

**PERTURBATIONS, DEFORMATIONS, AND VARIATIONS
(AND “NEAR-MISSES”) IN GEOMETRY, PHYSICS, AND
NUMBER THEORY**

B. MAZUR

In memory of René Thom

CONTENTS

1. Introduction	307
2. Universal unfoldings	309
3. Special, general, very general, generic...	313
4. Moduli: Parameters describing the deformations of a mathematical structure	314
(1). Classifying ellipses in the plane up to Euclidean congruence.	314
(2). Deforming Riemann surfaces.	315
5. Moduli and lattices	317
6. Lattices in real vector spaces	317
(1). Lattices in the plane.	318
(2). From lattices in the plane to elliptic curves over \mathbf{C} .	319
7. The modular curves $X_0(N)$ play a central role in three different classification problems	320
8. Deformations of group representations	321
9. Mirror symmetry – and back to deforming lattices in the plane	325
10. Quantization in physics, and noncommutative geometry	327
11. Near-misses in algebra and number theory (approximation of irrationalities)	329
References	334

1. INTRODUCTION

One can learn a lot about a mathematical object by studying how it behaves under small perturbations. The study of perturbations often leads to linear questions, which accounts for the success of much analysis, most notably the Calculus of Newton and Leibniz, the Calculus of Variations in differential geometry, and the perturbative methods in partial differential equations.

Received by the editors February 19, 2004.

2000 *Mathematics Subject Classification*. Primary 11Jxx, 14Bxx, 32Gxx.

I wish to thank Curt McMullen and Noam Elkies for helpful comments on an early draft, and Stephanie Yang and Ivan Petrakiev for providing the illustrations.

And the study of perturbations often does *not* lead to linear questions, which accounts for some fascinating aspects of the theory of dynamical systems.

Parts of quantum mechanics may be formulated as a noncommutative deformation of classical mechanics, and this has inspired the growth of analogous “quantum” deformations of some basic mathematical structures (*quantum groups, quantum cohomology, etc.*). One may hope that this interaction between mathematics and physics will lead to a thorough and clear understanding of the profound physical intuition coming from the study of Feynman integrals and quantum field theories that has recently been so effective in the treatment of a number of traditional mathematical problems and of some novel ones.

Whatever it leads to, *the perturbative strategy* is everywhere in mathematics and takes many forms.

Some questions become meaningful only when they are treated as specific instances within a field of closely related questions. Often the landscape of this larger field, its peculiar features, its ravines and gullies, holds the key to an appropriate understanding of any of the individual questions. Often that landscape becomes the focus of new questions.

This article is the amplified text of a one-hour talk I gave at the winter joint meetings of the MAA-AMS in New Orleans in 2000. It was given in the format, and spirit, of the weekly seminar held at Harvard University called the *Basic Notions Seminar*. The aim of that seminar, initiated by David Kazhdan, is to survey each week some central theme of mathematics – some idea that might have different manifestations as it crops up in different fields of mathematics – an idea, in short, that deserves to be contemplated by students not only in the context of its usefulness for this or that particular result but also because of its service as a unifying thread.

I would like to dedicate this article to René Thom, whose enormous influence on our subject has helped establish the bold unification that it presently enjoys and who taught us to be reflective about the central role played by the “basic notions” listed in the title.

Perturbation and deformation are terms that we will take as almost synonymous, with the slight difference that in the case of the latter term, we will be more acutely interested in the parameter space that tabulates the (allowed) deformations. One often considers *deformations* of a given structure with the ultimate goal of understanding the full repertoire of possible *variations* of the structure – of giving a complete classification of kindred structures: one starts modestly, considering only structures infinitesimally near the given one, as a prelude to the general problem of giving a total *classification*. This has historically been true of some of the examples to be discussed. When we manage to give a full classification – that is, when we have exhibited a family of such mathematical structure parametrized by a space such that every instance of that mathematical structure is accounted for by a unique point in the parameter space, and where the geometry of the parameter space faithfully reflects the possibilities of variation of this structure – only then can we truly say that we have provided a *universal* accounting for the possibilities of variation of the structure being studied. Such a parameter space is sometimes called the *moduli space* associated to the mathematical structure in question. For a very elementary example see 4.1 below, where we discuss the classification of *ellipses in the plane* up to congruence. For a twenty-four page annotated bibliography of literature about deformations, see [Do].

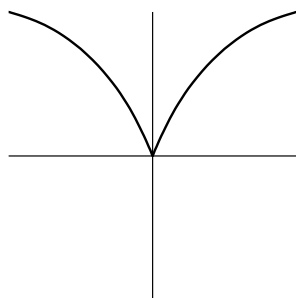
There are cases where the structure under study admits no variation, that is, where the structure is *stable* in the sense that any nearby structure is isomorphic to it. At times, though, it may be subtle to formulate the intended notion of “nearby”. Differentiable manifolds, for example, are stable: two compact differentiable manifolds that are “close” (with any reasonable *smooth* meaning of the word “close”) are isomorphic. The term *rigid* is sometimes reserved to refer to mathematical structures that are *stable* in some appropriately defined sense, such as the above, but also admit no nontrivial automorphisms. Compact (positive-dimensional) differentiable manifolds are stable but decidedly not *rigid*: they admit a rich group of self-diffeomorphisms.

The notion of “*near-miss*”, which we shall treat towards the end of this article, is meant to connote a modification of a problem which can be a good deal looser than a deformation.

2. UNIVERSAL UNFOLDINGS

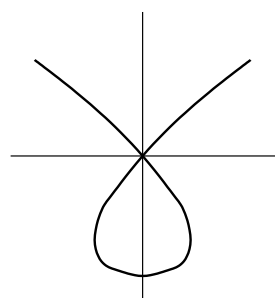
The theoretical biologist C.H. Waddington emphasized that the full study of any developing biological system requires an understanding of its *epigenetic landscape*, which is his term for the multi-dimensional phase space that contains both the one “standard” developmental pathway for the system together with a picture of the other nearby variant pathways. A thorough understanding of *development* will include knowledge of how deep the grooves are which form each of the possible pathways, i.e., how easy or difficult it would be for a disturbance to force a jumping from one course of development to another.

Waddington’s approach to morphogenesis is close, in spirit, to René Thom’s view that a topological singularity is best understood in the context of what he termed its *universal unfolding*. How, for example, does your specific geometrical object react to perturbation? Is it stable? Is it the generic member of its species? If not, can its particular specialties be viewed as coming about from an explicit process of “degeneration” from a more generic geometrical object? To illustrate what is meant here, consider a “starter example” from the theory of singularities of curves in the plane: the stability of nodal singularities (figure 1b) in contrast to the instability of a cuspidal singularity (figure 1a).



$x^2 = y^3$
(cusp)

Figure 1a



$x^2 = y^2(y + 1)$
(node)

Figure 1b

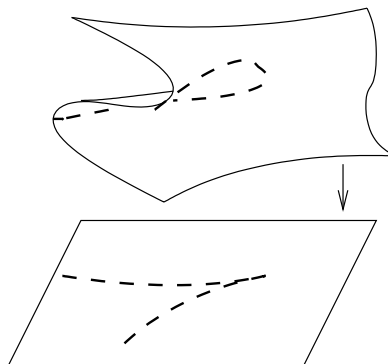


Figure 2

To put it in (outrageously) nontechnical language, suppose that you are trying to trace out the locus of zeroes about the origin of the equation $x^2 = y^3$ in the Cartesian plane, but you are doing this with pencil on a pad in a jiggly subway train. The equation itself has a *cuspidal* singularity at the origin, as in figure 1a, but very likely you will have drawn something like the *nodal* singularity as in figure 1b. In a word, the nodal singularity enjoys a type of stability that the cuspidal singularity does not. Of course, once we try to pin down what we mean by the question “is the nodal singularity stable?” we are faced with an intriguing array of candidates for our definition of stability, and the answer may depend upon the context we work in:

- Working in the space of parametrized smooth plane curves (e.g., the space of C^∞ immersions of the closed unit interval in \mathbf{R}^2 , two such immersions being deemed close if they are “close” in the C^1 -topology), a nodal singularity of such a curve is **stable** in the sense that if one such parametrized curve $C : t \mapsto (a(t), b(t))$ (for $t \in \mathbf{R}$) possesses a nodal singularity $s = (a(t_0), b(t_0)) \in \mathbf{R}^2$, any sufficiently close parametrized curve C' also must have a nodal singularity $s' = (a(t_1), b(t_1))$ at some time t_1 close to t_0 .

Moreover, a parametrized curve with a cuspidal singularity can be perturbed to one with a nodal singularity.

- Working in the space of C^∞ -functions $f(x, y)$ on \mathbf{R}^2 , then nodal singularities are *not* stable. For a smooth function $f(x, y)$ that cuts out a cuspidal singularity (say, a polynomial $y^2 = x^3$) or a nodal singularity (say, $y^2 = x^3 + x^2$) can be perturbed ever so slightly by adding terms to our equation consisting of the missing small powers of x with very small coefficients to make the singularity disappear.

For either of the above notions of stability, the cuspidal singularity is unstable. But there are arenas in which cuspidal singularities themselves achieve a stability, as anyone can attest who observes the caustic curve of light in their coffee cup. (If the projection of a smooth surface X in three-space to the horizontal plane has discriminant locus possessing a cuspidal singularity as in figure 2, then any nearby surface X' will still have a cuspidal singularity; cf. [G], [A2].)

René Thom’s emphasis on this type of question-asking is motivated by two further concerns. The first has to do with the experimental sciences and the fact that they are most naturally set up to study phenomena that are *stable*, or at least relatively stable. A “repeatable experiment” must, after all, be repeatable; its outcome

would be difficult to interpret if ever-so-minute changes in the laboratory could provoke qualitative fluctuations in its outcome. This attitude towards the notion of stability, closely related to questions regarding the sensitivity of the evolution of a trajectory in a dynamical system to initial data, is the subject of a vast and rich literature (including the work of Laplace, Poincaré, Kolmogoroff, Arnold, Moser, Sinai, Smale and many others).¹

The second is the intuition that unless you ask and answer deformational questions and determine what structures are close to the one you happen to be studying (with the word “close” taken in as many senses as is pertinent to the structure under scrutiny), you will miss the full story, the most powerful theory.

And even in the simplest geometric instances the “full story” has its surprises. Consider the stable surface in three-space dubbed by Thom the “swallowtail” (*la queue d’aronde*); see [T]. This is the discriminant locus in real (u, v, w) -space of the “general” quartic polynomial $x^4 + ux^2 + vx + w$ (we clear out the cubic term by appropriate translation of x). Think of this surface as being swept out by a family of plane curves $u = \text{constant}$. For negative u one has a curve with two cuspidal singularities and one ordinary singularity, all of these singularities annihilating themselves at $u = 0$ and emerging as a family of perfectly benign curves for positive u . Figure 3 shows the progression of plane curves:

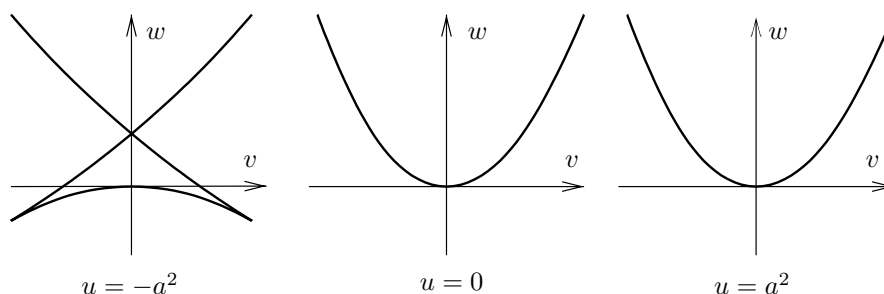


Figure 3

and figure 4 a picture of the surface these curves trace out in three-space. The singularity of the surface at the origin is *stable* (in Thom’s sense):

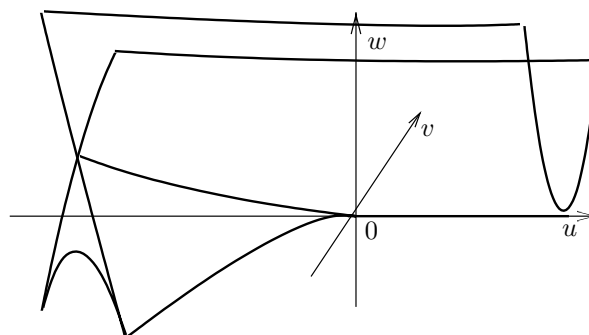


Figure 4

¹A short, conversational book, rich in polemic, that deals masterfully with some of the material we have just alluded to, and that covers much more besides, is V.I. Arnol’d’s *Catastrophe Theory*, in its 3rd, expanded and revised, edition [A1].

Or consider Oscar Zariski's discovery that there are two "types" of plane sextic curves with six cuspidal singularities. There are the ones where the six cuspidal singularities lie on a (plane) conic, as is the case for those sextic curves that are discriminant loci of general planar projections of generic cubic surfaces (figure 5):

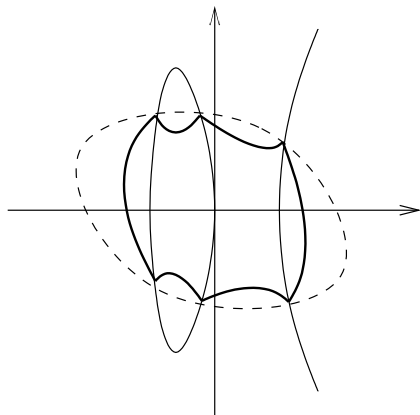


Figure 5

and there are the ones where the six cuspidal singularities *don't* lie on a plane conic (figure 6):

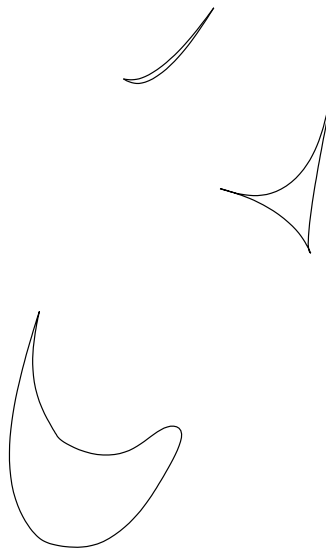


Figure 6

I am thankful to Ivan Petrakiev for providing these two figures, the second of which he calls the "laughing curve". You might guess that if you take any instance of the first type, a sextic whose six cusps lie on a conic, you could effect a minor perturbation of its defining equation keeping the cusps as cusps but forcing those six cusps to avoid a conic. Zariski's discovery is that this is not the case: there are two distinct components of the family of all plane sextic curves with six cusps. If you start with a (general) plane sextic curve with six cusps lying on a conic and *perturb* its defining equation to get a new plane sextic (retaining the fact that its

singularities are cusps), then that new plane curve will again have its six cusps lying on a conic: the general member of the first type of plane sextic is not a degeneration of members of the second type. One clean topological “distinguisher” here is the fact that if the six cusps of an irreducible sextic plane curve lie on a conic, the first betti number of the d -fold cyclic cover of the plane branched at the curve is nonzero when $d > 0$ is divisible by 6, whereas if the cusps do not lie on a conic, that first betti number is zero; see Zariski’s article “On the irregularity of cyclic multiple planes”, pp. 186-212 in [Z]. (For recent work on plane curves of minimal degree with prescribed singularities, see [G-L-S].)

3. SPECIAL, GENERAL, VERY GENERAL, GENERIC...

The notion of *general-ness* crops up everywhere in mathematical writing.

- For example, some questions are easily answered only when the objects under study are (or are perturbed so as to be) in “general position”. For this, we need only think of the various kinds of “intersection theories” in algebraic topology and algebraic geometry, where one strategy of associating a *number* that accounts for the (algebraic) number of intersections between two cycles is to “move” one of the cycles so that it is in general position with respect to the other, and then count.

- There are questions that are most easily answered when the objects under study are (or are perturbed so as to be) in “special position”. Some celebrated algebraic geometry depends upon a close examination of *the most degenerate cases* where the answers obtained then extend, by deformation, to all cases. For an instance of this, one might cite the famous assertion of Severi (in an appendix, “Anhang F”, of his monograph *Vorlesung über Algebraische Geometrie* [S]) claiming the irreducibility of the family of plane curves of a given degree n with at worst nodal singularities, where the strategy of the argument is to specialize to a union of n lines with no triple intersections. The claim is true, and the strategy (specialization to a “most degenerate case”) is a good one, but the first correct proof of the claim is due to J. Harris in 1985 (see [H]). Severi’s argument of it had an error;² for a fine exposition and discussion of the various solutions to this problem, see [Lo]; see this article as well for a fairly complete bibliography regarding this. One thing hanging open in the six decades between the publication of Severi’s “Anhang F” and the resolution of the problem by Harris is a famous application, due to Zariski, of Severi’s claim; namely: *the fundamental group of the complement of an irreducible plane curve possessing only nodes as singularities is cyclic*. (In contrast, this fundamental group, for irreducible curves with nonnodal singularities, can be quite complicated, or one should rather say it can be quite interesting, as already happens in the case of irreducible sextics with six cusps, the example described above.)

- Some questions themselves insist upon being perturbed, jiggled, before they have a simple answer. The great number of results about that elusive mathematical object *the general hypersurface of degree n* already attests to that. But there is also a rich store of simple, although fundamental, geometric questions that require *perturbation* of some of their elements before they can even be stated. As a random

²Severi argued, correctly, that every boundary component of the moduli space of plane curves of degree n with nodal singularities has a member consisting in n concurrent lines, and then (without sufficient argument) concluded that one could “generalize” this member to deduce that each component contained members consisting of configurations of lines with no triple intersection. A cautionary tale.

example, consider this problem in the geometry of surfaces in three-dimensional Euclidean space (discussed elaborately in [A1]): *a generic surface in three-space is tangent to no straight line in more than 4 points*. My brother and I have our favorite example where this maximum number of tangencies is achieved: think of a flowerpot packed with dirt, but not up to the rim, in which at least two bamboo stalks are growing. The surface in question is the idealized outer skin of this configuration or any mild deformation of it. Now take a thin metal rod (our straight line) and rest it on the rim of the flowerpot. This already gives two points of tangency between the surface and the line. Move the rod until it touches one of the bamboo stalks, (giving three points of tangency); swivel the rod so that it touches another as well (giving four).

Although there doesn't seem to be universal consensus, in algebraic geometry,

- the adjective *general* tends to mean that one has a parametrized family of objects in mind, the parameter space being irreducible, and the *general member* of the family is simply any member that corresponds to a parameter that avoids a specific proper closed subvariety in the parameter space (this “specific” *avoidance subvariety* is often not made explicit);

- the adjectival phrase *very general* tends to mean that one must avoid a specific countable union of proper closed subvarieties in the parameter space (also usually not made explicit);

- the adjective *generic* in Grothendieck's formulation is defined, in effect, to allow us to avoid *all* proper closed subvarieties.

4. MODULI: PARAMETERS DESCRIBING THE DEFORMATIONS OF A MATHEMATICAL STRUCTURE

Let us begin with a mock deformation problem that we are all (perhaps too) familiar with, and then pass, in earnest, to one of the most classical of deformation problems.

(1). Classifying ellipses in the plane up to Euclidean congruence. Any two ellipses in the plane are congruent if their minor axes have the same length and their major axes have the same length as well. An ellipse E is classified up to congruence, then, by two parameters: the two positive numbers (a, b) where a is the length of its major axis, and b the length of its minor axis. We might refer to (a, b) as the *moduli* of the ellipse E and form the manifold that collects all these pairs of parameters, or moduli, classifying our ellipses by considering the shaded octant in the (x, y) -plane in figure 7

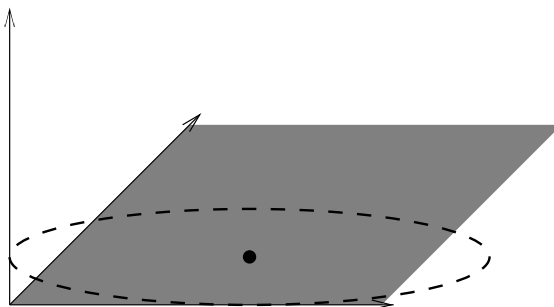


Figure 7

where each point $(x, y) = (a, b)$ in the interior of this octant corresponds to a unique congruence class of ellipses (with major and minor axes equal to a, b , respectively), where every congruence class of ellipses is indeed represented, but where something strange happens at the two boundary lines: the 45° -line corresponds to *circles*, which have significantly more symmetries than the generic ellipse, and the positive x -axis corresponds to degenerate ellipses, which have fewer. More germanely for what will follow, we might notice that the geometry of the open octant (its topology, its differential structure) reflects the various manners in which we can produce continuous or differentiable families of ellipses, that is, the ways in which we can deform ellipses (keeping, to be sure, the deformed structure *ellipses*). We might, then, dub this octant the *universal moduli space of true ellipses*, viewing its two boundary rays as the ways in which our ellipses may specialize, or degenerate, and recognizing as well that the topological or differential features of this moduli space are relevant to the manner in which ellipses may be deformed.

(2). Deforming Riemann surfaces. Take a closed Riemann surface S and consider the problem of “deforming it”. The underlying topology (and differentiable structure) of S is “stable”; i.e., as mentioned before, if you perturb S in any reasonably smooth way, you get the “same” differentiable manifold. The famous theorem of Riemann that the “Riemann sphere” is *stable, as a complex manifold* – i.e., that there is only one complex structure on the two-sphere – is the first fundamental result about classification of compact Riemann surfaces. And it was already known to Riemann (as exposed in his 1857 treatise *Theorie der Abel’schen Functionen* [R]) that there are $6g - 6$ real parameters of possible deformations of the complex analytic structure on a Riemann surface S of genus $g > 1$.

Riemann surfaces are complex manifolds, and one can try to deform them by “fixing” their underlying structure as differentiable manifold, and then perturbing their complex structure. But there are as many ways of effecting these perturbations as there are ways of presenting the Riemann surfaces. We shall present four different ways below. And each of these ways gives us another mode of “comparing different Riemann surfaces of the same genus,” e.g., judging how close one such surface is to another. Let us keep, for the moment, to the case of $g > 1$.

- If you think of S as a quotient of the hyperbolic plane by the action of a discrete group Γ , i.e., given by a homomorphism $\phi : \Gamma \hookrightarrow PGL_2(\mathbf{R})$, you would naturally try to deform S by perturbing this homomorphism ϕ in such a way as to, on the one hand, keep the action discrete and, on the other, “truly vary” the quotient Riemann surface.

- If you think of S as given by first stipulating a Riemannian metric on a (smooth) surface of genus g and then passing to its underlying conformal structure, you might think of deforming the Riemann surface by deforming the Riemannian metric. For example, take $g = 2$. Build a Riemann surface of genus two as follows. First form $Y =$ the Riemann surface (bounded by three circles) built by removing three small open discs with centers $0, 1,$ and ∞ from the Riemann sphere. Following Thurston

we can call this a *pair of pants* (see figure 8).

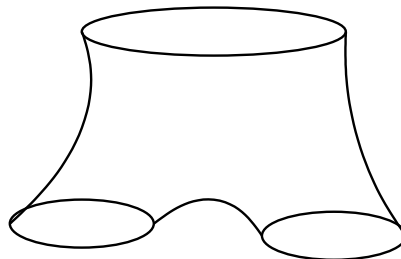


Figure 8

Then take two copies of Y and connect the boundary components of one of these copies with the corresponding boundary components of the other by gluing three complex annuli (as drawn in figure 9, they are cylinders). Note that there are two real parameters to describe each cylinder and its manner of attachment (i.e., the parameters are given by the *length* of the cylinder and the relative angle of twist of the two attaching maps to the bounding circles). All in all, we have six ($= 6 \cdot 2 - 6$) real parameters.³

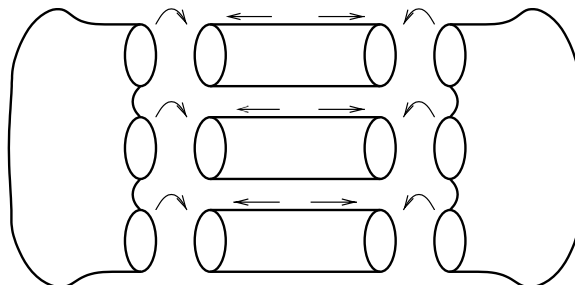


Figure 9

- Or you might perturb the conformal structure of the Riemann surface “directly” and compare two Riemann surfaces by measuring the distortion effected by *quasi-conformal mappings* between them.

- If you think of S as an algebraic curve defined over \mathbf{C} in projective space, you might try to deform S by perturbing the coefficients in its defining equations.

Each of these approaches yields different (yet intertwining) structures on the *moduli space* \mathcal{M}_g of “all” Riemann surfaces of genus g . Each manner of thinking of Riemann surfaces leaves its “imprint” on \mathcal{M}_g . For example, using the structure that \mathcal{M}_g inherits from the last of the bullets above, we may view \mathcal{M}_g as a complex algebraic variety of (complex) dimension $3g - 3$. Or, using the structure it inherits from the first, \mathcal{M}_g can be thought of as the quotient of the complex analytic Teichmüller space \mathcal{T}_g by the properly discontinuous action of the (discrete) mapping class group.

³More generally, for a surface S of genus $g > 1$ we can find a system of $3g - 3$ closed geodesics on S partitioning it into $2g - 2$ “pairs of pants”; given such a “pants-decomposition” of S , the *lengths* of the $3g - 3$ geodesics give half the parameters of deformation, while the other $3g - 3$ parameters come from considering how the $3g - 3$ hems and waists are sewn together. Appropriately normalized, these $6g - 6$ real numbers are called the *Fenchel-Nielsen parameters of the (pants-decomposed) surface S* .

5. MODULI AND LATTICES

It is usually a good thing when we manage to reduce a problem to a “linear” question: we stand a chance of dealing successfully with issues in linear algebra. First-order deformational problems are, in effect, linear to begin with, but we are often pleasantly surprised to find that certain more general variational questions relate (even if they don’t entirely reduce) to questions about linear groups and variations of lattices in these linear groups. A general format in which this happens is given by the theory of Griffiths spaces, which are particular quotient spaces of homogeneous spaces attached to Lie groups, the quotients being made via the discrete lattices acting on these homogeneous spaces. This rich structure comes about from the Hodge theory of algebraic varieties.

A classical example of this is given by the Hodge theory of Riemann surfaces. Attached to any Riemann surface S of genus g the complex vector space (of dimension g) of its holomorphic differential 1-forms is, as it turns out, a nondegenerate Hermitian space when endowed with the Hermitian form

$$(\omega, \eta) \longmapsto i \int_S \omega \wedge \bar{\eta},$$

where ω, η are holomorphic differential 1-forms. Its dual complex vector space is endowed with a rank $2g$ lattice formed by the linear functionals

$$\omega \longmapsto \int_{\gamma} \omega,$$

where γ runs through the integral 1-dimensional homology in the group $H^1(S, \mathbf{Z})$ (which is free of rank $2g$). This structure – the Hermitian structure on the complex g -dimensional vector space and the rank $2g$ lattice in its dual – is enough to determine the Riemann surface S up to isomorphism. That’s the good news. The (let us say) more interesting news is that it is quite subtle to determine which such structures actually represent Riemann surfaces. This latter question can be said to be still outstanding, even though there is much beautiful work, answering the question (how to pick out from among these structures those that come from Riemann surfaces) in different ways.

6. LATTICES IN REAL VECTOR SPACES

The previous discussion puts the spotlight on the question of classifying lattices in complex vector spaces. Here we will change our viewpoint a bit and consider classification of lattices in real vector spaces. Suppose that we are interested in lattices $L \subset \mathbf{R}^N$ in N -space such that the volume of their fundamental domain is 1, and we wish to consider them up to “rotational-equivalence”. That is, two such L ’s are viewed as *the same* if you can pass from one to another by an element in the group of rotations of N -space, i.e., $SO_N(\mathbf{R})$.

One way of organizing them is to start with the standard lattice $\mathbf{Z}^N \subset \mathbf{R}^N$ and consider our lattice to be the image of the standard lattice under a special linear transformation, i.e., an element of $SL_N(\mathbf{R})$. Of course two are considered equivalent if they are rotationally equivalent; i.e., we are really only dealing with the homogeneous space $X_N := SL_N(\mathbf{R})/SO_N(\mathbf{R})$, a point of which corresponds to a lattice of volume equal to one, with chosen basis, all taken up to rotational-equivalence. The action of $\Gamma_N := SL_N(\mathbf{Z})$ on X_N by left-multiplication removes the

choice of basis, so the quotient space of X_N under the action of Γ_N has the property that its points “classify” our equivalence classes of lattices. This quotient space has dimension $\frac{N^2+N-2}{2}$, and therefore there are generally these many “directions” to deform our lattices. For example, if $N = 3$ we have five possible parameters of deformation of a lattice (see figure 10).

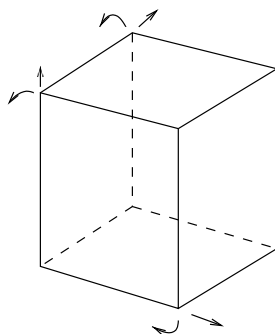


Figure 10

I once wanted to frame a series of (simple) mental exercises in geometric imagination as training for students of geometry. Prominent in this list of exercises was the one of making vivid (i.e., thinking through a “video”, to be played in one’s mind’s eye, which exhibited) the five parameters of variation of a parallelepiped of volume one. But visualizing the quotient space of X_N under the action of Γ_N is quite a bit harder (it would *not* have been on my list of simple exercises) and brings to the picture some extra surprises.

(1). Lattices in the plane. For example, if $N = 2$ such an L can be thought of as a tiling of the plane with parallelograms of area 1 (see figure 11).

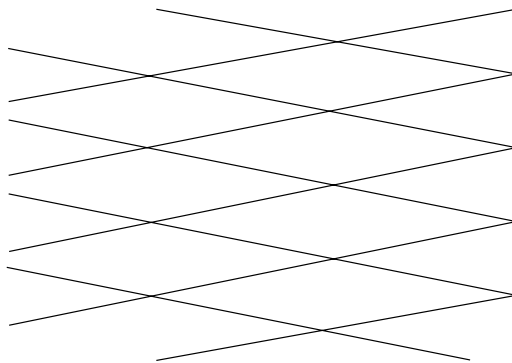


Figure 11

The famous answer to the question of classifying such lattices (up to rotation) is given by the *elliptic modular function* j , central to such diverse issues of mathematics ranging from the theory of elliptic functions and elliptic curves to the theory of the Monster group. To each lattice $L \subset \mathbf{R}^2$ one associates a complex number $j(L)$, and one has a one:one correspondence.

Lattices up to rotation and rescaling \longleftrightarrow complex numbers.

Here is the definition of the correspondence $L \longleftrightarrow j(L)$. View the Euclidean plane as the complex plane \mathbf{C} , and let $L \subset \mathbf{C}$ be a lattice. For k an even integer ≥ 4 denote by $G_k(L) \in \mathbf{C}$ the sum of the reciprocals of the k -th powers of nonzero elements in $L \subset \mathbf{C}$ (this infinite sum does converge when $k \geq 4$). The function $\{L \subset \mathbf{C}\} \mapsto G_k(L)$ is the classical *Eisenstein series of weight k* . The phrase “weight k ” refers to how this function behaves under rescaling the lattice: for z any nonzero complex number, $G_k(z \cdot L) = z^{-k}G_k(L)$. A standard way of renormalizing these Eisenstein series is to divide by the factor $2\zeta(k)$, where $\zeta(s)$ is the Riemann zeta-function, $E_k(L) := G_k(L)/2\zeta(k)$ (the Fourier series in the variable z of $E_k(L)$ for the lattice L generated by 1 and z in the upper half plane then has constant term equal to 1 and all other of its coefficients are rational numbers). Moreover, with this normalization the standard equations have fairly manageable coefficients. For example, the (normalized) *discriminant* of the lattice L is given by:

$$1728\Delta(L) = E_4(L)^3 - E_6(L)^2,$$

and the classical *j -invariant* of L is defined to be:

$$j(L) := E_4(L)^3/\Delta(L);$$

the j -invariant is indeed an “invariant” in that it is insensitive to change of scale of the lattice L .

The classical theory gives us that any complex number is the j -invariant of some lattice, and two lattices L (with fundamental parallelogram of area equal to one) have the same j -invariant if and only if they are rotations of one another. Weierstrass’s theory produces for us the famous doubly periodic meromorphic functions $x = x_L$ and $y = y_L$ with poles of order two and three, respectively, at lattice points in L which satisfy the formula

$$(*) \quad y^2 = 4x^3 - \frac{E_4(L)}{12}x - \frac{E_6(L)}{24}.$$

(2). From lattices in the plane to elliptic curves over \mathbf{C} . Every Riemann surface of genus 1 is isomorphic to the quotient $E = E_L := \mathbf{C}/L$ by a suitable lattice L . The doubly periodic meromorphic functions $x = x_L$ and $y = y_L$ become meromorphic functions on E_L and the equation $(*)$ serves to exhibit the Riemann surface E_L as a plane cubic curve $(*)$ with a flex point at infinity, i.e., as an *elliptic curve* over \mathbf{C} .

We have one:one correspondences between:

- Isomorphism classes of elliptic curves over \mathbf{C}
- Isomorphism classes of Riemann surfaces of genus 1
- lattices $L \subset \mathbf{C}$ up to (complex, nonzero) scalar multiplication
- complex numbers $j(L)$

In one sense we might say that there is a *single* parameter $j \in \mathbf{C}$ which answers the question of what isomorphism classes of elliptic curves there are and how elliptic curves deform. This is indeed roughly true, but with a proviso! The complex plane parametrized by j represents what people call the **coarse moduli space** for elliptic curves; the adjective *coarse* is put there to remind us that this moduli space doesn’t quite have all the features required of a bona fide moduli space, as described in the introduction. The problem here is that elliptic curves are *not* rigid: they have nontrivial automorphisms, and more importantly, some elliptic curves have more automorphisms than others: the square lattice $L \subset \mathbf{C}$ (j -invariant = 1728)

has an automorphism of order 4 given by multiplication by i , the hexagonal lattice (j -invariant = 0) has an automorphism of order 6 given by multiplication by a sixth root of unity, while all other lattices only have automorphisms given by multiplication by ± 1 . A consequence of this is that we cannot exactly provide a *universal family* of elliptic curves holomorphically parametrized by their j -invariant in \mathbf{C} . Here is a sketch of why: by “universality” there would have to be an automorphism η , say, of order 6 of the universal family which restricts to the automorphism of order six of the hexagonal lattice, and since none of the nearby members has such an automorphism, η must act nontrivially on the parameter space (of values of j), which contradicts the one:one correspondence between isomorphism classes of elliptic curves and their j -values. To take care of such limitations, the theory of *stacks* was invented, a theory that goes beyond the scope of our article.⁴

Just as we can parametrize elliptic curves by a single parameter (with the proviso of the previous paragraph!) forming a *coarse moduli space*, so too can we parametrize elliptic curves with various types of (discrete!) auxiliary structures by a single complex parameter (a single “*modulus*”). Here is an easy example of such an “auxiliary structure” (studied in depth, first by Legendre): consider elliptic curves E together with a specification of an ordering $\{e_1, e_2, e_3\}$ of the three points of order two in E . The parameter space classifying this problem may be identified with the complement of 0 and 1 in the complex plane. If we denote its parameter by $\lambda \in \mathbf{C} - \{0, 1\}$ (the symbol λ standing for *Legendre*), the parametrization is given by

$$E_\lambda : y^2 = x(x-1)(x-\lambda),$$

and the points of order two are given by $e_1 = 0$, $e_2 = 1$, and $e_3 = \lambda$.

In the same spirit, given any positive integer N , isomorphism classes of elliptic curves E given along with a chosen cyclic subgroup $C \subset E$ of order N are parametrized by the points on a complex curve; this curve is denoted $Y_0(N)$, and its completion (obtained by adding a finite set of *cusps*) is denoted $X_0(N)$, which has a nice “canonical” model as a smooth algebraic curve over \mathbf{Q} .⁵

7. THE MODULAR CURVES $X_0(N)$ PLAY A CENTRAL ROLE IN THREE DIFFERENT CLASSIFICATION PROBLEMS

As we have just mentioned, the (noncuspidal) points of $X_0(N)$ *classify* isomorphism classes of elliptic curves with chosen cyclic subgroup of order N . But the results of Wiles, Taylor, Diamond, Conrad, Breuil [B-C-D-T] tell us that the curves $X_0(N)$ play yet another, quite different, role in the classification of *elliptic curves*, or at least in the classification of elliptic curves defined by equations with coefficients in \mathbf{Q} , the field of rational numbers. (For a series of expository articles on this, see [M-F].) Namely, for every elliptic curve E *definable over* \mathbf{Q} there is a positive

⁴The concept of *moduli space* is adequate for the classification of mathematical objects that are rigid, i.e., have no nontrivial symmetries, or else all have the “same” symmetries. When there are abrupt changes in the structure of the group of symmetries as one passes continuously from object to object – when there is breaking, or generating, of symmetries – one needs a finer language. The mission of *the theory of stacks* is to provide such a language. These issues present themselves even in the most elementary circumstances: think of the boundary components of the octant classifying ellipses in the plane up to congruence, discussed in the beginning of section 4.

⁵The canonical model has the property that any noncuspidal rational point over a field K containing \mathbf{Q} can be represented by an elliptic curve E and cyclic subgroup $C_N \subset E$ defined over K .

integer N and a nonconstant mapping $X_0(N) \rightarrow E$. Moreover the minimal positive integer N for which this is true (the *conductor* of E) can be given by a (relatively) simple recipe if you are given the equations of E .

Let us refer to this fundamental result as the **modularity theorem for elliptic curves over \mathbf{Q}** , and take a moment to absorb the fact that this modularity theorem implies that these basic algebraic curves, $X_0(N)$, are crucially involved in the classification of elliptic curves in *two* rather incommensurate ways. To repeat:

- the points of $Y_0(N) \subset X_0(N)$ classify elliptic curves over \mathbf{C} with chosen cyclic subgroup of order N , while the complex analytic structure of the modular curves $X_0(N)$ give parameters for “all” deformations of elliptic curves over \mathbf{C} with this auxiliary data, and also

- all elliptic curves over \mathbf{Q} appear as quotient curves of the modular curves $X_0(N)$.

But these very modular curves seem to be the key to a third type of *universal* problem, namely universal deformations of Galois representations.

8. DEFORMATIONS OF GROUP REPRESENTATIONS

Most mathematicians have encountered variations of group representations in some form or other. The classical subject of Fourier transforms is already implicitly such a theory: the exponential function is all the equipment one needs to produce an elegant, canonical, parametrization of the “universal family” of one-dimensional continuous complex unitary representations of the real line, viewed as a Lie group. For each real number a , putting $\chi_a(x) := \exp(2\pi i ax)$ we have that the universal family of representations of the above type is given parametrically by

$$a \longmapsto \chi_a : \mathbf{R} \rightarrow \mathbf{C}^* = \mathrm{GL}_1(\mathbf{C})$$

for $a \in \mathbf{R}$. The miracle that this parameter space itself is again, canonically, the Lie group \mathbf{R} has repercussions throughout mathematics.

The “universal parametrization” of one-dimensional continuous complex unitary representations of any locally compact (commutative) topological group is treated by the theory of Pontrjagin. If G is a locally compact topological group, let

$$\hat{G} := \mathrm{Hom}_{\mathrm{cont}}(G, \mathbf{C}^*)$$

be the “Pontrjagin dual” of G ; that is, \hat{G} is the group parametrizing all degree 1 continuous unitary \mathbf{C} -valued representations of G . The fact that this parameter space of representations, \hat{G} , is again a (commutative) locally compact topological group is key to the further elaboration of Pontrjagin’s theory.

The more general question of appropriate parametrizations of finite, or infinite, dimensional linear representations of a given type, for a given group, is, of course, one of the great ongoing chapters of our subject. And the natural structure(s) that these parameter spaces come equipped with is, again, key to any further detailed study.

The moduli spaces of Riemann surfaces, which were the subject of section 3, admitted an interpretation as parameter spaces of actions of fundamental groups on the upper half plane; e.g., we might call these parameter spaces of *nonlinear* representations. In the remainder of this section, however, we will concentrate on deformations, and universal parametrizations, of certain types of *continuous linear representations of Galois groups of number fields*. (A reference for this material is my article “An introduction to the deformation theory of Galois representations”,

pp. 243-312 in [M-F]; for a study of moduli spaces of linear representations of fundamental groups of manifolds, see, for example, [K-M].)

Now, to talk of “the” Galois group G_K of a number field K is probably already slightly misleading. For to construct such a thing, we must first choose, somehow, an algebraic closure \bar{K} of K (one sane manner of doing this is to embed K in the field of complex numbers and then take \bar{K} to be the union of all algebraic numbers in \mathbf{C}). Then, letting L/K run through finite Galois extensions contained in \bar{K} , we form the projective limit of the finite groups $\text{Gal}(L/K)$,

$$G_K = \text{Gal}(\bar{K}/K) := \text{proj. lim. Gal}(L/K),$$

giving G_K the topology that derives from this (usually called the “Krull topology”); namely, a base of open subgroups for the Krull topology on G_K consists of the kernels of the natural surjections $\text{Gal}(\bar{K}/K) \rightarrow \text{Gal}(L/K)$ as L/K runs through, as above, all finite intermediate Galois extensions.

“The” Galois group $G_K = \text{Gal}(\bar{K}/K)$ of K depends upon a choice of algebraic closure \bar{K} . If we make a different choice of algebraic closure, \bar{K}' , we will not change the isomorphism type of the topological group G_K , and indeed any isomorphism $\bar{K} \cong \bar{K}'$ which is the identity on K will induce an isomorphism of topological groups $\text{Gal}(\bar{K}/K) \cong \text{Gal}(\bar{K}'/K)$. Following this line of reasoning, we see that, in fact, our Galois group G_K is well-defined, but *only* up to inner automorphisms.

How can one study a profinite group G that is well-defined *only* up to inner automorphisms?

One simple, but partial, answer is to retreat and just study the maximal abelian topological quotient of G , sometimes referred to as G^{ab} , the *abelianization* of G . For, if G is determined up to inner automorphism, G^{ab} is pinned down up to *unique* isomorphism; i.e., G^{ab} is as definitively pinned down as any mathematical object can be. Essentially equivalently, we may work with the continuous one-dimensional representations of G .

Of course it would be even better to classify *all* finite dimensional representations of G . This is a sensible thing to try to do, since a representation is given by a homomorphism up to inner automorphism, and therefore even though we may not be entitled to give particular elements of G actual names (only conjugacy classes), we can indeed refer, in a well-defined way, to specific representations of G despite the indeterminacy inherent in the fact that our G is itself well-defined only up to inner automorphism.

For K a number field, to study the abelianization G_K^{ab} (or equivalently, the collection of continuous one-dimensional representations of G_K), Class Field Theory provides us with a fairly concrete and sometimes manageable answer. For example, one of the initiating results of Class Field Theory, the Kronecker-Weber theorem, now over a century old, tells us that any finite abelian extension of \mathbf{Q} is contained in some cyclotomic extension, i.e., in an extension of the form $\mathbf{Q}(e^{2\pi ir})$ for some rational number r . Kronecker thought of this result in the following illuminating way: the abelian field extensions of \mathbf{Q} are systematically generated by the values of a *single analytic function*, the exponential function $\exp(2\pi iz)$, where z runs through the elements $r \in \mathbf{Q}$; a more modern paraphrase of this is that the abelian field extensions of \mathbf{Q} are generated by the coordinates of the points of finite order in the algebraic group \mathbf{C}^* . Kronecker subsequently found large caches of abelian extensions of quadratic imaginary fields K generated by the values of the elliptic modular function $j(z)$ where z runs through elements $\alpha \in K$ (a more modern

paraphrase being that these abelian extensions are generated by the coordinates of torsion points in certain elliptic curves – those admitting complex multiplication by K). Kronecker expressed the desire to extend this strategy to find all abelian extensions of more general number fields as the “dream of his youth.”

What about the more ambitious project of representations of G_K to GL_n for $n \geq 1$?

Since G_K is a projective limit of finite groups, any continuous representation to $\mathrm{GL}_n(\mathbf{C})$ will necessarily factor through a finite quotient group of G_K . To get more interesting, and more perspicacious, representations, we are led to replace the ring \mathbf{C} – the standard scalar ring of classical representation theory – by a wider, more receptive array of possible scalar rings, given the profinite nature of the group we wish to study. We find ourselves interested, therefore, in the continuous representations of G_K in $\mathrm{GL}_n(R)$ where R ranges through a wide class of topological rings, which include, for example, all finite commutative rings.

Once one agrees to allow such a large collection of rings as scalar rings, however, a fascinating *deformational issue* presents itself. To prepare to explain this, let $\mathbf{C}[t]$ denote the polynomial ring in the variable t over the field of complex numbers. Now, it is natural to view a representation of a group into $\mathrm{GL}_n(\mathbf{C}[t])$ as a parametrized (algebraically varying) family of representations of the group into $\mathrm{GL}_n(\mathbf{C})$, the parameter variable being t : specialize t to any complex number and the representation to $\mathrm{GL}_n(\mathbf{C}[t])$ specializes to a representation to $\mathrm{GL}_n(\mathbf{C})$. Following this line of thought, if $\mathbf{C}[[t]]$ denotes the ring of power series in the variable t with coefficients in \mathbf{C} , we may think of a continuous representation of our group into $\mathrm{GL}_n(\mathbf{C}[[t]])$ as a *formal deformation* of the complex representation we get by setting the variable t to 0. Indeed, just as it is natural to do *that*, it is reasonable to consider a continuous representation of a group, the group G_K for instance, into $\mathrm{GL}_n(\mathbf{Z}/p\mathbf{Z}[[t]])$ as a *formal deformation* of representation into $\mathrm{GL}_n(\mathbf{Z}/p\mathbf{Z})$ that we get by setting the variable t to 0. Here $\mathbf{Z}/p\mathbf{Z}[[t]]$ is the ring of power series in the variable t with coefficients in the finite field $\mathbf{Z}/p\mathbf{Z}$.

To give ourselves more flexibility in this notion of deformation, let us make the following definition. For a finite field k (say, $k = \mathbf{Z}/p\mathbf{Z}$) and a complete local noetherian ring R with residue field k , if

$$\bar{\rho} : G_K \rightarrow \mathrm{GL}_n(k)$$

is a continuous homomorphism, by a **deformation of $\bar{\rho}$ to R** let us mean a lifting

$$\rho : G_K \rightarrow \mathrm{GL}_n(R)$$

of the homomorphism $\bar{\rho}$, where ρ is taken only up to conjugation by an element of $\mathrm{GL}_n(R)$ trivial when reduced to k . The initial representation $\bar{\rho}$ we shall call the *residual representation* (its ring of scalars being the residue field), and the type of question we are led to ask is: *describe all deformations – or perhaps all deformations with a list of specific features – of the given residual representation.*

For simplicity let us assume that our residual representation $\bar{\rho}$ is absolutely irreducible (that means even after extending scalars from k to its algebraic closure, $\bar{\rho}$ is irreducible). To put some control on the collection of deformations of $\bar{\rho}$ to be studied, one fixes a specific finite set S of primes of the number field K which include all the primes dividing the characteristic of the finite field k and all the primes at which the residual representation is ramified; let us then consider only deformations of $\bar{\rho}$ which are unramified at all primes not in S .

It is a theorem that a *universal deformation* of $\bar{\rho}$ with the requested ramification conditions exists. This means that there is a (unique) complete noetherian local ring R with residue field k (R is dependent, of course, on $\bar{\rho}$ and S) with a continuous homomorphism $\rho : G_K \rightarrow \mathrm{GL}_n(R)$ which is a deformation of $\bar{\rho}$ unramified at primes outside S with the property that any other deformation ρ_o of $\bar{\rho}$ (unramified outside S) to some complete noetherian local ring R_o with residue field k is obtained from ρ by a unique ring homomorphism $R \rightarrow R_o$ (inducing the identity on residue fields).

The kinds of rings encountered as universal deformation rings can be quite interesting. Happily, often they are relatively benign.

Consider the following particular case, which is the focus of much arithmetic work in the classical theory of modular forms. Let $K = \mathbf{Q}$. Let $\bar{\rho} : G_{\mathbf{Q}} \rightarrow \mathrm{GL}_2(\mathbf{F}_p)$ be an absolutely irreducible residual representation in which complex conjugation does not act as a scalar, and S precisely the set of ramified primes of $\bar{\rho}$ (with the prime p thrown in). In this situation one frequently – but certainly not always – has that R is isomorphic to a power series ring in three variables over \mathbf{Z}_p . By specializing the three variables, one then gets from such an R a three-parameter family of representations to $\mathrm{GL}_2(\mathbf{Z}_p)$ which comprises, in fact, all deformations to $\mathrm{GL}_2(\mathbf{Z}_p)$ having the properties stipulated.

As a concrete example of this, consider the famous modular form (of weight 12) studied by Ramanujan (which already made an appearance in section 6 above) whose Fourier series expressed in the variable $q = e^{2\pi iz}$ is given by

$$\Delta = q \prod_{n=1}^{\infty} (1 - q^n)^{24} = \sum_{n=1}^{\infty} \tau(n) q^n.$$

For every prime number p there is a continuous representation $\bar{\rho}_{p,\Delta} : G_{\mathbf{Q}} \rightarrow \mathrm{GL}_2(\mathbf{F}_p)$ unramified at all primes $\ell \neq p$ and having the further property that the trace of the action of a Frobenius element for the prime ℓ in $G_{\mathbf{Q}}$ is given by the reduction mod p of the coefficient $\tau(\ell) \in \mathbf{Z}$. For all prime numbers $p \geq 29$, different from 691 and such that $\tau(p)$ is not divisible by p , one has that the corresponding universal deformation ring $R = R_{p,\Delta}$ classifying deformations of $\bar{\rho}_{p,\Delta}$ unramified outside p is isomorphic to a power series ring in three variables (see the proposition in section 11 of [M]); this gives quite a number of instances, for the only prime between 691 and 3 million with the property that $\tau(p) \equiv 0 \pmod{p}$ is $p = 2411$.

In the spirit of Kronecker (the “dream of his youth”) we might ask whether we can *explicitly generate* such universal deformation rings R for the $\bar{\rho}$'s described two paragraphs above and specific quotient rings of R universal for more stringent deformation problems associated to them (e.g., problems formulated by imposing a condition on the determinant, and local conditions at the finitely many primes in S). A celebrated conjecture of Serre tells us that we should at least find the residual representations described above in the natural action of the Galois group of \mathbf{Q} on p -torsion points in the jacobians of modular curves. Can we reconstruct the rings R , the finer quotients of R alluded to above, and their associated universal $G_{\mathbf{Q}}$ -representations, by suitably completing the correspondence rings of modular curves and suitably compiling representations of $G_{\mathbf{Q}}$ on p -power torsion points in the jacobians of modular curves?

Thanks to the magnificent modularity theorem due to Wiles, Taylor, Diamond, Conrad, and Breuil, we are significantly further along in our understanding of this question.

9. MIRROR SYMMETRY – AND BACK TO DEFORMING LATTICES IN THE PLANE

Mathematicians are quite adept at making good use of any symmetry satisfied by their objects of study. But some symmetries are more hidden than others. “Mirror symmetry” is a major (relatively new) development in mathematics, which mixes, and in a certain sense switches, *complex analytic geometry* and *symplectic geometry* (and deformational issues regarding these geometries). The key seems to be a quantum field theory underlying both kinds of geometry. A simple instance of this is given by elliptic curves (equivalently, lattices in the plane), and we will say a few words about that. There is an advantage and a disadvantage in choosing elliptic curves to illustrate mirror symmetry. The advantage is that it is a nice, relatively elementary, example, and most, perhaps, of what is expected to be true is actually proved. The disadvantage is that this case hardly begins to reveal the magic of mirror symmetry and its concurrent developments, which has transformed much of our thinking about even the most classical of topics: enumerative algebraic geometry. For example, who would have thought twenty years ago that the well known fact that a general cubic surface has 27 straight lines, any one of which meets 10 others, is equivalent to the associativity of the quantum cohomology ring on certain rational surfaces? (Cf. [Cr-M].) Or who would have expected that the fact that there are precisely 22930588887625 rational curves of degree 5 on a general quintic threefold can be understood from the vantage point of symplectic geometry?

A piece of “enumerative topology”: For almost everything that I will be saying in this section, the reference is Robbert Dijkgraaf’s fine article [Di]. Consider the function $Z(q, \lambda)$ of the variables q and λ which encodes the number of essentially distinct ways that a compact two-manifold of genus g can cover a torus by a mapping of degree d with ramification indices no greater than two. Here is its precise definition.

We let $\mathcal{N}_{g,d}$ denote the following “weighted number” of all not necessarily connected covers of an elliptic curve E with Euler characteristic $2 - 2g$, of degree d , possessing only ramification points of ramification degree 2; namely,

$$\mathcal{N}_{g,d} = \frac{\#\text{Hom}'(\Pi^{\{2g-2\}}, \mathcal{S}_d)}{d! \cdot (2g - 2)!},$$

where the group \mathcal{S}_d is the symmetric group on d letters, the group $\Pi^{\{2g-2\}}$ is the fundamental group of our (or any) elliptic curve with $2g - 2$ points removed, and the superscript $'$ is meant to signify that we consider only homomorphisms that have the property that the $2g - 2$ holonomies about the $2g - 2$ removed points map to cycles of length two in \mathcal{S}_d .

Define

$$(1) \quad Z(q, \lambda) := q^{-\frac{1}{24}} \sum_{g,d=1}^{\infty} \mathcal{N}_{g,d} \cdot q^d \lambda^{2g-2}.$$

Visibly $Z(q, \lambda)$ is defined by means of the topological structure of the two-manifold E . And this topological structure is *stable*, so our function $Z(q, \lambda)$ might, at first glance, seem the least likely candidate to have anything to do with deformations of complex or symplectic structures. Even if we gave E a complex structure or a symplectic structure, $Z(q, \lambda)$, in its definition above, would not *see* it. But there are surprises in store.

Giving E both a complex structure and a symplectic structure:

- Suppose that E is given a **complex structure**, which amounts to taking E to be equal to the elliptic curve E_L attached to a lattice $L \subset \mathbf{C}$, and here we will no longer require that the area of a fundamental parallelogram of L be 1. After multiplication by a suitable complex number, we can assume that L is generated by 1 and a complex number $\tau = \tau_1 + i\tau_2$ where $\tau_2 > 0$.

- Suppose also that we specify, on E , a $(1, 1)$ -form $\omega = \text{constant} \cdot dz \wedge d\bar{z}$ where if $t := \frac{1}{2\pi i} \int_E \omega = t_1 + it_2$, we have that $t_2 > 0$ is the area of the surface. Clearly ω is a closed two-form, which is nondegenerate in the usual sense and therefore may be thought of as a **(complexified) symplectic structure** on E .

Thus E , endowed with both structures, is dependent upon the *two* parameters, τ and t :

$$E = E_{\tau,t},$$

where, of course, the dependence on τ is really only on the lattice generated by 1 and τ . Therefore:

$$E_{\tau,t} = E_{\tau+1,t} = E_{-1/\tau,t}.$$

The relevance of this *combined* complex/symplectic structure on E to the study of the power series $Z(q, \lambda)$ is this: thanks to a beautiful idea of Kontsevich using the structure $E_{\tau,t}$, we may represent the coefficients of the power series in $\lambda Z(e^{2\pi it}, \lambda)$ as the *volumes* of suitable “moduli spaces” (the Hurwitz spaces $\mathcal{M}_g(E, d)$ of coverings of E of genus g and degree d) where the volume form on the Hurwitz space is normalized by the symplectic form ω . In a word, we have an “interpretation” of $Z(q, \lambda)$ as an integral, involving the symplectic structure of $E_{\tau,t}$, where the somewhat formal variable q has become $e^{2\pi it}$ for t the parameter describing the symplectic structure of E .

But now it turns out that the phenomenon of mirror symmetry (in this context) gives us a “functional equation”, if one may call it that, enjoyed by $Z(q, \lambda)$ under the transposition

$$E_{\tau,t} \longleftrightarrow E_{t,\tau}.$$

Here, one is willfully interchanging the parameter τ , which describes the variation of *complex structure*, with the parameter t of variation of *symplectic structure*, an exchange which crosses species of geometries. An underlying reason for this is that $Z(q, \lambda)$ has an interpretation in terms of the underlying quantum field theory for $E_{\tau,t}$, and this quantum field theory itself has an involution that makes the above interchange. The corollary of this functional equation is that since $E_{\tau,t}$ is unperturbed under the transformation $\tau \mapsto -1/\tau$, we get that $Z(e^{2\pi it}, \lambda)$ also satisfies a functional equation under the interchange $t \mapsto -1/t$. This allows one to deduce that the counting function $Z(q, \lambda)$ is, appropriately understood, (almost) a modular function of q . In fact, one has

$$(2) \quad Z(q, \lambda) := q^{-\frac{1}{24}} \oint \frac{dz}{2\pi iz} \prod_{w \in \mathbf{Z}_{\geq 0 + \frac{1}{2}}} (1 + zq^w e^{\lambda w^2/2})(1 + z^{-1}q^w e^{\lambda w^2/2}),$$

which presents our counting function $Z(q, \lambda)$ as some kind of generalized theta function; and, more specifically, writing $\log Z(q, \lambda)$ as a power series in λ ,

$$(3) \quad \log Z(q, \lambda) = \sum_{g=1}^{\infty} \lambda^{2g-2} F_g(q),$$

one gets that the coefficients $F_g(q)$ for $g > 1$ are “quasimodular” forms of level 1 and weight $6g - 6$ with rational coefficients. That is, they are in the ring $\mathbf{Q}[E_2, E_4, E_6]$ where E_k is the normalized Eisenstein series of weight k ,⁶ e.g.,

$$F_2(q) = \frac{1}{51840}(5E_2^3 - 3E_2E_4 - 2E_6).$$

For a different derivation of this formula, see [Ka] (noting that the two formulas differ by a factor of 2 because of differences in normalization). For a direct proof of the formula of Dijkgraaf for $F_g(q)$ (general g), see [K-Z]. For genus $g = 3$ one gets

$$F_3(q) = \frac{1}{35831808}(15E_2^4E_4 - 6E_2^6 - 12E_2^2E_4^2 + 7E_4^3 + 4E_2^3E_6 - 12E_2E_4E_6 + 4E_6^2).$$

(For more examples and discussion of $F_g(q)$ see [Ru] and [M-R-Y].⁷)

10. QUANTIZATION IN PHYSICS, AND NONCOMMUTATIVE GEOMETRY

Classical mechanics is often given the following concise mathematical formalization. Denoting by X the manifold of “positions”, one forms the manifold M whose points encode “position and momentum”; that is $M = T^*(X)$ is the total space of the cotangent bundle of X . Since the underlying differential equations in classical mechanics are of the second order, the geometrization of this theory will quite naturally⁸ take place on M .

We may think of the functions in $C^\infty(M)$ as the algebra of *observables* in our dynamics. On M we have a natural symplectic form which has local expressions $\omega = \sum_j dq_j \wedge dp_j$ where q_1, \dots, q_d are local coordinates for a chart on X , and p_1, \dots, p_d are the corresponding local *momentum variables*. The symplectic form ω gives rise to a Lie algebra structure on $C^\infty(M)$. Expressed in terms of coordinates, for $f, g \in C^\infty(M)$ we have

$$\{f, g\} := \sum_{i=1}^d \left(\frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i} \right),$$

and note that this operation is a special case of a Poisson bracket operation: a **Poisson bracket operation** on a commutative \mathbf{C} -algebra A_o is, by definition, a \mathbf{C} -bilinear operation $(f, g) \mapsto \{f, g\}$ which is a derivation in each variable separately, which satisfies $\{f, f\} = 0$ for all f and which satisfies the Jacobi identity. A general Poisson bracket operation on $C^\infty(M)$ where M is a manifold of dimension d is given, locally, as

$$\{f, g\}(x) = \sum_{i,j=1}^d \alpha^{i,j}(x) \partial_i(f) \partial_j(g),$$

⁶The reason why this is called “quasi-modular” is that its formula involves the Eisenstein series E_2 of weight two which is not *quite* a modular form.

⁷The displayed formula corrects a minor typo in the formula given in [Di]; cf. [M-R-Y].

⁸“Quite naturally” at least ever since Lagrange’s *Mécanique Analytique* [La], in which Lagrange emphasized the role of the momentum (which he called “la quantité de mouvement du corps”. See section 1 of part II of *Mécanique Analytique* for Lagrange’s discussion of *one* of Descartes’ mistaken views concerning this and his attribution of the idea to Wallis). Focussing not on the *geometric* points in physical space but rather on viewing *position* and *momentum* on an equal footing and working geometrically in this phase space may be considered “natural” only after someone has had the genius to have done so.

where $\alpha = (\alpha^{i,j})$ is a skew-symmetric bivector field on M obeying the Jacobi identity

$$\sum_{\ell=1}^d \alpha^{i,\ell} \partial_{\ell} \alpha^{j,k} + \alpha^{j,\ell} \partial_{\ell} \alpha^{k,i} + \alpha^{k,\ell} \partial_{\ell} \alpha^{i,j} = 0.$$

Call a manifold M with Poisson brackets structure on $C^{\infty}(M)$ a **Poisson manifold**, noting that symplectic manifolds are special cases of Poisson manifolds. A Poisson manifold does not yet have a specified *dynamics* until we supply one further ingredient, namely a function $H \in C^{\infty}(M)$ which plays the role of *energy*. Given H , one can then express the equations of time evolution of the corresponding dynamics quite simply as (Hamilton's equation)

$$\dot{f} = \{f, H\}.$$

We then have that $\dot{H} = \{H, H\} = 0$; i.e., *conservation of energy* holds.

Let us now pass from this formalization of “the algebra of observables” in classical mechanics to the corresponding formalization of “the algebra of observables” in quantum mechanics (and from Hamilton's equation to Schrödinger's equation). The underlying complex vector space of observables in the quantum mechanical model can be identified with the vector space of observables in the previously discussed classical model. But the algebra structure is now noncommutative, its “noncommutativity” being dependent upon the fact that *Planck's constant*, which will appear as the letter q below, is nonzero.⁹

The problem of “quantizing” a classical algebra of observables can be generalized and formalized as the problem of associating to a commutative Poisson algebra A_o over \mathbf{C} a *deformation* of A_o : that is, an associative (noncommutative) $\mathbf{C}[[q]]$ -algebra A which is flat over $\mathbf{C}[[q]]$, such that q is in the center of A , which is given with an isomorphism of algebras $A/qA \cong A_o$, and such that if $\{-, -\}$ denotes the Poisson brackets in A_o , and then if $a, b \in A$ with images $a_o, b_o \in A_o$,

$$\frac{1}{q}(ab - ba) \equiv \{a_o, b_o\} \pmod{q}.$$

Which Poisson algebras can be quantized in the above sense, and how unique is the quantization? In 1983 De Wilde and Lecomte, and also Deligne and Fedosov separately, showed that any symplectic manifold (M, ω) can be so quantized, the local problem for the standard symplectic structure on \mathbf{R}^{2n} having been settled earlier by Moyal, by an explicit formula (cf. [D-L], [F], [De]). Recently, guided very closely by intuition derived from Quantum Field Theory, Kontsevich [Ko 2] extended this result to provide such a quantum deformation of any Poisson manifold, i.e., any manifold M with a Poisson bracket on its algebra $A_o := C^{\infty}(M)$ of C^{∞} functions (see also [Ko 1]).

⁹L. Fadeev, referring to material, for example, in [B-F-F-L-S], writes in his introductory lecture on Quantum Field Theory in [Q-F], “Beautiful results, which I learned from A. Lichnerowicz, M. Flatro, and D. Sternheimer allow one to say that Classical Mechanics is unstable and that Quantum Mechanics is essentially a unique deformation of it into a nonequivalent stable structure. The degeneracy of Classical Mechanics (exactness of pure states) is intimately connected with its instability. So it is only natural that the passage from Classical Mechanics to Quantum Mechanics was prompted by the experimental activity of physicists. The stability of Quantum Mechanics, on the contrary, shows that an analogous modification of it is less feasible. In this case one has first to modify also all the general framework...”

The structural constants of Kontsevich’s deformation to noncommutative geometry which are given as concrete volumes reminiscent of integrals occurring in hyperbolic geometry and number theory, and which are organized according to a schema of delicate combinatorics coming from considerations in physics, suggest the emergence of some new chapter of mathematics. To be sure, a feature of the work that makes it enchanting is that its subtlety (including the issue of how unique the deformation is, given the Poisson manifold M) is governed by a “physical” (in some sense) model. More explicitly, Cataneo and Felder [Ca-F] suggested the following interpretation of Kontsevich’s deformation.¹⁰ Consider a topological quantum field theory on a disc D with fields $x : D \rightarrow M$ (differentiable maps of the disc to the Poisson manifold M) and differential 1-forms η on the disc with coefficients in x^*T^*M and which vanish on the boundary. Since dx is a differential 1-form on the disc with values in x^*TM , we view $dx \wedge \eta$ as a real-valued differential 2-form on the disc; similarly, if α is the bivector field on M that gives its Poisson structure, we form $\alpha(x)\eta$ as a differential 1-form on the disc with values in x^*TM and consider the real-valued differential 2-form $\alpha(x)\eta \wedge \eta$.

Define the action

$$S[x, \eta] := \int_D (dx \wedge \eta + \frac{1}{2}\alpha(x)\eta \wedge \eta).$$

Cataneo and Felder offer the following path integral interpretation of the multiplication law in the Kontsevich deformation of the Poisson algebra. If $m \in M$, and f, g are in $A_o = C^\infty(M)$, then:

$$f * g(m) = \int_{x(\infty)=m} f(x(1))g(x(0))e^{\frac{1}{q}S[x,\eta]} dx d\eta,$$

where $0, 1, \infty$ are (any) three fixed points on the boundary of D in counter-clockwise order. The integral is over all fields x with the constraint $x(\infty) = m$ and over all η ’s which vanish on the boundary. This star-multiplication law then extends naturally to a multiplication law on all of $A := A_o[[q]]$ (the element q being in the center). As with all “interesting” path integrals, the challenge is to stipulate (if possible) the measure with respect to which the integration is being made, and/or failing in that, to divine the coefficients of the power series in q that this integral points to. The article of Cataneo and Felder has an enlightening discussion of this, comparing $f * g$ to the formulas of Kontsevich; they also have quite a nice explanation for why this formula leads to an associative multiplication law.

11. NEAR-MISSES IN ALGEBRA AND NUMBER THEORY (APPROXIMATION OF IRRATIONALITIES)

It may seem entirely natural to ask for perturbational questions in analysis and geometry. What about algebra? Algebra tends to deal in *equalities*, in exact equations. Number theory often does. For example, in the Fermat problem, one asks for triples of perfect n -th powers a^n, b^n, c^n where the last perfect n -th power is *exactly* the sum of the first two, not just *approximately so*. You are not interested in “near-misses”.

Or are you? It may come as a surprise how many number-theoretic problems there are which ask for the solution to some “exact equation”—but which, when you jiggle these problems and ask for the structure of their “near-misses”, produce

¹⁰This hasn’t yet been turned into a rigorous alternative proof of Kontsevich’s result.

yet more profound problems with broader implications. One does not have to go far to see examples of this:

(1). How close can you get to \sqrt{d} by rationals whose denominators have restricted size? Consider the ancient theorem, ascribed in Plato's dialogue *Theaetetus*, to the mathematician (and later, Athenian general) Theaetetus, which asserts that if d is a natural number which is not a perfect square of an integer, then \sqrt{d} is irrational. That is, $X^2 - dY^2$ is never zero. If you ask "near-miss versions" of this – e.g., how close to zero can $X^2 - dY^2$ get for nonzero integral values of X and Y , and how often? – you find yourself with the famous problems first considered by Brahmagupta in the 7-th century or Bháskara in the 12-th (and, of course, the 17-th century European mathematicians). The smallest nonzero value you can possibly hope for is ± 1 . The resulting equation

$$X^2 - dY^2 = \pm 1,$$

which goes under the title *Pell's Equation*, has an infinite number of solutions (the group of units in the ring $\mathbf{Z}[\sqrt{d}]$). The full elucidation of this near-miss problem, which brings in the theory of continued fractions and of quadratic number fields and connects with the theory of "best rational approximants to irrational numbers", is still not entirely understood. For example, let $d > 1$ be a squarefree number and let (x, y) be the smallest pair of positive numbers solving the above displayed "Pell equation". Now an upper bound for x and y is known:

$$x + \sqrt{d}y < (e^2 \cdot d)^{2\sqrt{d}},$$

and Fermat knew instances of d 's which made that upper bound respectable (if $d = 109$, then $x = 158070671986249$ and $y = 15140424455100$), but we still do not know whether there is a positive number ϵ and an infinity of squarefree d 's for which

$$x + \sqrt{d}y > d^{d^\epsilon}.$$

(The answer to this question would be *yes* if, for example, there were an infinity of real quadratic fields of class number one.)

Nevertheless, the "partial elucidation" of this near-miss problem that is currently available to us has found immense applications.

(2). More generally, given a curve in the plane cut out by a polynomial, how near to it can a lattice point get and still miss? Consider the following problem that is superficially similar to (1). The polynomial $X^3 - Y^2$ can, of course, achieve the value zero when X is a perfect square. Suppose, however, that X is not a perfect square. How close to zero can $X^3 - Y^2$ get, for appropriate choice of integers Y ? In effect, how close to a whole number can the irrational number $\sqrt{X^3}$ get? This near-miss question, which a priori has only to do (again) with approximating

quadratic irrationalities, is intimately related to the arithmetic theory of elliptic curves over the rational numbers.¹¹

This link alone might make us want to understand the structure of the near-miss solutions to

$$X^3 - Y^2 = \text{zero}.$$

How do these solutions behave? A celebrated conjecture of Hall says that for any exponent $\alpha < 1/2$ there is a constant $C_\alpha > 0$ such that $|X^3 - Y^2|$ is greater than $C_\alpha \cdot X^\alpha$ for all X, Y such that $X^3 - Y^2$ is not zero. Hall's conjecture is now but a special case of some grand conjectures of Vojta (who made his conjectures by "unifying" the language of Nevanlinna's work in complex variables, with diophantine geometry). It is also directly related to the *ABC*-conjecture due to Masser and Oesterlé, which provides a "quantitative" version of the following qualitative assertion: *there is a strong inhibition for two (relatively prime) natural numbers which are highly divisible by perfect powers to have the property that their sum is also highly divisible by a perfect power.*

(3). More generally, given a smooth curve in the plane, how near to it can a point with rational coordinates get and still miss? "Ordinary geometry" seems to be inadequate for the purposes of physics-at-small-distances. For if you wish to measure, adequately, a phenomenon confined to a very small bit of space, you need very high energy to do it. So, *geometric points* would require infinite energy to be dealt with and therefore are inaccessible to the surveillance of the physicist. This type of "infinity" that crops up in physics is usually dealt with by a technique known as renormalization. Quantum mechanics, the quantum field theories, and the emerging noncommutative deformations of geometry are, perhaps, further manifestations of the ultimate unsuitability of using a geometry dependent upon a *substrate of points* to model the physical world.

Number theory has also felt, at times, the inadequacy of "ordinary space"; it is sometimes better to work in an adelic space, where the notion of "closeness" between two points is determined by a combination of shared congruences modulo ideals as well as distance in a more naively geometric sense. And sometimes the issue of "infinite energy" complicating the physics of small distance has a (remote) parallel in number theory: when you try to compute rational points on or "near" (say within distance δ of) a given hypersurface H in "ordinary" space, it is reasonable to search only for points whose coefficients have bounded numerator and denominator (say, $\leq N$); otherwise you will be faced with a process that takes infinite time. The algorithms used in searching for points are "lattice reduction algorithms". Therefore, to bring their running-time estimates to the forefront you are naturally led, as we shall see, from the given hypersurface H in "ordinary space" to consider a

¹¹The link being that any elliptic curve E over \mathbf{Q} is determined by the classical (integral) invariants c_4 and c_6 for the model of E over \mathbf{Z} with smallest (absolute value of) discriminant Δ where c_4 and c_6 are equal to the values of the normalized Eisenstein series introduced in section 6, $E_4(L)$ and $E_6(L)$, respectively, for an appropriate lattice L ; these invariants are yoked by the equation

$$c_4^3 - c_6^2 = 1728\Delta,$$

and "conversely" any integral solution to the equation

$$X^3 - Y^2 = 1728 \cdot Z,$$

with X and Y satisfying some mild congruence conditions, comes from an elliptic curve E over \mathbf{Q} , with its $c_4 = X$, $c_6 = Y$, and $\Delta = Z$.

certain family of mappings of H to the moduli space of three-dimensional lattices. Since all this is a fine mixture of practical numerical computation and theory, I think it is a good topic with which to end my talk. My reference here is Elkies' article [E].

That the procedure we will be alluding to below *has* practical applications is shown by its ability (more precisely, Elkies' ability) to work with rather large numbers – for example, to discover the following “near-miss” to Hall’s problem:

$$5853886516781223^3 - 447884928428402042307918^2 = 1641843.$$

Hall’s problem, which we discussed earlier, involves the inhomogeneous polynomial $x^3 - y^2$, and Elkies has an ingenious modification of his general algorithm to deal with that case. But, to describe the garden-variety version of the algorithm, consider the question of finding rational points of small height which are close to the image of a smooth mapping $\phi : [0, 1] \rightarrow \mathbf{R}^2$, a parametrization of a piece C of the real locus of a plane curve. More specifically, fix a large number N and a small number δ ; the game is to find rational points $(a/b, c/b)$ in the Euclidean plane \mathbf{R}^2 which are δ -close to the image of ϕ and such that the absolute values of a, b, c are $\leq N$. The two denominators are required to coincide because we are really working with the homogeneous coordinates (a, b, c) . How might we go about computing these near-misses?¹² The method to be described is just as useful, as Elkies points out, in cases where ϕ is not algebraic but just twice differentiable (e.g., Elkies will test his method out on fairly whimsical problems, getting, for example, near-misses of the form

$$\left[\frac{2063}{8128}\right]^\pi + \left[\frac{8093}{8128}\right]^\pi = 1.0000000000003607083-).$$

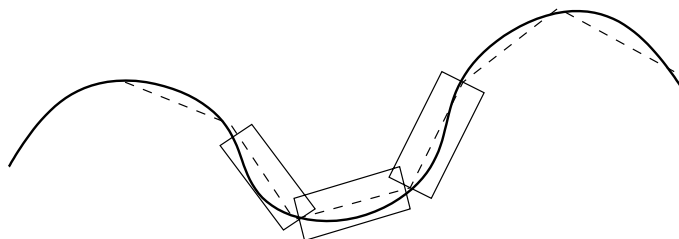


Figure 12

We begin by finding a piece-wise linear approximation to the arc C with small chord-length $\delta > 0$, as in figure 12, and where this δ is to be negotiated. Since C is smooth, it is entirely contained in the union of rectangles, as drawn, where the rectangles have length δ and width $O(\delta^2)$ where the constant in the “big O ” is determined by the size of the derivatives of ϕ and fixed. We search for rational points in each of these rectangles whose numerators and denominators are bounded as described above by N , and we do so “independently”, i.e., rectangle by rectangle. Choose one of these rectangles which correspond to a chord and call it Δ of length δ . Pass from projective two-space to the corresponding affine three-space and note that one can reconfigure this problem as the problem of finding integer lattice points in a parallelepiped of height, length, and width roughly proportional to

¹²This problem is more flexible than the somewhat analogous, but much more demanding, near-miss problem in “one lower dimension” answered by the classical theory of continued fractions; there the data is just a point $\gamma \in \mathbf{RP}^1$, and the object is to find rational points of small height which are *closest* to γ – closest in a certain technical sense, given their height.

$N, N\delta, N\delta^2$, which suggests that we keep $\delta \gg 1/N$ to make sure that the volume of these parallelepipeds are $\gg 1$, guaranteeing that they do indeed capture some points. Taking then

$$\delta = \text{fixed constant} \cdot 1/N,$$

where the fixed constant is $\gg 1$, we get a moving frame of parallelepipeds as in figure 13:

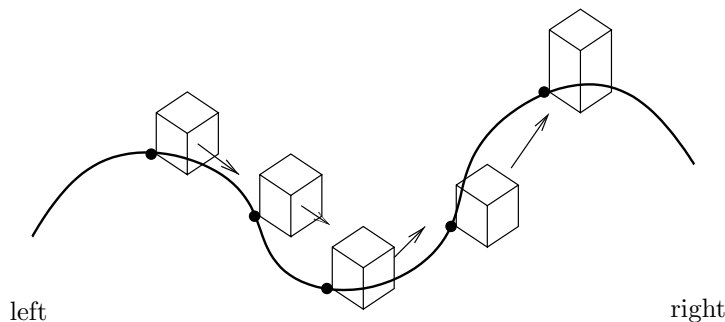


Figure 13

By multiplying by an appropriate element of $GL_3(\mathbf{R})$ which sends our parallelepiped to a fixed box of given dimensions in \mathbf{R}^3 and sends the integer lattice $\mathbf{Z}^3 \subset \mathbf{R}^3$ to some lattice L , we can invert our “reconfigured problem” to the problem of finding lattice points of L which lie in the standard box. One has excellent lattice reduction methods to deal with this, but if you wish to estimate in an a priori fashion the amount of computer time this will require, you must know something about the nature of the lattices L that arise; in particular you should have an estimate of the number and size of the *shortest vectors* that they possess.

Which L 's arise? These lattices depend upon the size N of the solution you wish to catch, the fixed constants (in the “big O” and the determination of δ), and an endpoint c of the chord Δ . Fixing the fixed constants and choosing the N to limit the rational points you are searching for, you get for $c \in C$ a lattice $\Lambda_N(c) \subset \mathbf{R}^3$. Some “a priori” knowledge of the structure of the mapping

$$\Lambda_N : C \longrightarrow \begin{array}{l} \text{the five-dimensional moduli space} \\ \text{of all lattices in } \mathbf{R}^3 \\ \text{taken up to rotation equivalence} \end{array}$$

influences our a priori estimate for the efficiency of near-miss computations. Relevant, for example, is the question of how close to a cusp does this image of C get? Does it fall into a restricted subspace? Regarding the latter of the two questions, a modification of this algorithm geared toward finding solutions of Hall’s equation has the unexpected property that the corresponding Λ_N ’s map to lattices which are symmetric squares of two-dimensional lattices, allowing for an extraordinary efficiency in lattice reduction.

For the general problem, controlling the structure of short vectors in the lattices of these ribbons, Elkies [E] tells us that:

Let C be a smooth arc in \mathbf{RP}^2 . Then for each $N > 1$ and $\delta \geq N^{-2}$ one can find $O(\delta N^3)$ rational points and $O(N)$ rational line segments each of length $O(1/N)$ in \mathbf{RP}^2 which together include all rational points of height $\leq N$ and distance $\leq \delta$ from C . The (line segments and) rational points can be computed in time $\ll \delta N^3 \log^{O(1)} N$. Ignoring the space used to record the line segments and rational points, the computation requires space $\ll \log^{O(1)} N$. All implied constants depend effectively on C .

* * *

In the opening pages of his book, *Structural stability and morphogenesis*, René Thom writes:

“The use of the term ‘qualitative’ in science, and above all in physics, has a pejorative ring. It was a physicist who reminded me, not without vehemence, of Rutherford’s dictum “qualitative is nothing but poor quantitative.”

If Henri Poincaré had been present in that conversation, surely he would have helped Thom out, reiterating his stance regarding the qualitative study of differential equations: questions about the general shape of curves are, “for certain applications, just as important as numerical calculations.” (Cf. Poincaré’s *Résumé Analytique*, written in 1901 at the request of Mittag-Leffler, a ninety-nine page extract of which is reprinted in [P]; see, in particular, loc. cit., p. 263.)

But I suspect that if the three (Poincaré, Rutherford’s friend, and Thom) were to have a chat today, the *qualitative* and the *quantitative* would be viewed a bit less as competitors and more as different facets of a fuller understanding of the mathematics in question. Moreover, as is evident in Elkies’ result and its formulation, we are in the era of high algorithmic awareness, on both a practical front (how long will it be until I get the results of my computation?) and a theoretical one (what are the asymptotics of time-cost in running this algorithm?). This nudges us to add to our (qualitative and/or quantitative) geometric analyses a study (qualitative and/or quantitative) of the behavior of the relevant algorithms. And, to be sure, of their perturbations.

REFERENCES

- [A1] Arnold, V.I.: *Catastrophe Theory* (transl. Wassermann, G.S.), Springer-Verlag, 3rd revised and expanded edition (1992) MR **93h**:58018
- [A2] Arnold, V.I., *Topological Invariants of Plane Curves and Caustics*, University Lecture Series, **5**, American Mathematical Society (1994) MR **95h**:57003
- [B-F-F-L-S] Bayen, F., Flato, M., Frönsdal, C., Lichnerowicz, A., Sternheimer, D.: Deformation Theory and Quantization. *Ann. Phys.* 111 (1978), 61-151 MR **58**:14737a
- [B-C-D-T] Breuil, C., Conrad, B., Diamond, F., Taylor, R.: On the modularity of elliptic curves over \mathbf{Q} : wild 3-adic exercises. *J. Amer. Math. Soc.* 14 (2001), no. 4, 843-939 MR **2002d**:11058
- [Ca-F] Cattaneo, A., Felder, G.: A path integral approach to the Kontsevich quantization formula, *Comm. Math. Phys.* 212 (2000), no. 3, 591-611 MR **2002b**:53141
- [Co] Connes, A.: A Short Survey of Noncommutative Geometry. *J. Math. Phys.* 41 (2000), 3832-3866 MR **2001m**:58016
- [Cr-M] Crauder, B., Miranda, R.: Quantum cohomology of rational surfaces, pp. 33-79 in *The Moduli Space of Curves*, Eds: Dijkgraaf, R., Faber, C., van der Geer, G., Birkhäuser (1995) MR **97i**:14033

- [De] Deligne, P: Déformations de l'algèbre d'une variété symplectique: comparaison entre Fedosov et De Wilde, Lecomte. (French), *Selecta Math.* 1 (1995), no. 4, 667-697 MR **97a**:58068
- [D-L] Dewilde, M., Lecomte, P.: Existence of Star-Products and of Formal Deformations of the Poisson Lie Algebra of Arbitrary Symplectic Manifolds. *Lett. Math. Phys.* 7 (1983), 487-496 MR **85j**:17021
- [Di] Dijkgraaf, R.: Mirror symmetry and elliptic curves, pp. 149-164 in *The Moduli Space of Curves*, Eds: Dijkgraaf, R., Faber, C., van der Geer, G., Birkhäuser (1995) MR **96m**:14072
- [Do] Doran, C.: Deformation Theory: An Historical Annotated Bibliography, www.math.columbia.edu/~doran/Hist%20Ann%20Bib.pdf
- [E] Elkies, N.: Rational points near curves and small nonzero $|x^3 - y^2|$ via lattice reduction, *Lecture Notes in Computer Science* 1838 (proceedings of ANTS-4, 2000; W. Bosma, ed.), Springer, Berlin, 2000, 33-63 (math.NT/0005139 on the arXiv) MR **2002g**:11035
- [E-L-S-V] Ekedahl, T., Lando, S., Shapiro, M., Vainshtein, A.: On Hurwitz numbers and Hodge integrals, *C. R. Acad. Sci. Paris Sér. I Math.* 328 (1999), no. 12, 1175-1180 MR **2001b**:14083
- [F] Fedosov, B. V.: A Simple Geometrical Construction of Deformation Quantization. *J. Diff. Geom.* 40 (1994), 213-238 MR **95h**:58062
- [G] Gibson, C. G., *Singular points of smooth mappings*, Research Notes in Mathematics, Pitman (1979) MR **80j**:58011
- [G-L-S] Greuel, G.-M., Lossen C., and Shustin, E.: Plane curves of minimal degree with prescribed singularities, *Invent. Math.* 133 (1998) 539-580 MR **99g**:14035
- [H] Harris, J.: On the Severi problem, *Invent. Math.* **84** (1986) 445-461 MR **87f**:14012
- [K-Z] Kaneko, M., Zagier, D.: A generalized Jacobi Theta Function and Quasi-modular Forms, pp. 165-172, in *The Moduli Space of Curves*, Eds: Dijkgraaf, R., Faber, C., van der Geer, G., Birkhäuser (1995) MR **96m**:11030
- [Ka] Kani, E.: The Number of Genus 2 Covers of an Elliptic Curve, preprint; see: www.ams.org/amsmtgs/210_abstracts/991-14-302.pdf
- [K-M] Kapovich, M., Millson, J. J.: On the deformation theory of representations of fundamental groups of closed hyperbolic 3-manifolds, *Topology*, Vol. 35, No. 4 (1996) 1085-1106 MR **97h**:57029
- [Ko 1] Kontsevich, M.: Enumeration of rational curves via torus actions, pp. 335-368, in *The Moduli Space of Curves*, Eds: Dijkgraaf, R., Faber, C., van der Geer, G., Birkhäuser (1995) MR **97d**:14077
- [Ko 2] Kontsevich, M.: Deformation Quantization of Poisson Manifolds, I. Preprint q-alg/9709040 (September 1997)
- [La] Lagrange, J.-L.: *Mécanique Analytique*, Gauthier-Villars, Paris (1888)
- [Lo] Loeser, F.: Déformations de courbes planes (d'après Severi et Harris), *Séminaire Bourbaki*, 1986/87, *Astérisque*, No. 152-153 (1987) **4** 187-205 MR **89d**:14038
- [M] Mazur, B.: An "infinite fern" in the deformation space of Galois representations, *Collect. Math.* **48**, 1-2 (1997) 155-193 MR **98g**:11053
- [M-F] *Modular Forms and Fermat's Last Theorem*, Eds: Cornell, G., Silverman, J., Stevens, G., Springer (1997) MR **99k**:11004
- [M-R-Y] McLellan, B., Roth, M., Yui, N.: Generating Functions Counting Simply Ramified Covers, preprint, Nov. 2003
- [P] *The Mathematical Heritage of Henri Poincaré*, Volume **39**, Part 2, Proceedings of Symposia in Pure Mathematics, American Mathematical Society (1983) MR **85h**:00010b
- [R] Riemann, B.: Theorie der Abel'schen Functionen, pp. 88-144 in *The Collected Works of Bernhard Riemann*, Ed: Weber, H., Dover (1953)
- [Ru] Rudd, R.: The string partition function for QCD on the Torus, hep-th/9407176
- [S] Severi, F.: Vorlesung über Algebraische Geometrie, Teubner (1921)
- [Q-F] *Quantum Fields and Strings: A course for mathematicians, vols. I, II*, Eds: Deligne, P. et al., A.M.S and I.A.S. (1999) MR **2000e**:81010

- [T] Thom, R.: Structural stability and morphogenesis, W. A. Benjamin, Inc., Reading, MA (1976) MR **58**:7722b
- [Z] Zariski, O.: Collected Papers, volume III: *Topology of curves and surfaces, and special topics in the theory of algebraic varieties*, Eds: Artin, M., and Mazur, B., M.I.T. Press (1978) MR **58**:21346c

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY, ONE OXFORD STREET, CAMBRIDGE,
MASSACHUSETTS 02138-2901

E-mail address: mazur@math.harvard.edu