Introduction and logistics

Seminar on Fermat's last theorem

1 Introduction and syllabus review

Welcome to the first meeting of the seminar! The idea for the day is this: in the first part of the class, we'll review the syllabus, and I'll attempt to sketch some general principles for how to give a mathematical talk of this kind (and how to attend one!). We'll take a short break; when we return, I'll give an introductory talk on the material, and you can assess for yourselves how well you think I did. Finally, we'll lay out the plan for the seminar and schedule the first round of talks.

If anyone has to leave early, talk to me at the break and we'll schedule you a talk then; we may not take the full time today, but best to be safe.

2 On giving (and attending) talks and writing

2.1 On attending talks

In this class, you'll be asked to fill out a feedback form on each presentation, which will also give you some guidelines for what to look for in a talk—what is working well? what isn't? what takeaways can you get? Beyond that, here are some general guidelines for following talks (not just in this class); none are required, but may be useful to you.

- The ideal is to follow the talk all the way through. This is not always possible, partially because many talks are bad and partially because for many you won't have the background. If you lose the thread, it's easy to just stop paying attention; try instead to find it again.
- It can be helpful or unhelpful to try to take notes during talks; see what works for you.
- Keep in mind some basic questions, and try to answer them for yourself by the end of the talk: what question is the speaker trying to answer? Why is it interesting or important? What is the answer, at least very roughly speaking? Is there a simple example where you can concretely understand the question and answer?
- Try to learn at least one (fairly concrete thing) from each talk (e.g. a definition, a theorem of interest, a proof idea). This is not a lofty goal, but it'll add up.
- Try to formulate at least one interesting question (perhaps write it down to remember it). If it isn't answered by the end of the talk, ask it.

For feedback, be precise, critical, and constructive. For example, "your talk sucked" is not useful feedback, but neither is "all in all pretty good!" If your classmates do something you liked in their presentations, it's worth pointing it out so they remember to keep doing it; conversely if they do something that makes it harder for you to follow, to help them improve it's important to tell them.

After each class, you'll fill out a feedback form (online) for each presentation (I recommend taking notes during each talk on what you'll put on your feedback forms). These will have your names on them; however your feedback will be passed along **anonymously**.

2.2 Giving presentations: general principles

This is fairly general advice, some of which will be more or less applicable to this class in particular.

- Structure your talks, similar to a paper: introduce your topic, outline the structure; at the end give some concluding remarks. (Like a paper, think of a talk as telling a story, and never stray far from the plot.) Here it can be reasonable for your conclusion to be more of a summary, as your audience may have forgotten and can't scroll back like in a paper. In general it can be worth referring to the bigger picture throughout the talk, to remind your audience why we're doing everything. In this class, your talk is part of a larger sequence of talks, so it's useful to relate your talk to whatever came before (and, if possible, what's coming next, to make it easier for your audience to keep track.
- However, avoid spending too much of your talk describing what you're going to do rather than doing it.
- Keep in mind your audience: lay? peers? experts? (In this case peers, but you likely have more background on your specific topic.)
- Often your audience will be mixed. Try to have something accessible and of interest to each (major) segment, e.g. a sketch for laypeople, a somewhat more technical discussion for non-expert peers, and deeper technical stuff for experts. (Less applicable in this class.)
- Prepare in advance—in this class you'll give a practice talk to me, and I'll ask you to have either slides or notes prepared. Your slides or notes will be posted on the class website after your talk as a reference for your classmates, so make sure they're readable. (Notes from this class will likewise be posted, after being edited to include the tentative schedule.)
- Engage the audience: gestures, motion, tone... Look at your audience as much as possible.
- Interact with the audience, if suitable: solicit questions, ask questions of the audience, etc. (Use your judgment here: in formal contexts this is sometimes frowned-upon.)
- Props: notes can be useful, but you should know your talk well enough that you only need to consult them occasionally to find your place, rather than doing the talk from them. (I haven't practiced this talk, so you'll probably notice I'm not doing this very well.)
- Slides are often good, especially if you need to show something—a graph, a figure, a dataset, a diagram. They are *not* good for a lot of text. Never read off your slides; talk about them!
- If you have slides, make sure they're readable. (You can write slides with LaTeX using Beamer.)

- If you're handwriting (e.g. on a blackboard), make sure your handwriting is readable (in particular large enough).
- Think about the layout of your boards, if there are several. Ideally practice in the same room you'll give your talk in, or use a similar board layout.
- Dually to how to attend a talk, think about what you want your audience to get out of it: ideally they'd follow the whole thing, but if not what's one takeaway they can get? What problem are you trying to solve, why should your audience care, what's your method, what's your answer, what are some examples in simple illustrative cases?
- For a one-off talk, it's usually best to avoid technical details as possible, at least for most of the talk. (You can have a more technical section if you want, but be aware much of your audience may tune out.) The important thing is the big picture: you won't be able to give every detail during a talk. In this class, since your talk is part of a program, you can go into more detail: we don't need every last technical detail of every proof, but you should at least provide the ideas, and be able to go through the details if asked. Important results should be fully proved in most cases.
- For all talks, including in this class: finishing *way* too early isn't great, since if you have the time you should use it, but finishing a little early can be good since it gives extra time for questions, and finishing late is very, very bad. (Starting on time is also important, though not quite as critical.)
- Omit extraneous detail, animations, distracting graphics, writing out overly long formulas; respect your audience's time (and your own).
- Don't talk too fast. (This is difficult when nervous.) Pacing is especially hard if you have slides: since everything is pre-written, it's easy to go through it way too fast. Slow down. It is much better to omit some details or not cover all of your material than to go too fast and leave your audience confused about everything.
- For questions: if you don't know the answer, that's okay and you can admit it; don't imply that you're stupid for not knowing. Conversely never imply that a question is stupid; you can however say something to the effect that it isn't closely related (if true). Partial answers or referrals to sources are also very reasonable.
- A formal academic talk should include at least your main sources (admittedly it can be more or less clear what this means), though not necessarily a full bibliography. A more informal talk can omit them, but if your work is very close to someone else's (or if you're presenting on someone else's work) you should certainly say so. In this class, we know at least your basic sources; if you heavily used another source that you found particularly useful, it's certainly not bad to mention.

A little more about slides vs. chalk talks: the main advantage and disadvantage of slides is that you can present more material. For things like showing off long formulas, graphs, charts, or other visuals, this can be very useful: drawing out your whole figure by hand might take a long time and be difficult to do accurately, but it can still be very helpful for your audience to follow (if it's a good graphic). Slides also act as built-in notes, helping you keep your pace, and they can help make your presentation look more polished. On the other hand, slides also enable you to go too fast, while having to write things by hand keeps you

at a slower pace, which is good for audience comprehension. Using a blackboard is also a little more flexible: you can amend things to clarify, answer audience questions more easily, etc., while it's hard to incorporate new stuff into your slides. Slides also take some time and technical background to prepare, though you shouldn't necessarily be preparing any less for chalk talks.

For a 45-50 minute talk, I tend to think chalk talks are usually better, if for no other reason than that you'd need a lot of slides. For shorter talks where you're more crammed for space, the ability of slides to condense your material is much more useful, e.g. I'd probably recommend them for 10-minute talks. In either case though it's a judgment call and you're free to do either; we should be able to borrow a projector from the department as needed.

2.3 On writing

Within this class, you'll have two kinds of writing assignments: homeworks dispersed throughout the semester, and a final project on a topic of your choice (related to the course material). You should think of these not only as about the mathematical material but also as exercises in effective mathematical communication, and they will be graded for both qualities.

Here are some general guidelines for technical writing (largely cribbed from Bjorn Poonen and a number of other sources I've since forgotten):

- General advice:
  - Above all else, you're trying to be clear.
  - Be as precise and unambiguous as possible. If you have to be imprecise about something, be transparent about it.
  - Use a relatively conversational tone—this is personal preference, but I find it makes it easier to read.
  - Define your terms before you use them, or refer to later definitions if necessary.
- Citations:
  - Use precise citations, e.g. page numbers or labels ("Conjecture 3," "Figure 4"). Include dates/versions for online materials or preprints in case of changes.
  - Don't cite forthcoming work unless a version is publicly available. However, feel free to cite private communications as needed, so long as it is clear this won't be a public work in the foreseeable future.
- The introduction is the part of your paper the most people will read. Write it last, and try to sell your reader on why they should care about your results, what they are, and how they're related to other work in the field, and explain how the paper is going to work (unless it's very short).
- A conclusion will not always be necessary; if it is, it should gesture towards potential future questions, implications of your work, possible criticisms, etc. It should not simply summarize the paper or rehash the introduction.

- If you're tempted to say that something is clear or trivial, make sure that you could quickly and completely explain it if asked; often things that are supposed to be clear hide errors.
- When possible, avoid beginning a sentence with a symbol or formula.
- Avoid contractions by default (they can however be used to make the tone feel lighter).
- Number your results ("Theorem 12.4") and refer to them by number rather than descriptively ("the previous theorem").
- Include examples of your results in simpler but still interesting cases, if possible.
- Keep in mind the big-picture story you are telling; sketch it out for yourself as an outline before writing.

3 Introduction to Fermat's last theorem

Fermat's last theorem refers to the following statement, first posed by Fermat: for any integer $n \geq 3$, there exist no triples of positive integers x, y, z such that

$$x^n + y^n = z^n.$$

There are a few things to observe. First, why $n \geq 3$? Well, we can look at smaller cases: the case $n = 0$ is trivial (this would then be the equation $x^0 + y^0 = z^0$, i.e. $1 + 1 = 1$, which clearly has no solutions), and the case $n = 1$ is not much harder ($x + y = z$ clearly has infinitely many solutions, by taking any x, y and setting $z = x + y$). The first interesting case is $n = 2$, where the equation is

$$x^2 + y^2 = z^2.$$

Note that, by the Pythagorean theorem, such a triple is equivalent to a right triangle with side lengths x and y and hypotenuse z, so solutions of this equation in positive integers are the same as right triangles with all sides having integer length.

Some experimentation quickly shows that there do exist at least a few solutions, e.g. $3^2 + 4^2 = 5^2$ or $5^2 + 12^2 = 13^2$. One can generate new solutions from old ones by multiplying by a fixed factor, e.g. for any positive integer x we get a solution

$$(3x)^2 + (4x)^2 = (5x)^2,$$

so there are infinitely many solutions; but this kind of seems like cheating. We could instead ask for *primitive* solutions, i.e. triples (x, y, z) which have no common factor. Euclid showed that in fact there are infinitely many of these as well: for any integers $m > n > 0$, the integers $x = m^2 - n^2$, $y = 2mn$, $z = m^2 + n^2$ give a solution (you can check this by hand), which is primitive if $m + n$ is odd (and in fact every primitive triple arises in this way). Thus there are infinitely many solutions in a strong sense.

So we're left with the case $n \geq 3$. Quick computations don't show any solutions, so we might guess, as Fermat did, that there are none. Proving this however is quite challenging even for small values of n.

This is an example of a Diophantine equation, i.e. a polynomial equation for which we are interested in integer solutions. There is no universal approach to Diophantine equations: this can actually be proven, in that no algorithm can exist which can solve any Diophantine equation (in a similar sense that no algorithm can exist which solves the halting problem, though much harder to prove); this is Hilbert's 10th problem, due to Matiyasevich. However there are many methods available which work on some Diophantine equations, some of which can be made to work for small values of n. No simple methods however seem to suffice to prove Fermat's last theorem in general. Instead of worrying about proving it for the moment, we make some observations about equivalent formulations.

First, the problem is formulated in terms of positive integers. However, for the purposes of applying algebraic methods it's often easier to work with arbitrary integers, positive or negative. Could we extend the statement to arbitrary integers?

Well, not immediately: there certainly are solutions in integers, e.g. $y = 0$ and $z = x$, since $x^n + 0^n = x^n$. We can avoid these "trivial" solutions by requiring that none of x, y, and z are equal to 0. Once this is done, we claim that if $x^n + y^n = z^n$ has no solutions in positive integers, then it has no solutions in integers at all (and obviously vice versa). Why?

If n is even, then changing the signs of x, y, or z doesn't affect the signs of x^n, y^n, or z^n respectively, so if any of them are negative we can just flip the sign(s) and get an equivalent solution in positive integers. Therefore assume that n is odd, and assume that we have a solution $x^n + y^n = z^n$ with x, y, z integers; we'll show that this gives a solution in *positive* integers.

If all of x, y, and z are positive, then we're done. If they're all negative, then $(-x)^n + (-y)^n = -(x^n + y^n) = -z^n = (-z)^n$ is a solution in positive integers.

Next, suppose that x is negative but y and z are positive. Then $y^n = z^n - x^n = z^n + (-x)^n$ is a solution, after shuffling the variables; the case where y is negative but x and z are positive is similar. The case where z is negative but x and y are positive is impossible since $z^n = x^n + y^n > 0$, so this resolves all of the cases where one of x, y, and z is negative and the rest are positive.

Finally, if two of the variables are negative and the remaining one is positive, we can multiply them all by -1 to get a solution of the previous type, which we know gives a solution in positive integers. Therefore the problem of solutions over positive integers and the problem of *nontrivial* solutions over *all* integers are equivalent.

Next, it's often convenient to be able to talk about rational numbers in place of integers. We claim that this framing, too, is equivalent: if $x^n + y^n = z^n$ has no nontrivial solutions in rational numbers, then it has no nontrivial solutions in integers. Indeed, write $x = a/b$, $y = c/d$, $z = e/f$; then this is

$$\frac{a^n}{b^n} + \frac{c^n}{d^n} = \frac{e^n}{f^n}.$$

Multiplying both sides by $(bdf)^n$ gives a solution in integers.

In fact, we can say more precisely that a nontrivial integral solution to $x^n + y^n = z^n$ is equivalent to a rational solution to $X^n + X^n = 1$: since the solution is nontrivial, $z \neq 0$ and so we can divide both sides by z^n, so letting $X = x/z$ and $Y = y/z$ gives a solution as claimed. To go in the reverse direction, clear denominators as above.

For our final reduction, we claim that it suffices to prove Fermat's last theorem for n

given by either an odd prime or $n = 4$. Why is this?

Suppose n is divisible by a prime number $p \geq 3$. A solution to $x^n + y^n = z^n$ is equivalent to $(x^{n/p})^p + (y^{n/p})^p = (z^{n/p})^p$, and so gives a solution to the Fermat equation for p; so if we can prove that Fermat's last theorem holds for every $p \geq 3$, then we have also proven it for every n divisible by an odd prime.

The only n which are not divisible by some odd prime are powers of 2. Indeed, there exists at least one n for which Fermat's last theorem fails, namely $n = 2$! However any higher power of 2 is also divisible by 4, so if we can prove Fermat's last theorem for $n = 4$ then it also holds for higher powers of 2, and therefore $n = 4$ together with odd primes proves the theorem for all n.

Let's now turn to the question of how one might go about proving such a thing. One observation is that $x^n + y^n = z^n$ is an example of a *smooth projective planar curve* defined over the rational numbers. A very difficult theorem of Faltings, the Mordell conjecture, states that any such curve defined by an equation of degree at least 4 has finitely many rational points. (One needs a different method for $n = 3$, but there are a few ways of handling this case separately.)

However, it is not enough to show that the Fermat equation has finitely many solutions: we want to show that it has no nontrivial ones at all. We'll need a different method.

There are a wide variety of approaches that can prove the claim in some cases; next week we'll see some methods for $n = 3$ and $n = 4$ due to Fermat and Euler, and a more powerful algebraic method due to Kummer which will prove the theorem for a large family of primes, called regular primes (it is conjectured that there are infinitely many, and in fact that most primes are regular, but this is not known). For a general approach, however, we'll need more machinery.

The eventual approach used by Wiles to prove Fermat's last theorem for all n goes back to an observation of Frey and Hellegouarch: if we knew a nontrivial solution to Fermat's equation, we could use it to produce certain elliptic curves (objects which we will encounter in the course of our seminar) with very strange properties. In particular, it appeared unlikely that they could be *modular* (a particular property of elliptic curves which we will also encounter in much greater detail). This would contradict the Shimura–Taniyama–Weil conjecture, which predicted that in fact *all* elliptic curves are modular.

Making this observation precise required some significant work, which eventually was accomplished by Ribet, Mazur, and Serre: the "Frey curves" associated to nontrivial solutions of the Fermat equation cannot be modular. This however is not obviously very helpful: not only is the Shimura–Taniyama–Weil conjecture very difficult, when formulated there was not much reason to believe it was even true. However, shortly thereafter Robert Langlands formulated what is now known as the Langlands program, a broad set of conjectures and philosophies which subsumes the Shimura–Taniyama–Weil conjecture.

To say a little more about this, we need to know more about modularity. The modularity of an elliptic curve is not really about the elliptic curve itself, but instead about an object that can be associated to it: its Galois representation (given by the Tate module). Eichler and Shimura showed that it is possible to associate Galois representations (which are algebraic, number-theoretic objects) to (certain) modular forms (which are analytic objects). If the Galois representation of an elliptic curve comes from a modular form, then we say that that elliptic curve is modular.

The more general story of the Langlands program focuses on Galois representations. Not all of these come from modular forms; for one thing, these all have "dimension 2" in a certain sense while Galois representations can have any dimension. The dimension 1 case is understood via class field theory, which is a main subject of a typical graduate-level algebraic number theory course; the higher-dimensional case is much harder. The Langlands conjectures can be thought of as describing how Galois representations (in any dimension) should come from higher-dimensional analogues of modular forms, called automorphic forms. So a piece of the 2-dimensional case of the Langlands program recovers the Shimura–Taniyama–Weil conjecture.

What Wiles proved was that the Shimura–Taniyama–Weil conjecture holds for all sufficiently "nice" elliptic curves, namely semistable ones. (The rest of the conjecture has since been proven, building on Wiles's methods, and is now known as the modularity theorem.) It turns out that the Frey curves are semistable, so this is good enough: they cannot be modular but are semistable, so they contradict Wiles's theorem and therefore cannot exist. The first goal of the class is to understand what this really means. There are a few parts to this:

- Understanding elliptic curves. To get here, we first want to understand something about:
  - elliptic functions, the analytic theory underlying the more algebraic/geometric theory of elliptic curves;
  - algebraic number theory, and in particular something about Galois representations.
- Understanding modular forms.
- The modularity theorem: what exactly does it say, in greater precision, and how does Fermat's last theorem follow?

I am guessing that this will take up a majority but not all of the semester. If we do have some more time, there are some further topics we can discuss depending on class and speaker interests, such as:

- related conjectures and applications of the modularity theorem
- some aspects of the proof of the modularity theorem: what ingredients go into it? what were the new contributions that let it be resolved?
- generalizations: the Langlands program, other questions about elliptic curves
- more recent developments of related methods
- other topics in the theories of elliptic curves, Galois representations, and modular forms

4 Assigning (some) talks

The most important talks to assign are next week's, which are on "classical" (i.e. pre-20th century) approaches to Fermat's last theorem:

- "Elementary" approaches: the cases $n = 3$ and $n = 4$, infinite descent, sums of two squares;
- Kummer's approach: the proof for "regular" primes.

This is a challenging pair of talks in that it requires going first and gives you the least time to prepare. On the other hand, the methods involved are generally a little less machinery-heavy than many of the others we'll have throughout the semester (especially the first; the second uses some ring theory).

Next week, we'll start down the road towards elliptic curves, taking a more historical perspective and starting with the theory of elliptic functions. (This is not strictly necessary to get a working theory of elliptic curves, but is interesting and will be helpful later.) Both talks on September 30th will be on elliptic functions.

After that, we'll take a detour to briefly study some algebraic number theory (number fields and their completions, Galois representations) before coming back to study elliptic curves; I think it's best to postpone assigning talks too far out in case our pace is different from expected in either direction, but if there's interest we can assign up through the expected two talks on number theory:

- Global fields, places, and local fields
- Galois theory and representations

Beyond that, you can start to think about which talks you might be interested in, and we'll assign further talks in a week or two. I am loosely expecting that we'll have around 13-14 talks on the textbook material and the remaining talks on further topics of interest, perhaps with some shorter talks at the end of the class on final project topics, so you should expect to give at least one talk on the material from the book, a second talk on either the book's material or further material, and possibly a mini-talk on your project.