WHY
CHARACTERISTIC p
IS
BETTER

Mel Hochster

INTRODUCTION

All given rings are commutative with 1. By and large, we restrict attention to rings that are generated by finitely many elements either over an algebraically closed field K or over the integers. Such rings may be thought of as having the form R/I where R is a polynomial ring in finitely many variables $x_1, \ldots, x_n$ over a field K or over $\mathbb{Z}$.

In particular, the rings we talk about are almost always NOETHERIAN, i.e., every ideal is finitely generated.

We shall eventually discuss the three problems mentioned in the abstract:

(1) What can one say about fixed rings (rings of invariants) of a linearly reductive algebraic group acting on a polynomial ring? (More about the notions involved later.)

(2) In a polynomial ring in two variables over a field, is it true that for any three elements f, g, h, one has that the product of their squares is in the ideal generated by their cubes, i.e., is $f^2g^2h^2 \in (f^3, g^3, h^3)$?

(3) For a polynomial ring R over a field K or $\mathbb{Z}$, if $S \supseteq R$ is a ring that is a finitely generated R-module (i.e., MODULE-FINITE over R) is it true that R is a direct summand of S (over R)?

(S is module-finite over R if there are finitely many elements $s_1, \ldots, s_t$ of S such that every element of S can be written as $r_1s_1 + \cdots + r_ts_t$ with the r_i in R. Note that such a representation is usually not unique.)

But before discussing these problems and the relevance of characteristic p, I want to recall a little bit of geometry.

ALGEBRA IS
GEOMETRY

Let K be an algebraically closed field. Then $\mathbb{A}^n_K$ is just another name for K^n, but thought of as an algebraic set or algebraic variety rather than as a vector space.

More precisely, a (closed) algebraic set X in $\mathbb{A}^n_K$ is just the set of solutions of some polynomial equations in n variables $x_1, \ldots, x_n$ over K: we write $V(f_1, \ldots, f_m)$ for the (simultaneous) solutions of the m equations

$f_i(x_1, \ldots, x_n) = 0,$

$1 \leq i \leq m.$

If an algebraic set is the union of two proper algebraic sets it is called REDUCIBLE; otherwise, it is IRREDUCIBLE. Every algebraic set is a finite irredundant union of IRREDUCIBLE algebraic sets in a unique way. Irreducible algebraic sets are called ALGEBRAIC VARIETIES.

Example 1:

$V(x_1x_2) = V(x_1) \cap V(x_2).$

Example 2:

Consider a 2 by 3 matrix of indeterminates (x_{ij}) and let Δ_j be the minor obtained by deleting the j th column. Thus, $\Delta_1 = x_{12}x_{23} - x_{21}x_{13}$, etc. Then:

$$V(\Delta_2, \Delta_3) = V(x_{11}, x_{12}) \cup V(\Delta_1, \Delta_2, \Delta_3)$$

Morphisms $X \to Y$ of algebraic sets (over K) are given by maps that can be expressed, coordinatewise, by polynomials when X and Y are thought of as embedded in, say, K^n and K^m respectively. X has a (ZARISKI) TOPOLOGY: the closed sets are precisely the subsets of X that are algebraic sets.

If one has an algebraic variety X, one can think about the ring obtained by restricting all the polynomial functions on $\mathbb{A}^n_K$ to X. This ring, denoted $K[X]$, is called the COORDINATE ring of X. Note the following:

(1) The coordinate ring of $\mathbb{A}^n_K$ is the polynomial ring $K[x_1, \ldots, x_n]$.

(2) Points of X correspond one-to-one with maximal ideals of $K[X]$. ($x \in X$ corresponds to $m_x = \{f \in K[X] : f(x) = 0\}$.

(3) X is a variety (irreducible) if and only if $K[X]$ is an integral domain.

Philosophically, one may want to think of any commutative ring as a ring of functions on a geometric object.

DIMENSION

Dimension (Krull dimension) can be defined in an arbitrary commutative ring with identity so that when $K = \mathbb{C}$ the dimension of $\mathbb{C}[X]$ is the same as the complex dimension of X, i.e., half the real dimension. One uses the supremum of lengths of chains of prime ideals. But since this may seem rather artificial on first glance, we simply note the following properties of dimension for algebraic sets:

(1) The dimension of $\mathbb{A}^n_K$ is n.

(2) The dimension of $X \cup Y$ is the supremum of $\dim X$ and $\dim Y$.

(3) If X is isomorphic to a Zariski open set in Y, a variety, then $\dim X = \dim Y$.

(4) If $X \to Y$ is a surjective map of varieties which has finite fibers, then $\dim X = \dim Y$.

INTERSECTIONS AND INTERSECTION MULTIPLICITIES

Two planes in three space may be parallel, but if they intersect at all, they must intersect in at least a line. More generally, vector subspaces V and W of K^n (they intersect, since 0 is a common point) must have an intersection of dimension at least $\dim V + \dim W - n$.

Less well known is that the same is true for varieties in $\mathbb{A}^n_K$ over an algebraically closed field K!

Now suppose that we have varieties V and W in $\mathbb{A}^n_K$ with the origin as an isolated point of intersection, that $\dim V = r$ and that $\dim W = s$ with $r + s = n$. For simplicity we consider the case where W is a vector space defined by the vanishing of r linear forms, i.e., $W = V(L_1, \ldots, L_r)$. If $K = \mathbb{C}$ we can define the intersection multiplicity of V and W at the origin as the number of points, near 0 in the usual (Hausdorff) topology on $\mathbb{C}^n \approx \mathbb{R}^{2n}$ in the intersection of V with a typical linear space $W_\epsilon = V(L_1 - \epsilon_1, \ldots, L_r - \epsilon_r)$ obtained by "perturbing" W slightly by the choice of $\epsilon = (\epsilon_1, \ldots, \epsilon_r)$. The answer should be the same for all choices of $\epsilon_1, \ldots, \epsilon_r$ close to 0 and off some proper closed set.

Example 3:

In $\mathbb{A}^2_{\mathbb{C}}$ with coordinate variables x, y the intersection multiplicity of $V(y)$ and $V(y-x^2)$ at the origin is two, because $V(y-\epsilon)$ and $V(y-x^2)$ have two points of intersection near the origin, $(\pm\sqrt{\epsilon}, \epsilon)$ for all small $\epsilon \neq 0$.

One can often compute intersection multiplicities more simply. First, the LOCAL RING of a variety at a point is the ring obtained by adjoining inverses for all functions in the ring that do not vanish at that point. Each element of the local ring does define a function on a (Zariski) open neighborhood of the point, and this is the analogue of the rings of germs of continuous or C^∞ or holomorphic functions at a point defined in various other kinds of geometry.

(More generally, a ring is called a LOCAL RING if it has a unique maximal ideal, and one can form the local ring R_P of R at a maximal or even prime ideal P by adjoining inverses for all elements not in P. In the local rings above, the maximal ideal consists of functions vanishing at the point under consideration.)

Then the intersection multiplicity of V and W can sometimes be computed by simply killing the equations of V and W in the local ring of $\mathbb{A}^n_K$ at the origin, and then taking the vector space dimension of the quotient over K.

Example 3 revisited:

When we kill the equations in Example 3 we get $\mathbb{C}[x,\ y]/(y-x^2, y) \cong \mathbb{C}[x]/(x^2) = \mathbb{C}+\mathbb{C}\overline{x}$, which is two dimensional.

BUT WHEN DOES THIS COMPUTATIONALLY SIMPLE PROCEDURE WORK?

(It is not necessary to invert elements that do not vanish at the origin in this particular example, because they become invertible in the quotient – this corresponds to the fact that the origin is the ONLY point of intersection of the two curves, not just an isolated point of intersection.)

GEOMETRY IS ALGEBRA: COHEN-MACAULAY RINGS

A Noetherian local ring R is called COHEN-MACAULAY if there is a sequence of elements $x_1, \ldots, x_d$ in the maximal ideal of R such that

(1) Every element of the sequence is a nonzerodivisor modulo its predecessors (we refer to elements generating a proper ideal and such that this condition holds as a REGULAR SEQUENCE).

(2) Every element of the maximal ideal has a power in $(x_1, \ldots, x_d)$.

In such a sequence one always has $d = \dim R$. When a local ring has dimension d, there are always sequences of elements $x_1, \ldots, x_d$ of the maximal ideal such that every element of the maximal ideal is nilpotent modulo the ideal $(x_1, \ldots, x_d)$. Such a sequence is called a SYSTEM OF PARAMETERS (s.o.p). One may also say:

A Noetherian local ring is Cohen-Macaulay if and only if some (equivalently, every) system of parameters is a regular sequence.

A Noetherian ring is COHEN-MACAULAY if all its local rings are. There are lots of other characterizations. In the $\mathbb{N}$-graded case, R is Cohen-Macaulay if and only if it is a finitely generated FREE module over a polynomial subring $K[u_1, \ldots, u_d]$ generated by homogeneous elements u_i (which can be taken of equal degree).

Example 4:

$K[x^2, xy, y^2] \subseteq K[x, y]$ IS Cohen-Macaulay: it's free over $K[x^2, y^2]$ (basis 1, xy).

$K[x^2, x^3, xy, y] \subseteq K[x, y]$ is NOT Cohen-Macaulay. E.g., it has generators 1, x^3, xy over $K[x^2, y]$, but these have the relation $y(x^3) = x^2(xy)$. (At (0,0), x^2, y is a s.o.p. but not a regular sequence.)

But one of the most geometrically significant facts about Cohen-Macaulay rings is the following:

At an isolated point x of intersection of two COHEN-MACAULAY varieties V, W in $\mathbb{A}^n_K$, if $\dim V + \dim W = n$, then one can find the intersection multiplicity of V and W at x by simply killing the defining ideals of both V and W (i.e., all functions that vanish on one or the other) in the local ring of $\mathbb{A}^n_K$ at x, and then the intersection multiplicity is simply the dimension of the quotient as a vector space over K.

This characterizes Cohen-Macaulay rings!

RINGS OF INVARIANTS

An ALGEBRAIC GROUP (Zariski closed subgroup of $GL(n, K)$) is called LINEARLY REDUCTIVE if every representation is completely reducible. In characteristic zero, these

include finite groups, products of $GL(1,K)$ (algebraic tori), and semi-simple groups. Over $\mathbb{C}$ such a group is the complexification of compact real Lie group. A key point is that when a linearly reductive algebraic group acts on a K-algebra R, if R^G is the ring of invariants or fixed ring $\{r \in R : g(r) = r \text{ for all } g \in G\}$ there is a canonical retraction map map $R \to R^G$, called the REYNOLDS OPERATOR, that is R^G-linear.

Example 5:

Let $G = GL(1,K)$, act on $R = K[x_1, x_2, y_1, y_2]$ such that if $a \in G$ then a sends each x_i to $x_i a^{-1}$ and each y_j to ay_j. $R^G = K[x_1y_1, x_1y_2, x_2y_1, x_2y_2] \cong K[u_{11}, u_{12}, u_{21}, u_{22}]/(\Delta)$ where $\Delta = \det(u_{ij})$. Note that R^G is not a UFD because of the relation $u_{11}u_{22} = u_{12}u_{21}$.

Example 6:

Let $G = GL(1,K)$, act on $R = K[x_1, \ldots, x_r, y_1, \ldots, y_s]$ such that if $a \in G$ then a sends each x_i to $x_i a^{-1}$ and each y_j to ay_j. $R^G = K[x_iy_j : i,j] \cong K[u_{ij}]/I_2(u_{ij})$ where (u_{ij}) is an $r \times s$ matrix of indeterminates and the ideal killed is generated by all 2×2 minors of (u_{ij}).

Example 7:

Let $G = GL(t,K)$, let X be an $r \times t$, Y a $t \times s$ and U an $r \times s$ matrix of indeterminates and let G act on $R = K[x_{ij}, y_{hk} : i,j,h,k]$ such that $A \in G$ sends the entries of X to those of XA^{-1} and the entries of Y to those of AY. Then $R^G = K[XY]$, the ring generated by the entries of XY over K, and $R^G \cong K[u_ij : i,j]/I_{t+1}(U)$, where the ideal killed is generated by the size $t+1$ minors of the matrix U. If $t = 1$ this is the previous example.

Example 8:

Let $G = SL(r,K)$, let X be an $r \times n$ matrix of indeterminates and let G act on $R = K[x_{ij} : i,j]$ such that $A \in G$ sends the entries of X to those of AX. Then R^G is the ring generated over K by the $r \times r$ minors of X over K, the homogeneous coordinate ring of a Grassmann variety. In general, there are certain quadratic relations on these minors.

What subtle property do all these rings have in common?

These rings of invariants are all Cohen-Macaulay!

There are few linearly reductive groups in char. $p > 0$, so that in a sense this is mainly a char. 0 theorem. The first proof, in [M. Hochster and J. L. Roberts, *Rings of invariants of reductive groups acting on regular rings are Cohen-Macaulay*, Advances in Math.

13 (1974), 115–175], involved a convoluted reduction to char. $p > 0$! Later, Boutot, in [J.-F. Boutot, *Singularités rationelles et quotients par les groupes réductifs*, Invent. Math. **88** (1987), 65–68] gave a proof using resolution of singularities and the Grauert-Riemenschneider vanishing theorem. But the development of tight closure theory, which systematically exploits char. p methods and obtains results in char. 0 by reduction to char. p, has provided the simplest proof.

PROPERTIES OF TIGHT CLOSURE

TIGHT CLOSURE is an operation on ideals of Noetherian rings containing a field. (It is also defined on submodules of modules but we largely ignore this here.) It is defined first in char. $p > 0$, then for finitely generated algebras over a field of char. 0 by REDUCTION TO CHARACTERISTIC p and finally for all Noetherian rings containing a field.

(The theory of tight closure was first developed explicitly by M. Hochster and C. Huneke, although it rests on ideas implicit in earlier work of C. Peskine, L. Szpiro, and P. Roberts, as well as earlier work of Hochster and Huneke. The first main paper developing the theory is [M. Hochster and C. Huneke, *Tight closure, invariant theory, and the Briançon-Skoda theorem*, J. Amer. Math. Soc. **3** (1990), 31–116.]

We skip the definition for the moment and focus on some important properties (some of these are valid in complete generality, while all are valid in considerable generality — certainly, for domains finitely generated over a field). We first recall that a Noetherian ring is REGULAR if its local rings are, and a local ring R is regular if its maximal is generated by dim R elements. When R contains a field this means that $\widehat{R}$ is isomorphic with a formal power series ring over a field. The local rings of a variety over $\mathbb{C}$ are regular iff it is a smooth analytic manifold. In particular, polynomial rings over a field (or the integers) are regular.

Let R, S be rings, $R \to S$ a homomorphism, and I, J ideals of R. We denote by I^* the tight closure of I.

(1) $I \subseteq I^* = (I^*)^*$ and if $I \subseteq J$ then $I^* \subseteq J^*$.

(2) $I^*S \subseteq (IS)^*$ (PERSISTENCE of tight closure).

(3) If R is regular, every ideal is tightly closed.

(4) If $R \subseteq S$ is a module-finite extension, $IS \cap R \subseteq I^*$.

(5) If R is a local domain and $x_1, \ldots, x_d$ is part of a system of parameters, then for each t the annihilator of x_{t+1} modulo $(x_1, \ldots, x_t)$ is contained in $(x_1, \ldots, x_t)^*$.

These basic properties lead to a host of theorems. First: a ring is called WEAKLY F-REGULAR ("F" stands for Frobenius here) if every ideal is tightly closed, and F-REGULAR if this holds for localizations as well. (3) implies that regular rings are F-regular, and (2) implies that direct summands of F-regular rings are F-regular. But then (5) implies that (weakly) F-regular rings are Cohen-Macaulay. It follows that direct summands of regular rings containing a field are Cohen-Macaulay! (This is an open question in general.) This proves that rings of invariants of reductive groups acting on polynomial rings are Cohen-Macaulay!

An element $x \in R$ is in the INTEGRAL CLOSURE of $I \subseteq R$ if for every homomorphism $h : R \to V$, where V is regular local of dimension at most 1 (a so-called DISCRETE VALUATION RING or DVR), $x \in IV$. It is immediate from (2) and (3) that the tight closure of an ideal is in the integral closure — the tight closure is usually much smaller. For example, in $K[x, y]$ every ideal is tightly closed but xy is in the integral closure of (x^2, y^2). (A necessary and sufficient condition for x to be in the integral closure of I in a Noetherian domain R is that there exist $c \neq 0$ such that $cx^n \in I^n$ for arbitrarily large values of n. In particular, if $x^k \in I^k$ then x is in the integral closure of I, since $1 \cdot x^{kh} \in I^{kh}$ for all $h \geq 1$.)

THE BRIANÇON-SKODA THEOREM

A remarkable theorem proved originally by analytic methods by Briançon and Skoda in characteristic 0 and by Lipman and Sathaye in general using ideas from duality theory asserts that in a regular ring of Krull dimension n, the integral closure of the nth power of an ideal generated by n elements is contained in the original ideal. Moreover, in a local ring of Krull dimension n, every ideal is contained in the integral closure of an ideal with at most n generators (assuming the residue field is infinite). Thus, in a regular ring of dimension n, the integral closure of the nth power of any ideal is contained in the ideal, since the issue is local. Tight closure yields one of the easiest proofs of this, and provides a generalization when the ring contains a field: one does not need the ring to be regular, but the conclusion is that the integral closure of the nth power of an n generator ideal is contained in the tight closure of the ideal (hence, in the ideal if the ring is regular).

Now, fgh is integral over (f^3, g^3, h^3) in any ring, and so its square is in the ideal if the ring is regular of dimension 2. I would be very curious to see an elementary proof of this for the ring $K[x, y]$. This answers the question raised in (2) of the abstract (and in the Introduction).

THE DEFINITION OF TIGHT CLOSURE

We shall stick with domains. Let R be a Noetherian domain of characteristic p. Then x is in the tight closure of the ideal $I = (u_1, \ldots, u_s)$ if there exists $c \in R\backslash\{0\}$ such that $cx^{p^e} \in (u_1^{p^e}, \ldots, u_s^{p^e})R$ for all $e \gg 0$. If we take p^e th roots this says that $c^{1/p^e} \in IR^{1/p^e}$ for all $e \gg 0$. In some "formal" sense one can think of c^{1/p^e} as approaching 1 as $e \to \infty$, which suggests why one should think of an element x that satisfies this condition as being "nearly" in I. The magic that underlies this definition is that in characteristic p one has that the map sending r to r^p is a ring homomorphism: the Frobenius endomorphism. The key point:

$$(x+y)^p = x^p + y^p$$

in rings of char. p.

The definition of tight closure is not easily swallowed in one gulp!

Example 9:

If K is any field of char. $p > 0$, $p \neq 3$, then in $R = K[X,Y,Z]/(X^3+Y^3+Z^3) = K[x,y,z]$ one has that $z^2 \in (x,y)^*$: in fact for any element $c \in (x,y,z)$ one has that $c(z^2)^{p^e} \in (x^{p^e}, y^{p^e})$ for all $e \geq 0$. This gives a hint of the char. 0 notion of tight closure: $z^2 \in (x,y)^*$ in this ring when K has char. 0 because the ring $\mathbb{Z}[x,y,z] \subseteq K[x,y,z]$ and z^2 is in the char. p tight closure of (x,y) when one works modulo m for m in a dense open subset of the maximal ideals of $\mathbb{Z}$ (i.e., modulo all sufficiently large prime integers.)

Another immediate application of the properties listed for tight closure is the following: recall that if $R \subseteq S$ is module-finite and I is an ideal of R, then $IS \cap R \subseteq I^*$. This implies that if R is weakly F-regular, then every ideal of R is contracted from S, and this implies that R is a direct summand of S under very mild hypotheses. In particular, regular rings containing a field are direct summands of their module-finite extensions, which is far from obvious in characteristic p. (In char. 0 a trace argument shows that rings that are integrally closed are direct summands of every module-finite extension.)

Oddly, the question of whether regular rings are direct summands is related to intersection theory: in the case of rings containing a field, it is equivalent to the following question (and one can frame a similar question for the case where the ring does not contain a field as well). Fix positive integers n, t, and r. Let $x = x_1, \ldots, x_n$, $y = y_1, \ldots, y_n$ and $z = z_1, \ldots, z_r$ be indeterminates. Let X be the variety corresponding to the ring $K[x,y,z]/(f)$ where $f = x_1^t \cdots x_n^t - \sum_{i=1}^n y_i x_i^{t+1}$, let $Y = V(x_1, \ldots, x_n) \subseteq X$, and suppose

that $Z \subseteq X$ is a variety meeting Y only at the origin. Is $\dim Y + \dim Z \leq \dim X$? This would follow at once if the origin, P, were a smooth point of X, which it is not — this is like a result mentioned earlier for varieties in $\mathbb{A}^n_K$. Results like this don't hold in general when X is not smooth, and it appears to be very hard to say when they do hold for a particular $P \in Y \subseteq X$, when Z is allowed to vary.

TIGHT CLOSURE AND BIG COHEN-MACAULAY ALGEBRAS

Let R be a local domain of char. $p > 0$ that is integrally closed in its field of fractions. Under very mild conditions (excellence) it is known that $x \in I^*$ in R if and only if R has a big Cohen-Macaulay algebra S such that $x \in IS$ (this is in [M. Hochster, *Solid Closure*, in Contemp. Math. **159** (1994) 103–172.]). It is closely related to a result of Huneke's and mine that under mild conditions on R as above the integral closure R^+ of R in an algebraic closure of its fraction field is a big Cohen-Macaulay algebra for R. This is false in char. 0 ! Characteristic p is better!

It is an open question whether $I^* = IR^+ \cap R$ in general: Karen Smith has shown this for parameter ideals.

QUESTIONS

(1) Can tight closure theory be extended to rings that do not contain a field, e.g., to finitely generated $\mathbb{Z}$-algebras?

(2) Is $\mathbb{Z}[x, y]$ a direct summand of all of its module-finite extensions?

(3) Does tight closure commute with localization?

OTHER APPLICATIONS

Some other applications of tight closure and areas where tight closure has made an appearance that I won't have time to talk about:

Local homological questions
Phantom homology
Macaulayfication

Uniform Artin-Rees theorems
Results related to the Kodaira vanishing theorem
Further connections with singularity theory
Rings of differential operators
More on big Cohen-Macaulay modules and algebras