Copyright notices and comments

If you find these lectures informative and interesting, and would like to compensate the author for his efforts in writing them, typesetting them and making them available to you, or if you would like simply to contribute to survival of an independent scholar, please send a donation to the following address:

Allan Adler
36 Rolens Drive, Apt. C4
Kingston, RI 02881

I will be able to receive mail at this address until August 1994. I will also update this notice regularly to reflect my current address. Finally, my current address will always be available from looking in the Combined Membership List of the American Mathematical Society, especially the electronic version at e-math.ams.com.

If on the other hand, you have criticisms and suggestions for improving these notes, please let me know.

Fermat's Last Theorem: Notes of Lecture 4

by Allan Adler

Mathematics Department
University of Rhode Island
Kingston, RI 02881

Mathematics Department
Brown University
Providence, RI 02912

July 28, Aug.5, Aug.12, 1993

Best way to contact me is email: ara@altdorf.ai.mit.edu

Deformations of Galois Representations

> Mathematicians of the world unite!
> You have nothing to lose but your cochains!!
> –Allan Adler

As we move towards the middle part of Rubin's sketch of Wiles' proof of Fermat's Last Theorem, we find that words such as "deformation" and "Selmer group" are starting to appear. In order to discuss these notions without excessive circumlocution, it is helpful to be familiar with some of the basic concepts of homological algebra. A brief introduction to the subject can be found in Lang's book Algebra, Ch. IV, where the only exercise given is to take any book on homological algebra and prove all the theorems in it without first looking at the proofs. One can also look in the book Homology, by S.MacLane, or in the books on algebra and on commutative algebra by N.Bourbaki. I will therefore present only a few of the basic notions here as they are needed, making no pretense at completeness.

The basic object of study in homological algebra is called a **chain complex**. A chain complex consists of a sequence of abelian groups

$$\dots A_{-2}, A_{-1}, A_0, A_1, A_2, \dots \tag{1}$$

indexed by the set of all integers. Furthermore, part of the structure is that for every integer n, there is a homomorphism

$$d_n : A_n \to A_{n-1}. \tag{2}$$

Finally it is assumed that the composition of any two consecutive homomorphisms is 0, i.e. for all n and for all $a \in A_n$, we have

$$d_{n-1}(d_n(a)) = 0. \tag{3}$$

When all these ingredients (1),(2) are present and satisfy the condition (3) for all n, we have a chain complex. It is simpler to denote it

$$\dots A_{n+1} \xrightarrow{d_{n+1}} A_n \xrightarrow{d_n} A_{n-1} \xrightarrow{d_{n-1}} \dots$$

or simply to describe the whole thing by a letter such as $\mathbf{A}$ which is also used to denote all the groups in the chain complex. Similarly, to avoid having to worry about subscripts all the time, sometimes we will simply use the same letter to denote all the maps, if no confusion is likely to occur.

Here are some simple examples of chain complexes:

(1) Let A_0 be the group $\mathbf{Z}$ of integers and let $A_n = 0$ for all $n \neq 0$. In this case, there is only one way to choose the maps d_n, namely $d_n = 0$ for all n. It is easy to see that this is a chain complex.

(2) Let A_1 be the group $\mathbf{Z}$ of integers and let $A_n = 0$ for all $n \neq 1$. This might appear to be the same as example (1), but it is not. The reason is that we really care what the indexing of the groups is. In Example (1), the group A_0 is $\mathbf{Z}$ and in this example A_0 is 0. Therefore these examples are considered different even though they are related by a simple shift of the indices.

(3) Let $A_0 = A_1 = \mathbf{Z}$, let $A_{-1} = \mathbf{Z}/2\mathbf{Z}$ and let $A_n = 0$ for all other values of n. Define $d_1 : A_1 \to A_0$ to be the homomorphism defined by $d_1(a) = 2a$ and let $d_0 : A_0 \to A_{-1}$

be reduction modulo 2. Let $d_n = 0$ for all other values of n. Then this defines a complex.

In a chain complex $\mathbf{A}$, the elements of the group A_n are sometimes called **n-chains**. That the composition of two consecutive homomorphisms in a chain complex is 0 can be expressed in another way: it is the same as saying that for every integer n, the image of the homomorphism d_{n+1} is contained in the kernel of the mapping d_n, or more symbolically,

$$im(d_{n+1}) \subseteq ker(d_n). \tag{4}$$

The situation in which we actually have $im(d_{n+1}) = ker(d_n)$ is of particular interest. When this occurs, we say that the chain complex $\mathbf{A}$ is **exact** at A_n. If a chain complex $\mathbf{A}$ is exact at A_n for every n, we simply say that the chain complex is exact. One way of measuring the exactness of a chian complex is to consider the groups $ker(d_n)/im(d_{n+1})$. These groups are called the **homology groups** of the complex. More precisely, we write $H_n(\mathbf{A})$ to denote the group

$$H_n(\mathbf{A}) = ker(d_n)/im(d_{n+1}),$$

which we call the n-homology group of $\mathbf{A}$.

In view of this, it is appropriate to give names to $im(d_{n+1})$ and to $ker(d_n)$ for every n. We call $ker(d_n)$ the group of **n-cycles** and we call $im(d_{n+1})$ the group of **n-boundaries**. The n in n-cycles and n-boundaries refers to the fact that both groups are subgroups of the group A_n. In other words, something in the kernel of one of these homomorphisms is called a "cycle" and if a cycle happens to be an n-chain, we call it an n-cycle. Similarly, if something happens to be in the image of one of these homomorphisms, we call it a boundary, and if a boundary happens to be an n-chain, we call it an n-boundary. The motivation for the terminology "cycles" and "boundaries" can perhaps be understood from the following example.

Fig.1

Consider the tetrahedron shown above. We are only considering the **surface** of the tetrahedron for the moment. Note that it has 4 vertices, which we have labeled $0, 1, 2, 3$. It also has 6 edges, which we denote respectively by $01, 02, 03, 12, 13, 23$. And finally, it has 4 faces, which we denote respectively by $012, 013, 023, 123$. [1] We associate to this situation the following chain complex K. If $n \neq 0, 1, 2$ then we let K_n=0. We let K_0 be a four dimensional real vector space with basis x_0, x_1, x_2, x_3. We let K_1 be a 6 dimensional real vector space with basis $x_{01}, x_{02}, x_{03}, x_{12}, x_{13}, x_{23}$. We let K_2 be a real four dimensional vector space with basis $x_{012}, x_{013}, x_{023}, x_{123}$. Thus the basis for K_n is indexed by the n-dimensional simplices in our tetrahedron. We define the mappings d_n as by the rule

$$d_n(x) = \begin{cases} x_j - x_i & \text{if } n = 1 \text{ and } x = x_{ij} \\ x_{jk} - x_{ik} + x_{ij} & \text{if } n = 2 \text{ and } x = x_{ijk} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

In this case d_n is called a **boundary operator** since it associates to each simplex a certain linear combination of the simplices that form the boundary of that simplex. Motivated by this example, one often refers to the maps d_n as boundary operators in any chain complex, although other terminology is also widely used.

For convenience, let us introduce additional symbols x_{ij} and x_{ijk} with i, j, k not necessarily in increasing order, with the convention that interchanging two of the subscripts changes the sign, e.g. $x_{21} = -x_{12}$ and $x_{231} = -x_{213} = x_{123}$. Thus, these new elements are expressed in terms of the old ones.

---

[1] As Euler observed, we have

$$V - E + F = 2,$$

where V is the number of vertices, E is the number of edges and F is the number of faces. This is true for any convex polyhedron, as Euler observed. In honor of these observation, one has named a very certain topological invariant, which is based on this observation, the **Euler characteristic**. Euler's observation may be expressed in more modern terminology by saying that the Euler characteristic of a sphere is 2.

In this example, we can see the origin of the term "cycle" to describe an element in the kernel of one of the mappings d_n. For example, any 1-cycle is a linear combination of elements of the form

$$x_{i_1 i_2} + x_{i_2 i_3} + \ldots + x_{i_{r-1} i_r} + x_{i_r i_1},$$

i.e. of elements corresponding to a closed path (i.e. a cycle) in the tetrahedron. Similarly, any 2-cycle is a scalar multiple of the sum $x_{012} - x_{013} + x_{023} - x_{123}$.

It is a simple exercise in linear algebra to compute the homology groups of the complex $\mathbf{K}$ we have just defined. And one finds that $H_n(\mathbf{K})$ is a vector space whose dimension is 0 for $n \neq 0, 2$ and 1 otherwise. The dimensions of these homology groups are called the **Betti numbers** of the chain complex. Often one denotes the n-th Betti number of a complex with a lower case h. If we take the alternating sum of the Betti numbers, we get

$$\sum_{n=0}^{\infty}(-1)^n dim\ H_n(K) = \sum_{n=0}^{\infty}(-1)^n h_n(K) = 1 - 0 + 1 = 2,$$

which is the same value we got from Euler's formula. It is true in general that one can compute the Euler characteristic either by taking the alternating sum of the dimensions of the chain groups or by taking the alternating sum of the dimensions of the homology groups.

What we have just done with the tetrahedron is typical of mathematical practice in applications of homology theory. We start with some object that we would like to study, e.g. the surface of the tetrahedron. We then torture the object until it gives us a chain complex. We then compute the homology of the chain complex and then we call those homology groups the homology groups of the object we started with. Thus, in the above example, we have found the homology groups of the surface of the tetrahedron to be $\mathbf{R}$ in dimensions 0 and 2 and to be 0 everywhere else.

It is important to emphasize that this terminology is inherently ambiguous. The reason is that there are often many ways to associate a chain complex to the object you wish to study. One might thereby wind up with various homology groups, depending on the method one chooses. In order to be unambiguous, when one speaks of the homology groups of the object one is studying, one really needs to specify also the method one used to obtain the chain complex. In practice, it is a lot of work to come up with an idea for producing a chain complex and one hopes that whatever idea one comes up with will work fairly generally on a wide class of objects. This is usually the case and one names the homology groups after the method one uses, or sometimes after the discoverer. The example we have just given for the tetrahedron is called **simplicial homology**.

Exercise 1: Let X be a finite set and let S be a collection of nonempty subsets of X such that if U, V are nonempty subsets of X with $U \in S$ and if $V \subseteq U$, then $V \in S$. Show that our labeling of the vertices, edges and faces of the tetrahedron is a special case of this construction, with $X = \{0, 1, 2, 3\}$ and S the set of all nonempty subsets of X other than X itself.

Exercise 2: In general, suppose that X has $n + 1$ elements, which we may suppose are the integers $0, 1, \ldots n$. Identify these integers with a standard basis for $\mathbf{R}^{n+1}$, say $x_0, \ldots, x_n$. If U is a subset of X, denote by $|U|$ the set of all linear combinations

$$\sum_{u \in U} c_u x_u \tag{6}$$

with all $c_u \geq 0$ and with $\sum_{u \in U} c_u = 1$. Denote by $|S|$ the union in $\mathbf{R}^{n+1}$ of all $|U|$ such that $U \in S$. We call $|S|$ the **geometric realization** of (X, S). Show that the geometric realization of the example given in the preceding exercise is the surface of a tetrahedron in $\mathbf{R}^3$. (Hint: According to the constructions we are given, we are actually supposed to work in $\mathbf{R}^4$. But note that for every $U \subset X$, the set $|U|$ is contained in the hyperplane $\sum_{t \in X} c_t x_t = 1$. This being a hyperplane in $\mathbf{R}^4$, we can identify it with a real Euclidean space of dimension 3. The set $|X|$ itself is a solid tetrahedron in the Euclidean space. Excluding X itself from S in effect says we only consider the surface of that tetrahedron.)

Exercise 3: Compute the simplicial homology groups of an icosahedron.

Exercise 4: Construct a pair (X, S) whose geometric realization is homeomorphic to the surface of a torus. Compute its simplicial homology groups.

Exercise 5: Returning to the case of the tetrahedron, take the chain complex we cosntructed and replace all the real vector spaces we constructed by complex vector spaces with the same bases, and define the maps between them the same way. Show that we still obtain a chain complex and compute the homology groups.

Remark. This last exercise illustrates one of the ways in which was can associate a different chain complex to the tetrahedron. Strictly speaking, when we originally introduced the simplicial homology groups, we should have called them the **simplicial homology groups with real coefficients**. What we have introduced in the preceding exercise are the homology groups with complex coefficients. The term "coefficients" refers to the fact that the elements of the chain groups are linear combination, with real coefficients, of the basis elements in one of them and with complex coefficients in the other. Similarly, one could have coefficients in the ratinal numbers, in the integers modulo 7, in the 5-adic integers, in the ring of entire functions, etc.

We also need the notion of a **cochain complex**. It is the same as a chain complex except that the homomorphisms between consecutive groups point in the direction of increasing numbers instead of descreasing numbers. For example, take the complex we considered, with real coefficients, for the tetrahedron. For each n, let L_n denote the dual space of the vector space K_n. The transpose of the mapping $d_n : K_n \to K_{n-1}$ is a homomorphism d'_n from L_n to L_{n+1}. Then the system of groups and mappings

$$\ldots L_n \xrightarrow{d'_{n+1}} L_{n+1} \xrightarrow{d'_{n+2}} L_{n+2} \ldots$$

is a cochain complex. Indeed, we have $d'_{n+1} \circ d'_{n+1} = (d_{n+1} \circ d_{n+2})' = 0' = 0$, where $'$ denotes transpose.

We obtain terminology for cochain complexes by prepending the prefix "co-" to the corresponding terminology for chain complexes. For example, the n-group in a cochain complex is called the group of n-cochains; the kernel of the mapping from the n-th group to the $n+1$-th group is called the group of n-cocycles; the image of the mapping from the $n-1$-th group to the n-th group is called the group of n-coboundaries; and the group of n-cocycles modulo the group of n-coboundaries is called the n-th cohomology group of the complex. It is denoted with upper letters, i.e. the n-cohomology group of the cochain complex $\mathbf{L}$ is denoted $H^n(\mathbf{L})$.

Here is another example of a cochain complex. We start with the 2-sphere S^2, which as we know by now is a differentiable manifold. For each integer n, denote by Λ^n the space of all $\mathcal{C}^\infty$ differential n forms on S^2. Thus, Λ^n is 0 unless $n = 0, 1, 2$. A 0-form is just a differentiable function on S^2. A 1-form is something which in local coordinate charts is of the form $adx_1 + bdx_2$, where a, b are functions of x_1, x_2. Similarly, a 2-form is something which in local coordinates has the form $cdx_1 \wedge dx_2$, where c is a function of x_1, x_2. The mapping from Λ^n to Λ^{n+1} is exterior derivative. In other words, for a 0-form f, it is the total differential $df = f_1 dx_1 + f_2 dx_2$, where f_i denotes the partial derivative with respect to x_i. For a 1-form $adx_1 + bdx_2$ it is $(b_1 - a_2)dx_1 \wedge dx_2$. In all other cases, it is the zero mapping. One can then verify that the cohomology groups are equal to $\mathbf{R}$ in dimensions 0 and 2 and are 0 in all other cases. This way of computing cohomology is due to de Rham and accordingly is called **de Rham cohomology**. It is explained in many places, including the book of Singer and Thorpe. Note that the final answer actually agrees with the result of simplicial cohomology of a tetrahedron and that the tetrahedron is topologically a sphere. One can define the de Rham cohomology groups on any manifold. The fact that one gets the same answer from simplicial cohomology as from de Rham cohomology is true for any compact manifold and is called the **de Rham theorem**.

In general, homology and cohomology provide a lot of useful information and that is why we study them. Here are some cute results one can prove using cohomology:

(i) The **Ham Sandwich Theorem**: Let $A_1, \ldots, A_n$ be measurable subsets of $\mathbf{R}^n$ such that each A_i has finite and positive measure. Then there exists a hyperplane H in $\mathbf{R}^n$ such that for each i, the measure of the part of A_i on one side of H is the same as the measure of the part of A_i on the other side of H.

(ii) **Brouwer Fixed Point Theorem** Any continuous mapping from the closed ball to itself in $\mathbf{R}^n$ has a fixed point.

Cohomology of Groups

The examples which are of concern to us in this lecture will not be based on the cohomology of topological spaces. They will be based on the cohomology of groups.

Let G be a group. By a **G-module** we mean an abelian group A together with a homomorphism $\rho : G \to Aut(A)$ from G to the group of automorphisms of the group A. For reasons of convenience, we will often write $g \cdot a$ instead of $\rho(g)(a)$ for $g \in G$ and $a \in A$. Given a group G and a G-module A, we can form a cochain complex $\mathcal{C}(G; A)$ as follows: for $n < 0$, we let $\mathcal{C}^n(G; A) = 0$. For $n \geq 0$, we let $\mathcal{C}^n(G; A)$ be the group of all functions

$$f : \underbrace{G \times \ldots G}_{n} \to A$$

from the product of n copies of G to A, i.e. all functions of n variables from G to A. We define the coboundary homomorphism δ as follows. Let f be an n-cochain of G with values in A. We define δf to the the $n+1$-cochain defined as follows. If $g_0, \ldots, g_n$ are elements of G, then we define

$$\begin{aligned}(\delta f)(g_0, \ldots, g_n) &= g_0 \cdot f(g_1, \ldots, g_n)\\ &- f(g_0 g_1, g_2, \ldots, g_n)\\ &+ f(g_0, g_1 g_2, g_3, \ldots, g_n)\\ &- \ldots \pm f(g_0, g_1, \ldots, g_{n-1} g_n)\\ &\mp f(g_0, g_1, \ldots, g_{n-1})\end{aligned}$$

I would be dishonest if I did not tell you that this formula looked as unpleasant to me when I first looked at it as it must now look to you. This is an example of why it is sometimes advantageous to strive for more general points of view, since from those points of view ugly formulas such as this one can fall out of simple reasoning, whereas otherwise one might never think of it in a milllion years.

At any rate, one can verify that this defines a cochain complex. The resulting groups are called the **cohomology groups of G with coeffients in A** and are denoted $H^n(G; A)$. Note that they depend not only on G and A but also on the homomorphism $\rho : G \to A$.

In view of the unnatural appearance of the mapping δ, it is a good idea to describe the n-cocycles and coboundaries for small values of n.

A 0-cochain is a function of no variables, hence is a constant function. Thus, a 0-cochain is simply an element of A. Let $a \in A$, viewed as a 0-cochain, and compute its coboundary δa. For $g \in G$, we have

$$(\delta a)(g) = g \cdot a - a.$$

Therefore, saying that a is a 0-cocycle is the same as saying that $g \cdot a = a$ for all elements g of G, i.e. that a is fixed by every element of G. Thus, when you study all the elements fixed by some group, you are studying cohomology. In particular, when you study symmetric polynomials, you are studying cohomology.

A 1-cochain is a function $f : G \to A$. The coboundary of f is the function $\delta f : G \times G \to A$ given by

$$(deltaf)(g_0, g_1) = g_0 \cdot f(g_1) - f(g_0 g_1) + f(g_0).$$

To say that f is a 1-cocycle is to say that for every pair g_0, g_1 of elements of G, we have

$$f(g_0 g_1) = f(g_0) + g_0 \cdot f(g_1).$$

This is rather like saying that f is a homomorphism, except for the slight modification of the last term of the right hand side. In fact, such mappings are called **crossed homomorphisms** from G to A. Actually, since G acts as automorphisms of A, in particular it acts as permutations of A. But A also acts as permutations of A, by translation. And therefore together A and G generate a larger group whose elements may be written as pairs (g, a) and where the operation of a pair (g, a) on A is by the rule

$$(g, a) \cdot b = a + g \cdot b$$

for all $b \in A$. The operation in this larger group is given by

$$(g_1, a_1)(g_2, a_2) = (g_1 g_2, a_1 + g_1 \cdot a_2).$$

We denote this larger group by $A \cdot G$. Suppose f is a 1-cocycle of G with values in A. Define the mapping $F : G \to A \cdot G$ by the rule $F(g) = (f(g), g)$. Then it is easy to see that F is a homomorphism of groups. Conversely, if $f : G \to A$ is a function such that this rule gives one a function gives one a homomorphism from G to $A \cdot G$, then f is a 1-cocycle. Thus, f is a 1-cocycle if and only if it allows G to act as a group of permutations of A by the rule

$$F(g)(a) = f(g) + g \cdot a.$$

As for 1-cocoundaries, we have in effect already described them when we defined the coboundary of a 0-cochain.

Exercise 6: Verify that if $a \in A$ then $\delta(\delta(a)) = 0$.

Deformations and group cohomology: Moduli of Riemann surfaces

In Rubin's sketch, Mazur is alleged to have constructed a deformation space for Galois representations and studied its properties. In order to motivate this study, I would first like to consider an example somewhat closer to home which illustrate certain important features.

Riemann showed that isomorphism classes of compact Riemann surfaces of genus $g > 1$ depend on $3g - 3$ complex parameters. I would like to try to provide some insight into this result in a way that will hopefully also shed light on the deformations of Galois representations.

Let X be a compact Riemann surface. Then the universal covering of X is analytically isomorphic to the upper half plane (UHP). This follows from the generalized Riemann mapping theorem. The fundamental group $\pi_1(X)$ acts as a group of covering transformations on UHP and the elments of this group are holomorphic. The group of all analytic automorphisms of UHP is the group $SL_2(\mathbf{R})$ modulo its center, which consists of $\pm 1_2$, i.e. the group of all linear fractional transformations with real coefficents and determinant 1. To simplify the discussion,[2] we will pretend that the group is actually $SL_2(\mathbf{R})$ and contained in the real vector space $M_2(\mathbf{R})$ of all 2×2 matrices. With this convention, we then have an injective homomorphism from $\pi_1(X)$ into $SL_2(\mathbf{R})$. Furthermore, the image of this homomorphism is a discrete subgroup and the orbit space $SL_2(\mathbf{R})/\pi_1(X)$ is compact.[3] Finally, if g is any element of $SL_2(\mathbf{R})$, we can conjugate by g to produce another representation, but the resulting Riemann surface will be conformally equivalent to X. In fact, the set of all compact Riemann surfaces of genus g can be identified with the set of all injective homomorphisms from $\pi_1(X)$ into $SL_2(\mathbf{R})$ such that the image is discrete and cocompact, modulo the action of $SL_2(\mathbf{R})$ by conjugation. The point of this observation is that all compact Riemann surfaces of genus g are homeomorphic and have the same fundamental group. Each one of them gives rise to an action of that group on UHP from which we can recover the Riemann surface, and in turn to a homomorphism of this same fundamental group into $SL_2(\mathbf{R})$. The structure of the fundamental group is known.

---

[2] We can realize the group of linear fractional transformations as a group of 3×3 matrices, but why introduce technical complications that don't really affect the main ideas?

[3] We express this by saying that the image of $\pi_1(X)$ in $SL_2(\mathbf{R})$ is cocompact.

Fig.2

There are $2g$ generators $\alpha_1, \dots, \alpha_g, \beta_1, \dots, \beta_g$, and the group $\pi_1(X)$ is the free group generated by these generators modulo the single relation

$$[\alpha_1, \beta_1][\alpha_2, \beta_2] \dots [\alpha_g, \beta_g] = 1,$$

where $[x, y] = xyx^{-1}y^{-1}$. A homomorphism from $\pi_1(X)$ into $SL_2(\mathbf{R})$ is determined by its values on the generators, hence by an element of the cartesian product of $2g$ copies of the group $SL_2(\mathbf{R})$. One can check that this product group is a manifold, since $SL_2(\mathbf{R})$ is a manifold.[4] Let Y denote the set of all $2g$-tuples of matrices $(M_1, \dots, M_{2g})$ such that each M_i is in $SL_2(\mathbf{R})$ and such that

$$[M_1, M_2][M_3, M_4] \dots [M_{2g-1}, M_{2g}] = 1.$$

The elements of Y are all of the homomorphisms of $\phi_1(X)$ into $SL_2(\mathbf{R})$. Weil proved in a general setting that the elements of Y that correspond to injective homomorphisms whose image is discrete and cocompact is an open subset Y_0 of Y. Let $y = (M_1, \dots, M_{2g})$ be an element of Y_0. We want to compute the tangent space to Y_0 at y. The tangent space to $SL_2(\mathbf{R})$ at the identity element is the space $sl_2(\mathbf{R})$ of matrices of trace 0, as one easily verifies. The nice thing about $SL_2(\mathbf{R})$ being a group is that multiplication by a group element g maps the tangent space at the identity element 1_2 to the tangent space at the element g. Now, consider a curve through y with a presecribed tangent vector at y. Say for definiteness that the tangent direction is $(A_1, \dots, A_{2g})$, where each A_i lies in $sl_2(\mathbf{R})$. And let us take the curve to be

$$t \mapsto (e^{A_1 t} M_1, \dots, e^{A_{2g} t} M_{2g}).$$

---

[4] Since it is a topological group, it is enough to check that it is locally euclidean at the identity element. But the exponential mapping $A \mapsto e^A$ from the space $sl_2(\mathbf{R})$ of matrices of trace 0 into $SL_2(\mathbf{R})$ has nonvanishing differential at 0 and is therefore locally invertible.

Applying the word to this curve, we obtain the function ϕ of t with values in $SL_2(\mathbf{R})$ given by

$$\phi(t) = [e^{A_1 t}M_1, e^{A_2 t}M_2] \ldots, [e^{A_{2g-1} t}M_{2g-1}, e^{A_{2g} t}M_{2g}].$$

We can also view ϕ as having values in the Euclidean space $M_2(\mathbf{R})$ and we compute its derivative at 0:

$$\begin{aligned}
\phi'(0) = {} & A_1[M_1, M_2][M_3, M_4] \ldots [M_{2g-1}, M_{2g}] \\
& + M_1 A_2 M_2 M_1^{-1} M_2^{-1} [M_3, M_4] \ldots [M_{2g-1}, M_{2g}] \\
& - M_1 M_2 M_1^{-1} A_1 M_2^{-1} [M_3, M_4] \ldots [M_{2g-1}, M_{2g}] \\
& - [M_1, M_2] A_2 [M_3, M_4] \ldots [M_{2g-1}, M_{2g}] \\
& \ldots \\
& + [M_1, M_2][M_3, M_4] \ldots [M_{2g-3}, M_{2g-2}] A_{2g-1} [M_{2g-1}, M_{2g}] \\
& + [M_1, M_2][M_3, M_4] \ldots [M_{2g-3}, M_{2g-2}] M_{2g-1} A_{2g} M_{2g} M_{2g-1}^{-1} M_{2g}^{-1} \\
& - [M_1, M_2][M_3, M_4] \ldots [M_{2g-3}, M_{2g-2}] M_{2g-1} M_{2g} M_{2g-1}^{-1} A_{2g-1} M_{2g}^{-1} \\
& - [M_1, M_2][M_3, M_4] \ldots [M_{2g-3}, M_{2g-2}][M_{2g-1}, M_{2g}] A_{2g}
\end{aligned} \tag{7}$$

The vanishing of the right hand side of (7) is precisely the condition for the $2g$-tuple $(A_1, \ldots, A_{2g})$ to be tangent to Y_0 at the point $y = (M_1, \ldots, M_{2g})$. Since y does belong to Y_0, we can rewrite this condition in the form:

$$0 = A_1 + M_1 A_2 M_1^{-1} - M_1 M_2 M_1^{-1} A_1 (M_1 M_2 M_1^{-1})^{-1} - [M_1, M_2] A_2 [M_1, M_2]^{-1} \ldots . \tag{8}$$

As Weil explains, if we write the word $[M_1, M_2][M_3, M_4] \ldots [M_{2g-1}, M_{2g}]$ out in full in the form

$$\delta_1 \delta_2 \ldots \delta_{4g}$$

where each δ_i is either an M or an M^{-1} and for each such i let U_i be the A, corresponding to that M, then the right hand of (8) is the sum of terms $\pm \eta_i U_i \eta_i^{-1}$, where the sign is +

if δ_i is an M and $-$ otherwise, and where η_i is the product $\delta_1 \dots \delta_{i-1}$ if δ_i is an M and is otherwise the product $\delta_1 \dots \delta_i$. Thus does one describe the tangent space to Y_0 at y.

On the other hand, consider the mapping ψ which associates to each of the generators γ_i of $\pi_1(X)$ the permutation

$$M \to A_i + \rho(\gamma_i) M \rho(\gamma)^{-1}$$

of $sl_2(\mathbf{R})$. In order for ψ to extend to a permutation representation of $\pi_1(X)$, it is necessary and sufficient that

$$[\psi(\gamma_1), \psi(\gamma_2)] \dots [\psi(\gamma_{2g-1}), \psi(\gamma_{2g})]$$

be the identity permutation. But if one expands this expression, one finds that this condition is precisely the same as equation (8). It follows that having a tangent vector to Y_0 at y is equivalent to having a 1-cocycle of $\pi_1(X)$ with coefficients in $sl_2(\mathbf{R})$.

On the other hand, the mapping $g \mapsto g \cdot y \cdot g^{-1}$ of $SL_2(\mathbf{R})$ into Y_0 maps the identity element of $SL_2(\mathbf{R})$ to y and the differential of that mapping sends the tangent space $sl_2(\mathbf{R})$ of $SL_2(\mathbf{R})$ at the identity to the space of 1-coboundaries. Explicitly, we start with the curve e^{At} through the identity element 1_2 of $SL_2(\mathbf{R})$ with tangent vector $A \in sl_2(\mathbf{R})$ at $t = 0$. Applying ψ we get

$$\psi(e^{At}) = e^{tA} \cdot y \cdot e^{-At} = (e^{tA} \cdot M_1 \cdot e^{-At}, \dots, e^{tA} \cdot M_{2g} \cdot e^{-At}).$$

Differentiating with respect to t and evaluating at $t = 0$ we get

$$([A, M_1], \dots, [A, M_{2g}],$$

which is a tangent vector to Y_0 at the point y. But in our earlier description of the tangent space at y, we used tangent vectors at the point $(1_2, \dots, 1_2)$ and moved them over to the point y by multiplication by y; that is implicit in our having written the curves thorugh y in the form $(e^{A_1 t} M_1, \dots, e^{A_{2g} t} M_{2g})$. Therefore, in order to compare the new tangent vector with the old one, we have to move the new one to the element $(1_2, \ldots, 1_2)$ by multiplying by $y^{-1} = (M_1^{-1}, M_2^{-1}, \ldots, M_{2g}^{-1}$. This transforms the tangent vector $([A, M_1], \ldots, [A, M_{2g}]$ at y to the tangent vector

$$(A - M_1 A M_1^{-1}, \ldots, A - M_{2g} A M_{2g}^{-1})$$

at $(1_2, \ldots, 1_2)$. The mapping which associates to a generator γ_i of $\pi_1(X)$ the matrix $A - M_i A M_i^{-1}$ does extend to a 1-cocycle of $\pi_1(X)$ with coefficients in $sl_2(\mathbf{R})$, namely the 1-cocycle

$$\gamma \mapsto A - \rho(\gamma) A \rho(\gamma)^{-1},$$

which is a coboundary, as we claimed.

Therefore, the tangent space to the orbit space $Y_0/SL_2(\mathbf{R})$ at the point represented by y is isomorphic to the cohomology group $H^1(\pi_1(X), sl_2(\mathbf{R}))$. The argument would work for any word or any collection of words, not just the special one we were considering. And Weil actually proves his theorem in that generality.

Analogies to Deformation of Galois Representations

We have persuaded ourselves that an elliptic curve E over $\mathbf{Q}$ has associated to it Galois representations of $G = Gal(\bar{\mathbf{Q}}/\mathbf{Q})$ into $GL_2(\mathbf{Z}_\ell)$ and into $GL_2(\mathbf{F}_\ell)$. We have also persuaded ourselves that a cusp form weight 2 and rational coefficients and corresponding to a Dirichlet series with Euler product gives rise to an elliptic curve over $\mathbf{Q}$ and therefore to Galois representations of G into $GL_2(\mathbf{Z}_\ell)$ and into $GL_2(\mathbf{F}_\ell)$. Let us agree to call a representation of G into $GL_2(\mathbf{Z}_\ell)$ an **ℓ-adic representation** and to call the corresponding representation into $GL_2(\mathbf{F}_\ell)$ a **residual representation**. As we explained in Lecture 1 in connection with Ribet's theorem, to prove Fermat's Last Theorem, we have to show that the ℓ-adic representation corresponding to E arises from a modular form. That is of course the hard part, but it is somewhat more tractable to show that the corresponding statement is true if we look at the residual representations instead. Suppose we succeeded in that more tractable task. Then we would have two ℓ-adic representations of G having the same residual representation. It therefore makes sense to consider all of the possible ℓ-adic representations that have that same residual representation. Thus we are led to consider the possible representations of G into $GL_2(\mathbf{Z}_\ell)$ and this appears to be quite similar in spirit to the problem we just considered in connection with representations of $\pi_1(X)$ into $SL_2(\mathbf{R})$. If that analogy holds, then we might expect a cohomology group $H^1(G, sl_2)$ to play a similar role in this situation. Furthermore, note that the action of the group $SL_2(\mathbf{R})$ on $sl_2(\mathbf{R})$ is a 3 dimension representation isomorphic to the representation of $SL_2(\mathbf{R})$ on the space of quadratic forms in 2 variables. The latter is referred to as ths "symmetric square" of the original representation of $SL_(\mathbf{R})$ on $\mathbf{R}^2$. We also refer to the corresponding representation of $\pi_1(X)$ on $sl_2(\mathbf{R})$ as the symmetric square of the representation ρ. We would expect similar considerations to apply to the situation of ℓ-adic representations of G. But if you now refer to Appendix 1 of Lecture 2, which contains Rubin's sketch of Wiles' proof, you will see that there is mention of the Selmer group of the symmetric square representation of the elliptic curve. So our speculative analogies do seem to be leading us to consider objects that sound like they are the relevant ones.

Our next goal is to explore these analogies further to see whether they can be placed on firmer ground. In order to achieve it, we first must understand the notion of a "profinite group" and the cohomology of profinite groups. A good reference for this material is the book Cohomologie Galoisienne, by Jean-Pierre Serre, Lecture Notes in Mathematics #5.

Profinite Groups

Definition. By a **profinite group**, we mean any group which arises as an inverse limit of finite groups.

In order to understand this definition, it is necessary to understand what an inverse limit is. I defined a weaker version of this concept in Lecture 2 when I defined the ℓ-adic integers. Please refer there for the details. I will now define the notion of an inverse limit in a somewhat more general setting. First we need the notion of an inverse system.

Definition. By a **directed set**, we mean a partially ordered set $(D, <)$ such that $\forall i, j \in D$ there is a $k \in D$ such that $k > i$ and $k > j$. By an **inverse system** of groups we mean a collection $(G_i)_{i \in D}$ of groups G_i indexed a directed set D, together with a homomorphism $\phi_{ij} : G_i \rightarrow G_j$ for all $i > j$ in D such that we have $\phi_{jk} \circ \phi_{ij} = \phi_{ik}$ whenever $i > j > k$.

Before completing the definition of the terminology "inverse limit", let us give some examples of directed sets and inverse limits.

(a) Let ℓ be a prime number. We take D to be the set of nonnegative integers and $>$ to have its usual meaning. For each i in D let G_i be the group $\mathbf{Z}/p^i\mathbf{Z}$ of integers modulo[5] p^i. Whenever $i > j$, let $\phi_{ij} : G_i \rightarrow G_j$ be reduction modulo p^j. In this case, the construction leads to the p-adic integers $\mathbf{Z}_p$, precisely in our earlier construction of them.

(b) Let H be any group and let D be a collection of normal subgroups of finite index in H. Assume that the intersection of any two subgroups in D always contains a third subgroup in D. Define $i > j$ in D if i is a subgroup of j. For every $i \in D$, let $G_i = H/i$. if $i > j$, define $\phi_{ij} : G_i \rightarrow G_j$ to be the canonical homomorphism from H/i onto H/j. Then the groups G_i and the homomorphisms ϕ_{ij} form an inverse system indexed by $(D, >)$. The inverse limit of that system is considered a completion of H. For example, if D runs over all normal subgroups of finite index, the inverse limit is called the **profinite completion** of H. If p is a prime and if D runs over all normal subgroups whose whose index is a power of p, then the inverse limit is called the **pro-p completion** of h. Similarly, one can define the **pro-nilpotent completion** and the **pro-solvable completion** of h.

A profinite group always has a natural topology. Indeed, since it is contained in the product of the finite groups G_i, each of which can be given the discrete topology and is

---

[5] For $i = 0$, the group G_i has only one element.

therefore compact and Hausdorff, the inverse limit is a closed subgroup of the cartesian product of the G_i. But that product is compact by the tychonoff theorem, so the inverse limit is compact and Hausdorff. Since the cartesian product is totally disconnected[6], the same is true of the inverse limit. conversely, one can show that any totally disconnected compact Hausdorff topological group H is a profinite group. To see this, in (b) above, let D be the collection of all open normal subgroups[7] of H. then one has a natural continuous mapping from H into the cartesian product of the groups G/i with $i \in D$, and one verifies that its image is precisely the inverse limit.

One important example of a profinite group is the Galois group of a normal extension[8] L of a field K. It is *not* assumed that L is a finite extension of K. Let G be the group of all automorhisms of L/K. as mentioned in lecture 2, we can give the group G the topology of pointwise convergence. This is the same as giving L the discrete topology and viewing G as a subset of the cartesian product

$$\prod_{i \in L} L,$$

which is Hausdorff and totally disconnected. So G inherits these properties. To see that G is actually compact, for every $i \in L$ denote by l_i the set of all conjugates[9] of i, then G leaves each of the sets l_i invariant. Therefore, instead viewing G as a subset of the product

---

[6] A topological space is said to be **totally disconnected** if no subset with more than one element is connected. some of these spaces are quite pathological. For example, the stone-čech compactification of an infinite discrete space is uncountable and totally disconnected and has the additional property that the closure of every open subset is open.

[7] An open subgroup of a compact group always has finite index.

[8] We say that L is **normal** over K if L is algebraic over K and if every irreducible polynomial $f(x) \in K[x]$ with one root in L has all its roots in L.

[9] If $f(x)$ is the irreducible monic polynomial with coefficients in K and such that $f(i) = 0$, then the roots of $f(x)$, including i, are called the conjugates of i.

of copies of L, it is better to view it as a subset of the product of copies of l_i indexed by the elements i of L. this makes G a closed subset of a cartesian product of finite sets and therefore compact. Since G is therefore compact, Hausdorff and totally disconnected, it follows that G is profinite. It is therefore the inverse limit of finite groups. this can be seen quite explicitly. For each extension E of K lying in L, let G_E denote the subgroup of G consisting of automorphisms σ of L/K fixing all the elements of E. According to the **infinite dimensional version of the fundamental theorem of Galois theory**, the mapping $E \mapsto G_E$ is a 1-1 correspondence between the intermediate extensions E/K and the *closed* subgroups of G. furthermore, if E/K is finite then G_E is an *open* subgroup of G of index equal to the degree of E over K. And E is normal over K if and only if G_E is a normal subgroup of G, with the group of automorphism of E/K being isomorphic to the group G/G_E. thus, G arises as the inverse limit of the groups $Gal(K/E) = G/G_E$ as E runs over finite normal extensions of K lying in L.

Here are some closed subgroups of $G = Gal(\bar{\mathbf{Q}}/\mathbf{Q})$ that are of particular interest in our studies:

(α) Let p be a prime number and let G_p denote the galois group of $\bar{\mathbf{Q}}_p/\mathbf{Q}_p$, where $\mathbf{Q}_p$ denotes the field of p-adic numbers and $\bar{\mathbf{Q}}_p$ denotes its algebraic closure. The field $\bar{\mathbf{Q}}$ of all algebraic numbers is a subfield of $\bar{\mathbf{Q}}_p$ and invariant under all of the elements of G_p. Therefore, each element of G_p determines an element of G and we therefore obtain a mapping (in fact a continuous homomorphism) of G_p into G. That homomorphism is injective, since $\bar{\mathbf{Q}}_p$ is generated over $\mathbf{Q}_p$ by the elements of $\bar{\mathbf{Q}}$. since an injective continuous mapping from a compact space to a Hausdorff space is a homeomorphism onto its image, it follows that G_p is topologically isomorphic, by a canonical homomorphism, to a closed subgroup of G which we also denote G_p. we call G_p the **decomposition group** of p.

(β) Let R_p denote the ring generated by $\mathbf{Z}_p$ and all roots of unity of order prime to p in

$\bar{\mathbf{Q}}_p$. Then R_p is invariant under all automorphisms σ in G_p. One can prove easily that R_p/pR_p is an algebraic closure[10] of $\mathbf{F}_p$. Furthermore, since pR_p is obviously invariant under G_p as well, it follows that G_p acts on $R_p/pR_p = \bar{\mathbf{F}}_p$. Thus, we have a homomorphism from G_p into $Gal(\bar{\mathbf{F}}_p/\mathbf{F}_p)$. The kernel of that homomorphism will be denoted I_p and is called the **inertia group** of p. According to our conventions, we can also identify I_p with a closed subgroup of G.

(γ) We can also describe I_p as the group of all σ in G_p such that $\sigma(\zeta) - \zeta \in pR_p$ for every root of unity ζ of order prime to p. we can generalize this construction by considering the set G^i of all $\sigma \in G_p$ such that the stronger condition $\sigma(\zeta) - \zeta \in p^i R_p$ holds for every root of unity ζ of order prime to p. For $i = 0$, we have $G^0 = G_p$. for $i = 1$, we have $G^1 = I_p$. for $i > 1$, we obtain the **higher ramification groups** of p. these can all be identified with closed subgroups of G. One can prove that G^{i+1} is a closed normal subgroup of G^i and that the quotient group G^i/G^{i+1} is abelian. This implies that G_p is pro-solvable. In particular, the symmetric group on $n \geq 4$ letters can never be the galois group of a finite extension of $\mathbf{Q}_p$. This is in marked contrast to the situation over $\mathbf{Q}$.

Let us now pursue further the analogy between the study of representations of $\pi_1(X)$ into $Sl_2(\mathbf{r})$ and the study of Galois representations of $G = \bar{\mathbf{Q}}/\mathbf{Q}$ into $Gl_2(\mathbf{Z}_\ell)$. In the former situation, we used the fact that $\pi_1(X)$ is finitely generated. Indeed, since we knew we had a finite set $\gamma_1, \ldots, \gamma_{2g}$ of generators in that case, we knew that a representation of $\pi_1(X)$ was determined by its values on those generators and could therefore be identified with a $2g$-tuple $(M_1, \ldots, M_{2g})$ of 2×2 matrices each of which was in $sl_2(\mathbf{R})$. This allowed

---

[10] The algebraic closure $\bar{\mathbf{F}}_p$ of $\mathbf{F}_p$ is the union of the fields $\mathbf{F}_{p^n}$. The multiplicative group of the latter is cyclic of order $p^n - 1$, so the nonzero elements of $\bar{\mathbf{F}}_p$ consist of roots of unity of order prime to p and one gets all such roots of unity. These are precisely the images modulo p of the corresponding roots of unity over $\mathbf{Q}_p$.

us to regard the set of representations of $\pi_1(X)$ into $Sl_2(\mathbf{R})$ as a subspace of the product of $2g$ copies of $Sl_2(\mathbf{R})$. If we wish to do something similar with the Galois representations of G into $Gl_2(\mathbf{Z}_\ell$, we need to have a corresponding notion of what it means for a profinite group to be "finitely generated". Since we would like to wind up with the representations being a subset of the product of a finite number of copies of $Gl_2(\mathbf{Z}_\ell)$, all we really require is the following definition.

Definition. *Let H be a profinite group and let S be a subset of H. we say that H is* **topologically generated** *by S if the subgroup of H generated by S is dense in H. We say that G is* **topologically finitely generated** *if H is topologically generated by a finite set.*

One can easily prove that if S generates H topologically then if every continuous homomorphism from H to another profinite group is uniquely determined by its values on S. In particular, let us tentatively make the following assumption:

Assumption 1. *The Galois group G of $\bar{\mathbf{Q}}/\mathbf{Q}$ is topologically finitely generated.*

With Assumption 1, we can conclude that the set of all Galois representations of G in