My Introduction to Schemes and Functors

David Mumford

February 24, 2009

1 Some recollections

I want to talk about how Grothendieck's revolution profoundly affected my own understanding of algebraic geometry. But to do that, I need to reconstruct for the reader the mathematical environment in which I grew up. When I started studying algebraic geometry around 1956, the Italian school was no longer active. André Weil and Oscar Zariski were carrying the ball and had together pioneered the extension of algebraic geometry to characteristic p. Weil was motivated by the idea of creating a merger of algebraic geometry with number theory to solve, e.g. Mordell's conjecture. Zariski was motivated by the need to make the work of the Italian school rigorous by using the new methods of commutative algebra. Everyone realized that the field needed better foundations to handle these new ideas, in which the explicit geometry of complex varieties was replaced by an abstract geometry based on algebra and number theory. Both Weil and Zariski cobbled together some tentative definitions to make discussions and papers possible, both written as books for the AMS colloquium series, though only Weil's was published. But their "Foundations" did not have the feeling of inevitability that one associated, for example, to Bourbaki and his treatise and were never widely used.

Zariski was highly conscious of the fact that the Italian literature was filled with deep ideas and that it was essential to mine them, to update them using the new perspectives. He was much more open to new techniques than Weil, who radiated cynicism about anyone else's abstractions. A key breakthrough occurred when Serre and Zariski's interests met in the attempt to understand the Italian work on calculating the dimensions of complete linear systems. In modern terms, this means calculating the dimension of $\Gamma(X, L)$, the global sections of a line bundle on a variety X. The Italians had worked from curves to surfaces to three-folds and so on, at each stage comparing the linear system on a variety X with its trace on a hypersurface section H. Again, putting this in modern terms means considering the map:

$$\Gamma(X, L) \to \Gamma(H, L \otimes o_H).$$

They wanted criteria for this map to be surjective and, if not, to know the dimension of the cokernel. Zariski published a paper *Complete Linear Systems on Normal Varieties and a Generalization of a Lemma of Enriques-Severi* in the 1952 Annals in which he began to reexamine their analysis. He proved that, when X is a normal variety, the above map was surjective if the degree of H is sufficiently big.

In many ways, the cohomology of sheaves was implicit in all the calculations of this sort made by the Italian school. Cokernels, like the one above, were theclassical way of dealing H^1's. H^2's came up also, for example in the Cayley-Bacharach theorem that a plane curve C of degree $n+m-3$ which passes through all but one of the nm points $A = D \cap E$ of intersection of curves D, E of degrees n and m must pass through the last point. This is a geometric way of saying that $H^2(o_{P^2}(-3))$ is one-dimensional: tensor the exact sequence:

$$0 \to o_{\mathbb{P}^2}(-D-E) \to o_{\mathbb{P}^2}(-D) \oplus o_{\mathbb{P}^2}(-E) \to o_{\mathbb{P}^2} \to o_A \to 0$$

with $o_{\mathbb{P}^2}(n+m-3)$ and work out the usual cohomology. A careful examination of the Italian work noting, line by line, the equivalent cohomological calculations would be an interesting study, but this has never been done.

It was J.-P. Serre who made the cohomological approach to algebro-geometric questions explicit. Sheaves and cohomology were the latest hot technique in the Cartan Seminar and Serre saw that this was exactly the formalism which made sense of the Italian work. His work appeared in his famous paper *Faisceaux Algébrique Cohérents* (FAC) in the 1955 Annals and had previously been the subject of his 1954 talk at the Amsterdam International Congress. Zariski embraced these ideas instantly and talked on *Algebraic Sheaf Theory* in the 1954 AMS Summer Institute. The main theorem of Zariski's 1952 paper now became $H^1(X, L(-H)) = (0)$ if the degree of H is large enough, which Serre reproved near the end of his paper FAC.

2 Continuous systems of curves on surfaces

Much of the Italian work on linear systems fell into place and could be extended amazingly once it was recast in terms of sheaves and cohomology. As students of Zariski, many of us exploited this golden opportunity over the next decade. Zariski asked us, in particular, if we could reprove and extend to characteristic p the wonderful synthesis of the Italian work on algebraic surfaces contained in Enriques's posthumous 1949 book *Le Superficie Algebriche*. But one fundamental issue remained mysterious.

I have to digress to introduce the Italian way of describing the key invariants of surfaces. In the early days of surface theory, everyone sought the natural generalization of *genus* from curves C to surfaces F. There were two natural definitions. Since the genus of a curve was the dimension of $\Gamma(\Omega^1_C)$ the space of 1-forms with no poles, one could take on a surface the dimension of $\Gamma(\Omega_F^2)$, the space of 2-forms with no poles. This they called p_g, the *geometric genus*. But one can also take the polynomial giving $\dim \Gamma(F, o_F(nH))$ for large n and evaluate it at $n = 0$. In modern terms this is the Euler characteristic $\chi(o_F)$ and, subtracting 1, the dimension of $\Gamma(o_F)$, they got the *arithmetic genus* p_a. In cohomology terms, it equals $h^2(o_F) - h^1(o_F)$. For the simplest examples, e.g. non-singular complete intersections and rational surfaces, they found that $p_g = p_a$.

But not always! Going back again to curves of genus g, the set of 0-cycles of degree n on C breaks up into a g-dimensional family of linear systems, the family being the Jacobian. In a similar way, Picard and the Italian school considered 'complete continuous systems of curves on the surface F', e.g. the biggest family of curves on F containing the hypersurface sections H of some high degree. Then they broke this family up into its component linear systems and defined what we call the *Picard variety* as the set of these. Although they avoided using divisors with negative coefficients, in effect, they defined the Picard variety $\mathrm{Pic}(X)$ as the group of divisors mod linear equivalence and called its dimension q, the *irregularity*. And then, experimentally, they found that $q = p_g - p_a$ always seemed to hold! They conjectured this must always be true[1]

Speaking loosely, they 'knew' a version of the fact that $\Gamma(\Omega_F^2)$ and $H^2(o_F)$ had the same dimension. Their argument can be caricatured by saying that for a high degree hypersurface H:

$$\begin{aligned} H^2(o_F) &\cong H^1(o_H(H.H)) \\ &\cong H^0(\Omega_H^1(-H.H))^* && \text{via Riemann-Roch on curves} \\ &\cong H^0(\Omega_F^2)^* && \text{via residues} \end{aligned}$$

Thus $p_g = h^2(o_F)$ and $p_g - p_a = h^1(o_F)$. Therefore their conjecture was that $h^1(o_F) = \dim \mathrm{Pic}(F)$.

Not only did these numbers always seem equal, there was a direct way of showing that *ought* to show they were equal. Take a complete continuous system of curves $\{H_t\}$ of high degree on F and intersect them with one member H_0 of the family, getting divisors $H_t.H_0$ on H_0. Let t tend to 0. Then these intersections tend to a linear system of divisors on H_0, which they called the *characteristic linear system* of the family $\{H_t\}$. If this was complete, i.e. equal to the full space of sections $\Gamma(o_{H_0}(H.H))$, then the desired equality could be proven. In modern terms, this is a consequence of the exact sequence:

$$0 \to \Gamma(o_F) \to \Gamma(o_F(H_0)) \to \Gamma(o_{H_0}(H.H)) \to H^1(o_F) \to 0$$

[1] A good deal of the history can be found in some half dozen papers all in the *Comptes Rendus de l'Academie des Sciences*, volume 140, 1905. The key paper of Enriques is in the *Rendiconti dell'Accademia delle Scienze di Bologna*, volume 9, 1904.

because, if the characteristic linear system is complete, the dimension of the continuous system divided by linear equivalence will be $\dim\left(\Gamma(o_{H_0}(H.H))/\Gamma(o_F(H_0))\right)$ or $h^1(o_F)$.

Enriques and Severi argued back and forth about whether they had a proof for this, but somehow each paper purporting to have a complete proof was answered with a *dubbio critico*. In fact, the equality was established first by Hodge using the analytic tool of harmonic differential forms. But this intervention was seen as a blemish on the Italian theory and, moreover, once Weil and Zariski constructed the theory in characteristic p, the question in the characteristic p case remained open.

3 Enter Grothendieck

Just as Zariski had welcomed Serre's introduction of sheaf cohomology, he welcomed Grothendieck's new schemes. He invited Grothendieck to Harvard in 1958 and tried to set up a regular visiting appointment. He didn't exactly tear up his foundational colloquium manuscript but he was deeply impressed by Grothendieck's new way of setting up algebraic geometry via schemes. One of Zariski's deepest theorems was that the inverse image of every normal point under a proper birational morphism from one variety onto another is connected. Then Grothendieck came along and he reproved this result now by a *descending* induction on an assertion on the higher cohomology groups with Zariski's theorem resulting from the H^0 case: this seemed like black magic.

Grothendieck was a hypnotizing presence at Harvard. He seemed to have infinite energy and was always willing to schedule another lecture, to explain yet another facet of his theory. It seemed to be advancing like a tidal wave. In staid God fearing Yankee country, when no other time could be found, he created consternation by proposing to hold a seminar at 11 o'clock on Sunday. He wrote so fast and fluidly on the blackboard, I thought it resembled the 'grass writing' that I had heard about in a lecture on Chinese calligraphy: writing like the waves in the grass as a gust of wind sweeps over it. The web with which Zariski had ensnared his students was now itself ensnared in a larger, stranger one.

My involvement came about because I had been studying the construction of varieties classifying families of algebraic structures, especially moduli spaces of vector bundles and of curves. Whereas I had thought loosely of such a classifying space as having a 'natural' one-one correspondence with the set of objects in question (just as Riemann and Picard had), Grothendieck expressed it with *functors*. This was clearly the right perspective. There were 'fine' moduli spaces which carried a universal family of objects, e.g. a universal family of curves from which all other families were unique pull-backs. Therefore they represented the functor of all such families. And there were also 'coarse' moduli spaces, the best possible representable approximation to the desired functor (the approximation being caused e.g. by the fact that some curves had automorphisms).

The most beautiful part of his formulation, however, seemed to me to be his 'reification' of infinitesimal deformations. In Kodaira and Spencer's work on analytic moduli spaces, they had introduced $H^1(\Theta_X)$, Θ_X the tangent bundle to X, to describe first order deformations of a compact complex analytic manifold. But now Grothendieck was saying these first order deformations *were actual families*, families whose parameter space was the embodied tangent vector $\mathrm{Spec}(k[\epsilon]/(\epsilon^2))$. And spanning the gap between families whose parameter space was a true variety and these first order families over the dual numbers were a whole stable of families over one point bases, spectra of all possible Artin rings. Not being a number theorist, this was the real punch of schemes, the really new thing for me.

Grothendieck came back to Harvard in 1961 and he, John Tate and I ran a seminar on existence theorems and the representability of various functors, especially Picard schemes. At about this time, I believe, the more comprehensive category of stacks emerged. Because, for instance, the quotient of a non-projective variety by a finite group need not exist as a scheme, it became clear that general existence theorems could only be true in a bigger category and stacks were the natural candidate. The definitive existence theorem for them came only later when M.Artin proved his Approximation theorem. It can be found in his 1971 book *Algebraic Spaces*.

Having an existence theorem for a Picard scheme of a variety X,which represented the functor of all families of line bundles, instantly solved the above completeness problem, the main conjecture of Italian algebraic geometry. It is immediate that the tangent space to the Picard scheme at the identity is $H^1(o_X)$ because, by the functorial definition of Pic, its tangent space gives the space of all line bundles over $\mathrm{Spec}(k[\epsilon]/(\epsilon^2))$ (or of infinitesimal deformations of any sufficiently ample divisor mod linearly equivalent deformations). So the key equality $q = h^1(o_X)$ is simply the statement that the Picard scheme is reduced. In characteristic zero, the exponential map shows immediately that all group schemes are reduced – consider the restriction maps:

$$\mathrm{Pic}(X \times \mathrm{Spec}(k[t]/(t^{n+1}))) \to \mathrm{Pic}(X \times \mathrm{Spec}(k[t]/(t^2))) \to \mathrm{Pic}(X).$$

Then, in characteristic zero, the exponential map defines the lifting

$$\{1 + ta_{ij}\} \mapsto \{1 + \cdots + t^n a_{ij}^n/n!\}$$

of the kernel of the right hand arrow to the big Pic on the left[2]. Thus Grothendieck's idea of *representing* the functor of families of line bundles over *all schemes* immediately gives a purely algebraic proof that $h^1(o_X) = q$.

[2]Recently Professor Donald Babbitt called my attention to Enriques's 1938 paper *Sulla proprietà caratteristica delle superficie algebriche irregolari* in the *Rendiconti della Accademia dei Lincei*, volume 27, pp.493-498. Although Enriques's 1905 paper on the completeness theorem missed the key issue, this paper does have the right idea. He speaks of the exponential map in the Picard variety and asserts that analogously higher order infinitely near curves

This seems to me the example par excellence of Grothendieck's basic philosophy – that if you analyze a question down to its simplest and most abstract components, answers to the most puzzling questions should fall out. Even nicer, the theorem turned out to be false in characteristic p and necessary and sufficient conditions for its truth can be given by asking that certain 'Bockstein operators' from $H^1(o_X)$ to $H^2(o_X)$ must be zero (see my 1966 book *Lectures on Curves on an Algebraic Surface* where I discuss many of Grothendieck's existence theorems in most constructive possible fashion). It shows the power of nilpotent schemes and the functorial point of view in the clearest possible light. Grothendieck, of course, went on to construct and prove many much sexier things for which he is better know. But to demonstrate the power of modern abstract ideas to solve older very concrete problems, I think that this example is unmatched.

can be generated from first order infinitely near ones. Unfortunately, he possesses no tools whatsoever for going beyond an intuitive description of why the method of higher order infinitesimals should work: the theory of schemes was clearly what he lacked.