$$\sigma_i = \sum_j g_{ij}\, \tau_j\,, \quad e_\beta^j = \sum_i e_\alpha^i\, g_{ij}\,,$$

where g_{ij} are the matrix elements of the transition function $g \equiv g_{\alpha\beta}$.

Let $J \in \mathfrak{so}_2$ be the rotation generator defined by $Je_1 = e_2$ and $Je_2 = -e_1$. Its matrix elements are $J_{11} = J_{22} = 0$ and $J_{21} = 1 = -J_{12}$. By using the data in the local expression ψ_α of the global angular form, we introduce a first-order differential operator ∇_α on U_α by

$$\nabla_\alpha s = \nabla_\alpha \left(\sum \sigma_i\, e_\alpha^i \right) := \sum\nolimits_i \left(d\sigma_i \otimes e_\alpha^i + 2\pi \xi_\alpha\, \sigma_i \otimes \sum\nolimits_k e_\alpha^k\, J_{ki} \right). \tag{2.5}$$

Lemma. On overlapping domains $U_\alpha \cap U_\beta$ one has $\nabla_\alpha = \nabla_\beta$.

Proof. The differential of the coefficient $\sigma_i = \sum g_{ij}\, \tau_j$ is

$$d\sigma_i = \sum \left(g_{ij}\, d\tau_j + \tau_j\, dg_{ij} \right)\,.$$

To compute dg_{ij} we adopt the viewpoint that the angle $\theta_\alpha - \theta_\beta$ of rotation between local frames is (or pushes down to) a function on $U_\alpha \cap U_\beta$. By taking the derivative and then matrix elements of the formula $g \equiv g_{\alpha\beta} = \mathrm{e}^{(\theta_\alpha - \theta_\beta)J}$ we obtain the expression

$$dg_{ij} = \left(d\, \mathrm{e}^{(\theta_\alpha - \theta_\beta)J} \right)_{ij} = \left(\mathrm{e}^{(\theta_\alpha - \theta_\beta)J} (d\theta_\alpha - d\theta_\beta) J \right)_{ij} = \eta_{\alpha\beta} \sum\nolimits_l g_{il}\, J_{lj}\,.$$

We also need the relation $\sum_{k,i} e_\alpha^k J_{ki}\, \sigma_i = \sum_{l,j} e_\beta^l J_{lj}\, \tau_j$, which results from applying $\phi_\alpha^{-1} \circ J$ to the identity $\sum_i e_i\, \sigma_i = \sum_{i,j} e_i\, g_{ij}\, \tau_j$ and using $\sum_i J_{ki}\, g_{ij} = \sum_l g_{kl}\, J_{lj}$. With all this information, the statement is verified by the following computation:

$$\begin{aligned}
\nabla_\alpha\, s &= \sum_{i,j} \left(d\tau_j \otimes e_\alpha^i\, g_{ij} + \tau_j\, \eta_{\alpha\beta} \otimes e_\alpha^i \sum\nolimits_l g_{il}\, J_{lj} + 2\pi \xi_\alpha\, \sigma_i \otimes \sum\nolimits_k e_\alpha^k J_{ki} \right) \\
&= \sum_j \left(d\tau_j \otimes e_\beta^j + 2\pi (\xi_\beta - \xi_\alpha) \tau_j \otimes \sum\nolimits_l e_\beta^l J_{lj} + 2\pi \xi_\alpha\, \tau_j \otimes \sum e_\beta^l J_{lj} \right) \\
&= \sum_j \left(d\tau_j \otimes e_\beta^j + 2\pi \xi_\beta\, \tau_j \otimes \sum\nolimits_l e_\beta^l J_{lj} \right) = \nabla_\beta\, s\,. \quad \square
\end{aligned}$$

The coincidence $\nabla_\alpha = \nabla_\beta$ means that there exists a globally defined differential operator $\nabla : \Gamma(E) \to \Gamma(T^*M \otimes E)$. It is easy to see that this operator is compatible with the exterior derivative d in the sense that $\nabla(fs) = df \otimes s + f\, \nabla s$ for any differentiable function f on M.

Definition. If E is a vector bundle over a manifold M, a *covariant derivative* (or *connection*) on E is a differential operator

$$\nabla :\; \Gamma(E) \to \Gamma(T^*M \otimes E)$$

which satisfies the Leibniz rule; i.e., if $s \in C^\infty(M)$ and $s \in \Gamma(E)$ then

$$\nabla(fs) = df \otimes s + f\,\nabla s\,.$$

Remark. A covariant derivative on E always exists. (In our rank-2 case we constructed ∇ from the geometric data of the vector bundle and a choice of partition of unity; more precisely, from the data of the global angular form ψ.) The space of covariant derivatives is an affine space modeled on the vector space $\Omega^1(M, \mathrm{End}(E))$. In other words, if both ∇ and ∇' are covariant derivatives on E, then $\nabla - \nabla' = \omega$ is a 1-form on M with values in $\mathrm{End}(E)$. $\square$

If ∇ is a covariant derivative and $X \in \Gamma(TM)$ a vector field, one gets a differential operator $\nabla_X : \Gamma(E) \to \Gamma(E)$ by contraction, i.e., if $\nabla s = \sum \omega^i \otimes s_i$ then $\nabla_X s = \sum \omega^i(X)\, s_i$ where $\omega^i(X)$ is the function on M made by pairing the vector field X with the 1-form ω^i.

Definition. The *curvature* of ∇ is the $\mathrm{End}(E)$-valued 2-form on M defined by

$$F^\nabla(X,Y) = \nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]}\,.$$

Remark. $[X,Y] = XY - YX$ is the commutator (or Lie bracket) of the two vector fields X, Y viewed as first-order differential operators $X : f \mapsto (d\,f)(X)$ on functions. Although $F^\nabla(X,Y)$ looks very much like a differential operator, it is in fact just a tensor field.

Problem. Use the Leibniz rule for ∇ to verify that $F^\nabla(X,Y)$ is a section of $\mathrm{End}(E)$; i.e., $F^\nabla(X,Y)\big|_x$ is a linear transformation (or endomorphism) of the fiber E_x. $\square$

We now return to our example of a rank-2 real vector bundle $E \to M$ with Euler form e and global angular form ψ. Define a tensor field $\mathcal{J} \in \mathrm{End}(E)$ by

$$\mathcal{J} e^1_\alpha = e^2_\alpha\,, \quad \mathcal{J} e^2_\alpha = -e^1_\alpha \quad (\text{on } E\big|_{U_\alpha}).$$

Thus $\mathcal{J}$ is the vector bundle analog of the rotation generator $J \in \mathfrak{so}_2$. Such a tensor field $\mathcal{J}$ is sometimes called an *almost complex structure* of E.

Problem. If ∇ is the covariant derivative determined by the global angular form ψ, show that the curvature of ∇ is given by the Euler form e :

$$F^\nabla = e \otimes 2\pi \mathcal{J}\,. \quad \square$$

Once the meaning of curvature is understood, the formula of the problem reinforces the interpretation of the Euler class $[e] \in H^2(M)$ as a measure of the twisting of the vector bundle $E \to M$.

2.2.2 Associated vector bundle

Since our explicit construction of ∇ in coordinates may appear ad hoc and unmotivated, we now wish to offer some perspective. We therefore embark on a brief detour into differential geometry. Apologies for the heavy-duty machinery introduced in the following subsection! (In fact, if you have no prior familiarity with the subject, you may have to consult a text on differential geometry to fully digest it. To protect you from getting overly worried: most of what comes afterwards will be independent of the material of Section 2.2.3.) The present investment will pay dividends in Section 2.3 on the Dirac quantization condition, where the Euler form e and global angular form ψ will take the roles of a magnetic field B and magnetic vector potential A, respectively.

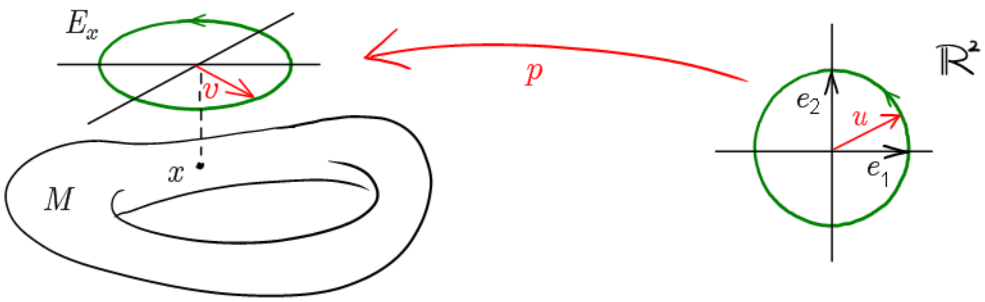
In working through the definitions of Section 2.2.3, the prime example to keep in mind is this. Given our oriented rank-2 real vector bundle $\pi: E \to M$ with Riemannian structure $\langle\cdot,\cdot\rangle$, let $\widetilde{\pi}: P \to M$ be the fiber bundle with fiber

$$\widetilde{\pi}^{-1}(x) = \mathrm{SO}(\mathbb{R}^2, E_x)\,,$$

i.e. an element $p \in \widetilde{\pi}^{-1}(x)$ is an orientation-preserving orthogonal transformation $p: \mathbb{R}^2 \to E_x$ from the oriented Euclidean plane $\mathbb{R}^2$ to the oriented fiber $E_x \simeq \mathbb{R}^2$. Notice that $\mathrm{SO}(\mathbb{R}^2, E_x)$ is in bijection with SO_2 as a set, but this is not an isomorphism of groups! Note also that each fiber $\widetilde{\pi}^{-1}(x)$ of P carries a right SO_2-action by composition:

$$p \mapsto p \cdot g := p \circ g \quad (g \in \mathrm{SO}_2).$$

Thus, as we shall learn presently, P is an example of a principal SO_2-bundle.



Moreover, the existence of the principal bundle P offers another vantage point on the vector bundle E as follows. Given a vector $v \in E_x$ and choosing some $p \in \mathrm{SO}(\mathbb{R}^2, E_x)$ we may express v as the image $v = p\,u$ of a vector $u \in \mathbb{R}^2$. This expression is not unique. Indeed, for any $g \in \mathrm{SO}_2$ we have $v = p\,u = (p \cdot g^{-1})(gu)$. Thus the attempt to factor v into p and u comes with a price: to achieve uniqueness, we need to identify the pair p, u with all pairs $(p \cdot g^{-1})(gu)$ and think of the vectors v as being in bijection with equivalence classes $[p\,;u]$:

$$P \times_{\mathrm{SO}_2} \mathbb{R}^2 \ni [p\,;u] \equiv [p \cdot g^{-1}; g\,u] \xrightarrow{1:1} p\,u = v \in E_x\,.$$

Thus there is an isomorphism $E_x \simeq \mathrm{SO}(\mathbb{R}^2, E_x) \times_{\mathrm{SO}_2} \mathbb{R}^2$ or, altogether,

$$E \simeq P \times_{\mathrm{SO}_2} \mathbb{R}^2\,.$$

One describes the situation by saying that the vector bundle E is *associated* to the principal bundle P and the vector space $\mathbb{R}^2$ (by the equivalence relation due to the joint SO_2-action). One calls $E = P \times_{\mathrm{SO}_2} \mathbb{R}^2$ an *associated vector bundle* for short.

2.2.3 Connection and curvature from principal fiber bundle

Definition 1. A *principal G-bundle* P (for a group G) over M is a fiber bundle $\widetilde{\pi}: P \to M$ carrying a right G-action $P \times G \to P$ which preserves the fibers of P and is free and transitive.

Remark 1a. ‘Free’ means that G acts without fixed points, while ‘transitive’ means that for any fixed $p_0 \in \widetilde{\pi}^{-1}(x)$ one has $p_0 \cdot G = \widetilde{\pi}^{-1}(x)$. These two properties imply that $\widetilde{\pi}^{-1}(x) \simeq G$ as sets (or topological spaces). Thus locally one has a factorization $P\big|_{U_\alpha} \simeq U_\alpha \times G$. Note, however, that the fiber $\widetilde{\pi}^{-1}(x)$ is *not* a group; in particular, there is no canonical choice of neutral element. This is evident from our example above, where we see clearly that there exists a priori no way of composing elements of the fiber $\widetilde{\pi}^{-1}(x) = \mathrm{SO}(\mathbb{R}^2, E_x)$.

Fact. A principal G-bundle $P \to M$ is trivial ($P \simeq M \times G$) iff there exists a global section.

Remark 1b. A major physics motivation for the notion of principal fiber bundle comes from gauge theory. In that setting, one identifies M with ordinary space (or space-time) and the fiber G acquires the physical meaning of gauge group; for example, $G = \mathrm{U}_1$ for electromagnetism, and $G = \mathrm{SU}_3$ for the strong interaction.

Definition 2. Let P be a principal G-bundle for a connected Lie group G. Inside the tangent space T_pP there exists a distinguished subspace V_p called the *vertical subspace* at p :

$$V_p := \left\{ \widehat{X}(p) \mid X \in \mathrm{Lie}(G) \right\}, \quad \widehat{X}(p) := \frac{d}{dt}\, p \cdot \mathrm{e}^{tX}\Big|_{t=0} .$$

A *principal connection* on P is a G-invariant $\mathrm{Lie}(G)$-valued 1-form ω on P with the property

$$\forall X \in \mathrm{Lie}(G)\,,\ \forall p \in P: \quad \omega_p(\widehat{X}(p)) = X\,, \quad \widehat{X}(p) \in V_p\,.$$

Remark 2a. The property of G-invariance of a principal connection ω means that

$$\forall p \in P,\ \ \forall v \in T_pP,\ \ \forall g \in G: \quad \omega_p(v) = \mathrm{Ad}(g)\, \omega_{p\cdot g}(dR_g(v)),$$

where $\mathrm{Ad}(g):\ \mathrm{Lie}(G) \to \mathrm{Lie}(G)$, $X \mapsto gXg^{-1}$ is the adjoint action, and dR_g is the differential of the right G-action $R_g(p) \equiv p \cdot g$. The mathematical raison d'etre for a principal connection is that it determines a G-invariant splitting $T_pP = V_p \oplus H_p$ (direct sum) where

$$H_p := \ker \omega_p = \{v \in T_pP \mid \omega_p(v) = 0\} \subset T_pP$$

is called the *horizontal subspace* at p.

Remark 2b. In the gauge theory context, the principal connection ω acquires the physical meaning of a gauge field. In more precise terms the statement is this. If $s:\ M \supset U \to P$ is a local section of the G-bundle $P \to M$, one defines a $\mathrm{Lie}(G)$-valued 1-form $A := s^*\omega$ by pullback along s. By definition, two different local sections s_1 and s_2 are related by

$$s_2(x) = s_1(x)\, g(x)\,,$$

where $g:\ U \to G$ is called a *gauge transformation*. The corresponding $\mathrm{Lie}(G)$-valued 1-forms $A^{(j)} = s_j^*\,\omega$ then are related by

$$A^{(2)} = g^{-1}dg + g^{-1}A^{(1)}g.$$

This is exactly the transformation law for a (non-abelian) gauge field A as known in physics. Thus from the present perspective, the freedom in choosing A comes from the freedom in choosing s.

The principal connection ω per se is the universal (or gauge-independent) object which arises by 'considering all gauges at once'.

Definition 3. Let $\widetilde{\pi}: P \to M$ be a principal G-bundle with principal connection ω. Given a curve $\gamma: [0, \epsilon] \to M$ one defines the *horizontal lift* $\widetilde{\gamma}: [0, \epsilon] \to P$ with initial point p_0, $\widetilde{\pi}(p_0) = \gamma(0)$, to be the curve determined by solving the first-order differential equation $(0 \le t \le \epsilon)$

$$\omega_{\widetilde{\gamma}(t)}\left(\frac{d}{dt}\widetilde{\gamma}(t)\right) = 0\,, \quad \widetilde{\pi} \circ \widetilde{\gamma} = \gamma\,, \quad \widetilde{\gamma}(0) = p_0\,.$$

Remark 3a. The differential equation amounts to saying that for all t the tangent vector $\frac{d}{dt}\widetilde{\gamma}(t)$ of the lifted curve lies in the horizontal subspace $H_{\widetilde{\gamma}(t)} = \ker\, \omega_{\widetilde{\gamma}(t)}$. By choosing a local section $s: U \to P$ one can express $\widetilde{\gamma}(t)$ by a mapping $g: [0, \epsilon] \to G$ as

$$\widetilde{\gamma}(t) = s(\gamma(t))\, g(t).$$

This ansatz takes care of the requirement $\widetilde{\pi} \circ \widetilde{\gamma} = \gamma$. The differential equation for the unknown gauge transformation function $g(t)$ then reads

$$\dot{g}(t)\, g(t)^{-1} + A_{\gamma(t)}(\dot{\gamma}(t)) = 0\,, \qquad A = s^*\omega\,.$$

For a closed curve $\gamma: [0,1] \to M$, $\gamma(0) = \gamma(1)$, the horizontal lift $\widetilde{\gamma}: [0,1] \to P$ will not be a closed curve in general. The element $g \in G$ determined by $\widetilde{\gamma}(1) = \widetilde{\gamma}(0)\, g$ is called the *holonomy* (along γ) of the principal connection ω.

Remark 3b. The celebrated *Berry phase* of quantum adiabatic dynamics is the holonomy of the so-called Berry (principal) connection of a principal U_1-bundle (see Section 2.4 below). The *Aharonov-Bohm effect* is another well-known example of holonomy due, in that case, to a flat connection on a non-simply connected domain.

Definition 4. In our special case of $E = P \times_{\mathrm{SO}_2} \mathbb{R}^2$ (or, more generally, for any associated vector bundle $E = P \times_G V$) the process of horizontal lifting of curves determines an isomorphism $\mathcal{T}_t: E_{\gamma(0)} \to E_{\gamma(t)}$ (referred to as *parallel transport* along γ) by

$$E_{\gamma(0)} \ni p_0\, v \mapsto \widetilde{\gamma}(t)\, v \in E_{\gamma(t)}\,. \quad \square$$

Finally, the notion of parallel transport gives rise to a covariant derivative ∇ on E as follows. If $Y \in \Gamma(TM)$ is a vector field, ones defines a differential operator $\nabla_Y: \Gamma(E) \to \Gamma(E)$ by

$$(\nabla_Y s)(x) := \lim_{t \to 0} \frac{\mathcal{T}_t^{-1} s(\gamma(t)) - s(x)}{t}\,,$$

where $\mathcal{T}_t: E_x \to E_{\gamma(t)}$ is parallel transport along a curve γ in M with $\gamma(0) = x$ and $\dot{\gamma}(0) = Y(x)$. One then gets $\nabla: \Gamma(E) \to \Gamma(T^*M \otimes E)$ by leaving the vector field argument Y in ∇_Y unspecified.

2.2.4 Covariant derivative from global angular form revisited

After this barrage of definitions, we return to our theme of Section 2.1.4: the covariant derivative corresponding to the global angular form $\psi \in \Omega^1(E^0)$ for an oriented real vector bundle $E \to M$ of rank 2. We aim for a more conceptual understanding of the origin of formula (2.5).

Notice that $\psi_\alpha = (2\pi)^{-1} d\theta_\alpha + \pi^*\xi_\alpha$ depends only on the angular coordinate θ_α (*not* the radial coordinate r_α) and is invariant under SO_2-rotations of each fiber. Hence, by the isomorphism

$$P \times_{\mathrm{SO}_2} \left(\mathbb{R}^2 \setminus \{0\}\right) \xrightarrow{\sim} E^0, \qquad [p\,; v] \equiv [p \cdot g^{-1}; gv] \mapsto p\, v,$$

the global angular form ψ on E^0 induces a principal connection ω on the principal SO_2-bundle P introduced in Section 2.2.2. This goes as follows. In the first step we pull back ψ to a 1-form $\tilde{\psi}$ on $P \times_{\mathrm{SO}_2} (\mathbb{R}^2 \setminus \{0\})$ along the isomorphism above. In the second step, we re-interpret $\tilde{\psi}$ as a principal connection ω on P as follows: if $v \in T_p P$ is a tangent vector at $p \in P$, we choose some curve $\gamma : \ (-\epsilon, \epsilon) \to P$ with $\gamma(0) = p$ and $\dot{\gamma}(0) = v$ and fix some non-zero vector $u \in \mathbb{R}^2 \setminus \{0\}$ to define

$$\omega_p(v) := \tilde{\psi}_{[p\,;u]} \left(\frac{d}{dt} \left[\gamma(t); u\right] \Big|_{t=0} \right) J,$$

where $J \in \mathfrak{so}_2$ is the generator determined by $\exp(2\pi J) = 1$ (and the orientation class of E).

Problem. Show that ω is well-defined, i.e. does not depend on the choice of u and γ. Show also that ω has the properties required of a principal connection. $\square$

By the general principles outlined in Section 2.2.3 the principal connection ω on P determines a connection ∇ on E. This, ultimately, is the rationale behind (2.5).

2.3 Application: Dirac monopole problem

We continue with a few words about a foundational theme of quantum mechanics: the quantization of electric charge. (This will be brief; for a more expansive and leisurely account, see my lecture notes on Advanced Quantum Mechanics.) It is an experimental finding that electric charge always occurs as an integer multiple $q_e = ne$ $(n \in \mathbb{Z})$ of a fundamental charge quantum e. Why nature has arranged for it to be that way is an open question of theoretical physics.

However, if magnetic monopoles exist, charge quantization can be understood by an argument due to Dirac (1931), who showed that quantum mechanics is consistent if and only if the product of any pair q_e, q_m of electric and magnetic charges is an integer multiple of $2\pi\hbar$:

$$q_e\, q_m \in 2\pi\hbar\,\mathbb{Z}\,.$$

This condition, known as the Dirac quantization condition, can be read in two directions. Given a smallest magnetic charge μ, it quantizes the electric charge according to $q_e \in (2\pi\hbar/\mu)\,\mathbb{Z}$. Conversely, given an electric charge quantum e, magnetic charge is quantized by $q_m \in (2\pi\hbar/e)\,\mathbb{Z}$.

The plan of this subsection is to give some indication of the mathematics behind the Dirac quantization condition. To begin, let us recall that in textbook versions of the Schrödinger quantum mechanics of a charged particle moving in a magnetic field $\vec{B}$, one is instructed to express $\vec{B}$ as the curl of $\vec{A}$ (for some choice of gauge) and take the Hamiltonian to be $H = (\vec{p} - e\vec{A})^2/2m$. This recipe fails in the presence of magnetic monopoles. Indeed, the total magnetic flux through a closed surface should be equal to the enclosed magnetic charge, but at the same it vanishes for

for $\vec{B} = \operatorname{curl} \vec{A}$ by Stokes' theorem. Thus the existence of magnetic monopoles is incompatible with the existence of a magnetic vector potential $\vec{A}$ of the usual kind.

One therefore has to proceed in a different fashion. The modern approach based on the fiber bundle concept goes a follows. Suppose there are (very massive, and hence static) magnetic monopoles at positions $p_1, \ldots, p_m$ in $\mathbb{R}^3$. The configuration space M for a charged particle then is defined as $\mathbb{R}^3$ with these points removed:

$$M = \mathbb{R}^3 \setminus \{p_1, \ldots, p_m\}\,.$$

The wave function of a particle with electric charge, say q_e, will be a section of some Hermitian line bundle $\pi :\ E \to M$. 'Hermitian' here means that the fiber $E_x \simeq \mathbb{C}$ carries a Hermitian structure, i.e. for every $x \in M$ there exists $\langle \cdot , \cdot \rangle_x :\ E_x \times E_x \to \mathbb{C}$. (Note that owing to $E_x \simeq \mathbb{C}$, wave functions can still be viewed as being locally complex-valued, just like in textbook quantum mechanics.) We assume that the structure group of E has been reduced to $G = \mathrm{U}_1$.

The information about the magnetic field due to the static monopoles (as well as any moving electric charges) is encoded in a principal connection 1-form ω on a principal U_1-bundle $P \to M$. This 1-form ω determines on $E \simeq P \times_{\mathrm{U}_1} \mathbb{C}$ a covariant derivative ∇, whose local expression is $\nabla = d - \mathrm{i} q_e A/\hbar$ for a choice of gauge potential $A := (\mathrm{i} q_m / 2\pi)\, s^* \omega$ and magnetic charge q_m. Its curvature $F^\nabla \equiv \nabla^2 = (q_e q_m / h)\, s^*(d\omega)$ is proportional to the magnetic field-strength 2-form $B = dA$. The physical meaning of the first-order differential operator $(\hbar/\mathrm{i})\nabla$ is that of quantum mechanical momentum of our particle with electric charge q_e. From this perspective, the Dirac quantization condition $q_e\, q_m / h \in \mathbb{Z}$ is simply a necessary and sufficient condition for the existence of the Hermitian line bundle E – and hence of wave functions $\Psi \in \Gamma(E)$; by the axioms of quantum theory the latter are required to be globally defined and single-valued.

Let us look at the computational details for the special case of a single magnetic monopole located at the origin of our coordinate system. Thus $M = \mathbb{R}^3 \setminus \{0\}$. We cover M by two open subsets $\{U_+\,, U_-\}$ where U_+ (U_-) is M with the negative (positive) z-axis removed. Then, using the standard system r, θ, ϕ of spherical polar coordinates, consider the transition function

$$g_{+-} \equiv g = \mathrm{e}^{\mathrm{i}\phi} :\ U_+ \cap U_- \to \mathrm{U}_1\,,$$

which is defined everywhere on $U_+ \cap U_- = \mathbb{R}^3 \setminus \{z\text{-axis}\}$. In the present setting, the principal connection 1-form ω on P has the local expressions

$$\omega\big|_{U_+} = \mathrm{i} d\psi_+ - (\mathrm{i}/2)(1 - \cos\theta) d\phi\,, \qquad \omega\big|_{U_-} = \mathrm{i} d\psi_- + (\mathrm{i}/2)(1 + \cos\theta) d\phi\,,$$

where $\psi_\pm$ are local coordinates for the U_1-fibers on $U_\pm$. These expressions match because $\mathrm{e}^{\mathrm{i}\psi_+} = g_{+-}\, \mathrm{e}^{\mathrm{i}\psi_-}$ and hence $d\psi_+ = d\psi_- + d\phi$. If we make the choice of gauge $d\psi_\pm = 0$ on $U_\pm$, the gauge potential for a magnetic monopole of charge μ is

$$A^\pm = (\mathrm{i}\mu/2\pi)\, s_\pm^* \,\omega = \pm(\mu/4\pi)(1 \mp \cos\theta)\, d\phi\,,$$

resulting in the same magnetic field strength $B = dA^{\pm} = (\mu/4\pi)\sin\theta\, d\theta \wedge d\phi$ for both domains, U_+ and U_-. Note that the magnetic flux totals

$$\int_\Sigma B = \mu$$

through any (suitably oriented) closed surface Σ surrounding the monopole at the origin.

We turn to the description of wave functions or sections $\Psi \in \Gamma(E)$. On $U_+ \cap U_-$ we have

$$\Psi = f_+ e_+ = f_- e_- \quad \text{where} \quad f_+ = g_{+-} f_- = \mathrm{e}^{\mathrm{i}\phi} f_- \quad \text{and} \quad e_- = e_+ g_{+-} = e_+ \mathrm{e}^{\mathrm{i}\phi},$$

and the covariant derivative is expressed by

$$\nabla\Psi = \big(df_+ - (\mathrm{i}/2)(1-\cos\theta) f_+ \, d\phi\big)\, e_+ = \big(df_- + (\mathrm{i}/2)(1+\cos\theta) f_- \, d\phi\big)\, e_- \,.$$

By writing this in the form $\nabla\Psi = \big(df_+ - \mathrm{i}qA^+/\hbar\big)\, e_+ = \big(df_- - \mathrm{i}qA^-/\hbar\big)\, e_-$ and comparing with the expression for $A^\pm$ above, we infer the relation $\mu/4\pi = \hbar/2q$ or

$$q\mu = h\,,$$

which is the Dirac quantization condition at the elementary level of one electric charge quantum q and one magnetic charge quantum μ.

For the fundamental case of $q\mu = h$ a pair of wave functions $\Psi^{(1)} = f_+^{(1)} e_+ = f_-^{(1)} e_-$ and $\Psi^{(2)} = f_+^{(2)} e_+ = f_-^{(2)} e_-$ (with minimal angular momentum) is given by

$$\begin{aligned} f_+^{(1)} &= f^{(1)}(r)\, \cos(\theta/2), & f_-^{(1)} &= f^{(1)}(r)\, \cos(\theta/2)\, \mathrm{e}^{-\mathrm{i}\phi}, \\ f_+^{(2)} &= f^{(2)}(r)\, \sin(\theta/2)\, \mathrm{e}^{\mathrm{i}\phi}, & f_-^{(2)} &= f^{(2)}(r)\, \sin(\theta/2)\,. \end{aligned}$$

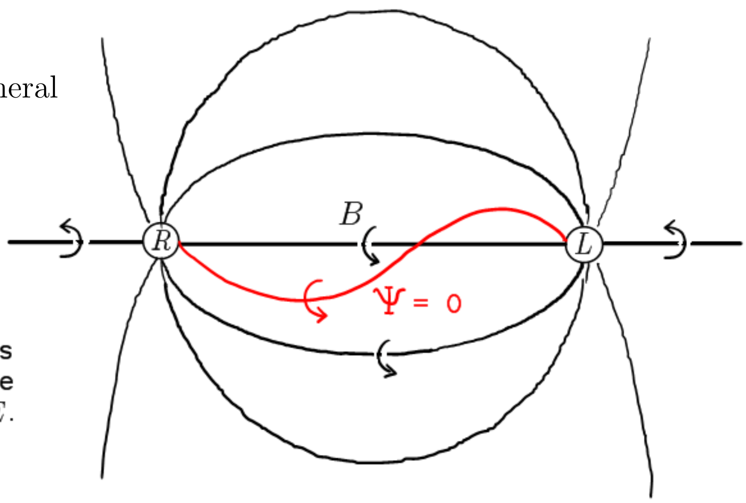
We observe that $\Psi^{(1)}$ has a nodal line emanating from the origin (where the magnetic monopole is located) along the negative z-axis. Similarly, $\Psi^{(2)}$ has a nodal line along the positive z-axis.

Problem. Construct the Euler class $[e]$ (actually, the first Chern class) of E and show that $\int_S e = 1$ for any surface S enclosing the magnetic monopole. □

A section is called *transversal* if it intersects the zero section transversally.

Fact. The first Chern class of a Hermitian line bundle $E \to M$ is Poincaré dual to the zero locus of a transversal section.

Remark. This is a special case of a more general theorem stated and proved in Bott & Tu. □



By Poincare duality, the nodal lines of a wave function Ψ represent the first Chern class associated with E.

2.4 Application: Berry phase

Let there be a family of Hamiltonians $H(x)$ depending on a set of parameters $x = (x_1, \ldots, x_m)$ in a parameter space M. Our object of interest is a system with parameters varying along a curve

$$\gamma : [0,1] \to M, \quad \tau \mapsto \gamma(\tau).$$

Let $h := H \circ \gamma$ be the 1-parameter family of Hamiltonians along this curve. Consider then for some large time parameter T the time-dependent Schrödinger equation

$$\frac{\partial}{\partial t} \Psi(t) = -\frac{\mathrm{i}}{\hbar} h(t/T) \Psi(t) \qquad (0 \leq t \leq T).$$

We wish to communicate a certain geometric fact (namely, Berry's phase) about the solutions of this equation in the adiabatic limit $T \to \infty$. For that purpose, it is convenient to make the substitutions $t = \tau T$ and $\Psi(t) = \Psi(\tau T) \equiv \Psi_T(\tau)$, bringing the equation to the form

$$\frac{\partial}{\partial \tau} \Psi_T(\tau) = -\mathrm{i}\frac{T}{\hbar} h(\tau) \Psi_T(t) \qquad (0 \leq \tau \leq 1).$$

In what follows we assume the spectrum of $H(x)$ to be discrete for all parameter values $x \in M$. We denote the ground state energy of $H(x)$ by $E_0(x)$, the energy of the first excited state by $E_1(x)$, and so on. We also assume that, along our curve γ, energy eigenvalues do not cross:

$$\forall \tau \in [0,1]: \quad E_0(\gamma(\tau)) < E_1(\gamma(\tau)) < \ldots < E_n(\gamma(\tau)) < \ldots,$$

and all multiplicities are equal to one. Thus, denoting by $V_n(x)$ the eigenspace of $H(x)$ with eigenvalue $E_n(x)$, we have

$$\forall \tau \in [0,1]: \quad \dim V_n(\gamma(\tau)) = 1 \qquad (n = 0, 1, 2, \ldots).$$

This type of situation is governed by the Quantum Adiabatic Theorem:

Fact (Born & Fock, 1928). If $\Psi(0) \in V_n(\gamma(0))$ then $\lim_{T\to\infty} \Psi_T(1) \in V_n(\gamma(1))$. $\square$

Remark. The quantum adiabatic theorem was proved by Born and Fock under some weak technical conditions on $H(x)$ not recorded here. In words it says that a quantum Hamiltonian system remains in its instantaneous energy eigenstate if a given perturbation is acting on it slowly enough and there is a gap between the energy eigenvalue and the rest of the energy spectrum. $\square$

Next, we take our curve in parameter space to be closed: $\gamma(1) = \gamma(0)$. Then by the quantum adiabatic theorem $\Psi(0)$ and $\lim_{T\to\infty} \Psi_T(1)$ lie in the same eigenspace $V_n(\gamma(0)) = V_n(\gamma(1))$, and it makes sense to compare phases. Removing the obvious dynamical phase $(T/\hbar) \int_0^1 E_n(\gamma(\tau))\, d\tau$ one expects

$$\lim_{T\to\infty} \mathrm{e}^{(\mathrm{i}T/\hbar) \int_0^1 E_n(\gamma(\tau))\, d\tau} \Psi_T(1) = \mathrm{e}^{-\mathrm{i}\alpha} \Psi(0).$$

You might have thought that the additional phase $\alpha \sim O(T^0)$ would be zero. However, this naive expectation is false, as was first explained to the physics community by M.V. Berry (1985). To write an expression for α, which is called Berry's geometric phase, one makes some choice of adiabatic eigenstates $\phi_n(\tau) \in V_n(\gamma(t))$ along the curve γ. (It turns out to be impossible in general to make a smooth choice $x \mapsto \phi_n(x)$ for all $x \in M$.) Doing so, the Berry phase is expressed as

$$\alpha = \frac{1}{\mathrm{i}} \int_0^1 d\tau \left\langle \phi_n(\tau), \frac{\partial}{\partial \tau} \phi_n(\tau) \right\rangle ,$$

which is independent of the choice of $\phi_n(\tau)$ if α is viewed as an angle, i.e., as representing an equivalence class $[\alpha] = [\alpha + 2\pi\mathbb{Z}] \in \mathbb{R}/2\pi\mathbb{Z}$. Indeed, if $\tilde{\phi}_n : \ \tau \mapsto V_n(\gamma(\tau))$ is another choice of adiabatic eigenstates, then there exists $\mathrm{e}^{\mathrm{i}\beta} : \ [0,1] \to \mathrm{U}_1$ such that $\tilde{\phi}_n(\tau) = \mathrm{e}^{\mathrm{i}\beta(\tau)}\, \phi_n(\tau)$ and

$$\frac{1}{\mathrm{i}} \int_0^1 d\tau \left\langle \tilde{\phi}_n(\tau), \frac{\partial}{\partial \tau} \tilde{\phi}_n(\tau) \right\rangle = \alpha + \int_0^1 \dot{\beta}(\tau)\, d\tau \in \alpha + 2\pi\mathbb{Z}.$$

Let us now verify Berry's formula for α. If $\Psi(0) \in V_n(\gamma(0))$, then for some choice of family $\phi_n(\tau)$ we make the ansatz

$$\Psi_T(\tau) = \mathrm{e}^{-\mathrm{i}\varphi(\tau)} \phi_n(\tau) + \dots ,$$

where the corrections (indicated by the dots) vanish in the adiabatic limit $T \to \infty$. By inserting this ansatz into the Schrödinger equation, we obtain

$$\mathrm{e}^{-\mathrm{i}\varphi(\tau)} \big(\dot{\varphi}(\tau)\, \phi_n(\tau) + \mathrm{i}\dot{\phi}_n(\tau) \big) + \dots = \mathrm{i} \frac{\partial}{\partial \tau} \Psi_T(\tau) = \frac{T}{\hbar} h(\tau) \Psi_T(\tau) = \frac{T}{\hbar} E_n(\gamma(\tau))\, \mathrm{e}^{-\mathrm{i}\varphi(\tau)} \phi_n(\tau) + \dots ,$$

and hence

$$\dot{\varphi}(\tau) + \mathrm{i} \left\langle \phi_n(\tau), \dot{\phi}_n(\tau) \right\rangle = \frac{T}{\hbar} E_n(\gamma(\tau)) + \dots .$$

By integrating this equation and passing to the limit $T \to \infty$, it follows that

$$\alpha = \lim_{T \to \infty} \left(\varphi(1) - \varphi(0) - \frac{T}{\hbar} \int_0^1 E_n(\gamma(\tau))\, d\tau \right) = \frac{1}{\mathrm{i}} \int_0^1 \left\langle \phi_n(\tau), \dot{\phi}_n(\tau) \right\rangle d\tau ,$$

which is Berry's formula.

We will now use the Berry phase to illustrate various notions and constructions of Sections 2.1 and 2.2. First of all, fixing some value of the quantum number n, we remove from the parameter space all points x where the energy level $E_n(x)$ becomes degenerate with another level. (By the so-called Wigner-von Neumann principle, such points typically form a submanifold of co-dimension three.) We still denote the resulting parameter space by M. Since $\dim V_n(x) = 1$ for all $x \in M$, we have a complex line bundle $\pi : \ E \to M$ with fiber $\pi^{-1}(x) = E_x \equiv V_n(x)$.

In each fiber E_x there exists the circle of unit vectors:

$$P_x := \{ \phi \in V_n(x) \mid \langle \phi, \phi \rangle = 1 \}.$$

The totality of these unit vectors form a principal U_1-bundle $\widetilde{\pi} : \ P \to M$ with fiber $\widetilde{\pi}^{-1}(x) = P_x \simeq \mathrm{U}_1$. Its total space P is canonically equipped with a principal connection 1-form, called the Berry connection. To describe it, let $\Phi : \ P \to \mathcal{H}$ be the tautological mapping which simply recalls what the points of P are, namely vectors in the Hilbert space $\mathcal{H}$ of the quantum Hamiltonian system.

By differentiating this mapping we get $d\Phi : TP \to \mathcal{H}$. The Hermitian scalar product of $d\Phi$ with Φ is the Berry connection 1-form,

$$\omega := \langle \Phi, d\Phi \rangle .$$

Problem. Show that ω has all the properties required of a principal connection. □

To link this up with Berry's formula, let $\phi_n : M \to P$, $x \mapsto \phi_n(x)$ be a local section and consider the pullback $A := \mathrm{i}^{-1} \phi_n^* \omega = \mathrm{i}^{-1} \langle \phi_n, d\phi_n \rangle$. The Berry phase along the closed curve γ is the line integral

$$\alpha(\gamma) = \oint_\gamma A = \frac{1}{\mathrm{i}} \int_0^1 \left\langle \phi_n(\gamma(\tau)), \frac{\partial}{\partial \tau} \phi_n(\gamma(\tau)) \right\rangle d\tau .$$

By Stokes' theorem, this can also be written as a surface integral over $\Sigma \subset M$ with $\partial\Sigma = \gamma$:

$$\alpha(\partial\Sigma) = \oint_{\partial\Sigma} A = \iint_\Sigma dA .$$

Another perspective on the Berry phase is offered by the procedure of using the principal connection to lift the curve $\gamma : [0,1] \to M$ to a horizontal curve $\widetilde{\gamma} : [0,1] \to P$. By going back to our verification of Berry's formula, we see that $\alpha = \varphi(1) - \varphi(0)$ arises as the solution of the differential equation

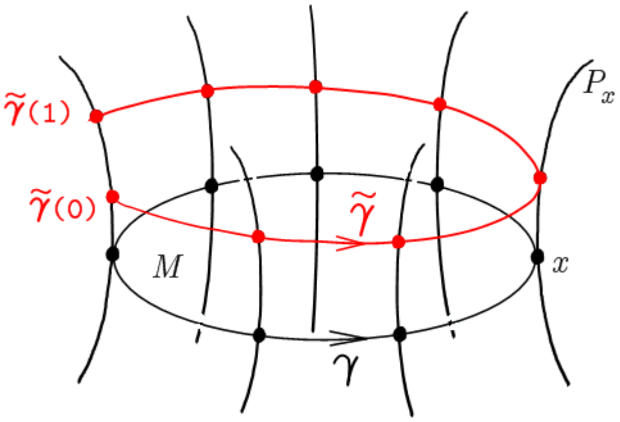
$$-\mathrm{i}\, \dot{\varphi}(\tau) + \left\langle \phi_n(\gamma(\tau)), \frac{\partial}{\partial \tau} \phi_n(\gamma(\tau)) \right\rangle = 0.$$

This is nothing but the equation

$$\omega_{\widetilde{\gamma}(\tau)} \left(\frac{d}{d\tau} \widetilde{\gamma}(\tau) \right) = 0$$

for the horizontal lift $\widetilde{\gamma}(\tau) = \mathrm{e}^{-\mathrm{i}\varphi(\tau)} \phi_n(\gamma(\tau))$ of γ.

Thus the Berry phase can be interpreted as the holonomy $\mathrm{e}^{-\mathrm{i}\alpha} = \widetilde{\gamma}(1)/\widetilde{\gamma}(0)$ of the lifted curve.



As usual, ω induces a covariant derivative ∇ on E. Its curvature $F := \mathrm{i}\nabla^2$ represents the first Chern class $[F/2\pi] \in H^2(M)$ associated with the vector bundle E. The class $[F/2\pi]$ is non-trivial in the presence of level crossings (called *diabolical points* in some of the older literature), as these make the vector bundle E twist.

Example. Consider a spin $S = 1/2$ particle with magnetic moment $\mu\,\vec{\sigma}$ in a magnetic field $\vec{B}$:

$$H(\vec{B}) = \mu \vec{B} \cdot \vec{\sigma} = \mu\, r \begin{pmatrix} \cos\theta & \sin\theta\, \mathrm{e}^{-\mathrm{i}\phi} \\ \sin\theta\, \mathrm{e}^{\mathrm{i}\phi} & -\cos\theta \end{pmatrix}, \qquad \vec{B} = r \begin{pmatrix} \sin\theta \,\cos\phi \\ \sin\theta\, \sin\phi \\ \cos\theta \end{pmatrix}.$$

There exist two energy levels, $\varepsilon_0 = -\mu\, r$ and $\varepsilon_1 = +\mu\, r$, which are non-degenerate for $\vec{B} \in M := \mathbb{R}^3 \setminus \{\vec{0}\}$. Fix either one of the two levels, say ε_0 for concreteness. A local coordinate description of the mapping $\Phi: P \to \mathcal{H} = \mathbb{C}^2$ for the corresponding principal bundle $P \to M$ of normalized eigenvectors, is given by

$$\Phi = \mathrm{e}^{\mathrm{i}\psi_+} \begin{pmatrix} \sin(\theta/2)\,\mathrm{e}^{-\mathrm{i}\phi} \\ -\cos(\theta/2) \end{pmatrix} = \mathrm{e}^{\mathrm{i}\psi_-} \begin{pmatrix} \sin(\theta/2) \\ -\cos(\theta/2)\,\mathrm{e}^{\mathrm{i}\phi} \end{pmatrix}.$$

The principal connection

$$\omega = \langle \Phi, d\Phi \rangle = \mathrm{i}d\psi_+ - \mathrm{i}\sin^2(\theta/2)\,d\phi = \mathrm{i}d\psi_- + \mathrm{i}\cos^2(\theta/2)\,d\phi$$

is seen to be exactly the same that we encountered in the context of the Dirac monopole problem with fundamental charges (Section 2.3). In fact, the two problems are mathematically equivalent, and one has (for the present example) the following precise correspondences:

Dirac monopole problem	$\leftrightarrow$	Berry phase problem
position space		parameter space
monopole		level degeneracy
complex lines $E_x \simeq \mathbb{C}$		instantaneous eigenspaces
gauge connection		Berry connection

Problem 1. Show that switching the two energy levels, $\varepsilon_0 \leftrightarrow \varepsilon_1$, corresponds to reversing the handedness of the magnetic monopole, $R \leftrightarrow L$.

Problem 2. For the case of arbitary spin S there exist $2S+1$ energy levels and hence $2S+1$ Berry line bundles over $M = \mathbb{R}^3 \setminus \{0\}$. Are all of these bundles and their Berry connections isomorphic to some Dirac monopole bundle? If so, what are the corresponding monopole charges?

3 Supersymmetry and Morse Theory

In 1982, E. Witten pointed out that the de Rham complex $\Omega(M)$ can be interpreted as the Hilbert space of a supersymmetric quantum mechanics [see J. Diff. Geo. 17, p. 661] and that by deforming the latter to a harmonic oscillator problem one can understand the so-called Morse inequalities, a classical result in the topology of compact manifolds. The purpose of the present chapter is to communicate this brilliant insight, which initiated a fruitful and lasting interaction between topology and the physics of supersymmetry. Among its many variations and ramifications, an outstanding result is the so-called heat kernel proof of the Atiyah-Singer index theorem.

3.1 Morse inequalities

We begin with a number of definitions. Let f be a differentiable function on a manifold M of dimension n. A point $x \in M$ with the property $(d\,f)_x = 0$ is referred to as a *critical* point of f. One calls such a point x *non-degenerate* if the Hessian $\mathrm{Hess}_x(f)$ is non-degenerate as a quadratic form. If all critical points of a function f are non-degenerate, then f is called a *Morse function.*

The *index,* $\mathrm{ind}_x(f)$, of a non-degenerate critical point x of f is defined as the number of negative eigenvalues of $\mathrm{Hess}_x(f)$. If the index is zero (maximal) one has a local minimum (resp. maximum). For an intermediate index, $0 < \mathrm{ind}_x(f) < n$, the point x is a saddle point of f.

One may want to know the number of critical points with index q of a function $f : M \to \mathbb{R}$. (Think, for example, of the Hamiltonian function of classical mechanics. Its critical points are the local equilibria of the Hamiltonian dynamics.) By the following statement, known as the Morse inequalities, this number is bounded below by the de Rham cohomology $\Omega^q(M)$.

Theorem. Let M be compact, and assume that $f \in \mathrm{C}^\infty(M)$ is a Morse function. Then if m_q denotes the number of critical points of f with index q, one has

1. $m_q \geq b_q \qquad (q = 0, 1, \ldots, n)$,
2. $m_q - m_{q-1} + \ldots \pm m_0 \geq b_q - b_{q-1} + \ldots \pm b_0 \qquad (q = 0, 1, \ldots, n)$,
3. $\sum_{q=0}^n (-1)^q \, m_q = \sum_{q=0}^n (-1)^q \, b_q$,

where $b_q \equiv \dim H^q(M)$ is the q^{th} Betti number.

Remark. Statements 1 and 2 are called the weak resp. strong form of the Morse inequalities. Statement 3 says that the Euler characteristic $\chi(M) = \sum_{q=0}^n (-1)^q \dim H^q(M)$ can be computed as an alternating sum of the number of critical points of a Morse function.

Problem. Use induction on q to show that the strong form of the Morse inequalities implies the weak form.

Example. The de Rham cohomology of the two-sphere S^2 is $b_0 = b_2 = 1$ and $b_1 = 0$. Thus a Morse function on S^2 must have at least one minimum and one maximum, and there need not be any saddle points. An example of such a function is the height function $f = \cos\theta$. For the two-torus $\mathrm{T}^2 = \mathrm{S}^1 \times \mathrm{S}^1$ one has $b_0 = b_2 = 1$ and $b_1 = 2$. In this case there must exist at least two saddle points in addition to the obligatory minimum and maximum.

Problem. Think of a Morse function for T^2. $\square$

Let us finish this introductory subsection by alerting the student to the following aspect. To define the Hessian of a function $f \in \mathrm{C}^\infty(M)$ in general, you must choose a covariant derivative: $\mathrm{Hess}(f) = \nabla(\mathrm{grad}\, f)$. (Indeed, if you try to define the Hessian as the matrix of second partial derivatives, you discover that your definition depends on the choice of local coordinates.) However, in the special case of a critical point this choice doesn't matter. The reason is that any two covariant derivatives ∇ and ∇' differ only by an $\mathrm{End}(TM)$-valued 1-form, $\nabla' - \nabla = \omega$, and hence

$$\big(\nabla(\mathrm{grad} f) - \nabla'(\mathrm{grad} f)\big)_x = \omega_x\big((\mathrm{grad} f)(x)\big) = 0 \, .$$

Problem. Show that the index of a critical point is well-defined, i.e. does not depend on the choice of local coordinates used to compute the Hessian as a matrix of partial derivatives.

3.2 Supersymmetric quantum mechanics

The Hilbert space, say V, of supersymmetric quantum mechanics comes with an orthogonal decomposition

$$V = V_0 \oplus V_1$$

into an 'even' subspace V_0 and an 'odd' subspace V_1. Such a decomposition is called a $\mathbb{Z}_2$-grading. One sometimes refers to V_0 and V_1 as the bosonic and fermionic subspaces respectively.

On V one is given a self-adjoint operator $Q = Q^\dagger$ which is a linear mapping between the two subspaces:

$$Q : V_0 \to V_1 \quad \text{and} \quad Q : V_1 \to V_0 .$$

One says that Q is odd with respect to the $\mathbb{Z}_2$-grading and calls it a 'fermionic charge'. The Hamiltonian of the supersymmetric quantum mechanics is taken to be the square $H = Q^2$. It is a pair of linear operators

$$H : V_0 \to V_0 \quad \text{and} \quad H : V_1 \to V_1 .$$

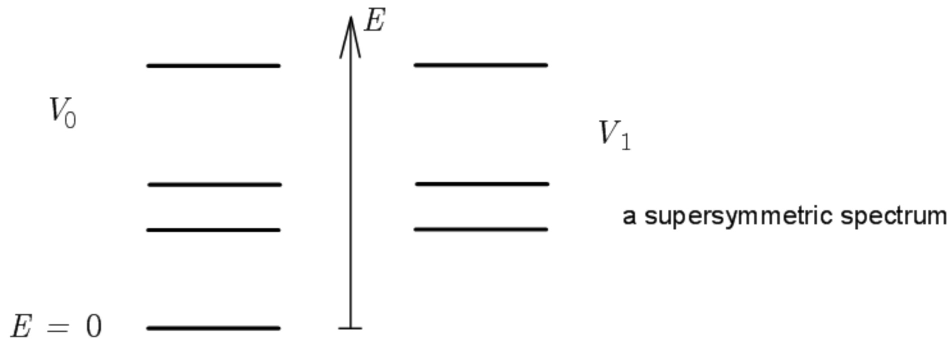
The spectrum of a supersymmetric Hamiltonian $H = Q^2$ has the following special properties. Let $\psi \in V_j$ for $j \in \{0, 1\}$ be an eigenvector: $H\psi = E\psi$. Then $Q\psi \in V_{1-j}$ is an eigenvector with the same eigenvalue:

$$H(Q\psi) = Q(H\psi) = E\,(Q\psi).$$

If $\psi \neq 0$ and $E \neq 0$ then $Q\psi$ is not the zero vector:

$$\langle Q\psi, Q\psi\rangle = \langle \psi, H\psi\rangle = E\,\langle \psi, \psi\rangle \neq 0 .$$

This means that $Q : V_0 \to V_1$ restricted and projected to the sector of excited states is a spectrum-preserving isomorphism. On the other hand, by the same argument one has $Q\psi = 0$ for $E = 0$. Thus the boson-fermion correspondence between excited states breaks down for the ground state sector.



Definition. The Witten index of a supersymmetric Hamiltonian H is

$$I_W(H) := \dim\ker H\big|_{V_0} - \dim\ker H\big|_{V_1} .$$

Thus $I_W(H)$ is the difference between the number of bosonic and fermionic ground states. □

An important property of the Witten index is that it does not change under continuous deformations of H. Indeed, although states may enter and leave the ground state sector $E = 0$ as some parameters in H are varied, they must do so as pairs of one bosonic and one fermionic state each. Thus the Witten index is a kind of topological invariant.

The following formula expresses the Witten index as a supertrace:

$$I_W(H) = \mathrm{Tr}_{V_0}\, \mathrm{e}^{-tH} - \mathrm{Tr}_{V_1}\, \mathrm{e}^{-tH} =: \mathrm{STr}_V\, \mathrm{e}^{-tH},$$

where t is any positive real number. In the limit of $t \to \infty$ one recovers the definition of $I_W(H)$.

Example. Let $V = L^2(\mathbb{R}) \otimes \mathbb{C}^2$ be the Hilbert space of a particle with spin $S = 1/2$ in one dimension, and take the Hamiltonian to be supersymmetric, $H = Q^2$, with

$$Q = \begin{pmatrix} 0 & \partial_x + f'(x) \\ -\partial_x + f'(x) & 0 \end{pmatrix}, \quad \partial_x \equiv \frac{\partial}{\partial x},$$

where f is a differentiable function on $\mathbb{R}$ with enough growth at infinity so that $\int_{\mathbb{R}} \mathrm{e}^{-2f(x)} dx < \infty$.

Problem. Compute the Witten index $I_W(H) = \mathrm{STr}\, \mathrm{e}^{-tH}$ of this problem for $t \to \infty$ and $t \to 0$.

3.2.1 De Rham complex and supersymmetric quantum mechanics

Let M be a Riemannian manifold M of dimension n and consider its space of differential forms, $\Omega(M)$. The Riemannian structure of M determines a Hodge star operator,

$$\star: \ \Omega^k(M) \to \Omega^{n-k}(M, L),$$

by linear extension of the formula

$$\star \left(dx^{i_1} \wedge \ldots \wedge dx^{i_k}\right) = \sum g^{i_1 j_1} \cdots g^{i_k j_k} \, \iota(\partial_{j_k}) \cdots \iota(\partial_{j_1}) \, \mathrm{dvol}_M,$$

where g^{ij} are the components of the metric tensor and $\mathrm{dvol}_M = \sqrt{g}\,|dx^1 \wedge \cdots \wedge dx^n|$ with $\sqrt{g} \equiv \sqrt{\det(g_{ij})}$ is the Riemannian volume density. The star operator in turn determines an L^2-space of differential forms by the scalar product

$$\langle \alpha, \beta \rangle := \int_M \alpha \wedge \star\beta.$$

The $\mathbb{Z}$-grading $\Omega(M) = \oplus_{k \geq 0} \Omega^k(M)$ is orthogonal with respect to this scalar product.

Remark. The constitutive laws of Maxwell electrodynamics are $D = \varepsilon_0 \star E$ and $B = \mu_0 \star H$. The electromagnetic field energy is expressed as

$$\frac{1}{2} \int_{\mathbb{R}^3} (E \wedge D + B \wedge H) = \frac{\langle D, D \rangle}{2\varepsilon_0} + \frac{\langle B, B \rangle}{2\mu_0}. \quad \square$$

Next, define the co-derivative $\delta: \ \Omega^{k+1}(M) \to \Omega^k(M)$ by

$$\delta = (-1)^k \star^{-1} d \star .$$

For a compact manifold M this operator satisfies

$$\langle \alpha, d\beta \rangle = \int_M \alpha \wedge \star d\beta = \int_M d\beta \wedge \star\alpha = -\int_M \beta \wedge \star\delta\alpha = -\langle \delta\alpha, \beta \rangle,$$

so $-\delta \equiv d^\dagger$ is the adjoint of d. Note that $\delta^2 = 0$.

Problem. By using the coordinate expression for the Hodge star operator, verify the following formula for $\delta: \Omega^{k+1}(M) \to \Omega^k(M)$:

$$\delta: \sum_{i_0 < \ldots < i_k} \omega_{i_0 i_1 \ldots i_k} \, dx^{i_0} \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k} \mapsto \sum_{i_1 < \ldots < i_k} (\delta\omega)_{i_1 \ldots i_k} \, dx^{i_1} \wedge \cdots \wedge dx^{i_k} \quad \text{where}$$

$$(\delta\omega)_{i_1 \ldots i_k} = \sum g_{i_1 j_1} \cdots g_{i_k j_k} \frac{1}{\sqrt{g}} \partial_j \left(\sqrt{g}\, \omega^{j j_1 \ldots j_k}\right), \quad \omega^{j j_1 \ldots j_k} = \sum g^{jl} g^{j_1 l_1} \cdots g^{j_k l_k} \omega_{l l_1 \ldots l_k}. \quad \square$$

To set up a supersymmetric quantum mechanics, we take V_0 (resp. V_1) to be the space of square-integrable differential forms of even (resp. odd) degree, and let $V = V_0 \oplus V_1$ and $Q := d - \delta = d + d^\dagger$. The supersymmetric Hamiltonian is

$$\mathcal{H} = Q^2 = (d-\delta)^2 = -(\delta d + d\delta) \equiv -\triangle\,,$$

where $\triangle \equiv \delta d + d\delta$ is called the Laplacian on differential forms.

Problem. Show that by the isomorphism $\mathcal{I}(dx^i) = \sum g^{ij}\partial_j$ sending 1-forms to vector fields, one has the following correspondences for the case of $M = \mathbb{R}^3$:

$$\begin{aligned} \triangle\big|_{\Omega^0(\mathbb{R}^3)} &= \delta d = \mathrm{div} \circ \mathrm{grad}\,, \\ \triangle\big|_{\Omega^1(\mathbb{R}^3)} &= \mathcal{I}^{-1}\left(\mathrm{grad} \circ \mathrm{div} - \mathrm{curl} \circ \mathrm{curl}\right) \circ \mathcal{I}\,. \end{aligned}$$

3.3 Hodge theorem

Let M be a Riemannian manifold with Laplacian $\triangle = \delta d + d\delta$.

Definition. A *harmonic* differential form $\omega \in \Omega(M)$ is a form in the kernel of the Laplacian:

$$\triangle \omega = 0.$$

We denote the space of harmonic k-forms on M by $\mathrm{Harm}_k(M)$.

Fact. A harmonic differential form ω is both exact and co-exact:

$$d\omega = 0 = \delta\omega\,.$$

Proof. From $\triangle\omega = 0$ one has

$$0 = \langle \omega, -\triangle\omega\rangle = \langle d\omega, d\omega\rangle + \langle \delta\omega, \delta\omega\rangle\,,$$

and the statement follows because $\langle\cdot,\cdot\rangle$ is positive semi-definite.

Theorem (Hodge). Let M be a compact n-dimensional Riemannian manifold. Then $H^k(M)$ is in bijection with $\mathrm{Harm}_k(M)$, i.e., every closed k-form $\omega \in \Omega^k(M)$ is cohomologous to one and only one harmonic k-form on M.

Proof. The operator $-\triangle$ is elliptic and by the compactness of M its set of eigenvalues is discrete. Because different eigenspaces are orthogonal to each other, every k-form ω has a unique orthogonal decomposition $\omega = \omega' + \triangle\omega''$ where ω' is harmonic. Introducing two operators $H : \omega \mapsto \omega'$ (harmonic projection) and $G : \omega \mapsto \omega''$ (Green operator of the Laplacian) we write

$$\omega = H\omega + \triangle G\omega\,.$$

Now the exterior derivative d commutes with the Laplacian. (Indeed, $d\triangle = d\delta d = \triangle d$.) Therefore, d commutes with the harmonic projection H and also with $\Delta G = G\Delta = \mathrm{Id} - H$.

Moreover, it follows that d commutes with the Green operator G. From this we infer that a closed form differs from its harmonic projection only by an exact form:

$$\omega = H\omega + (\delta d + d\delta)G\omega = H\omega + d(\delta G\omega),$$

which has the consequence that $[\omega] = [H\omega]$.

To show that $H\omega$ is the unique harmonic form representing the cohomology class $[\omega]$, one observes that if an exact form $\alpha = d\beta$ is harmonic, then α is the zero form. Indeed, since a harmonic form is co-closed,

$$\langle \alpha, \alpha \rangle = \langle d\beta, \alpha \rangle = -\langle \beta, \delta\alpha \rangle \quad \overset{\delta\alpha=0}{\Longrightarrow} \quad \alpha = 0\,.$$

Corollary. For a compact manifold M the number of linearly independent harmonic q-forms equals the q^{th} Betti number $b_q(M) = \dim H^q(M)$. Owing to $b_0(M) = 1$ for a connected manifold, it follows that the solution space of Laplace's equation for functions, $\triangle f = 0$, is one-dimensional. (The solution space consists of the constant functions.)

Remark. The compactness of M is crucial in order for Hodge's Theorem to hold. Indeed, in the case of, say, $M = \mathbb{R}^2 \simeq \mathbb{C}$, one has $b_0 = 1$ but there is a very large supply of harmonic functions. (The real or imaginary part of any analytic function $z \mapsto f(z)$ is harmonic.)

3.4 Weak form of the Morse inequalities

Let M be a compact n-dimensional manifold. Fixing some Morse function $f \in \mathrm{C}^\infty(M)$, consider the deformed Cartan derivative

$$d_s := \mathrm{e}^{-sf} d \circ \mathrm{e}^{sf} = d + s\,\varepsilon(d\,f) \qquad (s \in \mathbb{R})\,,$$

where $\varepsilon(\alpha)$ denotes the operation of exterior multiplication: $\varepsilon(\alpha)\,\omega := \alpha \wedge \omega\,.$ This operator still satisfies

$$d_s^2 = \left(\mathrm{e}^{-sf} d \circ \mathrm{e}^{sf}\right)^2 = \mathrm{e}^{-sf} d^2 \circ \mathrm{e}^{sf} = 0\,.$$

Thus one may consider the cohomology of the deformed de Rham complex $\Omega(M)$ with differential operator d_s :

$$b_k(s) := \dim\left(\Omega^k(M) \cap \ker d_s\right) - \dim\left(\Omega^k(M) \cap \mathrm{im}\, d_s\right)\,.$$

Since M is assumed to be compact, the function f and its derivatives are bounded. By a standard argument of operator analysis it then follows that the $b_k(s)$ are continuous functions of s. This means that the $b_k(s)$ are in fact independent of s, as the only way for a integer-valued function to change is to make a jump. Thus in particular $b_k(s) = b_k(0)$.

The idea of the following is to analyze the situation in the limit where the deformation parameter s is sent to infinity.

3.4.1 Witten Laplacian

Now fix a Riemannian structure on M (and thus a co-derivative δ) and introduce the deformed co-derivative

$$\delta_s := \mathrm{e}^{sf}\delta \circ \mathrm{e}^{-sf}.$$

The deformed de Rham operator $d_s - \delta_s$ is still self-adjoint, and its square

$$\mathcal{H}_s := (d_s - \delta_s)^2 = -(\delta_s\, d_s + d_s\, \delta_s)\,,$$

is a supersymmetric Hamiltonian commonly referred to as *Witten's Laplacian.*

By the same argument that was used in the proof of the Hodge Theorem, the dimension of the zero-energy eigenspace of $\mathcal{H}_s$ in $\Omega^k(M)$ is equal to the Betti number $b_k(s)$:

$$b_k(s) = \dim\ker \mathcal{H}_s\Big|_{\Omega^k(M)}.$$

Thus our attention now turns to the ground states of $\mathcal{H}_s$. As we shall see, their number is relatively easy to compute in the limit of $s \to \infty$.

Problem. Recalling the L^2-scalar product $\langle \alpha, \beta\rangle = \int_M \alpha \wedge \star\beta$ determined by the Riemannian structure of M, show that the operator of exterior multiplication $\varepsilon(d\,f) :\ \Omega^k(M) \to \Omega^{k+1}(M)$ is adjoint to the operator of contraction $\iota(\mathrm{grad}\, f) :\ \Omega^{k+1}(M) \to \Omega^k(M)$. $\square$

By using the result of the problem we have

$$-\delta_s = \mathrm{e}^{sf} d^\dagger \circ \mathrm{e}^{-sf} = (d_s)^\dagger = \big(d + s\,\varepsilon(d\,f)\big)^\dagger = -\delta + s\,\iota(\mathrm{grad} f)\,,$$

and hence

$$\mathcal{H}_s = \big(-\delta + s\,\iota(\mathrm{grad} f)\big)\big(d + s\,\varepsilon(d\,f)\big) + \big(d + s\,\varepsilon(d\,f)\big)\big(-\delta + s\,\iota(\mathrm{grad} f)\big)\,.$$

By a relation known as the *canonical anti-commutation relations* [for fermion creation operators $c_j^\dagger \equiv \varepsilon(dx^j)$ and annihilation operators $c_j \equiv \iota(\partial_j)$] one has

$$\varepsilon(d\,f)\,\iota(\mathrm{grad} f) + \iota(\mathrm{grad} f)\,\varepsilon(d\,f) = (d\,f)(\mathrm{grad} f) = |d\,f|^2 = \sum g^{ij}(\partial_i f)(\partial_j f)\,.$$

Thus the expression for $\mathcal{H}_s$ can be reorganized as

$$\mathcal{H}_s = -\triangle + s^2|d\,f|^2 + s\big(\mathcal{L}_{\mathrm{grad} f} + \mathcal{L}^\dagger_{\mathrm{grad} f}\big),$$

where $\mathcal{L}_X$ is the so-called Lie derivative in the direction of the vector field X :

$$\mathcal{L}_X = \iota(X) \circ d + d \circ \iota(X).$$

Note also that $\mathcal{L}^\dagger_{\mathrm{grad} f} = -\,\varepsilon(d\,f) \circ \delta - \delta \circ \varepsilon(d\,f)$.

Problem. If you have some background in Riemannian geometry, you may find it a rewarding exercise to verify the following statements about the 1-form sector:

1. On 1-forms one has the relation $\mathcal{L}_{\mathrm{grad}f} = \nabla_{\mathrm{grad}f} + \mathrm{Hess}(f)$ where $\mathrm{Hess}(f) = \nabla(\mathrm{grad}f)$ and ∇ is the Levi-Civitá covariant derivative for the Riemannian manifold M.

2. The Weitzenböck formula for the Laplacian $\triangle = \delta d + d\delta$ on 1-forms states that
$$-\triangle = \nabla^{\dagger}\nabla + \mathrm{Ric}\,,$$
where $\mathrm{Ric} \in \Gamma(\mathrm{End}(T^*M))$ is the Ricci curvature, and $\nabla^{\dagger} : \Gamma(T^*M \otimes T^*M) \to \Gamma(T^*M)$ is the adjoint of $\nabla : \Gamma(T^*M) \to \Gamma(T^*M \otimes T^*M)$. Similarly, the Witten Laplacian on 1-forms can be expressed as
$$\mathcal{H}_s = \mathrm{e}^{sf}\nabla^{\dagger} \circ \mathrm{e}^{-2sf}\nabla \circ \mathrm{e}^{sf} + \mathrm{Ric} + 2s\mathrm{Hess}(f)\ .$$

3. It follows from Weitzenböck's formula that the de Rham cohomology $H^1(M)$ is trivial for any manifold which admits a Riemannian structure with positive Ricci curvature.

3.4.2 Deformation to a harmonic-oscillator problem

What makes Witten's idea of deformation $(d \to d_s)$ so useful is the observation that the eigenvalue problem for $\mathcal{H}_s$ reduces to a harmonic-oscillator problem for $s \to \infty$. In that limit, the potential term $s^2|d\,f|^2$ grows beyond all bounds everywhere on M with the exception of the set of critical points of f, where $d\,f = 0$. As a consequence, the zero-energy eigenfunctions become localized at the critical points, and one can do the analysis by expanding around their local data.

For the following, fix a point $p \in M$, $(d\,f)_p = 0$, in the critical set of f. In a neighborhood of p let the metric tensor be expanded in a system of Riemann normal coordinates $\{x^1, \ldots, x^n\}$ as
$$g_{il} = \delta_{il} + \frac{1}{3}\sum R_{ikjl}(p)\, x^k x^j + \ldots\,,$$
where $R_{ikjl}(p)$ are the covariant components of the Riemann curvature tensor evaluated at p, and $x^1(p) = \ldots = x^n(p) = 0$. (Note that in a normal coordinate system, the Christoffel symbols of the connection vanish at the point p.)

Problem. Verify the given expansion of g_{ij} by computing from it the Riemannian curvature $R(\partial_i, \partial_j)\,\partial_l = (\nabla_{\partial_i}\nabla_{\partial_j} - \nabla_{\partial_j}\nabla_{\partial_i})\,\partial_l\,.$ □

Now consider the Taylor expansion of f around the critical point p:
$$f = f(p) + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x^i \partial x^j}(p)\, x^i x^j + \ldots\,.$$
By an orthogonal rotation of the coordinates we may assume the Hessian to be of diagonal form:
$$\frac{\partial^2 f}{\partial x^i \partial x^j}(p) = a_i\,\delta_{ij}\,.$$
The Lie derivative and its adjoint expand as
$$\mathcal{L}_{\mathrm{grad}f} = (\partial^i f)\,\iota(\partial_i)\,d + d \circ (\partial^i f)\,\iota(\partial_i) = \sum_i a_i\,\varepsilon(dx^i)\,\iota(\partial_i) + \ldots\,,$$
$$\mathcal{L}^{\dagger}_{\mathrm{grad}f} = -(\partial_i f)\,\varepsilon(dx^i)\,\delta - \delta \circ (\partial_i f)\,\varepsilon(dx^i) = -\sum_i a_i\,\iota(\partial_i)\,\varepsilon(dx^i) + \ldots\,.$$

By scaling the coordinates as $x^i = y_i/\sqrt{s}$ this gives the following expansion for the Witten Laplacian:

$$\mathcal{H}_s = s \sum_{i=1}^{n} \left(-\frac{\partial^2}{\partial y^i \partial y^i} + a_i^2 y_i^2 + a_i \left[\varepsilon(dy_i), \iota(\partial/\partial y_i)\right] \right) + \mathcal{O}(\sqrt{s}) \, .$$

We see that in leading approximation, the problem separates into n one-dimensional problems, each with Hamiltonian sh where

$$h = -\frac{\partial^2}{\partial y^2} + a^2 y^2 + a \left[\varepsilon(dy), \iota(\partial_y)\right].$$

It is easy to write down the ground state of the harmonic oscillator Hamiltonian h on $\Omega(\mathbb{R})$. For $a > 0$ the ground state ψ is found in the functions, $\psi = \mathrm{e}^{-ay^2/2}$. On the other hand, for $a < 0$ the ground state is found in the 1-forms, $\psi = \mathrm{e}^{+ay^2/2} dy$.

Problem. Verify that in both cases the ground state energy is zero. □

The leading-order approximation to the ground state wave function of $\mathcal{H}_s$ (localized at p) is now obtained by multiplying together all of the n ground state wave functions for the one-dimensional problems with Hamiltonians of the type of h above. Since every negative eigenvalue a_i of $\mathrm{Hess}_p(f)$ makes for a coordinate differential dy_i in the ground state, the product wave function is a differential form with degree k equal to the index of p. Thus, every critical point p of index k contributes a (perturbative) zero-energy ground state of $\mathcal{H}_{s\to\infty}$ in the space of k-forms, $\Omega^k(M)$.

On putting together the full chain of arguments,

$$b_k \equiv b_k(0) = b_k(s) = \lim_{s\to\infty} \dim\ker \mathcal{H}_s \Big|_{\Omega^k(M)},$$

it might now appear that b_k would have to be equal to the number m_k of of critical points of index k. Such a conclusion, however, is premature and in fact false in general because the neglected terms in $\mathcal{H}_s$ may cause some of the perturbative ground states to acquire a non-zero energy.

Nevertheless, the $s \to \infty$ perturbative analysis of $\mathcal{H}_s$ does show irrefutably that if there are to be b_k linearly independent ground states in $\Omega^k(M)$, then there cannot be less than b_k critical points of index k. Thus a safe conclusion is that

$$b_k \leq m_k \quad (k = 0, \ldots, n),$$

which establishes the weak form of the Morse inequalities.

3.5 Strong form of the Morse inequalities

We have yet to understand in quantitative terms how the number m_k of perturbative ground states is reduced to the number b_k of true ground states. To that end, the previous calculation needs to be improved. For a first idea, one might try to use perturbation theory to compute corrections in the small parameter $1/s$. Alas, this doesn't lead to anything: the corrections turn out to be identically zero to all orders (!) of the perturbation expansion in $1/s$. (The reason for this can be traced back to the supersymmetry of the problem.)

Notwithstanding the absence of perturbative corrections, there *must* exist *some* corrections that remove $m_k - b_k$ of the perturbative ground states from the zero-energy sector. The lift in energy is expected to depend on s like e^{-s}, which is not analytic in $1/s$ at $s = \infty$ and hence invisible in perturbation theory. Corrections of this kind are indeed caused by processes of quantum tunneling between the perturbative ground states. Pursuing this line of reasoning, our task would seem to become that of computing the ground-state energy splittings due to tunneling processes.

While that task can be accomplished in principle by using semiclassical methods (namely the WKB approximation or its field-theory analog, the so-called instanton method), Witten proposed an even better idea, which is this: just like one uses Hodge theory for $s = 0$ to pass from the de Rham cohomology groups to the ground states of $\mathcal{H} = -\triangle$, we can now use the ideas of Hodge theory in the limit of $s \to \infty$ to revert from the ground states of $\mathcal{H}_s = -\triangle_s$ to the cohomology groups of a certain differential complex, X. This new line of attack has several advantages. For one, the differential complex X that emerges for $s \to \infty$ is quite simple; in fact, its vector spaces X^k are finite-dimensional – they are spanned by the perturbative ground states. For another, the cohomology side computes quantized numbers (namely, dimensions) instead of exponentially small energy splittings. Hence there is no need to compute with total precision; an approximate calculation will suffice if it only captures the leading behavior for $s \to \infty$.

Thus Witten's strategy is to reduce the deformed de Rham complex

$$\ldots \longrightarrow \Omega^k(M) \xrightarrow{d_s} \Omega^{k+1}(M) \longrightarrow \ldots$$

to a finite-dimensional complex in the limit of $s \to \infty$, say

$$\ldots \longrightarrow X^k \xrightarrow{\delta} X^{k+1} \longrightarrow \ldots .$$

The task then is to construct the differential operator δ which corresponds to d_s in this limit.

We now proceed in three steps. First, we show that the very existence of a finite-dimensional complex X with differential operator δ already implies the strong form of the Morse inequalities with $m_k \equiv \dim X^k$. Second, we work through the simple example of the circle $M = \mathrm{S}^1$ to get a feeling for the differential operator δ. Third, we recount Witten's sketch of the semiclassical construction of δ in the general case.

3.5.1 Strong Morse inequalities for a differential complex

Suppose that we are given a differential complex

$$\ldots \longrightarrow X^k \xrightarrow{\delta} X^{k+1} \longrightarrow \ldots$$

of vector spaces X^k $(k \geq 0)$ of finite dimension $m_k = \dim X^k$ with differential operator δ (not to be confused with the $\delta = -d^\dagger$ of earlier) and cohomology

$$\dim \big((X^k \cap \ker\delta) \,/\, (X^k \cap \mathrm{im}\delta) \big) =: b_k \,.$$

Lemma. For all $k \geq 0$ one has the inequality

$$m_k - m_{k-1} + \ldots + (-1)^k m_0 \geq b_k - b_{k-1} + \ldots + (-1)^k b_0.$$

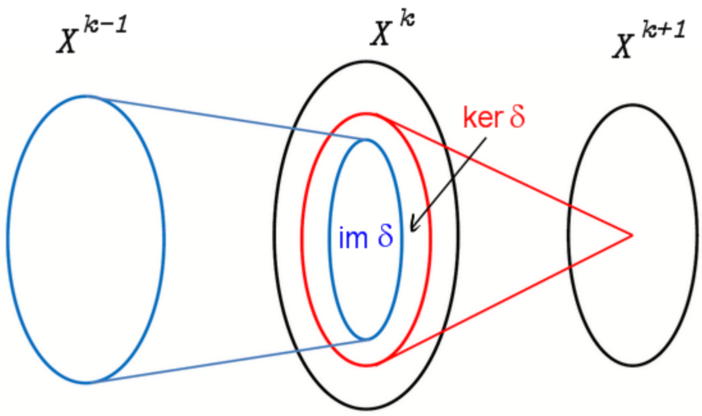
Proof. To simplify the notation, we write $X^k \cap \ker\delta \equiv \ker\delta$ and $X^k \cap \mathrm{im}\delta \equiv \mathrm{im}\delta$ whenever the degree k is clear from the context. Now let $c_k := \dim(X^k \cap \mathrm{im}\delta)$ for $k \geq 1$ and $c_0 := 0$. Then

$$m_k = c_k + b_k + c_{k+1}.$$

Indeed, we may decompose X^k as a direct sum,

$$X^k \simeq \ker\delta \oplus (X^k / \ker\delta) \simeq \mathrm{im}\delta \oplus (\ker\delta \,/\, \mathrm{im}\delta) \oplus (X^k / \ker\delta),$$

and, since $\delta :\ X^k/\ker\delta \to X^{k+1} \cap \mathrm{im}\delta$ is an isomorphism, the claimed formula follows by taking dimensions on both sides. The situation is depicted in the following diagram.



Next we take differences:

$$m_k - m_{k-1} = c_{k+1} + b_k - b_{k-1} - c_{k-1},$$

and by iteration we obtain

$$m_k - m_{k-1} + \ldots + (-1)^k m_0 = c_{k+1} + b_k - b_{k-1} + \ldots + (-1)^k b_0\,.$$

The claimed inequality now follows because c_{k+1} is a non-negative number.

3.5.2 The example of $M = \mathrm{S}^1$

Consider the circle $M = \mathrm{S}^1$ with angular coordinate $\theta \in [0, 2\pi)$ and let $f :\ \mathrm{S}^1 \to \mathbb{R}$ be a Morse function with n minima $p_1, \ldots, p_n$ and n maxima $q_1, \ldots, q_n$. We may assume that these critical points are ordered by

$$0 \leq \theta(p_1) < \theta(q_1) < \ldots < \theta(p_n) < \theta(q_n) < 2\pi.$$

The perturbative ground state spaces (of $\mathcal{H}_s = -\triangle_s$) due to them are

$$X^0 \simeq \mathrm{span}\left\{\psi^{(p_i)}\right\}_{i=1,\ldots,n}, \quad \psi^{(p_i)} = \mathrm{e}^{-sf''(p_i)\,(\theta-\theta(p_i))^2/2},$$
$$X^1 \simeq \mathrm{span}\left\{\psi^{(q_i)}\right\}_{i=1,\ldots,n}, \quad \psi^{(q_i)} = \mathrm{e}^{+sf''(q_i)\,(\theta-\theta(q_i))^2/2} d\theta\,.$$

The '$\simeq$' signs remind us that the left-hand sides emerge in the limit of $s \to \infty$ while the right-hand sides are for very large but finite s.

Now fix some minimum p_i for $1 < i < n$. We wish to compute the quantum tunneling matrix elements between the corresponding ground state wave function $\psi^{(p_i)}$ and the 1-form ground states associated with the neighboring maxima, $\psi^{(q_{i-1})}$ and $\psi^{(q_i)}$. To do so, one makes the WKB ansatz $\psi_{\text{WKB}} = A\,\mathrm{e}^{-sW}$ and solves the Hamilton-Jacobi equation

$$-|s\,dW|^2 + s^2|d\,f|^2 = E$$

for the unknown function W on the interval between q_{i-1} and q_i. The zero-energy solution ($E = 0$) satisfying the condition $\psi_{\text{WKB}}(p_i) = \psi^{(p_i)}(p_i) = 1$ is

$$\psi_{\text{WKB}}^{(p_i)} = \mathrm{e}^{-sf+sf(p_i)}\,\chi_{[q_{i-1},\,q_i]}\,.$$

As usual, the WKB approximation procedure needs to be terminated at the neighboring critical points, where the Hamilton-Jacobi equation becomes singular. To deal with these singular points of the WKB approximation, we have cut off $\psi_{\text{WKB}}^{(p_i)}$ by a smooth regularization of the characteristic function for the interval $[q_{i-1}, q_i]$. Doing the same for $\psi^{(q_i)}$, say, we get

$$\psi_{\text{WKB}}^{(q_i)} = \mathrm{e}^{+sf-sf(q_i)}\,\chi_{[p_i, p_{i+1}]}\,d\theta\,.$$

An improved approximation for the vector spaces X^0 and X^1 is obtained by taking them to be the span of these WKB-ground states instead of the original perturbative ground states.

We are now ready to compute the quantum tunneling matrix elements we need. First, consider

$$\begin{aligned}\left\langle \psi_{\text{WKB}}^{(q_i)} \middle|\, d_s\,\psi_{\text{WKB}}^{(p_i)} \right\rangle &= \int_{S^1} \mathrm{e}^{+sf-sf(q_i)}\chi_{[p_i,p_{i+1}]}\;d_s\;\mathrm{e}^{-sf+sf(p_i)}\;\chi_{[q_{i-1},\,q_i]} \\ &= \mathrm{e}^{-sf(q_i)+sf(p_i)}\int_{S^1}\chi_{[p_i,\,p_{i+1}]}\,d\,\chi_{[q_{i-1},\,q_i]}.\end{aligned}$$

Contributions to the integral come only from the close vicinity of the point $q_i \in [p_i, p_{i+1}]$, where $\chi_{[p_i,p_{i+1}]} \equiv 1$ and the value of $\chi_{[q_{i-1},\,q_i]}$ drops from 1 to 0. Hence

$$\left\langle \psi_{\text{WKB}}^{(q_i)} \middle|\, d_s\,\psi_{\text{WKB}}^{(p_i)} \right\rangle = -\,\mathrm{e}^{-s(f(q_i)-f(p_i))}.$$

In a similar manner one finds

$$\left\langle \psi_{\text{WKB}}^{(q_{i-1})} \middle|\, d_s\,\psi_{\text{WKB}}^{(p_i)} \right\rangle = +\,\mathrm{e}^{-s(f(q_{i-1})-f(p_i))}.$$

Due to the vanishing overlap of the characteristic functions for distinct indices i, it is clear that there are non-zero matrix elements beyond those given.

Notice that our WKB wave functions do not quite have the correct L^2-norm. However, this slight error won't make any difference for our goal of computing the cohomology of the reduced differential complex emerging for $s \to \infty$. Changing the normalization we now set

$$\begin{aligned}|p_i\rangle &\equiv \mathrm{e}^{+sf(p_i)}\psi_{\text{WKB}}^{(p_i)} = \mathrm{e}^{-sf}\;\chi_{[q_{i-1},\,q_i]} \quad \text{and} \\ |q_i\rangle &\equiv \mathrm{e}^{-sf(q_i)}\psi_{\text{WKB}}^{(q_i)} = \mathrm{e}^{+sf}\;\chi_{[p_i,\,p_{i+1}]}\,d\theta.\end{aligned}$$

Then our calculation shows that $d_s:\ \Omega^0(\mathrm{S}^1) \to \Omega^1(\mathrm{S}^1)$ restricts to the vector spaces

$$X^0 = \mathrm{span}\{|p_i\rangle\}_{i=1,\dots,n} \quad \text{and} \quad X^1 = \mathrm{span}\{|q_i\rangle\}_{i=1,\dots,n}$$

as the differential operator $\delta:\ X^0 \to X^1$ given by

$$\delta\,|p_i\rangle = -|q_i\rangle + |q_{i-1}\rangle \quad (i = 2, \dots, n), \qquad \delta\,|p_1\rangle = -|q_1\rangle + |q_n\rangle.$$

It is easy to compute the cohomology of δ. Its kernel is one-dimensional, being spanned by

$$\sum_{i=1}^{n} |p_i\rangle = \mathrm{e}^{-sf} \sum_{i=1}^{n} \chi_{[q_{i-1},\, q_i]} \qquad (q_0 \equiv q_n),$$

the symmetric linear combination of all the (properly normalized) bump functions concentrated at the minima of f. If the χ's are taken to be a partition of unity, i.e. $\sum \chi_{[q_{i-1},\, q_i]} \equiv 1$, then this linear combination is simply e^{-sf}. By recalling the expression $\mathcal{H}_s = -\triangle_s = d_s^\dagger d_s$ of the Witten Laplacian on functions, we immediately see that $\mathcal{H}_s\,\mathrm{e}^{-sf} = 0$. Thus the WKB <u>approximation</u> has done the curious trick of producing the <u>exact</u> result for the ground state! (It is a well-known hallmark of supersymmetric Hamiltonians that a semiclassical *approximation* actually gives the *exact* answer when the right question is asked.)

Turning to the image of δ in $X^1 \simeq \mathbb{R}^n$, we see that this has codimension one. The missing direction is spanned by the symmetric linear combination

$$\sum_{i=1}^{n} |q_i\rangle = \mathrm{e}^{+sf} \sum_{i=1}^{n} \chi_{[p_i,\, p_{i+1}]}\, d\theta \qquad (p_{n+1} \equiv p_1).$$

If $\sum \chi_{[p_i,\, p_{i+1}]} = 1$, this is the exact ground state of $\mathcal{H}_s = d_s d_s^\dagger$ in $\Omega^1(M)$. Notice that, of course, these results for the cohomology of δ are in agreement with $b_0(\mathrm{S}^1) = b_1(\mathrm{S}^1) = 1$.

Problem. Carry out a similar WKB calculation for $M = \mathrm{S}^2$ with Morse function f given in terms of a complex stereographic coordinate $z = \tan(\theta/2)\,\mathrm{e}^{\mathrm{i}\phi}$ by

$$f = \frac{1 - |w|^2}{1 + |w|^2}\,, \quad w = \sum_{k=0}^{3} \frac{\bar{z}_k z + 1}{z - z_k}, \quad z_k = \sqrt{2}\,\mathrm{e}^{2\pi\mathrm{i}k/3} \quad (k = 1, 2, 3), \quad z_0 = 0.$$

(Notice that f has the symmetry of a tetrahedron.)

3.5.3 Witten's narrative

We finish the story with a brief account of the outcome of Witten's analysis for the case of a general compact manifold M. The method he primarily invokes is instanton calculus for a $(0+1)$-dimensional supersymmetric field theory with Euclidean (i.e., imaginary-time) action functional

$$\begin{aligned} S_E = \frac{1}{2} \int dt\, \Big(& g_{ij} \frac{d\phi^i}{dt} \frac{d\phi^j}{dt} + g_{ij}\, \bar{\psi}^i \Big(\frac{d\psi^j}{dt} + \Gamma^j_{kl} \frac{d\phi^l}{dt} \psi^k \Big) \\ & + \frac{1}{4} R_{ijkl}\, \bar{\psi}^i \psi^k \bar{\psi}^j \psi^l + s^2 g^{ij} \frac{\partial f}{\partial \phi^i} \frac{\partial f}{\partial \phi^j} + s \Big(\frac{\partial^2 f}{\partial \phi^i\, \partial \phi^j} - \Gamma^k_{ij} \frac{\partial f}{\partial \phi^k} \Big) \bar{\psi}^i \psi^j \Big), \end{aligned}$$

where the summation convention is assumed. By the procedure of canonical quantization, this classical action functional yields the quantum Hamiltonian $\mathcal{H}_s = -\triangle_s$. The bosonic fields ϕ^i arise from a choice of local coordinates x^i on M via pull back $\phi^i = x^i \circ \phi$ by the bosonic field mapping $\phi: \mathbb{R} \to M$. The fermionic fields ψ^i correspond to the coordinate differentials dx^i. The Γ^j_{kl} are the Christoffel symbols expressing the Levi-Civitá connection of the Riemannian manifold M in the system of coordinates x^i. The Riemann curvature tensor is expressed by its components R_{ijkl}. The last term of the actional functional arises by pull back of the Hessian tensor of f,

$$\mathrm{Hess}(f) = \nabla(d\,f) = \left(\frac{\partial^2 f}{\partial x^i\,\partial x^j} - \Gamma^k_{ij}\frac{\partial f}{\partial x^k}\right) dx^i \otimes dx^j.$$

Based on instanton calculus for this supersymmetic field theory (augmented by some considerations involving the WKB method), Witten arrives at the following prescription for constructing the differential complex (X, δ).

With each critical point p of the Morse function f one associates a vector $|p\rangle$, and one sets

$$X^k = \mathrm{span}\{|p\rangle\}_{p\,:\,\mathrm{ind}_p(f) = k}\,.$$

The differential operator $\delta: X^k \to X^{k+1}$ is expressed by

$$\delta\,|p\rangle = \sum |q\rangle\, n(q,p),$$

where the sum runs over the critical points q of index $k+1$, and the coefficients $n(q,p)$ are integers. They are computed as follows.

For each critical point p let V_p denote the negative eigenspace of the Hessian of f at p. This vector space V_p is oriented by the choice of a state vector $|p\rangle$, i.e. by a choice of ordering for the product of coordinate differentials in the differential form $|p\rangle$.

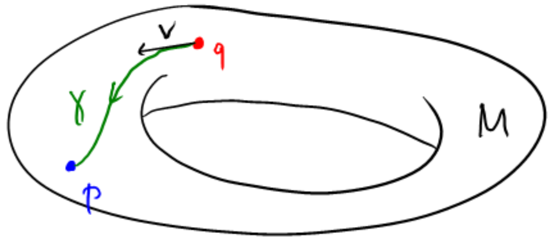
For definiteness, let now p be a critical point of index k. If q is any critical point of index $k+1$, one is to use the Riemannian geometry of M to solve the gradient flow equation

$$\dot\gamma(t) = -(\mathrm{grad} f)(\gamma(t))$$

or in coordinates,

$$\dot x^i = -\sum g^{ij}\frac{\partial f}{\partial x^j}\,,$$

for a path $\gamma: \mathbb{R} \to M$ of steepest descent beginning at $\gamma(-\infty) = q$ and ending at $\gamma(+\infty) = p$.



Let $v \in V_q$ be the initial direction of the path γ and define $\widetilde{V}_q$ to be the orthogonal complement of v in V_q :

$$V_q = \widetilde{V}_q \oplus \mathbb{R}v.$$

The given orientation of V_q induces an orientation of $\widetilde{V}_q$ by the canonical process of taking the inner product with $v \neq 0$.

Now the geometric data associated with the steepest descent path γ connecting q with p determines an isomorphism τ between $\widetilde{V}_q$ and V_p. The definition of τ needs perhaps some explanation. For simplicity, assume the situation of a generic Morse function f, where the k lowest eigenvalues of the transverse part of the Hessian $\nabla(d\,f)$ are separated from the rest of the spectrum by a finite gap all along γ. The eigenvectors associated with eigenvalues then span a rank-k vector bundle, say $\mathcal{V}$, over γ. As a sub-bundle of the tangent bundle TM, this vector bundle is equipped with a connection in the obvious manner (by restriction of the Levi-Civitá connection of TM). Thus we have an operation of parallel transport in $\mathcal{V}$. Since $\mathcal{V}$ connects $\widetilde{V}_q$ with V_p, this operation results in an isomorphism $\tau: \widetilde{V}_q \to V_p$.

If τ is orientation preserving one sets $n_\gamma = +1$, otherwise $n_\gamma = -1$. If there is more than one path γ of steepest descent between q and p, then one takes the sum over them:

$$n(q,p) := \sum_\gamma n_\gamma .$$

This concludes the description of δ. While it may not be clear from it that $\delta^2 = 0$, this property is in fact inherited by δ from its parent d_s.

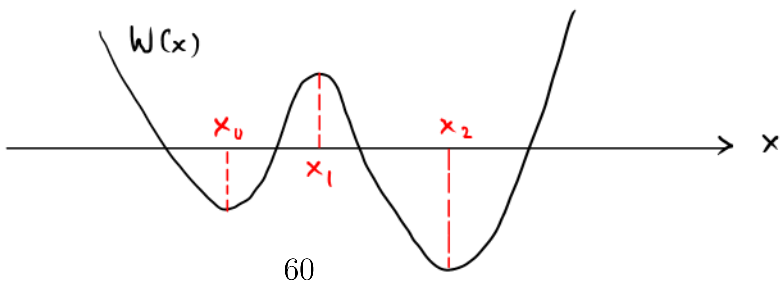
3.6 Escape over a barrier: Kramers' formula

We now apply the formalism developed in this chapter to a problem of non-equilibrium statistical physics, namely that of calculating the rate of escape over an energy barrier by thermal activation of a statistical population. An explicit expression for the escape rate in the limit of a high barrier or low temperature is provided by Kramers' formula. In the sequel, we will derive it by using supersymmetry and the Witten Laplacian.

For simplicity we consider the situation in one space dimension. We assume that the population dynamics is overdamped and governed by a Fokker-Planck equation with a diffusion term and a drift term. Denoting the space coordinate by x we write the Fokker-Planck equation for the time evolution of the population density $P(x,t)\,dx$ as

$$\frac{\partial P}{\partial t} = D\frac{\partial^2 P}{\partial x^2} + \beta\frac{\partial}{\partial x}(W'P).$$

Here D is the diffusion constant, and β is given by a so-called fluctuation-dissipation relation: $\beta = D/k_B T$ with temperature T. The function $x \mapsto W(x)$ is a potential energy function for the stochastic dynamics. We assume W to have three extrema: a metastable minimum at x_0, a maximum (the peak of the 'barrier') at x_1, and a global minimum at x_2. The graph of W is sketched below.



In such a situation, any initial distribution $P(x,0)\,dx$ is sure to converge to the following equilibrium distribution function:

$$\lim_{t\to\infty} P(x,t) = P_\infty(x) \propto \mathrm{e}^{-W(x)/k_BT}, \qquad \int_{\mathbb{R}} P_\infty(x)\,dx = 1\,,$$

whose population is concentrated near the absolute minimum x_2.

We now ask how long it takes for the population to reach this equilibrium state. More precisely, let the population be initially localized at the metastable minimum x_0. The question then is: how long does it take for the population to escape over the energy barrier peaking at x_1?

This question has a simple explicit answer for low temperatures, $k_BT \ll W(x_1) - W(x_0)$ and $k_BT \ll W(x_0) - W(x_2)$. The answer for the escape rate in this limit is given by Kramers' formula (Kramers, 1940):

$$\frac{1}{\tau} = D\,\frac{\sqrt{-W''(x_1)W''(x_0)}}{2\pi k_BT}\,\mathrm{e}^{-(W(x_1)-W(x_0))/k_BT}\,.$$

This result and its generalizations to higher dimension (see, e.g., arXiv.org/pdf/1106.5799.pdf) have been applied to numerous problems of non-equilibrium statistical physics. An early application was the calculation of fission rates of atomic nuclei heated by neutron capture.

In the following we present a rather straightforward derivation of Kramers' formula by using the supersymmetry of the Witten Laplacian which is associated with the generator of the Fokker-Planck dynamics. The first step is to convert the Fokker-Planck equation into a Schrödinger-like equation in imaginary time by substituting

$$P(x,t) = \sqrt{P_\infty(x)}\,\psi(x,t).$$

The resulting imaginary-time Schrödinger equation for the 'wave function' ψ reads

$$\frac{\partial\psi}{\partial t} = -\mathcal{H}_0\psi$$

with 'Hamiltonian'

$$\mathcal{H}_0 = D\left(-\frac{\partial}{\partial x} + \frac{W'}{2k_BT}\right)\left(\frac{\partial}{\partial x} + \frac{W'}{2k_BT}\right),$$

which we recognize as the bosonic part of a supersymmetric Laplacian of Witten type. Note that the imaginary-time Schrödinger dynamics does not conserve the quantum probability $\int_{\mathbb{R}} |\psi(x,t)|^2 dx$ but rather the integral $\int_{\mathbb{R}} \sqrt{P_\infty(x)}\,\psi(x,t)dx = \int_{\mathbb{R}} P(x,t)dx = 1$.

As usual, the Schrödinger equation can be solved formally in terms of the eigenvalues $0 = \omega_0 < \omega_1 < \omega_2 \le \ldots$ and the orthonormal system of eigenfunctions $\psi_0, \psi_1, \psi_2, \ldots$ of $\mathcal{H}_0$:

$$\psi(x,t) = \sqrt{P_\infty(x)} + \sum_{n=1}^{\infty} \psi_n(x)\,\mathrm{e}^{-\omega_n t}\langle\psi_n|\psi(\cdot,0)\rangle.$$

Here $\langle\psi_n|\psi(\cdot,0)\rangle = \int_{\mathbb{R}} \psi_n(x)\psi(x,0)\,dx$, and $\psi_0(x) \equiv \sqrt{P_\infty(x)}$ is the normalized ground state wave function with eigenvalue ω_0 . Notice that $\langle\psi_0|\psi(\cdot,0)\rangle = \int_{\mathbb{R}} P(x,0)\,dx = 1$.

Now, based on our experience with the Witten Laplacian, we anticipate that the eigenfunction ψ_1 comes from a perturbative ground state by the inclusion of 'quantum tunneling' corrections.

Thus we expect the corresponding eigenvalue ω_1 to be exponentially small (for small temperatures T) and separated by a large gap from all higher eigenvalues. If so, then for times $t \gg 1/\omega_2$ the distribution function reduces to a sum of just two terms,

$$\psi(x,t) = \sqrt{P_\infty(x)} + \mathrm{e}^{-\omega_1 t} a_1 \psi_1(x), \qquad a_1 = \langle \psi_1 | \psi(\cdot, 0) \rangle.$$

The first summand is the ground state corresponding to the Fokker-Planck equilibrium. The second one is contributed by the first excited state ψ_1, which decays at the very slow rate of $1/\tau = \omega_1$. This rate is the desired rate $1/\tau$ for escape over the barrier.

Hence our second step is to calculate the relevant frequency ω_1. This is done by using the WKB-improved perturbation theory described in Section 3.5.2. The difference is that, this time, we aren't after topological invariants and we will have to be a little more accurate to arrive at the correct form of Kramers' result for the escape rate. We introduce the following notation:

$$\mathcal{H}_0/D = d_T^\dagger d_T\,, \quad d_T = \mathrm{e}^{-W/2k_BT} d\, \mathrm{e}^{+W/2k_BT}, \quad d_T^\dagger = \mathrm{e}^{+W/2k_BT} d^\dagger\, \mathrm{e}^{-W/2k_BT}, \quad \mathcal{H}_1/D = d_T d_T^\dagger\,.$$

The (negative of the) Witten Laplacian of the present situation is

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1 = DQ^2, \qquad Q = d_T + d_T^\dagger.$$

We now use supersymmetry in a powerful way: instead of calculating the eigenvalue ω_1 of $\mathcal{H}_0$ directly, we do something equivalent but simpler: we calculate the corresponding eigenvalue, E, of the fermionic charge Q. Kramers' escape rate will then be given by $1/\tau = \omega_1 = DE^2$.

We begin by writing down the (WKB-improved) perturbative ground states associated with the extrema x_0, x_1, x_2 :

$$\psi^{(x_0)} = c_0\, \mathrm{e}^{-(W-W(x_0))/2k_BT} \chi_{(-\infty,\, x_1]}\,, \quad c_0 = \left(\frac{W''(x_0)}{2\pi k_B T}\right)^{1/4}$$

$$\psi^{(x_1)} = c_1\, \mathrm{e}^{+(W-W(x_1))/2k_BT} \chi_{[x_0,\, x_2]}\, dx\,, \quad c_1 = \left(\frac{-W''(x_1)}{2\pi k_B T}\right)^{1/4},$$

$$\psi^{(x_2)} = c_2\, \mathrm{e}^{-(W-W(x_2))/2k_BT} \chi_{[x_1, +\infty)}\,, \quad c_2 = \left(\frac{W''(x_2)}{2\pi k_B T}\right)^{1/4}.$$

These are normalized in such a way that $\langle \psi^{(x_k)} | \psi^{(x_k)} \rangle \approx 1$ approximately for small T. Note that the overlaps $\langle \psi^{(x_k)} | \psi^{(x_l)} \rangle$ for $k \neq l$ are negligibly small.

Next, following the blueprint of degenerate perturbation theory, we compute the transition matrix elements:

$$\begin{aligned}
&\langle \psi^{(x_1)} | Q | \psi^{(x_0)} \rangle = \langle \psi^{(x_1)} | d_T | \psi^{(x_0)} \rangle = \langle \psi^{(x_0)} | d_T^\dagger | \psi^{(x_1)} \rangle \\
&= \int c_1\, \mathrm{e}^{+(W-W(x_1))/2k_BT} \chi_1\, d_T\, c_0\, \mathrm{e}^{-(W-W(x_0))/2k_BT} \chi_0 \\
&= c_1 c_0\, \mathrm{e}^{-(W(x_1)-W(x_0))/2k_BT} \int \chi_1\, d\, \chi_0 = -c_1 c_0\, \mathrm{e}^{-(W(x_1)-W(x_0))/2k_BT},
\end{aligned}$$

and similarly,

$$\langle \psi^{(x_2)} | Q | \psi^{(x_0)} \rangle = \langle \psi^{(x_2)} | d_T | \psi^{(x_0)} \rangle = \langle \psi^{(x_0)} | d_T^\dagger | \psi^{(x_2)} \rangle = +c_1 c_2\, \mathrm{e}^{-(W(x_1)-W(x_2))/2k_BT}.$$

All other matrix elements vanish on simple grounds of degree. Thus the matrix of Q with respect to the three perturbative ground states $\psi^{(x_0)}, \psi^{(x_1)}, \psi^{(x_2)}$ has the form

$$\begin{pmatrix} 0 & \langle \psi^{(x_0)} | d_T^\dagger | \psi^{(x_1)} \rangle & 0 \\ \langle \psi^{(x_1)} | d_T | \psi^{(x_0)} \rangle & 0 & \langle \psi^{(x_1)} | d_T | \psi^{(x_2)} \rangle \\ 0 & \langle \psi^{(x_2)} | d_T^\dagger | \psi^{(x_1)} \rangle & 0 \end{pmatrix} .$$

It is an easy exercise in linear algebra to compute the eigenvalues and eigenvectors of this 3×3 matrix. First of all, we have an eigenvalue zero with (unnormalized) eigenvector

$$\begin{aligned} \psi_{\text{g.s.}} &= -\psi^{(x_0)} / \langle \psi^{(x_1)} | d_T | \psi^{(x_0)} \rangle + \psi^{(x_2)} / \langle \psi^{(x_1)} | d_T | \psi^{(x_2)} \rangle \\ &= c_1^{-1} \, \mathrm{e}^{-(W - W(x_1))/2k_B T} \left(\chi_{(-\infty, x_1)} + \chi_{(x_1, +\infty)} \right) \propto \sqrt{P_\infty(x)}. \end{aligned}$$

As expected, this is the ground state wave function corresponding to the equilibrium distribution function P_∞ by the substitution $P = \sqrt{P_\infty}\, \psi$. The two excited states are

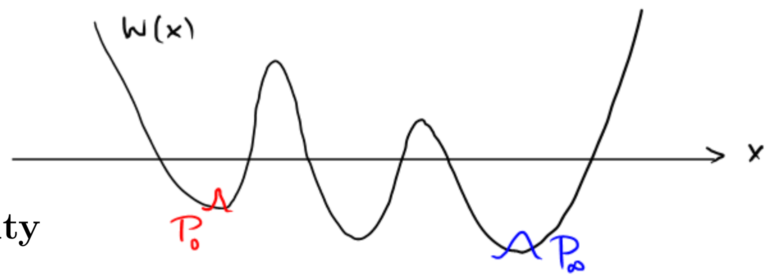
$$\begin{aligned} \psi_{\text{exc.}}^{\pm} &= (d_T^\dagger \pm E) \psi^{(x_1)} = \psi^{(x_0)} \, \langle \psi^{(x_0)} | d_T^\dagger | \psi^{(x_1)} \rangle + \psi^{(x_2)} \, \langle \psi^{(x_2)} | d_T^\dagger | \psi^{(x_1)} \rangle \pm E \psi^{(x_1)} \, , \\ E &= \sqrt{ \left| \langle \psi^{(x_1)} | d_T | \psi^{(x_0)} \rangle \right|^2 + \left| \langle \psi^{(x_1)} | d_T | \psi^{(x_2)} \rangle \right|^2 }. \end{aligned}$$

Assuming the inequality $k_B T \ll W(x_0) - W(x_2)$, the expression for E simplifies to

$$E = \left| \langle \psi^{(x_1)} | d_T | \psi^{(x_0)} \rangle \right| = c_1 c_0 \, \mathrm{e}^{-(W(x_1) - W(x_0))/2k_B T}.$$

It is now readily seen that the square of this, $1/\tau = \omega_1 = D E^2$, is indeed equal to Kramers' expression for the escape rate.

Problem. Can you predict the rate of escape over two one-dimensional barriers in sequence?



3.7 Brascamp-Lieb inequality

Let us finish this chapter by mentioning another use of the Witten Laplacian, this time in the realm of equilibrium statistical physics. Consider the Boltzmann-Gibbs distribution

$$Z^{-1} \mathrm{e}^{-H(x)/k_B T} d^N x$$

for N real degrees of freedom $x_1, \ldots, x_N$. The expectation value of an observable f is defined by

$$f_{\text{av}} = \frac{1}{Z} \int_{\mathbb{R}^N} f(x) \, \mathrm{e}^{-H(x)/k_B T} d^N x.$$

We are interested in the fluctuations of f as given by its variance:

$$\text{var}(f) = \frac{1}{Z} \int_{\mathbb{R}^N} \left(f(x) - f_{\text{av}} \right)^2 \mathrm{e}^{-H(x)/k_B T} d^N x.$$

The goal is to derive a useful upper bound for var(f).

As before, we set up a (fictitious) quantum mechanics with ground state wave function

$$\psi_0(x) \equiv \mathrm{e}^{-h(x)} = \frac{1}{\sqrt{Z}}\,\mathrm{e}^{-H(x)/2k_BT}, \qquad \int_{\mathbb{R}^N} \psi_0(x)^2 d^N x = \int_{\mathbb{R}^N} \mathrm{e}^{-2h(x)} d^N x = 1,$$

which allows us to view $\mathrm{var}(f)$ as a ground state expectation value:

$$\mathrm{var}(f) = \big\langle \psi_0 | (f - f_{\mathrm{av}})^2 | \psi_0 \big\rangle \,.$$

Note that $\langle \psi_0 | (f - f_{\mathrm{av}}) | \psi_0 \rangle = f_{\mathrm{av}} - f_{\mathrm{av}} = 0$. Thus $(f - f_{\mathrm{av}})|\psi_0\rangle$ is orthogonal to the ground state.

Now ψ_0 may be viewed as the ground state (in the space of functions, or 0-forms) of a Witten Laplacian

$$-\triangle_h = d_h^\dagger d_h + d_h d_h^\dagger \,, \qquad d_h = \mathrm{e}^{-h} d\, \mathrm{e}^{+h}.$$

Since $(f - f_{\mathrm{av}})|\psi_0\rangle$ is orthogonal to the kernel of $-\triangle_h$, we can write

$$\begin{aligned}(f - f_{\mathrm{av}})|\psi_0\rangle &= (-\triangle_h)\,(-\triangle_h)^{-1}(f - f_{\mathrm{av}})|\psi_0\rangle \\ &= d_h^\dagger(-\triangle_h)^{-1} d_h (f - f_{\mathrm{av}})|\psi_0\rangle = d_h^\dagger(-\triangle_h)^{-1}(d\, f)|\psi_0\rangle,\end{aligned}$$

which results in the following expression for the variance:

$$\mathrm{var}(f) = \big\langle \psi_0 |\, \iota(\mathrm{grad}\, f)(-\triangle_h)^{-1}(d\, f)|\psi_0\big\rangle = \int_{\mathbb{R}^N} d^N x\, \mathrm{e}^{-h}\iota(\mathrm{grad}\, f)(-\triangle_h)^{-1}\mathrm{e}^{-h} d\, f \,.$$

Next we observe that

$$\begin{aligned}\mathrm{e}^{+h}(-\triangle_h)^{-1}\mathrm{e}^{-h} &= \big(-\,\mathrm{e}^{+h}\triangle_h \circ \mathrm{e}^{-h}\big)^{-1} = \big(d_{2h}^\dagger d + d\, d_{2h}^\dagger\big)^{-1} \\ &= \big(d^\dagger d + d d^\dagger + \iota(\mathrm{grad}\, 2h)\circ d + d \circ \iota(\mathrm{grad}\, 2h)\big)^{-1} = \big(-\triangle + \mathcal{L}_{\mathrm{grad}\, 2h}\big)^{-1}.\end{aligned}$$

By using the relations $\mathcal{L}_X = \nabla_X + (\nabla X)$ and $-\triangle = \nabla^\dagger \nabla$ (in Euclidean space) we then obtain

$$\mathrm{e}^{+h}(-\triangle_h)^{-1}\mathrm{e}^{-h} = \big(\nabla^\dagger\nabla + \nabla_{\mathrm{grad}\, 2h} + \mathrm{Hess}(2h)\big)^{-1} = \big(\mathrm{e}^{2h}\nabla^\dagger \mathrm{e}^{-2h}\nabla + \mathrm{Hess}(2h)\big)^{-1}.$$

This yields the identity

$$\mathrm{var}(f) = \int_{\mathbb{R}^N} d^N x\, \mathrm{e}^{-2h}\iota(\mathrm{grad}\, f)\big(\mathrm{e}^{2h}\nabla^\dagger\mathrm{e}^{-2h}\nabla + \mathrm{Hess}(2h)\big)^{-1} d\, f.$$

Now the operator $\mathrm{e}^{2h}\nabla^\dagger\mathrm{e}^{-2h}\nabla$ is self-adjoint w.r.t. $\langle f | \, g\rangle = \int f(x)\, g(x)\, \mathrm{e}^{-2h(x)} d^N x$ and positive. If $\mathrm{Hess}(2h) > 0$, then we may drop that operator in the denominator to arrive at the so-called *Brascamp-Lieb inequality*:

$$\mathrm{var}(f) \le \int_{\mathbb{R}^N} d^N x\, \mathrm{e}^{-2h}\iota(\mathrm{grad}\, f)\big(\mathrm{Hess}(2h)\big)^{-1} d\, f = \left(\sum_{i,j=1}^N (\partial_i f)\big(\mathrm{Hess}(2h)\big)^{-1}{}_{ij}(\partial_j f)\right)_{\mathrm{av}} .$$

Remark. This inequality can be used, for example, to prove exponential decay of the correlation functions for a (lattice) scalar field ϕ with energy function $H = (\phi, -\triangle\phi) + V(\phi)$, where $\triangle$ is the (lattice) Laplacian and $V(\phi) = \sum_x V(\phi_x)$ a convex interaction potential.