Algebra & Number Theory

Volume 3
2009
No. 8

mathematical sciences publishers

Algebra & Number Theory

www.jant.org

See inside back cover or www.jant.org for submission instructions.

The subscription price for 2009 is US $140/year for the electronic version, and $200/year (+$35 shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA.

Algebra & Number Theory (ISSN 1937-0652) at Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

ANT peer review and production are managed by EditFLOW™ from Mathematical Sciences Publishers.

PUBLISHED BY
mathematical sciences publishers
http://www.mathscipub.org
A NON-PROFIT CORPORATION
Typeset in LaTeX

On coproducts in varieties, quasivarieties and prevarieties

George M. Bergman

If the free algebra F on one generator in a variety $\mathbf{V}$ of algebras (in the sense of universal algebra) has a subalgebra free on two generators, must it also have a subalgebra free on three generators? In general, no; but yes if F generates the variety $\mathbf{V}$.

Generalizing the argument, it is shown that if we are given an algebra and subalgebras, $A_0 \supseteq \cdots \supseteq A_n$, in a prevariety ($\mathbb{SP}$-closed class of algebras) $\mathbf{P}$ such that A_n generates $\mathbf{P}$, and also subalgebras $B_i \subseteq A_{i-1}$ $(0 < i \leq n)$ such that for each $i > 0$ the subalgebra of A_{i-1} generated by A_i and B_i is their coproduct in $\mathbf{P}$, then the subalgebra of A generated by $B_1, \ldots, B_n$ is the coproduct in $\mathbf{P}$ of these algebras.

Some further results on coproducts are noted:

If $\mathbf{P}$ satisfies the amalgamation property, then one has the stronger "transitivity" statement, that if A has a finite family of subalgebras $(B_i)_{i \in I}$ such that the subalgebra of A generated by the B_i is their coproduct, and each B_i has a finite family of subalgebras $(C_{ij})_{j \in J_i}$ with the same property, then the subalgebra of A generated by all the C_{ij} is their coproduct.

For $\mathbf{P}$ a residually small prevariety or an arbitrary quasivariety, relationships are proved between the least number of algebras needed to generate $\mathbf{P}$ as a prevariety or quasivariety, and behavior of the coproduct operation in $\mathbf{P}$.

It is shown by example that for B a subgroup of the group $S = \mathrm{Sym}(\Omega)$ of all permutations of an infinite set Ω, the group S need not have a subgroup isomorphic over B to the coproduct with amalgamation $S \amalg_B S$. But under various additional hypotheses on B, the question remains open.

MSC2000: primary 08B25, 08B26, 08C15; secondary 03C05, 08A60, 08B20, 20M30.

Keywords: coproduct of algebras in a variety or quasivariety or prevariety, free algebra on n generators containing a subalgebra free on more than n generators, amalgamation property, number of algebras needed to generate a quasivariety or prevariety, symmetric group on an infinite set.

arXiv reference: 0806.1750.

After publication of this article, updates, errata, related references etc., if found, will be recorded at http://math.berkeley.edu/~gbergman/papers.

1. Prologue, for the nonexpert

It is well known that the free group on two generators contains a subgroup free on three generators. Can one deduce, from this alone, that it contains a subgroup free on four generators?

It is unfair to say "from this alone" without indicating what facts about groups are to be taken for granted. So suppose we want to use only the fact that groups form a *variety* of algebras in the sense of universal algebra — a class of structures consisting of all sets with a family of operations of specified arities, satisfying a specified list of identities. Then we can ask, for $\mathbf{V}$ any variety and n any positive integer:

(1) If, in $\mathbf{V}$, the free algebra on n generators has a subalgebra free on $n+1$ generators, must it have a subalgebra free on $n+2$ generators?

Our first result will be a negative answer to this question, in the most extreme case, $n = 1$.

On the other hand, a fact that is second nature to combinatorial group theorists is that if G_1 and G_2 are overgroups of a common group H, and one forms $G_1 \amalg_H G_2$, their coproduct with amalgamation of H (in group theorists' notation and language $G_1 *_H G_2$, their free product with amalgamation of H), then the canonical maps of G_1 and G_2 into that coproduct are embeddings. This says that the variety of all groups has "the amalgamation property"; and we shall see in Section 6 that if a variety $\mathbf{V}$ has this property, then it also has the property that for any algebras A_1, A_2 in $\mathbf{V}$ and subalgebras $B_1 \subseteq A_1$, $B_2 \subseteq A_1$, the coproduct $A_1 \amalg A_2$ contains the coproduct $B_1 \amalg B_2$. From this it is not hard to show that for any such $\mathbf{V}$, (1) has an affirmative answer.

However, the amalgamation property is relatively rare. For instance, though it is satisfied by the variety of all groups, and by all varieties of abelian groups, it does not seem to be satisfied by most other varieties of groups — in fact, it is a longstanding open question whether it is satisfied by any variety of groups other than those just mentioned [Neumann 1967, Problem 6; Kovács and Newman 1974, page 422].

But in Section 2, after finding our counterexample to (1), we shall see that a different condition, more common than the amalgamation property, implies a positive answer to (1); namely, that the free algebra of rank n in $\mathbf{V}$ *generate* $\mathbf{V}$, that is, not lie in any proper subvariety thereof. (As, for example, the free group of rank 2 generates the variety of all groups.)

In Sections 3 and 4 we shall generalize this to a result about when an algebra (not necessarily free) containing a coproduct of subalgebras, some of which in turn contain coproducts of subalgebras, will itself contain an "obvious" iterated coproduct. The condition that a certain one of our algebras generate the class we are working in will again be a key assumption; not, this time, for generation as a variety, but as a *prevariety,* which means, roughly, a class of algebras determined by identities and universal *implications*. (For example, the class of torsion-free groups, that is, groups satisfying $(\forall x)\ x^n = 1 \Rightarrow x = 1$ for each $n > 0$, is a prevariety.) The definition of prevariety, and of the related concept of *quasivariety*, are recalled in Section 3.

We end with some further results on quasivarieties and prevarieties, and a brief section on subgroups of infinite symmetric groups.

Acknowledgments and Reader's Advisory. I am indebted to the referee for several helpful suggestions, and to the editorial staff of the journal for requesting that I write this introduction for the general reader.

Carrying that suggestion further, I have added, as Section 11, a quick summary of some common terminology which should make this note readable (if not light reading) by anyone for whom this prologue was. Readers not familiar with the basic language of universal algebra might start with that section.

2. Free subalgebras of free algebras

The original question that led to this investigation [Bergman 2007, Question 4.5] was whether an algebra A in a variety $\mathbf{V}$ which contains a subalgebra isomorphic to the coproduct in $\mathbf{V}$ of two copies of itself, $A \amalg_{\mathbf{V}} A$, must also contain a copy of the three-fold coproduct $A \amalg_{\mathbf{V}} A \amalg_{\mathbf{V}} A$. As indicated above, this can fail even for A free of rank 1: a free algebra of rank 1 in a variety $\mathbf{V}$ may have a subalgebra free of rank 2 but fail to have any subalgebra free of rank 3. Let us begin by examining how we might concoct such an example.

To do so, we must "foil" the obvious ways one would expect a free three-generator subalgebra to arise. If $\langle x \rangle$ is free on x and contains a subalgebra $\langle y, z \rangle$ free on y and z, then $y = px$ and $z = qx$ for some derived unary operations p, q of $\mathbf{V}$. Since $\langle qx \rangle$ is isomorphic to $\langle x \rangle$, its subalgebra corresponding to $\langle y, z \rangle$, namely $\langle pqx, qqx \rangle$, will be free on those two generators, and one might expect $\langle px, pqx, qqx \rangle$ to be free on the three indicated generators. (If it seems to the reader that it *must* be free on those elements, he or she may be implicitly assuming that $\mathbf{V}$ has the amalgamation property, to be discussed in Section 6.)

For this to fail, there must be some ternary relation T in the operations of $\mathbf{V}$ such that $T(px, pqx, qqx)$ is an identity in one variable x, but $T(u, v, w)$ is not satisfied by all 3-tuples of elements of algebras in $\mathbf{V}$. On the other hand, since $y = px$ and $z = qx$ generate a free algebra, the relation $T(px, pqx, qqx)$ implies that $T(y, pz, qz)$ *is* an identity in two variables y and z in $\mathbf{V}$.

Let us pause to note that if we construct a variety with such an identity T, we will have eliminated one possibility for a free three-generator subalgebra of $\langle x \rangle$ of rank 3; but every 3-tuple of expressions obtained from y and z using the operations of $\mathbf{V}$ represents another potential generating set for a free subalgebra. In principle, we might use different relations to exclude different 3-tuples; but let us see whether we can make do with just one such relation T, such that $T(u, v, w)$ holds for all 3-tuples (u, v, w) of elements of $\langle x \rangle$. Note that in this case, since $\langle x \rangle$ contains a free algebra of rank two, $\mathbf{V}$ must satisfy identities saying that $T(u, v, w)$ holds for any elements u, v, w of any $\mathbf{V}$-algebra that lie in a common two-generator subalgebra.

In testing out this approach, let us temporarily allow structures involving primitive *relations* as well as operations. Then we could let $\mathbf{V}$ be the class of objects defined by two primitive unary operations, p and q, and one primitive ternary relation, T, subject only to the countable family of "identities"

(2) $\qquad T(a(y, z), b(y, z), c(y, z)),$

one for each 3-tuple of words a, b, c in two variables y, z and the operations p, q. (Of course, since p and q are unary, each of a, b, c really just involves one of y or z.) In an object of $\mathbf{V}$ generated by ≤ 2 elements, T thus holds identically, so in describing the structures of ≤ 2-generator objects, we can ignore the relation T, and simply specify the actions of p and q. Since the family of identities (2) by which we have defined $\mathbf{V}$ includes no identities in the operations p and q alone, the possible structures of such objects are simply the structures of M-set, for M the free monoid on generators p and q. In this monoid M, the left ideal generated by p and q is free on those two elements; hence in the free $\mathbf{V}$-object on one generator x, the elements px and qx satisfy no relations in p and q; so with T, as noted, also contributing no information, px and qx indeed generate a free subobject. On the other hand, if we take the free M-set on three generators x, y, z, and define T to hold precisely on those 3-tuples thereof in which all three components lie in a subalgebra generated by two elements, we see that this satisfies the definition of a free $\mathbf{V}$-object on three generators, and that $T(x, y, z)$ does *not* hold. Hence the free object on one generator does not contain a copy of the free object on three generators.

Let us now try to mimic the above behavior in a variety of genuine algebras. In addition to two unary operations p and q, let us introduce a 0-ary operation 0 and a ternary operation t, with the idea that the relation $T(u, v, w)$ will be the condition $t(u, v, w) = 0$. To keep our new operations from complicating our structures more than necessary, let us introduce some "nonproliferation" identities:

(3) $\qquad p0 = q0 = pt(x, y, z) = qt(x, y, z) = 0.$

(4) $\qquad t(u, v, w) = 0$ whenever any of u, v, w is either 0, or is itself of the form $t(u', v', w')$.

Finally, we impose the identities corresponding to (2):

(5) $t(a(x,y), b(x,y), c(x,y)) = 0$ for all derived operations a, b, c in two variables.

In a free $\mathbf{V}$-algebra, the elements $t(u,v,w)$ that are not 0 may be thought of as "tags", showing that certain 3-tuples (u,v,w) obtained from the generators using p and q alone do not have the form indicated in (5). By (3) and (4), these elements have essentially no other effect. By the same reasoning as for structures with a primitive relation T, we get:

Proposition 1. *Let $\mathbf{V}$ be the variety defined by a 0-ary operation 0, two unary operations p and q, and a ternary operation t, subject to identities (3), (4), (5).*

Then in the free $\mathbf{V}$-algebra $F_{\mathbf{V}}(x)$ on one generator x, the subalgebra generated by px and qx is free on those generators; but $F_{\mathbf{V}}(x)$ (and hence also the free algebra on two generators) has no subalgebra free on three or more generators. □

The above result was based on $F_{\mathbf{V}}(x)$ satisfying an identity (namely, $t(x,y,z) = 0$) that did not hold in all of $\mathbf{V}$; and we might hope that if $\mathbf{V}$ is a variety where this does not happen, but which, as above, has unary derived operations p and q such that px and qx are free generators of the subalgebra $\langle px, qx\rangle \subseteq F_{\mathbf{V}}(x)$, then the subalgebra $\langle px, pqx, qqx\rangle$ will have to be free on the indicated three generators. To investigate this question, consider a ternary relation T in the operations of $\mathbf{V}$ about which we now merely assume that $T(px, pqx, qqx)$ holds in $F_{\mathbf{V}}(x)$, and let us see whether we can deduce that T holds for all 3-tuples of elements of $F_{\mathbf{V}}(x)$.

As noted earlier, the conditions that $T(px, pqx, qqx)$ holds in $F_{\mathbf{V}}(x)$, and that $y = px$ and $z = qx$ generate a free algebra $\langle y, z\rangle$, show that in that free algebra, $T(y, pz, qz)$ holds, hence that in any $\mathbf{V}$-algebra, T holds on any 3-tuple whose last two terms are obtained from a common element by applying p, respectively, q to it. Let us now apply this observation to a 3-tuple in $F_{\mathbf{V}}(x)$ of the form $(a(px,qx), px, qx)$ where a is any derived operation of $\mathbf{V}$, and use the independence of px and qx a second time. We conclude that $T(a(y,z), y, z)$ holds for every such a. In other words, in any $\mathbf{V}$-algebra, T holds on every 3-tuple whose first term lies in the subalgebra generated by the last two terms.

But there is no evident way to carry this process further. And in fact, we can again get a negative result by the same technique of realizing T as $t(u,v,w) = 0$, embodying the conditions that we have found T must satisfy this time, in the system of identities:

(6) $t(u, pv, qv) = 0$ for all u, v.

(7) $t(a(u,v), u, v) = 0$ for all u, v, and all binary terms a.

The one tricky point is to show that the variety so defined now has the property that there are no identities satisfied by the free algebra on one generator that are not identities of the whole variety. In contrast to the earlier example, our development has not called on any such identities; but neither has it shown that none exist. With some work, one can prove this; but an easier approach, which we will follow, is to let $\mathbf{V}_0$ denote the variety defined by the identities discussed above, and let our $\mathbf{V}$ be the subvariety of $\mathbf{V}_0$ generated by the free algebra on one generator therein. Here are the details.

Proposition 2. *Let $\mathbf{V}_0$ be the variety defined by a 0-ary operation 0, two unary operations p and q, a ternary operation t, and the identities (3), (4), (6), and (7); and let $\mathbf{V}$ be the subvariety of $\mathbf{V}_0$ generated by the free algebra $F_{\mathbf{V}_0}(x)$ on one generator. Thus, $F_{\mathbf{V}}(x) = F_{\mathbf{V}_0}(x)$, so $\mathbf{V}$ is generated by $F_{\mathbf{V}}(x)$.*

In this situation, the subalgebra $\langle px, qx\rangle \subseteq F_{\mathbf{V}}(x)$ is free in $\mathbf{V}$ (and in fact in $\mathbf{V}_0$) on the two generators px and qx; but the subalgebra $\langle px, pqx, qqx\rangle$ is not free in $\mathbf{V}$ (and hence not in $\mathbf{V}_0$) on px, pqx and qqx.

Proof. The last sentence of the first paragraph is clear in the general context of a subvariety generated by a free algebra in any variety.

We shall next show that $\langle px, qx\rangle \subseteq F_{\mathbf{V}}(x)$ is free on px and qx in $\mathbf{V}_0$. Since it is a subalgebra of $F_{\mathbf{V}}(x)$ and hence belongs to $\mathbf{V}$, it will then *a fortiori* be free on those generators in that subvariety.

To do this, we need to prove that any relation satisfied in $F_{\mathbf{V}}(x)$ by px and qx also holds between y and z in $F_{\mathbf{V}_0}(y, z)$. Let M again denote the free monoid on the two symbols p and q. We know as before that the elements of $F_{\mathbf{V}}(x)$ obtained from x using p and q alone form a free M-set on one generator, and hence that the sub-M-set $M\{px, qx\}$ is free as an M-set on px and qx. Thus, if the given relation satisfied by px and qx involves only the operations p and q, it will indeed be satisfied by y and z. Hence in what follows, we may assume the relation involves t and/or 0.

Now by (3) and (4), if either side of our relation in px and qx involves 0 or t other than in the outermost position, that side equals 0, and the corresponding expression in y and z does as well; so we can replace that side by 0 in our relation. Given the forms of the identities (6) and (7), it is not hard to see that to complete our proof it will suffice to show that if

$$t(b(px, qx), c(px, qx), d(px, qx)) = 0 \tag{8}$$

is an identity of $\mathbf{V}_0$, where

$$b(px, qx), c(px, qx), d(px, qx) \in M\{px, qx\}, \tag{9}$$

then $\mathbf{V}_0$ also satisfies the identity

$$t(b(y, z), c(y, z), d(y, z)) = 0. \tag{10}$$

Moreover, (3) and (4) yield no relations of the form (8) satisfying (9), so we need only look at relations (8) of the forms (6) and (7).

An instance of (7) can have the form (8) only if the given u and v have the forms $c(px, qx)$ and $d(px, qx)$; but then putting y and z in place of px and qx in that instance of (7) again gives an instance of (7), and hence a relation in $F_{\mathbf{V}_0}(y, z)$, as required. If an instance of (6) has the form (8), then we have $u = b(px, qx)$, but there are two possibilities for the element v: it can either be x, or of the form $e(px, qx)$. In the former case, this instance of (6) is also an instance of (7), and the preceding argument applies. In the latter case, the relation has the form $t(b(px, qx), pe(px, qx), qe(px, qx)) = 0$, and we see that $t(b(y, z), pe(y, z), qe(y, z)) = 0$ is again an instance of (6), and hence a relation in $F_{\mathbf{V}_0}(y, z)$. This completes the proof that $\langle px, qx\rangle$ is free on px and qx.

To see, finally, that $\langle px, pqx, qqx\rangle$ is not free on the indicated generators in $\mathbf{V}$, we note that $F_{\mathbf{V}_0}(x)$, which generates $\mathbf{V}$, has 3-tuples of elements of $M\{x\}$ to which neither (6) nor (7) applies, for example, (x, qx, px). Hence $t(x, y, z) = 0$ is not an identity of $\mathbf{V}$; hence the elements px, pqx, qqx, which *do* satisfy that relation, cannot be free generators of a free subalgebra. □

After obtaining the above result, I wondered whether for *every* 3-tuple (ax, bx, cx) in $M\{x\}$, one could find a ternary relation $T_{a,b,c}$ on $M\{x\}$ that could be embodied in a construction like the above, giving an algebra in which that 3-tuple was not a free generating set. If so, then it would seem that by defining a variety with operations 0, p, and q and countably many ternary operations $t_{a,b,c}$, one for each such choice of a, b and c, one should be able to get an example where, as above, $F_{\mathbf{V}}(x)$ generated $\mathbf{V}$ and $\langle px, qx\rangle \subseteq F_{\mathbf{V}}(x)$ was free on px, qx, but where $F_{\mathbf{V}}(x)$ contained no subalgebra free on three generators.

But just a bit more experimentation revealed 3-tuples (ax, bx, cx) for which no $T_{a,b,c}$ with the desired property exists. Translating the resulting obstruction into a proof of a positive statement, this is:

Proposition 3. *Let* $\mathbf{V}$ *be a variety of algebras such that the free algebra* $F_{\mathbf{V}}(x)$ *on one generator generates* $\mathbf{V}$ *as a variety, and contains a subalgebra free of rank* 2 *in* $\mathbf{V}$, *say on generators* px *and* qx, *where* p *and* q *are derived unary operations of* $\mathbf{V}$. *Then the subalgebra* $\langle px, pqx, pqqx\rangle \subseteq F_{\mathbf{V}}(x)$ *is free in* $\mathbf{V}$ *on the indicated three generators.*

Proof. It will suffice to show that for any three elements

(11) $\qquad ax, bx, cx \in F_{\mathbf{V}}(x),$

there exists a homomorphism $\langle px, pqx, pqqx\rangle \to F_{\mathbf{V}}(x)$ carrying px, pqx, $pqqx$ to ax, bx, cx respectively, since this will show that every relation satisfied by px, pqx and $pqqx$ is an identity of $F_{\mathbf{V}}(x)$, and hence, by hypothesis, of $\mathbf{V}$.

Given elements (11), let us first use the freeness of $\langle px, qx\rangle$ to get a homomorphism $f : \langle px, qx\rangle \to F_{\mathbf{V}}(x)$ carrying px to $aqqx$, and qx to x. Thus the image of $(px, pqx, pqqx)$ under this map is $(aqqx, px, pqx)$. Since this 3-tuple, and hence the subalgebra it generates, again lies in $\langle px, qx\rangle$, we can compose this homomorphism with another homomorphism, $g : \langle px, qx\rangle \to F_{\mathbf{V}}(x)$; let this take px to bqx and qx to x. This takes the preceding 3-tuple to (aqx, bqx, px). Finally, mapping $\langle px, qx\rangle$ to $F_{\mathbf{V}}(x)$ by the homomorphism h sending px to cx and qx to x, we get the desired 3-tuple (ax, bx, cx). Hence, the composite hgf : $\langle px, pqx, pqqx\rangle \to F_{\mathbf{V}}(x)$ acts as required. □

In the question we have answered, the choice of ranks one, two and three was, of course, made to give a concrete test problem. This, and the restriction to free algebras rather than coproducts of general algebras, make our counterexamples, Propositions 1 and 2, formally stronger, but our positive result, Proposition 3, weaker than the corresponding result without those restrictions. We shall generalize Proposition 3 in the next two sections so as to remove these restrictions.

3. Prevarieties and quasivarieties

In the proof of Proposition 3, we used the fact that if $\mathbf{V}$ is the variety generated by an algebra A, then a $\mathbf{V}$-algebra generated by a family of elements, $B = \langle\{x_i \mid i \in I\}\rangle$, is free on those generators if and only if there exist homomorphisms $B \to A$ taking the x_i to all choices of I-tuples of elements of A. For our generalization, we would like to say similarly that if an algebra B is generated by a family of subalgebras B_i ($i \in I$), then it is their coproduct if and only if every system of homomorphisms from the algebras B_i to our given algebra A extends to a homomorphism $B \to A$. We shall see that this is true for coproducts, not in the *variety* generated by A, but in the *prevariety* so generated (definition recalled below).

There are a few points of notation and terminology in which usage is not uniform; we begin by addressing these.

First, we admit the empty algebra when the operations of our algebras include no 0-ary operations.

Second, note that the operators $\mathbb{H}$, $\mathbb{S}$ and $\mathbb{P}$ on classes of algebras that appear in Birkhoff's Theorem and related results each come in two slightly different flavors. One may associate to a class $\mathbf{X}$ of algebras the class of all *factor algebras* of members of $\mathbf{X}$ by congruences, or the class of algebras *isomorphic* to such factor algebras, that is, the homomorphic images of members of $\mathbf{X}$. Likewise, one may associate to $\mathbf{X}$ the family of *subalgebras* of members of $\mathbf{X}$, or the family of algebras *isomorphic* to such subalgebras; that is, algebras *embeddable* in members of $\mathbf{X}$. And finally, we may associate to $\mathbf{X}$ the class of *direct product* algebras constructed from members of $\mathbf{X}$ in the standard way as algebras of tuples, or the class of algebras *isomorphic* to algebras so constructed, that is, algebras P that admit a family of maps to the indicated members of $\mathbf{X}$ giving P the universal property of their direct product. It is probably an accident of history that the symbols $\mathbb{H}$, $\mathbb{S}$ and $\mathbb{P}$ were assigned, in two cases (subalgebras and products) to particular explicit constructions, but in the remaining case (homomorphic images) to the isomorphism-closed concept. The standard remedy is to introduce an operator $\mathbb{I}$, taking every class $\mathbf{X}$ of algebras to the class of algebras isomorphic to members of $\mathbf{X}$, and apply $\mathbb{I}$ in conjunction with $\mathbb{S}$ and $\mathbb{P}$ when the wider construction is desired. But that wider construction usually *is* what is desired, so, following [McKenzie et al. 1987], we will use the less standard definitions:

Definition 4. *If $\mathbf{X}$ is a class of algebras of the same type, then $\mathbb{S}\mathbf{X}$ will denote the class of algebras isomorphic to subalgebras of algebras in $\mathbf{X}$, and $\mathbb{P}\mathbf{X}$ the class of algebras isomorphic to direct products of algebras in $\mathbf{X}$ (including the direct product of the empty family, the one-element algebra). As usual, $\mathbb{H}\mathbf{X}$ will denote the class of homomorphic images of algebras in $\mathbf{X}$.*

A third point on which terminology is divided concerns the definition of "quasivariety". Both usages agree that this means a class of algebras A determined by a set of conditions of the form

$$(12) \qquad (\forall\, x \in A^I)\ (\textstyle\bigwedge_{j\in J} a_j(x) = b_j(x)) \implies c(x) = d(x),$$

where I and J are sets, and the a_j and b_j and c and d are I-ary terms in the algebra operations. (J may be empty, in which case (12) represents an ordinary identity.) The point of disagreement is whether I and J are required to be finite. The more standard usage, which, somewhat reluctantly, I will follow, assumes this; a class of algebras defined by sentences (12) where I and J are not required to be finite is then called a *prevariety*. The other usage is exemplified by [Adámek and Sousa 2004], where "quasivariety" is defined with no finiteness restriction on I and J, while "prevariety" is used for a still more general sort of class of algebras (typified by monoids in which every element is invertible; that is groups regarded as monoids).

(My discomfort with the standard usage is that the prefix "pre-" suggests a concept used mainly for technical purposes in the development of another concept, as in "preorder", "presheaf" and "precategory". Also, the relationship between "prevariety" and "quasivariety" is not mnemonic, as "quasivariety" and "elementary quasivariety" would be. Incidentally, if $\mathbf{X}$ is a finite set of finite algebras, the prevariety and the quasivariety that it generates are the same, so works like [Clark and Davey 1998] don't have to distinguish the concepts.)

The concept of quasivariety is a natural one only for finitary algebras. (The constructions of reduced products and ultraproducts, occurring in standard characterizations of quasivarieties, are not in general defined on infinitary algebras.) Most of our results on prevarieties will not require finitariness; so algebras comprising prevarieties will not be assumed finitary unless this is explicitly stated.

We summarize this and some related conventions in:

Definition 5. *A* prevariety *will mean a class* $\mathbf{P}$ *of algebras of a given* (*not necessarily finitary*) *type that can be defined by a class of conditions of the form* (12); *equivalently that is closed under the operators* $\mathbb{S}$ *and* $\mathbb{P}$.

A prevariety $\mathbf{P}$ *which is finitary* (*that is*, *every primitive operation of which has finite arity*), *will be called a* quasivariety *if*

(13) $\mathbf{P}$ *can be defined by conditions* (12) *in each of which* I *and* J *are finite*;

equivalently,

(14) $\mathbf{P}$ *is closed under ultraproducts*;

equivalently,

(15) $\mathbf{P}$ *is closed under reduced products.*

If $\mathbf{X}$ *is a class of algebras of a given type*, *the least prevariety containing* $\mathbf{X}$, *namely*, $\mathbb{S}\mathbb{P}\mathbf{X}$, *will be called the* prevariety generated by $\mathbf{X}$. *Likewise*, *if the type is finitary*, *the least quasivariety containing* $\mathbf{X}$, *namely*, $\mathbb{S}\mathbb{P}\mathbb{P}_{\mathrm{ult}}\mathbf{X} = \mathbb{S}\mathbb{P}_{\mathrm{red}}\mathbf{X}$, *where* $\mathbb{P}_{\mathrm{ult}}$ *and* $\mathbb{P}_{\mathrm{red}}$ *denote*, *respectively*, *the constructions of ultraproducts and reduced products* (*and algebras isomorphic thereto*), *will be called the* quasivariety generated by $\mathbf{X}$. *Again without the assumption of finitariness*, *the least variety containing* $\mathbf{X}$, *namely*, $\mathbb{H}\mathbb{S}\mathbb{P}\mathbf{X}$, *will be called the* variety generated by $\mathbf{X}$.

In any prevariety, one has algebras presented by arbitrary systems of generators and relations. In particular, every family of algebras has a coproduct. A useful characterization of these is

Lemma 6. *Let* $\mathbf{X}$ *be a class of algebras of a given type*, *let* $\mathbf{P} = \mathbb{S}\mathbb{P}\mathbf{X}$ *be the prevariety generated by* $\mathbf{X}$, *let* B *be an algebra in* $\mathbf{P}$, *and let* $f_i : B_i \to B$ $(i \in I)$ *be a family of maps from algebras in* $\mathbf{P}$ *into* B.

Then the algebra B *is a coproduct of the* B_i *in* $\mathbf{P}$, *with the* f_i *as the coprojection maps*, *if and only if the following two conditions are satisfied*:

(16) B *is generated as an algebra by the union of the images* $f_i(B_i)$.

(17) *For every* A *in our generating class* $\mathbf{X}$, *and every choice of a family of maps* $g_i : B_i \to A$ $(i \in I)$, *there exists a homomorphism* $g : B \to A$ *such that* $g_i = g f_i$ *for all* $i \in I$.

Sketch of proof. "Only if" is straightforward: the necessity of (16) is shown, as usual, by applying the universal property of B as a coproduct to the maps f_i, regarded as taking the B_i into the subalgebra C of B that they together generate (which belongs to $\mathbf{P}$, since $\mathbf{P}$ is closed under taking subalgebras); while the necessity of (17) is a case of that universal property.

Conversely, assuming (16) and (17), let us show that B and the f_i satisfy the universal property of the coproduct. Let C be any algebra in $\mathbf{P}$, given with homomorphisms $a_i : B_i \to C$.

If there exists a homomorphism $a : B \to C$ with $a_i = af_i$ for all i, then by (16) it will be unique.

To see that such a map exists, we write C as a subalgebra of a direct product $\prod_{j\in J} A_j$ with all A_j in $\mathbf{X}$. Then for each $j \in J$, the composites of the given maps $a_i : B_i \to C$ with the jth projection $C \to A_j$ give a system of maps $a_{ij} : B_i \to A_j$ $(i \in I)$. By (17), for each j the a_{ij} are induced by a single map $a_{*j} : B \to A_j$; doing this for all $j \in J$ gives a map $a : B \to \prod_J A_j$, whose restriction to each $f_i(B_i) \subseteq B$ lies in $C \subseteq \prod_{j\in J} A_j$. Hence $a(B)$ lies in C by (16). The relations $a_{ij} = a_{*j} f_i$ now show that $a_i = af_i$, as required. □

Remarks. We shall see in Section 7 that it can happen that though each B_i lies in $\mathbf{P} = \mathbb{SP}\mathbf{X}$, no $A \in \mathbf{X}$ simultaneously admits maps from all B_i. In that case, condition (17) is vacuous, and the lemma says that (16) characterizes the coproduct $\coprod_{\mathbf{P}} B_i$. Though implausible-sounding, this is correct: in that case an algebra B with maps of the B_i into it, to belong to $\mathbf{P} = \mathbb{SP}\mathbf{X}$, must embed in the product of the vacuous family of members of $\mathbf{X}$, hence can have at most one element, so there is hardly any way it can differ from the desired universal object; (16) merely guarantees that if all B_i are empty, B is also.

In a different direction, taking $I = \varnothing$ in the above result and recalling that a coproduct of the empty family of objects in a category is an initial object of the category, the result says that an algebra B is initial in $\mathbf{P}$ if and only if it is generated by the empty set and admits a homomorphism into each $A \in \mathbf{X}$.

4. P-independent subalgebras

Definition 7. *If A is an algebra in a prevariety $\mathbf{P}$, we shall call a family of subalgebras $B_i \subseteq A$ $(i \in I)$* $\mathbf{P}$-independent *if the subalgebra $B \subseteq A$ that they generate, given with the system of inclusion maps $B_i \to B$, is a coproduct of the B_i in $\mathbf{P}$.*

Here, finally, is the promised generalization of Proposition 3.

Theorem 8. *Suppose that $\mathbf{P}$ is a prevariety of algebras and A_0 an algebra in $\mathbf{P}$, and that for some natural number n we are given subalgebras $A_1, \dots, A_n$ and $B_1, \dots, B_n$ of A_0, such that for $i = 1 \dots n$, A_i and B_i are $\mathbf{P}$-independent, and are both contained in A_{i-1}. Assume, moreover, that $\mathbf{P}$ is generated as a prevariety by A_n.*

Then $B_1, \dots, B_n$ are $\mathbf{P}$-independent.

Proof. Let us prove by induction on $i = 0, \dots, n$ a statement a little stronger than what we will need for $i = n$, namely that for every system of homomorphisms

$f_j : B_j \to A_i$ $(j = 1, \ldots, i)$, there exists a unique homomorphism f from the subalgebra of A_0 generated by $B_1, \ldots, B_i$ and A_i into A_i which acts on each B_j $(j = 1, \ldots, i)$ as f_j, and which acts as the identity on A_i.

This is clear for $i = 0$. Let $0 < i \leq n$, inductively assume the result for $i-1$, and suppose we are given $f_j : B_j \to A_i$ $(j = 1, \ldots, i)$. Since $A_i \subseteq A_{i-1}$, our inductive hypothesis gives us a homomorphism g from the subalgebra of A_0 generated by $B_1, \ldots, B_{i-1}$ and A_{i-1} into A_{i-1} which agrees with f_j for $j = 1, \ldots, i-1$, and is the identity on A_{i-1}. Note that g will carry the subalgebra generated by $B_1, \ldots, B_i$ and A_i into the subalgebra generated by A_i (into which it carries $B_1, \ldots, B_{i-1}$ and A_i) and B_i (which is contained in A_{i-1}, and so is left fixed).

But by assumption, that subalgebra is the coproduct of A_i and B_i, so we can map it into A_i by a homomorphism h which acts as the identity on A_i and as f_i on B_i. Now $f = hg$ clearly has the property required for our inductive step.

Taking the $i = n$ case of our result, and ignoring the condition that f be the identity on A_n, we see that the subalgebra $B \subseteq A_0$ generated by $B_1, \ldots, B_n$ satisfies (17) for $\mathbf{X}$ the singleton family $\{A_n\}$. Since by assumption A_n generates $\mathbf{P}$, Lemma 6 tells us that B is the coproduct of the B_i in $\mathbf{P}$. $\square$

Remark. We might call a family of subalgebras B_i of an algebra A in a prevariety $\mathbf{P}$, given with a distinguished member B_0 which generates $\mathbf{P}$, "almost $\mathbf{P}$-independent" if every family of homomorphisms $f_i : B_i \to B_0$ such that f_0 is the identity map of B_0 can be realized by a homomorphism on the subalgebra generated by the B_i. We see from the proof of Theorem 8 that that theorem remains true if the $\mathbf{P}$-independence hypothesis is weakened to say that each pair (A_i, B_i), with A_i taken as the distinguished member, is almost $\mathbf{P}$-independent, and the conclusion strengthened to say that the $n+1$-tuple $(A_n, B_1, \ldots, B_n)$, with A_n as distinguished member, is almost $\mathbf{P}$-independent. The condition of almost $\mathbf{P}$-independence seemed too technical to use in the formal statement of the theorem; but one might keep it in mind. It is interesting that while Proposition 2 showed that in the situation described there, the subalgebras $\langle px \rangle$, $\langle pqx \rangle$ and $\langle qqx \rangle$ of $F_{\mathbf{V}}(x)$ were not $\mathbf{V}$-independent, the above proof shows that, with the last of them taken as distinguished, they are *almost* $\mathbf{V}$-independent.

Note that Theorem 8 holds even in the case $n = 0$: If A_0 generates $\mathbf{P}$, then the subalgebra of A_0 generated by the empty set is the initial object of $\mathbf{P}$.

A case of Theorem 8 with a simpler hypothesis is

Corollary 9. *Suppose A and $B_1, \ldots, B_n$ are algebras in a prevariety $\mathbf{P}$, such that A generates $\mathbf{P}$, and such that for each i, A contains an isomorphic copy of $A \amalg_{\mathbf{P}} B_i$. Then A contains an isomorphic copy of $\coprod_{\mathbf{P}}^{i=1,\ldots,n} B_i$.* $\square$

Recall next that a free algebra in a prevariety $\mathbf{P}$ is also free on the same generators in the variety $\mathbf{V}$ generated by $\mathbf{P}$. Hence we can apply the above results to free algebras in a variety, and obtain the following result extending Proposition 3 (though we omit, for brevity, the explicit description of the free generators).

Corollary 10. *Suppose* $\mathbf{V}$ *is a variety and* $m < n$ *are positive integers such that the free* $\mathbf{V}$*-algebra* $F_{\mathbf{V}}(x_1, \ldots, x_m)$ *on* m *generators has a subalgebra free on* n *generators, and such that* $\mathbf{V}$ *is generated as a variety by* $F_{\mathbf{V}}(x_1, \ldots, x_m)$*. Then for every natural number* N*,* $F_{\mathbf{V}}(x_1, \ldots, x_m)$ *has a subalgebra free on* N *generators.*

Proof. A free $\mathbf{V}$-algebra $F_{\mathbf{V}}(x_1, \ldots, x_n)$ has subalgebras free on all smaller numbers of generators; so the above hypothesis implies that $F_{\mathbf{V}}(x_1, \ldots, x_m)$ has a subalgebra free on $m+1$ generators. This is a coproduct of a free algebra on m generators and a free algebra on one generator, so we get the hypothesis of Corollary 9 with $\mathbf{P}$ the prevariety generated by $A = F_{\mathbf{V}}(x_1, \ldots, x_m)$, the n of that corollary taken to be N, and each B_i taken to be free on one generator. The conclusion shows that $F_{\mathbf{V}}(x_1, \ldots, x_m)$ has a subalgebra free on N generators in the prevariety it generates. As noted, a free algebra in a prevariety $\mathbf{P}$ is also free in the variety $\mathbf{V}$ generated by $\mathbf{P}$. $\square$

Can we strengthen this result to give free subalgebras of countably infinite rank? Yes if our algebras are finitary. We will need:

Lemma 11. *Let* $\mathbf{P}$ *be a prevariety of* finitary *algebras,* A *an algebra in* $\mathbf{P}$*, and* $(B_i)_{i\in I}$ *a family of subalgebras of* A*, such that every finite subset* $I_0 \subseteq I$ *is contained in a subset* $I_1 \subseteq I$ *such that the family of subalgebras* $(B_i)_{i\in I_1}$ *is* $\mathbf{P}$*-independent. Then* $(B_i)_{i\in I}$ *is* $\mathbf{P}$*-independent.*

Proof. We need to show that the map $f_I : \coprod_{\mathbf{P}}^{I} B_i \to A$ whose composite with each coprojection $q_j : B_j \to \coprod_{\mathbf{P}}^{I} B_i$ is the inclusion of B_j in A is one-to-one. By finitariness of $\mathbf{P}$, every element of $\coprod_{\mathbf{P}}^{I} B_i$ lies in the subalgebra generated by finitely many of the B_i, hence it will suffice to show that for any finite subset $I_0 \subseteq I$, the restriction of f_I to the subalgebra of $\coprod_{\mathbf{P}}^{I} B_i$ generated by $\{B_i \mid i \in I_0\}$ is one-to-one. By assumption, I_0 is contained in a subset I_1 such that the family $(B_i)_{i\in I_1}$ is $\mathbf{P}$-independent; hence the canonical map $f_{I_1} : \coprod_{\mathbf{P}}^{I_1} B_i \to A$ is one-to-one; but that map factors through f_I, so f_I is one-to-one on its image, the subalgebra of $\coprod_{\mathbf{P}}^{I} B_i$ generated by $\{B_i \mid i \in I_1\}$, hence on the smaller subalgebra generated by $\{B_i \mid i \in I_0\}$, as required. $\square$

(We shall see in Sections 6 and 8 respectively that if a prevariety $\mathbf{P}$ either satisfies the amalgamation property (which is not in general the case in the situation we are interested in here) or is generated as a prevariety by a single algebra (which is true in the situation to which we are about to apply the above lemma) then any subfamily of a $\mathbf{P}$-independent family of subalgebras is $\mathbf{P}$-independent; so in such cases, the hypothesis of the above lemma can be simplified merely to say that every finite subset of I is $\mathbf{P}$-independent. But in a general prevariety $\mathbf{P}$, a subfamily of a

P-independent family need not be **P**-independent, hence that simplified statement does not carry the full force of the lemma. For an example of **P**-independence not carrying over to subfamilies, take for **P** the variety **V** of monoids with two distinguished elements x and y, let A be the **V**-algebra generated by a universal two-sided inverse to x, denoted x^{-1}, let B_1 and B_2 both be the subalgebra of A generated by $u = x^{-1}y$, which is a free monoid on two generators u and x, regarded as a member of **V** by setting $y = xu$, and let B_3 be the whole algebra A. It is not hard to verify that in $B_1 \amalg_{\mathbf{V}} B_2$, the images u_1, u_2 of the copies of u from B_1 and B_2 are distinct (though they satisfy $xu_1 = xu_2$). Since in A itself, in contrast, their images are equal, B_1 and B_2 are not **V**-independent subalgebras of A. But in $B_1 \amalg_{\mathbf{V}} B_2 \amalg_{\mathbf{V}} B_3$, the properties of two-sided inverses force the generators of B_1 and B_2 to fall together with the corresponding elements of B_3, so the family consisting of these three subalgebras satisfies the definition of **V**-independence.)

Combining the above lemma with our earlier results, we get:

Corollary 12. *Let* **P**, *in* (i) *and* (ii) *below, be a prevariety of* finitary *algebras, and* **V**, *in* (iii), *a variety of such algebras. Then*:

(i) *If* $A_0 \supseteq A_1 \supseteq \cdots \supseteq A_i \supseteq \ldots$ *are algebras in* **P** *such that every* A_i *generates* **P** *as a prevariety*; *and if for each* $i > 0$, B_i *is a subalgebra of* A_{i-1} *such that* A_i *and* B_i *are* **P**-*independent*, *then the countable family* $(B_i)_{i>0}$ *is* **P**-*independent.*
Hence:

(ii) *If* A *is an algebra which generates* **P** *as a prevariety*, *and we are given a countable family of algebras* $(B_i)_{i>0}$ *in* **P**, *such that for each* i, A *has a subalgebra isomorphic to* $A \amalg_{\mathbf{P}} B_i$, *then* A *has a subalgebra isomorphic to* $\coprod_{\mathbf{P}}^{i>0} B_i$.
Hence:

(iii) *If for some positive integer* m *the free algebra* $F_{\mathbf{V}}(x_1, \ldots, x_m)$ *generates* **V** *as a variety*, *and contains a subalgebra free on* $> m$ *generators*, *then it contains a subalgebra free on countably many generators.* □

Lemma 11 and Corollary 12 both fail if the assumption that our algebras are finitary is deleted. To see this for the lemma, let **V** be the variety determined by one operation a of countably infinite arity, and identities saying that whenever two of $x_0, x_1, \ldots$ are equal, we have

$$a(x_0, x_1, \ldots) = x_0. \tag{18}$$

Let A be a countably infinite set, on which a is defined by letting (18) hold for *all* $x_0, x_1, \ldots$. Then every finite subset of A is a free subalgebra on that set, from which one sees that any finite family of distinct singleton subsets is an independent set of subalgebras; but the set of all of these is not independent, because their coproduct in the variety **V**, the free **V**-algebra on countably many generators, does not satisfy (18) identically.

To show that the statements of Corollary 12 all need the finitariness condition, it suffices to give a counterexample to statement (iii) in the absence of that condition. The idea will be the same as above, but the details are more complicated, and I will be a little sketchy.

The variety in question will have countably many 0-ary operations $c_0, c_1, \ldots,$ two unary operations p and q, an operation a of countable arity, and an additional 0-ary operation 0, satisfying the analogs of (3) and (4) with a in place of t. As in Section 2, let M denote the free monoid on the symbols p and q. Let $\mathbf{V}_0$ be defined by the abovementioned analogs of (3) and (4), together with the (uncountable) family of identities saying that

(19) $a(x_0, x_1, \ldots) = 0$ if infinitely many of the x_i belong to $M\{u\}$ for some common element u.

These identities do not imply $a(c_0, c_1, \ldots) = 0$, so $a(x_0, x_1, \ldots) = 0$ is not an identity in any free algebra in $\mathbf{V}_0$. Once again, let $\mathbf{V}$ be the subvariety of $\mathbf{V}_0$ generated by the free algebra $F_{\mathbf{V}}(x)$ on one generator.

One finds that the subalgebra $\langle px, qx\rangle \subseteq F_{\mathbf{V}}(x) = F_{\mathbf{V}_0}(x)$ is free on px and qx. The key point is that if an element $a(x_0, x_1, \ldots)$ with $x_0, x_1, \ldots \in \langle px, qx\rangle$ equals 0 in $F_{\mathbf{V}}(x)$ by an application of (19), and the element u of the hypothesis of (19) is x, then the infinite family of elements of $M\{u\}$ in question will be the union of a family of elements of $M\{px\}$ and a family of elements of $M\{qx\}$, one of which must still be infinite; so the relation $a(x_0, x_1, \ldots) = 0$ still follows from the expressions for the x_i in terms of px and qx.

However, I claim that $F_{\mathbf{V}}(x)$ contains no subalgebra free on countably many generators. For note that a family of independent elements of $F_{\mathbf{V}}(x)$ cannot include the value of any primitive or derived 0-ary operation (since their behavior under homomorphisms is not free), nor any element obtained with the help of a, by the analogs of (3) and (4); hence such a family must lie entirely in $M\{x\}$. But by (19) (with $u = x$), any infinite family $x_0, x_1, \ldots$ of elements of $M\{x\}$ satisfies the relation $a(x_0, x_1, \ldots) = 0$, which we have seen is not an identity of $\mathbf{V}$; so no infinite family of elements of $F_{\mathbf{V}}(x)$ is independent.

5. Some questions

Proposition 1 shows that Corollary 10 becomes false if we delete the assumption that $F_{\mathbf{V}}(x_1, \ldots, x_m)$ generates $\mathbf{P}$. In the absence of that assumption, it is not clear what forms the relation of mutual embeddability can assume.

Question 13. For $\mathbf{V}$ a variety, let us say that two natural numbers m and n are $\mathbf{V}$-equivalent (with respect to embeddability of free algebras) if $F_{\mathbf{V}}(x_1, \ldots, x_m)$

and $F_{\mathbf{V}}(x_1, \ldots, x_n)$ each contain an isomorphic copy of the other. Clearly, the **V**-equivalence classes are blocks of consecutive integers. Which decompositions of the natural numbers into blocks can be realized in this way?

More generally, given a prevariety **P** and algebras $A_1, \ldots, A_r$ in **P**, let us define a preorder $\preceq$ on r-tuples of natural numbers by writing $(m_1, \ldots, m_r) \preceq (n_1, \ldots, n_r)$ if the coproduct in **P** of m_1 copies of A_1, m_2 copies of A_2, etc., through m_r copies of A_r, is embeddable in the coproduct of n_1 copies of A_1 etc., through n_r copies of A_r. What preorderings on ω^r can be realized in this way? In particular, what equivalence relations on ω^r can be the equivalence relation determined by such a preorder? Do these answers change if one requires **P** to be a variety?

In [Zaĭtsev 1992], for certain varieties **V** of Lie algebras over a field of characteristic 0, bounds are obtained on n/m for any m, n equivalent under the relation of the first paragraph of Question 13 above. The idea is to note that if $F_{\mathbf{V}}(x_1, \ldots, x_m)$ and $F_{\mathbf{V}}(x_1, \ldots, x_n)$ are mutually embeddable, they must have the same Gelfand–Kirillov dimension (a measure of growth rate). Upper and lower bounds are obtained for the Gelfand–Kirillov dimension of $F_{\mathbf{V}}(x_1, \ldots, x_n)$, leading to the asserted conclusions. However, it seems most likely that for such **V**, the Gelfand–Kirillov dimension of $F_{\mathbf{V}}(x_1, \ldots, x_n)$ will grow with n in a "smooth" fashion; if so, one should in fact be able to prove that no free algebras of distinct ranks in **V** are mutually embeddable, in which case such varieties will not give interesting examples relevant to Question 13.

For results on *isomorphisms* and *surjections* among free algebras, rather than embeddings, see [Świerczkowski 1961; Clark 1969; Cohn 1966]. The last of these shows that all consistent cases are realized by module-varieties $\mathbf{Mod}_R$ for rings R.

In generalizing Proposition 3 from free algebras to general coproducts, we found that the context that made the argument work was that of coproducts in a *prevariety*. Theorem 8 does not give us the corresponding statement for general $A_0, \ldots, A_n, B_1, \ldots, B_n$ with the prevariety **P** replaced by a variety **V**, and the condition that **P** be generated by A_n as a prevariety replaced by the condition that **V** be generated by A_n as a variety, since for a variety **V** and an algebra $A \in \mathbf{V}$, the condition that **V** be generated by A as a variety is weaker than the condition that it be generated by A as a prevariety. (For example, the variety of abelian groups is generated by the infinite cyclic group as a variety, but not as a prevariety, since all groups in the prevariety it generates must be torsion-free.) But I don't have a counterexample to the modified statement.

Question 14. (i) Does Theorem 8 remain true if "prevariety" is everywhere replaced by "variety"?

If not, or if the question proves difficult, one might examine some special cases; for example:

(ii) If in Corollary 9 we replace "prevariety" by "variety", and add the assumption that A is free of rank 1 in that variety (but not that the B_i are free), does the statement still hold?

(iii) If $\mathbf{V}$ is a variety, and A an algebra that generates $\mathbf{V}$ as a variety, and that contains as a subalgebra a coproduct of two copies of itself in $\mathbf{V}$, must it contain a coproduct of three copies of itself in $\mathbf{V}$?

More likely to have positive answers, since quasivarieties are more like prevarieties than varieties are, is:

Question 15. Same questions (i), (ii), (iii) as above, but with "variety" everywhere replaced by "quasivariety" (necessarily, of finitary algebras).

Looking back further, to Section 2, the mixture of positive and negative results there suggests:

Question 16. Is there a nice criterion for whether a 3-tuple (a, b, c) of monoid words in two letters p, q has the property proved in Proposition 3 to hold for the 3-tuple (p, pq, pqq), and in Proposition 2 *not* to hold for the 3-tuple (p, pq, qq), namely, of witnessing the existence of subalgebras free on three generators in all relatively free one-generator algebras $\langle x\rangle$ that contain free two-generator subalgebras $\langle px, qx\rangle$ in the varieties they generate?

More generally, given $n > 1$ and $N > 1$, one may ask for a criterion for an N-tuple $(a_1, \dots, a_N)$ of words in n letters $p_1, \dots, p_n$ to witness the existence of a free subalgebra on N generators in any relatively free algebra on one generator that contains a free subalgebra $\langle p_1x, \dots, p_nx\rangle$ on n generators in the variety it generates.

(Still more generally, for $n > m > 0$ and $N > 1$, one may ask how to decide whether a given N-tuple of terms in m variables and n operation symbols each of arity m witnesses the result of Corollary 10. But since terms in operation symbols of arity > 1 are more complicated than words in unary operation symbols, there seems to be less likelihood of a simple answer.)

The next four sections are related to, but do not depend on, the material above, except for the definitions. Section 6 recalls what it means for a category of algebras to have the *amalgamation property*, obtains some equivalent statements, and then shows that for prevarieties with that property, one has stronger results on when a family of subalgebras of an algebra generates a subalgebra isomorphic to their coproduct than those that we have seen to hold in general. In a different direction, motivated by the fact that the prevarieties considered in Section 4 were by hypothesis each generated by a single algebra, Sections 7-9 show that the number of algebras needed to generate a prevariety has important consequences for the behavior of coproducts therein. The brief section Section 10, which is included in this note only for convenience, answers a different question about coproducts, also raised in [Bergman 2007], concerning subgroups of the full symmetric group on an infinite set.

6. The amalgamation property, and its consequences for P-independence

In any class of algebras that admits coproducts with amalgamation (pushouts), it is well known and easy to verify that the *amalgamation property* (definition recalled in (20) below) is equivalent to the condition that for all pairs of one-to-one maps with common domain, $A \to B$ and $A \to C$, the coprojection maps of B and C into the coproduct with amalgamation $B \amalg_A C$ are also one-to-one. The next lemma gives some further consequences of that property, in the same vein. We formulate it in a context more general than that of categories of algebras, though less sophisticated than that of [Kiss et al. 1982, Section 6].

In that lemma, the functor $U : \mathbf{C} \to \mathbf{Set}$ plays the role of the underlying set functor of a category of algebras, but we shall not need to assume it faithful, as one does when defining the concept of a concrete category.

One other notational remark: So far, I have generally written $\coprod_{\mathbf{P}}$ for "coproduct in the category $\mathbf{P}$"; but when discussing coproducts with amalgamation of an object, we will use the subscript position for that object, leaving the category to be understood from the context. I will follow this mixed practice for the rest of the paper, showing the category when no amalgamation is involved. (The superscript position, which might otherwise be assigned to one of these, is used here for index sets over which coproducts are taken. If there were danger of ambiguity, we could write $\coprod_{\mathbf{P},A}$ rather than $\coprod_A$, or regard coproducts with amalgamation as coproducts in a comma category $(A \downarrow \mathbf{P})$ and so write $\coprod_{(A\downarrow\mathbf{P})}$.)

Lemma 17. *Let* $\mathbf{C}$ *be a category and* $U : \mathbf{C} \to \mathbf{Set}$ *a functor, and let us call a morphism* f *in* $\mathbf{C}$ *"one-to-one" if* $U(f)$ *is a one-to-one set map, and emphasize this by indicating such morphisms using tailed arrows:* $\rightarrowtail$.

Assume that $\mathbf{C}$ *admits pushouts of pairs of one-to-one morphisms; that is, that if* $f : S \rightarrowtail A$ *and* $g : S \rightarrowtail B$ *are one-to-one, then the coproduct with amalgamation* $A \amalg_S B$ *exists.* (*But we do not assume at this point that the maps of* S, A *and* B *to that coproduct are one-to-one.*)

Then the following three conditions are equivalent:

(20) $\mathbf{C}$ *has the amalgamation property* [Kiss et al. 1982, page 82]. *That is, given objects* A, B, C *of* $\mathbf{C}$, *and one-to-one morphisms* $f : A \rightarrowtail B$, $g : A \rightarrowtail C$, *there exists an object* D, *and one-to-one morphisms* $f' : B \rightarrowtail D$, $g' : C \rightarrowtail D$, *such that* $f'f = g'g$.

(21) *For all objects* S, T, A, B *and one-to-one morphisms* $S \rightarrowtail T$, $S \rightarrowtail A$,

and $f : A \rightarrowtail B$ *in* $\mathbf{C}$, *the induced morphism* $f \amalg_S T : A \amalg_S T \to B \amalg_S T$ *is one-to-one.*

(22) *For all objects* S, *positive integers* n, *and finite families of objects and one-to-one morphisms* $S \rightarrowtail A_i$ *and* $f_i : A_i \rightarrowtail B_i$ *in* $\mathbf{C}$ $(i{=}1, \dots, n)$, *the induced morphism* $\coprod_S^{i=1,\dots,n} f_i : \coprod_S^{i=1,\dots,n} A_i \to \coprod_S^{i=1,\dots,n} B_i$ *is one-to-one.*

Moreover, *if* $\mathbf{C}$ *also admits direct limits* (*colimits over directed partially ordered sets*), *and if* U *respects these* (*for example*, *if* $\mathbf{C}$ *is a quasivariety of finitary algebras*, *and* U *its underlying set functor*), *then* $\mathbf{C}$ *has coproducts with amalgamation of possibly infinite families of one-to-one maps* $S \rightarrowtail A_i$ (*for fixed* S, *and* i *ranging over a possibly infinite set* I); *and* (22) *goes over to such coproducts. That is*, (20)–(22) *are also equivalent to*:

(23) *For all objects* S, *nonempty sets* I, *and families of objects and one-to-one morphisms* $S \rightarrowtail A_i$ *and* $f_i : A_i \rightarrowtail B_i$ *in* $\mathbf{C}$ $(i \in I)$, *the induced morphism* $\coprod_S^{i\in I} f_i : \coprod_S^{i\in I} A_i \to \coprod_S^{i\in I} B_i$ *is one-to-one.*

Proof. (20)⇒(21): Given objects and maps as in (21), the amalgamation property implies, as mentioned above, that the coprojection $A \to A \amalg_S T$ is one-to-one. From this and the assumed one-to-oneness of the map $A \to B$ we similarly get one-to-oneness of the coprojection $A \amalg_S T \to B \amalg_A (A \amalg_S T) = B \amalg_S T$, as desired.

(21)⇒(20): Given objects and maps as in (20), apply (21) with A and its identity map in the role of S and its map to A, and with C in the role of T, noting that the domain of the resulting map, $A \amalg_A C$, can be identified with C. This gives one-oneness of the coprojection $C \to B \amalg_A C$. By symmetry one also has one-oneness of the coprojection $B \to B \amalg_A C$. Taking $D = B \amalg_A C$, we get (20).

(21)⇒(22): The case $n = 1$ of (22) is trivial. To get the case $n = 2$ we make a double application of (21), first getting one-oneness for $f_1 \amalg_S A_2 : A_1 \amalg_S A_2 \rightarrowtail B_1 \amalg_S A_2$ and then for $B_1 \amalg_S f_2 : B_1 \amalg_S A_2 \rightarrowtail B_1 \amalg_S B_2$. Composing, we get one-oneness of the desired map.

This shows that two-fold coproducts over S respect one-to-oneness of maps among objects having one-to-one maps of S into them. Induction now gives the corresponding result for n-fold coproducts.

(22)⇒(21): Given objects and maps as in (21), apply the $n = 2$ case of (22) with $A \rightarrowtail B$ in the role of $f_1 : A_1 \rightarrowtail B_1$ and the identity map of C in the role of $f_2 : A_2 \rightarrowtail B_2$.

Under the additional assumptions about direct limits, one notes that for infinite I, one can obtain $\coprod_S^{i\in I} A_i$ as the direct limit, over the directed system of all finite subsets $I_0 \subseteq I$, of the objects $\coprod_S^{i\in I_0} A_i$. Since by (22), the indicated maps among these finite coproducts are one-to-one, and by assumption direct limits respect U

(and hence one-oneness), the corresponding maps among the coproducts over I are also one-to-one. The converse is immediate: (23) includes (22). $\square$

Let us now note how the amalgamation property implies conditions on independent subalgebras stronger than those of Section 4. In considering categories of algebras, we shall take the functor U of Lemma 17 to be the underlying set functor. Thus, "one-to-one", in our formulation (20) of the amalgamation property and our statements of conditions equivalent thereto, has its usual meaning for algebras.

Corollary 18. *Suppose that* $\mathbf{P}$ *is a prevariety having the amalgamation property, that* A *is a* $\mathbf{P}$*-algebra, that* $(B_i)_{i\in I}$ *is a finite* $\mathbf{P}$*-independent family of subalgebras of* A*, and that for each* $i \in I$*,* $(C_{ij})_{j\in J_i}$ *is a finite* $\mathbf{P}$*-independent family of subalgebras of* B_i*. Then* $(C_{ij})_{i\in I,\, j\in J_i}$ *is a* $\mathbf{P}$*-independent family of subalgebras of* A*. (In particular, in such a prevariety, examples like those of Propositions 1 and 2 cannot occur.)*

If $\mathbf{P}$ *is in fact a* quasivariety *having the amalgamation property, then the above result holds without the finiteness restrictions on* I *and the* J_i*.*

Proof. Since all the algebras named are subalgebras of A, the unique homomorphic image of the initial algebra of $\mathbf{P}$ in all of them is the same; let us call this S. Because S is a homomorphic image of the initial algebra of our category, the operator $\coprod_S$ on nonempty families of algebras containing S is just $\coprod_{\mathbf{P}}$.

We now apply the implication (20)⇒(22) of Lemma 17 (or if $\mathbf{P}$ is a quasivariety, the stronger implication (20)⇒(23)), taking for the S of (22) and (23) the S of the preceding paragraph, and for the maps $A_i \rightarrowtail B_i$ the inclusions $\coprod_{\mathbf{P}}^{j\in J_i} C_{ij} \subseteq B_i$. We conclude that the natural map from

$$\coprod_{\mathbf{P}}^{i\in I}\Big(\coprod_{\mathbf{P}}^{j\in J_i} C_{ij}\Big) = \coprod_{\mathbf{P}}^{\substack{i\in I\\ j\in J_i}} C_{ij}$$

to $\coprod_{\mathbf{P}}^{I} B_i$ is one-to-one. Identifying the latter algebra with its embedded image in A, we get the desired conclusion.

To get the parenthetical remark about examples like those of Propositions 1 and 2, we take $I = \{0, 1\}$, $J_0 = \{0\}$, $J_1 = \{0, 1\}$, and let A, the B_i and the C_{ij} all be free of rank 1. $\square$

Remarks. Lemma 17 was a compromise between proving the minimum we needed to get the above corollary — that (20) implies the special case of (22) where S is the image of the initial object of $\mathbf{C}$ in A, and so can be ignored in forming coproducts (and if $\mathbf{C}$ has, and U respects, direct limits, the corresponding case of (23)) — and digressing to state and prove a more complete statement. That statement would involve the versions of conditions (20) and (22) for κ-fold families for any cardinal κ, would establish the equivalence between those two conditions for each such κ, would note that the statements for larger κ imply those for smaller κ, and would verify that the statements for *finite* $\kappa \geq 2$ are all equivalent, and also equivalent to (21). The reader should not find it hard to work out the details.

The reason we brought S into (22), though the only case of (22) that our application needed was where S was a homomorphic image of the initial object and so had no effect, was so as to get an if-and-only-if relation between (22) and (20), the amalgamation property. (The latter is a well-known property, satisfied by the categories of groups, semilattices, lattices, and commutative integral domains, and many others. See the first column of the table in [Kiss et al. 1982, pages 98–107] for more results, positive and negative.) That equivalence fails if S in (22) is restricted to homomorphic images of the initial object. For instance, the normal form for coproducts of monoids shows that the variety **Monoid** satisfies the cases of (21)–(23) where S is the initial (trivial) monoid. However **Monoid** does not satisfy the amalgamation property (20); for example, letting $A = \langle x \rangle$, the free monoid on one generator, and letting B and C be the overmonoids of A gotten by adjoining a left inverse y, respectively a right inverse z, to x, one finds that in $B \amalg_A C$, the elements xy of B and zx of C fall together with 1; so the maps from B and C to this algebra are not one-to-one. On the other hand, because the special case of (21)–(23) which we have seen suffices for Corollary 18 holds, **Monoid** does satisfy the conclusion of that corollary.

Here is another result (alluded to in the discussion following Lemma 11) of a sort similar to the above, which for simplicity of wording we will again state in terms of the amalgamation property, though again, only the cases of (21)–(23) where S is a homomorphic image of the initial object of $\mathbf{P}$ are needed.

Corollary 19. *Suppose* $\mathbf{P}$ *is a prevariety having the amalgamation property, and* A *a* $\mathbf{P}$*-algebra. Then every nonempty subfamily of a* $\mathbf{P}$*-independent family of subalgebras of* A *is* $\mathbf{P}$*-independent.*

Proof. Given a $\mathbf{P}$-independent family of subalgebras B_i $(i \in I)$ of A, their $\mathbf{P}$-independence says that the subalgebra of A that they generate is isomorphic to their coproduct, which we see coincides with their coproduct over the common image S in all these algebras of the initial algebra of $\mathbf{P}$. For any nonempty subset $J \subseteq I$, the coproduct of the B_i for $i \in J$ likewise coincides with their coproduct over S. We now apply (21) with this algebra S for both the S and A of that condition, with $\coprod_S^J B_i$ for T, and with $\coprod_S^{I-J} B_i$ for B, and then bring in the assumed $\mathbf{P}$-independence of the whole family. We thus get one-oneness of the natural maps shown by the first arrow in

$$\text{(24)} \qquad \textstyle\coprod_{\mathbf{P}}^J B_i \cong \coprod_S^J B_i \rightarrowtail \coprod_S^J B_i \amalg_S \coprod_S^{I-J} B_i \cong \coprod_S^I B_i \cong \coprod_{\mathbf{P}}^I B_i \rightarrowtail A,$$

and the above arrows and isomorphisms compose to the map we wished to show one-to-one. $\square$

7. P-compatible algebras

The prevarieties considered in Section 4 were each generated by a single algebra. Although any *variety* of algebras can be generated as a *variety* by a single algebra (namely, by a free algebra on sufficiently many generators), prevarieties generated as *prevarieties* by a single algebra are rather special. This was shown by Mal'cev for quasivarieties, in a result that we will generalize in the next section. In this section we shall see that the size of the collection of algebras needed to generate $\mathbf{P}$ as a prevariety is a nontrivial and interesting invariant of $\mathbf{P}$, even if $\mathbf{P}$ happens to be a variety.

Definition 20. *Let* $\mathbf{P}$ *be a prevariety. Then a set* $\mathbf{X}$ *of* $\mathbf{P}$*-algebras will be called* $\mathbf{P}$-compatible *if for every* $A_0 \in \mathbf{X}$, *the coprojection map* $A_0 \to \coprod_{\mathbf{P}}^{A\in\mathbf{X}} A$ *is one-to-one*; *equivalently*, *if there exists an algebra* B *in* $\mathbf{P}$ *admitting one-to-one homomorphisms* $A \to B$ *for all* $A \in \mathbf{X}$.

Theorem 21. *Suppose* $\mathbf{P}$ *is a prevariety that is residually small* (*that is*, *that can be generated as a prevariety by a* set *of algebras*) *and* κ *is a cardinal. Then condition* (25) *below implies condition* (26); *and if* $\mathbf{P}$ *is a quasivariety* (*in which case*, *we recall*, *our algebras are assumed finitary*), *the two conditions are equivalent.*

(25) $\mathbf{P}$ *can be generated*, *as a prevariety*, *by a set of* $\leq \kappa$ *algebras.*

(26) *Every set* $\mathbf{X}$ *of subdirectly irreducible algebras in* $\mathbf{P}$ *can be written as the union of* $\leq \kappa$ *subsets* $\mathbf{X}_\alpha$ $(\alpha \in \kappa)$, *each of which is* $\mathbf{P}$*-compatible.*

Proof. (25)⇒(26): Suppose $\mathbf{P}$ is generated by a set of $\leq \kappa$ algebras, $\mathbf{Y} = \{B_\alpha \mid \alpha \in \kappa\}$, and that as in (26), $\mathbf{X}$ is a set of subdirectly irreducible algebras in $\mathbf{P}$. Each $A \in \mathbf{X}$ is embeddable in a direct product of copies of the B_α, hence, being subdirectly irreducible, in one of the B_α. Letting $\mathbf{X}_\alpha$ be the set of members of $\mathbf{X}$ embeddable in B_α, we get the conclusion of (26) (using the second formulation in the definition of $\mathbf{P}$-compatibility).

To prove that when $\mathbf{P}$ is a quasivariety, (26)⇒(25), note that by our residual smallness hypothesis, there is a set $\mathbf{X}$ of subdirectly irreducible algebras in $\mathbf{P}$ which contains, up to isomorphism, all such algebras. By (26) we may write $\mathbf{X} = \bigcup_{\alpha\in\kappa} \mathbf{X}_\alpha$ where each $\mathbf{X}_\alpha$ is $\mathbf{P}$-compatible. Hence for each α, we can choose an algebra A_α in $\mathbf{P}$ in which all members of $\mathbf{X}_\alpha$ can be embedded. Since a quasivariety is generated as a prevariety by its subdirectly irreducible algebras [Gorbunov 1998, Theorem 3.1.1], the prevariety generated by $\{A_\alpha \mid \alpha \in \kappa\}$ is all of $\mathbf{P}$. □

To get easy examples showing that the least κ for which (26) holds can be, inter alia, any natural number, consider algebras with a single unary operation a, and for each positive integer d, let C_d be the algebra of this type consisting of d elements, $x, ax, \dots, a^{d-1}x$, cyclically permuted by a.

Now let n be any natural number, and let $d_1, \ldots, d_n$ be positive integers none of which is the least common multiple of any subset of the others. (In particular, none of them is 1, since 1 is the least common multiple of the empty set.) Let $\mathbf{P}$ be the prevariety generated by the n algebras $C_{d_1}, \ldots, C_{d_n}$. Since this is generated by finitely many finite finitary algebras, it is a quasivariety. From the description $\mathbf{P} = \mathbb{S}\mathbb{P}\{C_{d_1}, \ldots, C_{d_n}\}$ we see that all algebras in $\mathbf{P}$ satisfy

(27) $(\forall x)\ a^{\mathrm{lcm}(d_1,\ldots,d_n)} x = x,$

(28) $(\forall x, y, z)\ ax = x \implies y = z,$

(29) For all x, the least positive integer d such that $a^d x = x$ is the least common multiple of some subset of $\{d_1, \ldots, d_n\}$,

(30) $(\forall x, y)\ a^d x = x \implies a^d y = y.$

From (27)–(30) and our assumption that none of the d_i is the least common multiple of a subset of the rest, one can verify that the subdirectly irreducible objects of $\mathbf{P}$ are precisely the n algebras C_{d_i}; and by (30) these are pairwise incompatible; so for this quasivariety, the least κ as in Theorem 21 is n.

For $d_1, \ldots, d_n$ as above, consider next, for contrast, the quasivariety $\mathbf{P}$ generated by a single algebra, the disjoint union $C_{d_1} \sqcup \cdots \sqcup C_{d_n}$. This will still satisfy (27)–(29), but not (30). The algebras $C_{d_1}, \ldots, C_{d_n}$ will still be subdirectly irreducible in $\mathbf{P}$, but they are no longer incompatible. Indeed, since $\mathbf{P}$ is generated by a single algebra, the least cardinal κ as in Theorem 21 is now 1.

For an intermediate case, given d_1, d_2, d_3 as above, let $\mathbf{P}$ be generated by the three disjoint unions $C_{d_1} \sqcup C_{d_2}$, $C_{d_1} \sqcup C_{d_3}$ and $C_{d_2} \sqcup C_{d_3}$. Since none of these generating algebras contains copies of all three C_{d_i}, these algebras, and hence all algebras in $\mathbf{P}$, satisfy the implication

(31) $(\forall x_1, x_2, x_3, y, z)\ (a^{d_1} x_1 = x_1) \wedge (a^{d_2} x_2 = x_2) \wedge (a^{d_3} x_3 = x_3) \implies y = z.$

Hence, though any two of C_{d_1}, C_{d_2} and C_{d_3} are $\mathbf{P}$-compatible, the set consisting of all three is not.

If we take an infinite sequence of integers $d_1, d_2, \ldots$, none of which divides any of the others (for instance, the primes), and let $\mathbf{P}$ be the prevariety generated by all *finite* disjoint unions of the C_{d_i}, this will no longer be a quasivariety. For it will satisfy the sentence

(32) $(\forall x_1, \ldots, x_n, \ldots; y; z)\ \left(\bigwedge_{i=1}^{\infty} a^{d_i} x_i = x_i \right) \implies y = z,$

so the direct limit, as sets with one unary operation, of the above generating family of finite unions of C_i (mapped into one another by inclusion) does not lie in $\mathbf{P}$; their direct limit in $\mathbf{P}$ is, by (32), the trivial (one-element) algebra. This example also shows that in our earlier result, Lemma 17, the added direct limit hypothesis was indeed needed to get from (20)–(22) to (23). For it is easy to see that $\mathbf{P}$

satisfies (20), while to see that it does not satisfy (23), we may take for S and the A_i the empty algebra, and for the B_i the above algebras C_{d_i}.

To see that the implication (26)⇒(25) of Theorem 21, which we proved for quasivarieties, does not hold for general prevarieties, let us construct a prevariety not having "enough" subdirectly irreducible algebras: Let p be a prime, let G be the additive group of an infinite-dimensional vector space over the field of p elements, and let $\mathbf{P}$ be the prevariety consisting of all G-sets A such that if an element of A is fixed by an element of G, then all elements of A are fixed by that element, and if an element of A is fixed by infinitely many elements of G, then all elements of A are equal. Then $\mathbf{P}$ is residually small: the set of G-sets G/H for all finite subgroups $H \subset G$ generates $\mathbf{P}$. But any finite-dimensional subspace of G is an intersection of two properly larger finite dimensional subspaces, hence any nontrivial algebra in $\mathbf{P}$ can be decomposed as a subdirect product of algebras with larger pointwise stabilizers; so $\mathbf{P}$ has no subdirectly irreducible algebras, so it satisfies (26) for $\kappa = 0$. On the other hand, no two nonempty algebras in $\mathbf{P}$ having different pointwise stabilizers are compatible, so (25) does not hold for any finite κ.

8. All under one roof: prevarieties where all algebras are P-compatible

For prevarieties that can be generated by *one* algebra, a stronger result can be proved than the $\kappa = 1$ case of (26); moreover, we can weaken the above assumption "generated by one algebra" to a condition that is necessary as well as sufficient for our strengthened conclusion.

We need the following definition. (Recall that a preordering $\preceq$ on a set or class means a reflexive, transitive, but not necessarily antisymmetric binary relation.)

Definition 22. *A preordered* class $(\mathbf{K}, \preceq)$ *will be called* absolutely directed *if every* set *of elements of* $\mathbf{K}$ *is majorized by some element of* $\mathbf{K}$.

In particular, a preordered *set* is absolutely directed if and only if it has a greatest element (an element $\succeq$ all elements).

In the next result, condition (33) can be seen to be a strengthening of the $\kappa = 1$ case of (26) (with "nontrivial" replacing "subdirectly irreducible"), while (34) is a weakening of the condition that $\mathbf{P}$ be generated as a prevariety by a single algebra. The equivalence of (35) and (37) for quasivarieties is due to Mal'cev.

By the *trivial* algebra we will always mean the one-element algebra. (So trivial algebras and empty algebras are never the same thing.)

Theorem 23 [Mal'cev 1966; Gorbunov 1998, Proposition 2.1.19]. *Let* $\mathbf{P}$ *be a prevariety, and let* $\preceq$ *be the preordering "is embeddable in" among algebras in* $\mathbf{P}$. *Then the following conditions are equivalent.*

(33) *Every set of nontrivial algebras in* $\mathbf{P}$ *is* $\mathbf{P}$*-compatible.*

(34) $\mathbf{P}$ *is generated as a prevariety by a class of algebras absolutely directed under* $\preceq$.

Moreover, if $\mathbf{P}$ *is a quasivariety* (*so that, again, our algebras are assumed finitary*), *then those conditions are also equivalent to each of the following*:

(35) *Every pair of nontrivial algebras in* $\mathbf{P}$ *is* $\mathbf{P}$*-compatible.*

(36) $\mathbf{P}$ *is generated as a* quasivariety *by a class of algebras absolutely directed under* $\preceq$.

(37) $\mathbf{P}$ *is generated as a quasivariety by a single algebra.*

Proof. (33) says that the class of nontrivial algebras in $\mathbf{P}$ is absolutely directed under $\preceq$. But $\mathbf{P}$ is generated as a prevariety by its nontrivial algebras (the trivial algebra being the direct product of the empty family thereof), so this implies (34).

To show (34)⇒(33), let $\mathbf{X}$ be an absolutely directed class of algebras generating $\mathbf{P}$, and $\mathbf{Y}$ any set of nontrivial algebras in $\mathbf{P}$. Since $\mathbf{Y}$ is a set, we can find a set $\mathbf{X}_0 \subseteq \mathbf{X}$, homomorphisms into members of which separate points of algebras in $\mathbf{Y}$, and by the absolute directedness of $\mathbf{X}$, some one algebra $A \in \mathbf{X}$ contains embedded images of all members of $\mathbf{X}_0$; hence homomorphisms into A separate points of algebras in $\mathbf{Y}$. Hence if we form a direct product A^I of sufficiently many copies of A, then for each nonempty $B \in \mathbf{Y}$, we can use maps to some coordinates of A^I to separate points of B; and since B is nontrivial and nonempty, the set of maps so used will be nonempty, and we can repeat some of them to fill in the remaining coordinates if any; thus we can embed B in A^I. The same conclusion is vacuously true if B is empty, so A^I has subalgebras isomorphic to all $B \in \mathbf{Y}$, proving that $\mathbf{Y}$ is $\mathbf{P}$-compatible.

Now let $\mathbf{P}$ be a quasivariety.

Clearly, (33)⇒(35). The converse holds because we can go from pairwise coproducts to finite coproducts by induction, while coproducts of infinite families are direct limits of coproducts of finite families, and in a quasivariety, direct limits respect the underlying set functor. Thus, (33)–(35) are equivalent.

(34)⇒(36) is trivial, since the quasivariety generated by a class of algebras contains the prevariety generated by the same class. We shall now show (36)⇒(37), then note two alternative ways of getting back from (37): to (34), or to (35).

Given (36), let $\mathbf{X}$ be an absolutely directed class of algebras generating $\mathbf{P}$ as a quasivariety. Since the *finite* sentences (12) form (modulo notation) a *set*, if we choose for each such sentence not satisfied by $\mathbf{P}$ a member of $\mathbf{X}$ for which it fails, we get a set of algebras $\mathbf{X}_0 \subseteq \mathbf{X}$ which again generates $\mathbf{P}$. By assumption, $\mathbf{X}_0$ is majorized by an algebra $A \in \mathbf{X}$, and this will likewise generate $\mathbf{P}$, proving (37).

Now assume (37), and let A be an algebra that generates $\mathbf{P}$ as a quasivariety.

On the one hand, one can deduce (34) from the fact that $\mathbf{P}$ is generated as a prevariety by the class of ultrapowers of A ([Gorbunov 1998, Corollary 2.3.4(i)];

compare last paragraph of Definition 5 above) by verifying that that class is absolutely directed under $\preceq$. The idea is that given a set of ultrafilters $\mathcal{U}_j$ ($j \in J$), each on a set I_j, these yield a "product" filter on $\prod^J I_j$, and any ultrafilter $\mathcal{U}$ containing this will have the property that all the ultrapowers $A^{\mathcal{U}_j}$ embed in the ultrapower $A^{\mathcal{U}}$.

To get (35), on the other hand, suppose by way of contradiction that B_0 and B_1 were a non-$\mathbf{P}$-compatible pair of nontrivial algebras in $\mathbf{P}$. Without loss of generality, suppose B_0, has non-one-to-one coprojection into $B_0 \amalg_{\mathbf{P}} B_1$; let elements $x \neq y$ of B_0 fall together there. Since $\mathbf{P}$ is determined by finite sentences (12), the conjunction of finitely many of these universal sentences with finitely many equations holding among finitely many elements of B_0 and B_1 must imply $x = y$. But every finite system of relations among elements of each of the B_i is realizable by relations among some family of elements of A (otherwise A, and hence $\mathbf{P}$, would satisfy an implication (12) saying that the conjunction of such a system of relations implies that all elements are equal, contradicting our assumption that the B_i are nontrivial). On the other hand, since $\mathbf{P}$ does not satisfy an implication forcing $x = y$ to hold in B_0, the above equations involving elements of B_0 must be satisfiable by a family of elements of A with *distinct* elements representing x and y. But combining this family with the family of elements of A chosen above to satisfy our finitely many relations holding in B_1, we see that the sentences (12) defining $\mathbf{P}$ imply that those two elements *are* equal, giving the required contradiction. $\square$

In the above theorem we had to exclude the trivial algebra from certain statements. The following addendum to that theorem shows that in many prevarieties, not only is that restriction unnecessary, but trivial algebras can be used in formulating a very simple criterion, (39), for the equivalent conditions of the theorem to hold.

Corollary 24. *In the context of Theorem 23, suppose that*

(38) $\quad$ $\mathbf{P}$ *has at least one* nontrivial *algebra with a* trivial *subalgebra*

(*that is, a nontrivial algebra with an element idempotent under all the algebra operations*).

Then (33)–(34), and, if $\mathbf{P}$ *is a quasivariety, (35)–(37) are equivalent to the condition obtained by deleting the word "nontrivial" from (33); and also to*

(39) $\quad$ *Every algebra in* $\mathbf{P}$ *is* $\mathbf{P}$*-compatible with the trivial algebra.*

Proof. It is easy to deduce from (38) that each of (33) and (35) is equivalent to the formally strengthened version of itself gotten by deleting the restriction "nontrivial": given a set $\mathbf{X}$ of algebras (respectively, a pair of algebras) of $\mathbf{P}$ including the trivial algebra, which we want to embed simultaneously in some algebra, we "sneak the trivial member of our set in" by hiding it in a nontrivial algebra as in (38), then apply (33) (respectively, (35)) to the resulting family of nontrivial algebras. As a special case of this version of either condition, we have (39).

On the other hand, given (39), we can get the strengthened form of (33) by a version of the construction by which one embeds a family of groups in their direct product group. Let $\{B_i \mid i \in I\}$ be any set of algebras in $\mathbf{P}$. By (39), embed each B_i in an algebra A_i containing an idempotent element e_i. Taking $A = \prod^I A_i$, we can embed each B_j in A by using the inclusion map at the jth component, and mapping to every other component by collapsing everything to a trivial subalgebra. □

Clearly *every* prevariety of groups, monoids, or lattices satisfies (39), hence satisfies (33) with the nontriviality condition deleted, (34), and, if it is a quasivariety, (35)–(37).

On the other hand, the variety $\mathbf{V}$ of unital associative (or unital associative commutative) algebras over any field satisfies (33) (and hence (34)–(37)), by the standard description of coproducts of such algebras, but not (38) or the version of (33) with "nontrivial" deleted; rather, the trivial algebra (with $1 = 0$) is not $\mathbf{V}$-compatible with any other algebra. Hence in the absence of (38), the exclusion of the trivial algebra in (33) and (35) is indeed needed to make Theorem 23 hold. Our constructions in the preceding section with unary algebras also illustrate this: in the prevariety generated by a single algebra C_d ($d > 1$), the conditions of Theorem 23 hold, but C_d satisfies $(\forall x, y, z)\ ax{=}x \Rightarrow y{=}z$, so the trivial algebra is not $\mathbf{P}$-compatible with any nontrivial algebra.

There are also examples where (38) holds, but where the equivalent conditions of Theorem 23 and Corollary 24 do not; for example, the variety of groups or monoids with one distinguished element, or of lattices with two distinguished elements: a counterexample to (39) is given by any group or monoid with distinguished element that is not the identity, or any lattice with a pair of distinguished elements that are not equal.

These same examples also show that the analog of the implication (37)⇒(33) does not hold for varieties, with "generated as a variety" in place of "generated as a quasivariety", since every variety satisfies the analog of (37).

The next corollary is a result promised in the comment following Lemma 11.

Corollary 25. *Suppose* $\mathbf{P}$ *is a prevariety generated by a single algebra, or, more generally, satisfying* (34), *and let* $(B_i)_{i\in I}$ *be a family of nontrivial algebras in* $\mathbf{P}$. (*Again, if* $\mathbf{P}$ *satisfies* (38), *the restriction "nontrivial" can be dropped.*) *Then*

(40) *for every* $J \subseteq I$, *the natural map* $\coprod_{\mathbf{P}}^{i\in J} B_i \to \coprod_{\mathbf{P}}^{i\in I} B_i$ *is one-to-one.*

Hence

(41) *any subfamily of a* $\mathbf{P}$*-independent family of subalgebras of a nontrivial algebra* A *is* $\mathbf{P}$*-independent.*

Proof. To see (40), note that $\coprod_{\mathbf{P}}^{i\in I} B_i \cong \big(\coprod_{\mathbf{P}}^{i\in J} B_i\big) \amalg_{\mathbf{P}} \big(\coprod_{\mathbf{P}}^{i\in I-J} B_i\big)$, with the natural map $\coprod_{\mathbf{P}}^{i\in J} B_i \to \coprod_{\mathbf{P}}^{i\in I} B_i$ corresponding to the first coprojection under this decomposition. By the implication (34)⇒(33) (or its modified version if $\mathbf{P}$ satisfies (38)), the indicated coproducts over J and $I-J$ are $\mathbf{P}$-compatible, hence that coprojection map is one-to-one, as claimed. (A slight hiccup in this argument: If J or $I-J$ is empty, can we be sure the coproduct over that subset, namely the initial algebra, is nontrivial? No, but if it is trivial, and if $\mathbf{P}$ is not the trivial prevariety, then since the initial algebra can be mapped into every algebra, (38) holds, and so we are in the case where we don't need nontriviality.)

To get (41), recall that the statement that $(B_i)_{i\in I}$ is an independent family of subalgebras of A means that the subalgebra of A generated by these subalgebras can be identified with their coproduct. If none of the B_i is trivial, then this observation together with (40) immediately gives the desired conclusion. If at least one of the B_i is trivial, then since by assumption A is not, (38) holds, and by the parenthetical addendum to the first part of this corollary, we again have (40) and can proceed as before. □

On a different topic, let us note the extent to which Theorem 21 does and does not go over from prevarieties to quasivarieties.

Corollary 26. *If* $\mathbf{P}$ *is a* quasivariety, *then* (*even without the residual smallness assumption of Theorem 21*), *condition* (26) *implies*

(42) $\mathbf{P}$ *can be generated as a quasivariety by* $\leq \kappa$ *algebras.*

The reverse implication holds if κ *is finite*, *but not for any infinite* κ.

Proof of (26)⇒(42). A quasivariety $\mathbf{P}$ is generated as a prevariety, and hence as a quasivariety, by its subdirectly irreducible algebras [Gorbunov 1998, Theorem 3.1.1], hence, as in the proof of Theorem 23, we can find a *set* $\mathbf{X}$ of these that generates it as a quasivariety. By (26) we can write $\mathbf{X}$ as $\bigcup_{\alpha\in\kappa} \mathbf{X}_\alpha$ where each $\mathbf{X}_\alpha$ is $\mathbf{P}$-compatible. If for each $\alpha \in \kappa$ we let A_α be an algebra in $\mathbf{P}$ containing embedded images of all members of $\mathbf{X}_\alpha$, then $\mathbf{P}$ is generated as a quasivariety by this set of κ algebras.

For the converse assertion when κ is a natural number n, let $\mathbf{P}$ be generated as a quasivariety by $A_1, \ldots, A_n$. Then $\mathbf{P} = \mathbb{S}\mathbb{P}\mathbb{P}_{\mathrm{ult}}\{A_1, \ldots, A_n\}$, where $\mathbb{P}_{\mathrm{ult}}$ denotes closure under ultraproducts; thus, each $\mathbf{P}$-subdirectly irreducible object of $\mathbf{P}$ is embeddable in a member of $\mathbb{P}_{\mathrm{ult}}\{A_1, \ldots, A_n\}$. Moreover, the operator $\mathbb{P}_{\mathrm{ult}}$ respects finite decompositions; that is, any ultraproduct of a family of structures indexed by a finite union of sets, $(A_i)_{i\in I_1\cup\cdots\cup I_n}$, can be written as an ultraproduct of one of the subfamilies $(A_i)_{i\in I_m}$. Hence $\mathbb{P}_{\mathrm{ult}}\{A_1, \ldots, A_n\} = \mathbb{P}_{\mathrm{ult}}\{A_1\} \cup \cdots \cup \mathbb{P}_{\mathrm{ult}}\{A_n\}$. The class of subdirectly irreducible algebras in $\mathbf{P}$ that are embeddable in members of a given $\mathbb{P}_{\mathrm{ult}}\{A_i\}$ is contained in the one-generator quasivariety $\mathbb{S}\mathbb{P}\mathbb{P}_{\mathrm{ult}}\{A_i\}$, hence by the implication (37)⇒(33), each of these n classes has the property that all its subsets are **P**-compatible. This gives (26).

On the other hand, given any infinite κ, let **V** be the variety of sets given with a κ-tuple of 0-ary operations (constants) c_α $(\alpha \in \kappa)$. Since as a quasivariety, **V** is generated by its finitely presented objects [Gorbunov 1998, Proposition 2.1.18], and there are only κ of these, it satisfies (42). On the other hand, since all operations are 0-ary, every equivalence relation on a **V**-algebra is a congruence, so the subdirectly irreducible algebras are precisely the two-element algebras. We have one of these for each 2-class equivalence relation on κ, and one more corresponding to the partition of that set into κ and $\varnothing$. This gives 2^κ subdirectly irreducible algebras, no two of which are **V**-compatible. □

9. Afterthoughts on P-compatible algebras

Perhaps the concept of "**P**-compatible algebras" is not the best handle on the phenomena we have been examining; or at least should be complemented by another way of looking at them. Suppose that for algebras A and B in **P**, we say that A is "comfortable" with B in **P** if the coprojection $A \to A \amalg_{\mathbf{P}} B$ is one-to-one; equivalently, if A is **P**-compatible with some homomorphic image of B in **P**. This relation is not in general symmetric; for example, in the variety of associative unital rings, each ring $\mathbb{Z}/n\mathbb{Z}$ is comfortable with $\mathbb{Z}$, but not vice versa. Algebras A and B are **P**-compatible if and only if each is comfortable with the other. (So the relations of **P**-compatibility and of being comfortable in **P** may be characterized in terms of one another.) More generally, an arbitrary family of algebras is **P**-compatible if and only if each is comfortable with the coproduct of the others.

If **P** is generated as a prevariety by a class of algebras **X**, we see that an algebra A is comfortable in **P** with an algebra B if and only if homomorphisms into members of **X** that contain homomorphic images of B separate points of A. Hence, if we classify algebras $B \in \mathbf{P}$ according to which algebras A are comfortable with them, then algebras B_1 and B_2 will belong to the same equivalence class under this relation if the subclass of **X** consisting of algebras containing homomorphic images of B_1 coincides with the subclass of those containing images of B_2. (We do not assert the converse.) In particular, if **P** is residually small, so that **X** can be taken to be a set, the number of these equivalence classes has the cardinality of a set. More generally, if **P** is generated by the union of κ classes of algebras, each absolutely directed under the relation $\preceq$ of Theorem 23, the same reasoning shows that it will have at most 2^κ equivalence classes under this equivalence relation. On the other hand, if we classify algebras according to which other algebras they *are comfortable with*, we may, so far as I can see, get up to 2^{2^κ} classes.

For any algebra A in $\mathbf{P}$, the class of algebras which are comfortable with A forms a subprevariety of $\mathbf{P}$. The class of algebras that A is comfortable with likewise yields a subprevariety on throwing in the trivial algebra. (A stronger statement, also easy to see, is that this class is closed under taking subalgebras and under taking products with arbitrary algebras in $\mathbf{P}$; equivalently, that if this class contains an algebra B, then it contains every algebra in $\mathbf{P}$ admitting a homomorphism to B).

10. On infinite symmetric groups: an answer and a question

This last section does not depend on any of the preceding material.

It was shown in [de Bruijn 1957] (see also [Bergman 2007]) that for Ω an infinite set, the group $S = \mathrm{Sym}(\Omega)$ of all permutations of Ω contains a coproduct of two copies of itself (from which it was deduced by other properties of that group that it contains a coproduct of $2^{\mathrm{card}(\Omega)}$ copies of itself). In [Bergman 2007, Question 4.4], I asked, inter alia, whether, for every subgroup B of S, if we regard S as a member of the variety of groups given with homomorphisms of B into them, S contains a coproduct of two copies of itself *in that variety*.

The answer is negative. To see this, pick any $x \in \Omega$ and let B be the stabilizer in S of x. Writing elements of S to the left of their arguments and composing them accordingly, we see that the partition of S into left cosets of B classifies elements according to where they send x, and that for each $y \in \Omega$, the coset sending x to y has elements of finite order; for example, if $y \neq x$, the 2-cycle interchanging x and y.

On the other hand, I claim that if S_1 and S_2 are any two groups with a common subgroup B proper in each, then in the coproduct with amalgamation $S_1 \amalg_B S_2$ there are left cosets of B containing no elements of finite order. Indeed, the standard normal form in that coproduct shows that each left coset is generated by a possibly empty alternating string of left coset representatives of B in S_1 and S_2. When that string is nonempty and has even length, one sees that elements of finite order cannot occur. Hence for $S = \mathrm{Sym}(\Omega)$ and B as above, S, as a group containing B, cannot contain a copy of $S_1 \amalg_B S_2$.

So let us modify our earlier question.

Question 27. For Ω an infinite set, what nice conditions, if any, on a subgroup $B \subseteq S = \mathrm{Sym}(\Omega)$ will imply that S has a subgroup containing B and isomorphic over B to $S \amalg_B S$?

For instance, will this hold if B is equal to, or contained in, the stabilizer of a subset of Ω having the same cardinality as Ω? If B is finite?

In that same question in [Bergman 2007], I asked whether for any submonoid B of the monoid $\mathrm{Self}(\Omega)$ of self-maps of Ω, the monoid $\mathrm{Self}(\Omega)$ must contain, over B, a coproduct of two copies of itself with amalgamation of B. It seems likely that the subgroup $B \subset \mathrm{Sym}(\Omega) \subset \mathrm{Self}(\Omega)$ used above also gives a counterexample to this part of the question. This will be so if we can show that the subgroup of invertible elements of the monoid coproduct of two copies of $\mathrm{Self}(\Omega)$ with amalgamation of B is isomorphic to the group coproduct of two copies of $\mathrm{Sym}(\Omega)$ with amalgamation of B, since we have seen that this is not embeddable over B in the group $\mathrm{Sym}(\Omega)$ of invertible elements of $\mathrm{Self}(\Omega)$. But the analysis of coproducts of monoids with amalgamation, even when the submonoid being amalgamated is a group, seems difficult.

The final part of that question posed the same problem for the endomorphism algebra of an infinite-dimensional vector space over a field. To this I also do not know the answer; and in view of the results of [Wehrung 2007], it is natural to ask the same question for lattice of equivalence relations on an infinite set.

11. Glossary for the nonexpert in universal algebra

I indicate below the meanings of some basic concepts of universal algebra, though more briefly and informally than would be done in a textbook presentation. (Definitions of some other concepts are recalled in the sections where they are used. I do not define concepts of category theory, such as coproduct; or of set theory, such as ultraproduct, and the distinction between sets and proper classes. For these, see standard references such as [Mac Lane 1971; Chang and Keisler 1990].)

An *n-ary operation* on a set X means a function $X^n \to X$; here n is called the *arity* of the operation. An *algebra* is a set given with a family of operations of specified arities. The list of operation-symbols and their arities is the *type* of the algebra (used here only in the phrase "algebras of the same type"). Constants in the definition of an algebra structure (for example, the 0 and 1 of a ring structure) are in this note regarded as 0-*ary* operations; indeed, X^0 is a one-element set, so a map $X^0 \to X$ specifies an element of X. Given a subset S of an algebra X, the *subalgebra* of X generated by S is here denoted $\langle S \rangle$.

The given operations of an algebra are called *primitive* operations. Expressions in a family of variable-symbols and iterated applications of the primitive operations determine *derived operations*. Such expressions are themselves called *terms*. For instance, $(xy)z$ and $x(yz)$ are distinct ternary terms in the operations of a group. (They must be considered distinct so that they can be used to write the group identity of associativity.) The variable-symbols are also considered terms; they are the starting-point for the recursive construction of all terms. This technical sense of *term* will not stop us from using the word in other ways, for example, in referring to the mth term of a sequence.

An algebra all of whose primitive operations have finite arity is called *finitary*. (This does not preclude there being infinitely *many* primitive operations; for example, we have this for modules over an infinite ring.)

As indicated in Section 1, a *variety* of algebras is the class of all algebras of a given type satisfying a given set of identities. In any variety $\mathbf{V}$, one can construct a *free* algebra on any set, satisfying the usual universal property.

The above concepts are assumed from Section 2 on. Starting with Section 3, we also refer to the variety of algebras *generated* by a family $\mathbf{X}$ of algebras of a given type, that is, the least variety containing $\mathbf{X}$. This is clearly the class of all algebras that satisfy all identities satisfied by all members of $\mathbf{X}$. *Birkhoff's Theorem* states that it is also the class of all homomorphic images of subalgebras of (generally infinite) direct products of members of $\mathbf{X}$, abbreviated $\mathbb{H}\mathbb{S}\mathbb{P}(\mathbf{X})$. (Definitions of *prevariety* and *quasivariety*, and results for these analogous to Birkhoff's Theorem, are recalled in Section 3.)

To motivate a concept used from Section 7 on, note that if an algebra A is *embedded* in a direct product $\prod_I A_i$, by a homomorphism with components $f_i : A \to A_i$, then $A \cong f(A) \subseteq \prod_I f_i(A)$. Modeled on the properties of this subalgebra, one defines a *subdirect product* of a family of algebras $(B_i)_{i\in I}$ to be a subalgebra of $\prod_I B_i$ which projects surjectively to each B_i. An algebra that, up to isomorphism, cannot be so expressed without one of the projection maps being an isomorphism is called *subdirectly irreducible*.

References

[Adámek and Sousa 2004] J. Adámek and L. Sousa, "On reflective subcategories of varieties", *J. Algebra* **276**:2 (2004), 685–705. MR 2005c:18003 Zbl 1056.08003

[Bergman 2007] G. M. Bergman, "Some results on embeddings of algebras, after de Bruijn and McKenzie", *Indag. Math.* (*N.S.*) **18**:3 (2007), 349–403. MR 2008m:08016 Zbl 1130.08002

[de Bruijn 1957] N. G. de Bruijn, "Embedding theorems for infinite groups", *Nederl. Akad. Wetensch. Proc. Ser. A.* **60** = *Indag. Math.* **19** (1957), 560–569. MR 20 #4589

[Chang and Keisler 1990] C. C. Chang and H. J. Keisler, *Model theory*, 3rd ed., Studies in Logic and the Foundations of Mathematics **73**, North-Holland, Amsterdam, 1990. MR 91c:03026 Zbl 0697.03022

[Clark 1969] D. M. Clark, "Varieties with isomorphic free algebras", *Colloq. Math.* **20** (1969), 181–187. MR 39 #5445 Zbl 0191.00902

[Clark and Davey 1998] D. M. Clark and B. A. Davey, *Natural dualities for the working algebraist*, Cambridge Studies in Advanced Mathematics **57**, Cambridge University Press, Cambridge, 1998. MR 2000d:18001 Zbl 0910.08001

[Cohn 1966] P. M. Cohn, "Some remarks on the invariant basis property", *Topology* **5** (1966), 215–228. MR 33 #5676 Zbl 0147.28802

[Gorbunov 1998] V. A. Gorbunov, *Algebraic theory of quasivarieties*, Siberian School of Algebra and Logic, Consultants Bureau, New York, 1998. MR 2001a:08004 Zbl 0986.08001

[Kiss et al. 1982] E. W. Kiss, L. Márki, P. Pröhle, and W. Tholen, "Categorical algebraic properties. A compendium on amalgamation, congruence extension, epimorphisms, residual smallness, and injectivity", *Studia Sci. Math. Hungar.* **18**:1 (1982), 79–140. MR 85k:18003 Zbl 0549.08001

[Kovács and Newman 1974] L. G. Kovács and M. F. Newman, "Hanna Neumann's problems on varieties of groups", pp. 417–431 in *Proceedings of the Second International Conference on the*

Theory of Groups (Canberra 1973), edited by M. F. Newman., Lecture Notes in Math. **372**, Springer, Berlin, 1974. MR 50 #4750 Zbl 0306.20028

[Mac Lane 1971] S. Mac Lane, *Categories for the working mathematician*, Graduate Texts in Mathematics **5**, Springer, New York, 1971. MR 50 #7275 Zbl 0705.18001

[Mal'cev 1966] A. I. Mal'cev, "Several remarks on quasivarieties of algebraic systems", *Algebra i Logika Sem.* **5**:3 (1966), 3–9. In Russian. MR 34 #5728

[McKenzie et al. 1987] R. N. McKenzie, G. F. McNulty, and W. F. Taylor, *Algebras, lattices, varieties, I*, Wadsworth & Brooks/Cole, Monterey, CA, 1987. MR 88e:08001 Zbl 0611.08001

[Neumann 1967] H. Neumann, *Varieties of groups*, Ergebnisse der Math. **37**, Springer, New York, 1967. MR 35 #6734 Zbl 0251.20001

[Świerczkowski 1961] S. Świerczkowski, "On isomorphic free algebras", *Fund. Math.* **50** (1961), 35–44. MR 25 #2017 Zbl 0104.25601

[Wehrung 2007] F. Wehrung, "Embedding coproducts of partition lattices", *Acta Sci. Math.* (*Szeged*) **73**:3-4 (2007), 429–443. MR 2008m:06011 Zbl 05368543

[Zaĭtsev 1992] M. V. Zaĭtsev, "Embeddability of relatively free Lie algebras", *Uspekhi Mat. Nauk* **47**:4(286) (1992), 191–192. In Russian; translated in *Russian Math. Surveys* **47** (1992), 236-237. MR 94c:17013 Zbl 0954.17500

Communicated by Ehud Hrushovski
Received 2008-06-10 Revised 2009-11-23 Accepted 2009-11-26

gbergman@math.berkeley.edu *University of California, Berkeley, Department of Mathematics, Berkeley CA 94720-3840, United States*

Exponential sums nondegenerate relative to a lattice

Alan Adolphson and Steven Sperber

Our previous theorems on exponential sums often did not apply or did not give sharp results when certain powers of a variable appearing in the polynomial were divisible by p. We remedy that defect in this paper by systematically applying *p-power reduction*, making it possible to strengthen and extend our earlier results.

1. Introduction

In the papers [AS 1987a; 1987b; 1989; 1990a; 1990b] we established properties of the L-functions of exponential sums on affine space $\mathbb{A}^n$ and the torus $\mathbb{T}^n$. The purpose of this article is to prove a general result that leads to a sharpening of the theorems of those papers.

Let p be a prime, let $q = p^r$, and let $\mathbb{F}_q$ be the field of q elements. Let $f \in \mathbb{F}_q[x_1^{\pm 1}, \dots, x_n^{\pm 1}]$ be a Laurent polynomial, say,

$$f = \sum_{j \in J} a_j x^j, \tag{1.1}$$

where $a_j \in \mathbb{F}_q^\times$ and J is a finite subset of $\mathbb{Z}^n$. Let $\mathbb{Z}\langle J\rangle$ be the subgroup of $\mathbb{Z}^n$ generated by the elements of J. By the basic theory of abelian groups, there exists a basis $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ for $\mathbb{Z}^n$ and integers $d_1, \dots, d_k$ such that $d_1\boldsymbol{u}_1, \dots, d_k\boldsymbol{u}_k$ is a basis for $\mathbb{Z}\langle J\rangle$. After a coordinate change on $\mathbb{T}^n$, we may assume that $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ is the standard basis. The Laurent polynomial f may then be written in the form

$$f = g(x_1^{d_1}, \dots, x_k^{d_k}),$$

for some $g \in \mathbb{F}_q[x_1^{\pm 1}, \dots, x_k^{\pm 1}]$. Write $d_i = p^{b_i} e_i$ for each i, where $b_i \geq 0$ and $(e_i, p) = 1$. Since raising to the p-th power is an automorphism of $\mathbb{F}_q$, the exponential sums associated to the polynomials f and $g(x_1^{e_1}, \dots, x_k^{e_k})$ are identical. Furthermore, the theorems in the aforementioned papers generally produce sharper results when applied to $g(x_1^{e_1}, \dots, x_k^{e_k})$ than when applied to f. (Thus there is no

MSC2000: primary 11L07; secondary 11T23, 14F30.
Keywords: exponential sums, p-adic cohomology.

improvement over our earlier work if $p \nmid [\mathbb{Z}^k : \mathbb{Z}\langle J\rangle]$.) We refer to $g(x_1^{e_1}, \dots, x_k^{e_k})$ as the *p-power reduction* of f.

Over $\mathbb{A}^n$, the technique of p-power reduction is less versatile because one cannot make the same sorts of coordinate changes. One has a standard toric decomposition $\mathbb{A}^n = \bigcup_{A \subseteq \{1,\dots,n\}} \mathbb{T}_A$, where $\mathbb{T}_A$ denotes the $|A|$-dimensional torus with coordinates $\{x_i\}_{i\in A}$. Given $f \in \mathbb{F}_q[x_1, \dots, x_n]$, one can try to analyze the corresponding exponential sum on $\mathbb{A}^n$ by analyzing its restriction to each of these tori, but the picture is complicated by the fact that p-power reduction may require different coordinate changes on different tori. It thus seems worthwhile to generalize our previous results to apply directly to the polynomial as given, to avoid the task of performing p-power reduction on a case-by-case basis.

Let M_J be the *prime-to-p saturation* of $\mathbb{Z}\langle J\rangle$,

$$M_J = \{u \in \mathbb{Z}^n \mid ku \in \mathbb{Z}\langle J\rangle \text{ for some } k \in \mathbb{Z} \text{ satisfying } (k, p) = 1\},$$

and let $\mathbb{R}\langle J\rangle$ denote the subspace of $\mathbb{R}^n$ spanned by the elements of J. We will get a strengthening of our earlier results when M_J is a proper subset of $\mathbb{Z}^n \cap \mathbb{R}\langle J\rangle$. Let

$$[\mathbb{Z}^n \cap \mathbb{R}\langle J\rangle : \mathbb{Z}\langle J\rangle] = p^a e,$$

where $a \geq 0$ and $(e, p) = 1$. Then

$$[\mathbb{Z}^n \cap \mathbb{R}\langle J\rangle : M_J] = p^a, \tag{1.2}$$

so $M_J \neq \mathbb{Z}^n \cap \mathbb{R}\langle J\rangle$ if and only if $a > 0$.

Part of the motivation for this work was a desire to understand Theorems 3.6.5 and 3.6.7 from [Katz 2005] from our point of view. Suppose that $f \in \mathbb{F}_q[x_1, \dots, x_n]$ is a homogeneous polynomial of degree $d = p^k e$, $(e, p) = 1$, $k \geq 1$. Katz showed that if $f = 0$ defines a smooth hypersurface in $\mathbb{P}^{n-1}$, then the L-function associated to the exponential sum defined by f (see Section 2 for the definition) is a polynomial (n odd) or the reciprocal of a polynomial (n even) of degree

$$\frac{1}{p^k}((d-1)^n + (-1)^n(p^k - 1)),$$

all of whose reciprocal roots have absolute value $q^{n/2}$. Note that in this situation $[\mathbb{Z}^n : M_J] = p^k$. Our results in [AS 1989] do not apply to polynomials of degree divisible by p. However, we show here that when M_J is a proper subset of $\mathbb{Z}^n$ one can weaken the definition of nondegeneracy used in that article and still deduce analogous conclusions. In particular, we show that the above theorem of Katz is true as well for nonhomogeneous polynomials, provided that the homogeneous part of highest degree defines a smooth hypersurface in $\mathbb{P}^{n-1}$ and $[\mathbb{Z}^n : M_J] = p^k$. In other words, the conclusion remains true when one perturbs the smooth homogeneous polynomial by adding arbitrary terms of degrees $p^k e'$, $e' < e$. (In earlier work, analogous results for exponential sums involving polynomials of degree divisible by p were proved under the additional assumption that the homogeneous form of second highest degree "behaved nicely" relative to the leading form: see [AS 2000; 2009; Rojas-León 2006].)

This generalization of Katz's theorem (Proposition 5.1 below) will be derived as a consequence of Theorem 4.17. Another consequence of that theorem is the following result. Consider the Dwork family of hypersurfaces

$$x_1^n + \cdots + x_n^n + \lambda x_1 \ldots x_n = 0$$

in $\mathbb{P}^{n-1}$. If $n = p^k e$, where $k \geq 1$ and $(p, e) = 1$, and $\lambda \neq 0$, this hypersurface is singular (except for $n = 2, 3$). We show (Corollary 5.9 below) that the zeta function of this hypersurface has the form

$$Z(t) = \frac{R(t)^{(-1)^{n-1}}}{(1-t)(1-qt)\ldots(1-q^{n-2}t)},$$

where $R(t)$ is a polynomial of degree

$$(p^k - 1)e^{n-1} + e^{-1}((e-1)^n + (-1)^n(e-1)),$$

all of whose reciprocal roots have absolute value $q^{(n-2)/2}$. (Zeta functions of the Dwork family have also been studied recently in [Rojas-León and Wan 2007] and [Katz 2007].)

As another example, we strengthen the classical theorem of Chevalley–Warning. Let $f = \sum_{j \in J} a_j x^j \in \mathbb{F}_q[x_1, \ldots, x_n]$ and let $N(f)$ denote the number of solutions of $f = 0$ with coordinates in $\mathbb{F}_q$. Let $\mathbb{N}$ denote the nonnegative integers, let $\mathbb{N}_+$ denote the positive integers, and let $J' = \{(j, 1) \in \mathbb{N}^{n+1} \mid j \in J\}$. Let $\Delta(J')$ denote the convex hull of $J' \cup \{(0, \ldots, 0)\}$ in $\mathbb{R}^{n+1}$.

Theorem 1.3. *Let μ be the smallest positive integer such that $\mu\Delta(J')$, the dilation of $\Delta(J')$ by the factor μ, contains a point of $M_{J'} \cap (\mathbb{N}_+)^{n+1}$. Then* $\operatorname{ord}_q N(f) \geq \mu - 1$, *where* ord_q *denotes the p-adic valuation normalized by* $\operatorname{ord}_q q = 1$.

For example, the equation $\sum_{i=1}^n x_i^{p^{k_i}} = 0$ has q^{n-1} solutions: since raising to the p-th power is an automorphism of $\mathbb{F}_q$, one can assign arbitrary values to $x_1, \ldots, x_{n-1}$ and there will be a unique value of x_n satisfying the equation. Since $M_{J'} = \mathbb{Z}\langle J'\rangle$ is the lattice generated by the $\{(0, \ldots, 0, p^{k_i}, 0, \ldots, 0, 1)\}_{i=1}^n$, $\mu = n$ and Theorem 1.3 gives the precise divisibility by q.

For a more subtle example, let $p = 3$, $n = 3$, and consider the polynomial

$$f = x_1 x_2^2 + x_2 x_3^2 + x_1^2 x_3.$$

The lattice $M_{J'} = \mathbb{Z}\langle J'\rangle$ is the rank-three sublattice of $\mathbb{Z}^4$ with basis the vectors

$$\boldsymbol{u}_1 = (1, 2, 0, 1), \boldsymbol{u}_2 = (0, 1, 2, 1), \boldsymbol{u}_3 = (2, 0, 1, 1).$$

The only point of $\Delta(J')\cap(\mathbb{N}_+)^4$ is $(1,1,1,1)$ and one has

$$(1,1,1,1)=\tfrac{1}{3}(\boldsymbol{u}_1+\boldsymbol{u}_2+\boldsymbol{u}_3), \tag{1.4}$$

thus $(1,1,1,1)\notin M_{J'}$. It follows that $\mu>1$, so Theorem 1.3 implies that $N(f)$ is divisible by 3^r. (In fact, $\boldsymbol{u}_1+\boldsymbol{u}_2\in M_{J'}\cap(\mathbb{N}_+)^4$, so $\mu=2$.) On the other hand, since the degree of f equals the number of variables, the classical Chevalley–Warning Theorem does not predict the divisibility of $N(f)$ by 3. If we take the same polynomial f but assume $p\neq 3$, then (1.4) shows that

$$(1,1,1,1)\in M_{J'},$$

so $\mu=1$ and Theorem 1.3 does not predict any divisibility by p.

Theorem 1.3 is a special case of Theorem 3.3, which we prove by the method of [Ax 1964], as applied in [AS 1990a]. Ax expresses an exponential sum as a sum of certain products of Gauss sums; Stickelberger's Theorem computes the valuation of each Gauss sum, so one must determine which of these products of Gauss sums has minimal valuation. This minimum is in general difficult to calculate directly, so one replaces the set of valuations by a larger set whose minimum is easier to calculate. We used a convexity argument in [AS 1990a, Lemma 1], which is the approach we take here. Another method for estimating this minimum is via the "p-weight" of the polynomial: see [Moreno et al. 2004, Section 4] for a description of this approach and references to related work. The results obtained from these two approaches do not seem comparable, that is, neither implies the other as far as we know.

The first main idea of this paper is that when computing the action of Dwork's Frobenius operator, which gives the L-function of the exponential sum on the torus, one can ignore the action of Frobenius on power series whose exponents lie outside of M_J since such power series contribute nothing to the spectral theory of Frobenius. This idea is explained in Section 2. The second main idea is the notion of nondegeneracy relative to a lattice, which is introduced in Section 4. It guarantees that the p-power reduction of f will be nicely behaved. This leads to precise formulas for the degree of the L-function and the number of roots of a given archimedean weight.

2. Trace formula

Let $\Psi:\mathbb{F}_q\to\mathbb{Q}(\zeta_p)$ be a nontrivial additive character and define

$$S_m(\mathbb{T}^n,f)=\sum_{x\in\mathbb{T}^n(\mathbb{F}_{q^m})}\Psi(\mathrm{Tr}_{\mathbb{F}_{q^m}/\mathbb{F}_q}(f(x))),$$

where $\mathrm{Tr}_{\mathbb{F}_{q^m}/\mathbb{F}_q}$ denotes the trace map. In the special case where $f \in \mathbb{F}_q[x_1, \dots, x_n]$, we can also define

$$S_m(\mathbb{A}^n, f) = \sum_{x \in \mathbb{A}^n(\mathbb{F}_{q^m})} \Psi(\mathrm{Tr}_{\mathbb{F}_{q^m}/\mathbb{F}_q}(f(x))).$$

There are corresponding L-functions

$$L(\mathbb{T}^n, f; t) = \exp\Bigl(\sum_{m=1}^{\infty} S_m(\mathbb{T}^n, f)\frac{t^m}{m}\Bigr)$$

and

$$L(\mathbb{A}^n, f; t) = \exp\Bigl(\sum_{m=1}^{\infty} S_m(\mathbb{A}^n, f)\frac{t^m}{m}\Bigr).$$

Let $\mathbb{Q}_p$ denote the field of p-adic numbers and $\mathbb{Z}_p$ the ring of p-adic integers. Set $\Omega_1 = \mathbb{Q}_p(\zeta_p)$. Then Ω_1 is a totally ramified extension of $\mathbb{Q}_p$ of degree $p-1$. Let K denote the unramified extension of $\mathbb{Q}_p$ of degree r and set $\Omega_0 = K(\zeta_p)$. The Frobenius automorphism $x \mapsto x^p$ of $\mathrm{Gal}(\mathbb{F}_q/\mathbb{F}_p)$ lifts to a generator τ of $\mathrm{Gal}(\Omega_0/\Omega_1)$ by setting $\tau(\zeta_p) = \zeta_p$. Let Ω be the completion of an algebraic closure of Ω_0. Let “ord” denote the additive valuation on Ω normalized by $\mathrm{ord}\, p = 1$ and let “ord_q” denote the additive valuation normalized by $\mathrm{ord}_q\, q = 1$.

Let $E(t) = \exp(\sum_{i=0}^{\infty} t^{p^i}/p^i)$ be the Artin–Hasse exponential series. Let $\gamma \in \Omega_1$ be a solution of $\sum_{i=0}^{\infty} t^{p^i}/p^i = 0$ satisfying $\mathrm{ord}\,\gamma = 1/(p-1)$ and set

$$\theta(t) = E(\gamma t) = \sum_{i=0}^{\infty} \lambda_i t^i \in \Omega_1[\![t]\!].$$

The series $\theta(t)$ is a splitting function in Dwork’s terminology and its coefficients satisfy

$$\mathrm{ord}\,\lambda_i \geq \frac{i}{p-1}. \tag{2.1}$$

Define the *Newton polyhedron* of f, written $\Delta(f)$, to be the convex hull in $\mathbb{R}^n$ of the set $J \cup \{(0, \dots, 0)\}$. Let $C(f)$ be the cone in $\mathbb{R}^n$ over $\Delta(f)$, that is, $C(f)$ is the union of all rays in $\mathbb{R}^n$ emanating from the origin and passing through $\Delta(f)$. For any lattice point $u \in C(f) \cap \mathbb{Z}^n$, let $w(u)$, the *weight* of u, be defined as the smallest positive real number (necessarily rational) such that $u \in w(u)\Delta(f)$, where $w(u)\Delta(f)$ denotes the dilation of $\Delta(f)$ by the factor $w(u)$. Then

$$w : C(f) \cap \mathbb{Z}^n \to \frac{1}{N}\mathbb{Z}$$

for some positive integer N. We fix a choice $\tilde{\gamma}$ of N-th root of γ and set $\tilde{\Omega}_0 = \Omega_0(\tilde{\gamma})$, $\tilde{\Omega}_1 = \Omega_1(\tilde{\gamma})$. We extend $\tau \in \mathrm{Gal}(\Omega_0/\Omega_1)$ to a generator of $\mathrm{Gal}(\tilde{\Omega}_0/\tilde{\Omega}_1)$ by setting $\tau(\tilde{\gamma}) = \tilde{\gamma}$. Let $\tilde{\mathcal{O}}_0$ be the ring of integers of $\tilde{\Omega}_0$.

Let M be a lattice such that $M_J \subseteq M \subseteq \mathbb{Z}^n \cap \mathbb{R}\langle J\rangle$, let $L = \mathrm{Hom}_{\mathbb{Z}}(M, \mathbb{Z})$, and let $\ell \in L$. We extend ℓ to a function on $\mathbb{Z}^n \cap \mathbb{R}\langle J\rangle$ as follows. For $u \in \mathbb{Z}^n \cap \mathbb{R}\langle J\rangle$ we have $p^a u \in M$ by (1.2), so we may define

$$\ell(u) = p^{-a}\ell(p^a u).$$

This definition identifies L with a subgroup of $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}^n \cap \mathbb{R}\langle J\rangle, p^{-a}\mathbb{Z})$. Define

$$M_0(f) = \{u \in \mathbb{Z}^n \cap C(f) \mid \operatorname{ord} \ell(u) \geq 0 \text{ for all } \ell \in L\}.$$

Note that $M_0(f) = M \cap C(f)$. For $i > 0$ let

$$M_i(f) = \{u \in \mathbb{Z}^n \cap C(f) \mid \inf_{\ell \in L}\{\operatorname{ord} \ell(u)\} = -i\}.$$

Note that since L has finite rank, the infimum over L always exists. Furthermore, we have $M_i(f) = \varnothing$ for $i > a$ and

$$\mathbb{Z}^n \cap C(f) = \bigcup_{i=0}^{a} M_i(f).$$

We consider the following spaces of power series (where $b \in \mathbb{R}$, $b \geq 0$, $c \in \mathbb{R}$, and $0 \leq i \leq a$):

$$\begin{aligned}
L_i(b,c) &= \left\{ \sum_{u \in M_i(f)} A_u x^u \mid A_u \in \Omega_0,\ \operatorname{ord} A_u \geq bw(u) + c \right\},\\
L_i(b) &= \bigcup_{c \in \mathbb{R}} L_i(b,c),\\
B_i &= \left\{ \sum_{u \in M_i(f)} A_u \tilde{\gamma}^{Nw(u)} x^u \mid A_u \in \tilde{\mathcal{O}}_0,\ A_u \to 0 \text{ as } u \to \infty \right\},\\
B'_i &= \left\{ \sum_{u \in M_i(f)} A_u \tilde{\gamma}^{Nw(u)} x^u \mid A_u \in \tilde{\Omega}_0,\ A_u \to 0 \text{ as } u \to \infty \right\}.
\end{aligned}$$

We also define $L(b,c)$, $L(b)$, B, B' as the unions of these spaces for $i = 0, \ldots, a$. Note that if $b > 1/(p-1)$, then $L_i(b) \subseteq B'_i$ and for $c \geq 0$, $L_i(b,c) \subseteq B_i$. Similarly $L(b) \subseteq B'$ and for $c \geq 0$, $L(b,c) \subseteq B$. Define a norm on B_i, $i = 0, \ldots, a$, as follows. If

$$\xi = \sum_{u \in M_i(f)} A_u \tilde{\gamma}^{Nw(u)} x^u,$$

then set

$$\|\xi\| = \sup_{u \in M_i(f)} |A_u|.$$

One defines a norm on B in an analogous fashion.

Let $\hat{f} = \sum_{j \in J} \hat{a}_j x^j$ be the Teichmüller lifting of f, that is, $\hat{a}_j^q = \hat{a}_j$ and the reduction of $\hat{f}$ modulo p is f. Set

$$F(x) = \prod_{j \in J} \theta(\hat{a}_j x^j), \qquad F_0(x) = \prod_{i=0}^{r-1} F^{\tau^i}(x^{p^i}).$$

The estimate (2.1) implies that $F(x)$ and $F_0(x)$ are well-defined and satisfy

$$F(x) \in L_0\Big(\frac{1}{p-1}, 0\Big), \quad F_0(x) \in L_0\Big(\frac{p}{q(p-1)}, 0\Big).$$

We define the operator ψ on series by

$$\psi\Big(\sum_{u \in \mathbb{Z}^n} A_u x^u\Big) = \sum_{u \in \mathbb{Z}^n} A_{pu} x^u.$$

Clearly, $\psi(L(b, c)) \subseteq L(pb, c)$.

Lemma 2.2. *For* $1 \leq i < a$ *we have*

$$\psi(L_i(b, c)) \subseteq L_{i+1}(b, c),$$

and for $i = a$ *we have*

$$\psi(L_a(b, c)) = 0.$$

Furthermore, the same assertions hold with $L_i(b, c)$ *replaced by* B_i'.

Proof. Let $\ell \in L$ and $pu \in M_i(f)$. Since $\operatorname{ord} \ell(pu) \geq -i$, it follows that $\operatorname{ord} \ell(u) \geq -i-1$. By definition of $M_i(f)$ the first inequality is an equality for some $\ell \in L$. The second inequality is then an equality also for that ℓ, hence $u \in M_{i+1}(f)$. □

The operator $\alpha = \psi^r \circ F_0$ is

- an $\widetilde{\Omega}_0$-linear endomorphism of the space B', and
- an Ω_0-linear endomorphism of $L(b)$ for $0 < b \leq p/(p-1)$.

The operator $\alpha_0 = \tau^{-1} \circ \psi \circ F_0$ is

- an $\widetilde{\Omega}_1$-linear endomorphism of B',
- an Ω_1-linear endomorphism of $L(b)$ for $0 < b \leq p/(p-1)$,
- an $\widetilde{\Omega}_0$-semilinear endomorphism of B', and
- an Ω_0-semilinear endomorphism of $L(b)$ for $0 < b \leq p/(p-1)$.

It follows from [Serre 1962] that the operators α^m and α_0^m acting on B' and $L(b)$ for $0 < b \le p/(p-1)$ have well defined traces. In addition, the Fredholm determinants $\det(I - t\alpha)$ and $\det(I - t\alpha_0)$ are well defined and p-adically entire. The Dwork trace formula asserts

$$S_m(\mathbb{T}^n, f) = (q^m - 1)^n \mathrm{Tr}(\alpha^m), \tag{2.3}$$

where α acts either on B' or on some $L(b)$, $0 < b \le p/(p-1)$, and the nontrivial additive character implicit on the left-hand side is given by

$$\Psi(x) = \theta(1)^{\mathrm{Tr}_{\mathbb{F}_q/\mathbb{F}_p}(x)}.$$

Let δ be the operator on formal power series with constant term 1 defined by $g(t)^\delta = g(t)/g(qt)$. Using the relationship $\det(I - t\alpha) = \exp(-\sum_{m=1}^\infty \mathrm{Tr}(\alpha^m)t^m/m)$, Equation (2.3) is equivalent to

$$L(\mathbb{T}^n, f; t)^{(-1)^{n-1}} = \det(I - t\alpha)^{\delta^n}. \tag{2.4}$$

Let Γ be the map on power series defined by

$$\Gamma\Big(\sum_{u\in\mathbb{Z}^n} A_u x^u\Big) = \sum_{u\in M_0(f)} A_u x^u.$$

Define $\tilde{\alpha} = \Gamma \circ \alpha$, an endomorphism of B_0' and $L_0(b)$ for $0 < b \le p/(p-1)$. The main technical result of this paper is the following.

Theorem 2.5. *If $M_J \subseteq M$, then as operator on B_0' and $L_0(b)$ for $0 < b \le p/(p-1)$ the map $\tilde{\alpha}$ satisfies*

$$S_m(\mathbb{T}^n, f) = (q^m - 1)^n \mathrm{Tr}(\tilde{\alpha}^m).$$

Equivalently,

$$L(\mathbb{T}^n, f; t)^{(-1)^{n-1}} = \det(I - t\tilde{\alpha})^{\delta^n}.$$

Proof. To fix ideas, we work with the space B'. Note that if $u \in M_0(f)$ and $v \in M_i(f)$, $1 \le i \le a$, then $u + v \in M_i(f)$. This shows that multiplication by F and F_0 are stable on B_i' for $i = 1, \dots, a$. Lemma 2.2 then implies that $\alpha(B_i') \subseteq B_{i+1}'$ for $i = 1, \dots, a-1$ and $\alpha(B_a') = 0$. It follows that any power of α acting on $\bigcup_{i=1}^a B_i'$ has trace 0, so on $\bigcup_{i=1}^a B_i'$ we have $\det(I - t\alpha) = 1$. From [Serre 1962, Proposition 9] we then get

$$\det(I - t\alpha \mid B') = \det\big(I - t\alpha \mid B'/\textstyle\bigcup_{i=1}^a B_i'\big).$$

Under the Banach space isomorphism $B_0' \cong B'/\bigcup_{i=1}^a B_i'$, the operator $\tilde{\alpha}$ is identified with the operator induced by α on $B'/\bigcup_{i=1}^a B_i'$. This proves the theorem. $\square$

3. First applications

To improve the results of [AS 1987a], one can replace the space $L(p/(p-1))$ and its associated counting function $W(k)$ used there by the space $L_0(p/(p-1))$ for the lattice M_J and its associated counting function

$$W_0(k) = \mathrm{card}\{u \in M_J \cap C(f) \mid w(u) = k/N\}. \tag{3.1}$$

But since the main results of [AS 1987a] are concerned with the n-torus $\mathbb{T}^n$, it is simpler to just replace f by its p-power reduction as described in the introduction. For example, the first inequality of Theorem 1.8 of that paper becomes the one in Theorem 3.2 below.

Theorem 3.2. *We have* $0 \le \deg L(\mathbb{T}^n, f; t)^{(-1)^{n-1}} \le n!\, V(f)/[\mathbb{Z}^n : M_J]$, *where* $V(f)$ *denotes the volume of* $\Delta(f)$ *relative to Lebesgue measure on* $\mathbb{R}^n$.

The second inequality of [AS 1987a, Theorem 1.8] can be similarly improved.

Suppose that $f \in \mathbb{F}_q[x_1, \dots, x_n]$ and let $\omega(f)$ be the smallest positive real (hence rational) number such that $\omega(f)\Delta(f)$, the dilation of $\Delta(f)$ by the factor $\omega(f)$, contains a point of $M_J \cap (\mathbb{N}_+)^n$. We prove the following strengthening of [AS 1987b, Theorem 1.2].

Theorem 3.3. *If* f *is not a polynomial in some proper subset of* $\{x_1, \dots, x_n\}$, *then*

$$\mathrm{ord}_q\, S_1(\mathbb{A}^n, f) \ge \omega(f).$$

As an example of Theorem 3.3, consider the polynomial

$$f(x_1, x_2) = x_1 x_2^4 + x_1^7 x_2^3 + x_1^{13} x_2^2.$$

If $p \ne 5$, then $M_J = \mathbb{Z}^2$; so $\omega(f) = 7/25$, which gives the estimate of [AS 1987b, Theorem 1.2]. Theorem 3.3 gives an improvement when $p = 5$. In this case,

$$M_J = \{(u_1, u_2) \in \mathbb{Z}^2 \mid u_1 + 6u_2 \text{ is divisible by } 25\}$$

so $\omega(f) = 1$.

Proof of Theorem 3.3. Let Φ_0 be the set of all functions $\phi : J \to \{0, 1, \dots, q-1\}$ such that

$$\frac{1}{q-1} \sum_{j \in J} \phi(j) j \in (\mathbb{N}_+)^n.$$

For $\phi \in \Phi_0$ define $\phi' \in \Phi_0$ by

$$\phi'(j) = \begin{cases} 0 & \text{if } \phi(j) = 0, \\ \text{least positive residue of } p\phi(j) \text{ modulo } q-1 & \text{if } \phi(j) \ne 0. \end{cases}$$

We denote the i-fold iteration of this operation by $\phi^{(i)}$. Note that since $q = p^r$, one has $\phi^{(r)} = \phi$. By [AS 1990a, Equation 13] we have

$$\operatorname{ord}_q S_1(\mathbb{A}^n, f) \geq \min_{\phi\in\Phi_0} \left\{ \frac{1}{r} \sum_{i=0}^{r-1} \sum_{j\in J} \frac{\phi^{(i)}(j)}{q-1} \right\}. \tag{3.4}$$

Clearly $\sum_{j\in J} \phi^{(i)}(j) j \in \mathbb{Z}\langle J\rangle$ for all i, so

$$\frac{1}{q-1} \sum_{j\in J} \phi^{(i)}(j) j \in M_J \cap (\mathbb{N}_+)^n.$$

If we define Φ_1 to be the set of all functions $\phi : J \to \mathbb{R}_{\geq 0}$ such that

$$\frac{1}{q-1} \sum_{j\in J} \phi(j) j \in M_J \cap (\mathbb{N}_+)^n,$$

then $\frac{1}{q-1}\Phi_0 \subseteq \Phi_1$, so Equation (3.4) implies

$$\operatorname{ord}_q S_1(\mathbb{A}^n, f) \geq \min_{\phi\in\Phi_1} \left\{ \sum_{j\in J} \phi(j) \right\}.$$

The assertion of Theorem 3.3 then follows from [AS 1990a, Lemma 1] by taking the set L of that lemma equal to $M_J \cap (\mathbb{N}_+)^n$. (Theorem 3.3 can also be proved by repeating *mutatis mutandis* the argument of [AS 1987b, Section 4] with $L(p/(p-1))$ replaced by $L_0(p/(p-1))$.) $\square$

We derive a generalization of Theorem 1.3 from Theorem 3.3. Let $f_1, \ldots, f_r \in \mathbb{F}_q[x_1, \ldots, x_n]$ and let $N(f_1, \ldots, f_r)$ denote the number of solutions in $\mathbb{F}_q$ to the system $f_1 = \cdots = f_r = 0$. Let $y_1, \ldots, y_r$ be additional variables and set

$$F = \sum_{i=1}^{r} y_i f_i \in \mathbb{F}_q[x_1, \ldots, x_n, y_1, \ldots, y_r].$$

It is easily seen that

$$S_1(\mathbb{A}^{n+r}, F) = q^r N(f_1, \ldots, f_r).$$

Applying Theorem 3.3 to F gives the following result, of which Theorem 1.3 is the special case $r = 1$.

Corollary 3.5. $\operatorname{ord}_q N(f_1, \ldots, f_r) \geq \omega(F) - r$.

4. Nondegeneracy relative to a lattice

The results of [AS 1989; 1990b] are cohomological in nature and require a more detailed development. Suppose that $\mathbb{Z}\langle J\rangle$ has rank k. Let M be a lattice, $\mathbb{Z}\langle J\rangle \subseteq M \subseteq \mathbb{Z}^n \cap \mathbb{R}\langle J\rangle$, and set $L = \mathrm{Hom}_{\mathbb{Z}}(M, \mathbb{Z})$. For $\ell \in L$ we define a "differential operator" E_ℓ on the ring $\mathbb{F}_q[x^u \mid u \in M]$ by linearity and the formula

$$E_\ell(x^u) = \ell(u)x^u.$$

This definition is motivated by the fact that if we write

$$\ell(u_1, \dots, u_n) = \sum_{j=1}^{n} a_j u_j,$$

where $u = (u_1, \dots, u_n) \in M \subseteq \mathbb{Z}^n$ and the a_j are rational numbers, and put $E_\ell = \sum_{j=1}^{n} a_j x_j \partial/\partial x_j$, then in characteristic 0,

$$E_\ell(x^u) = \sum_{j=1}^{n} a_j x_j \frac{\partial}{\partial x_j}(x^u) = \ell(u)x^u.$$

Let f be given by (1.1) and let σ be a subset of $\Delta(f)$. Define

$$f_\sigma = \sum_{j \in J \cap \sigma} a_j x^j.$$

We say that f is *nondegenerate relative to* $(\Delta(f), M)$ if for every face σ of $\Delta(f)$ that does not contain the origin, the Laurent polynomials $\{E_\ell(f_\sigma)\}_{\ell \in L}$ have no common zero in $(\bar{\mathbb{F}}_q^\times)^n$, where $\bar{\mathbb{F}}_q$ denotes an algebraic closure of $\mathbb{F}_q$. Note that the condition $\mathbb{Z}\langle J\rangle \subseteq M$ guarantees that all f_σ lie in $\mathbb{F}_q[x^u \mid u \in M]$, so the $E_\ell(f_\sigma)$ are defined. Note also that to check this condition, it suffices to check it on a set of the form $\{E_{\ell_i}\}_{i=1}^k$, where $\{\ell_i\}_{i=1}^k$ is any basis of L. (We remark that this idea, to replace the differential operators $x_i\partial/\partial x_i$ by certain linear combinations with coefficients that are not p-integral, appears in nascent form in [Dwork 1962], where it was needed to calculate the p-adic cohomology of smooth hypersurfaces of degree divisible by p.)

The condition used in [AS 1989], that f be nondegenerate relative to $\Delta(f)$, is equivalent to the condition that f be nondegenerate relative to $(\Delta(f), \mathbb{Z}^n \cap \mathbb{R}\langle J\rangle)$ in the sense of the present definition. We make the relationship between this definition and our earlier one more explicit. There is a basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ for $\mathbb{Z}^n$ and positive integers $d_1, \dots, d_k$, $k \le n$, such that $d_1\mathbf{e}_1, \dots, d_k\boldsymbol{e}_k$ is a basis for M. After a coordinate change on $\mathbb{T}^n$, we may take $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ to be the standard basis for $\mathbb{Z}^n$.

This implies that there exists a Laurent polynomial

$$g = \sum_{c \in C} b_c x^c \in \mathbb{F}_q[x_1^{\pm 1}, \dots, x_k^{\pm 1}],$$

where C is a finite subset of $\mathbb{Z}^k$, such that

$$f(x_1, \dots, x_n) = g(x_1^{d_1}, \dots, x_k^{d_k}). \tag{4.1}$$

Note that (4.1) implies

$$[\mathbb{Z}\langle C\rangle : \mathbb{Z}\langle J\rangle] = d_1 \cdots d_k \ (= [\mathbb{Z}^n \cap \mathbb{R}\langle J\rangle : M]). \tag{4.2}$$

Remark. When we choose $M = M_J$, it follows from Equation (1.2) that each d_i is a power of p. In this case, the exponential sums associated to f and g are identical.

Proposition 4.3. *The Laurent polynomial f is nondegenerate relative to the pair $(\Delta(f), M)$ if and only if g is nondegenerate relative to $(\Delta(g), \mathbb{Z}^k)$.*

Proof. Equation (4.1) implies that there is a one-to-one correspondence between the faces of $\Delta(f)$ and the faces of $\Delta(g)$. Specifically, the face σ of $\Delta(f)$ corresponds to the face σ' of $\Delta(g)$ defined by

$$\sigma' = \{(d_1^{-1}u_1, \dots, d_k^{-1}u_k) \in \mathbb{R}^k \mid (u_1, \dots, u_k) \in \sigma\}.$$

Furthermore, we have

$$f_\sigma(x_1, \dots, x_k) = g_{\sigma'}(x_1^{d_1}, \dots, x_k^{d_k}).$$

Using $u_1, \dots, u_k$ as coordinates on $\mathbb{Z}^k$, we may take as basis for $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}^k, \mathbb{Z})$ the linear forms $\{\ell'_i\}_{i=1}^k$ defined by

$$\ell'_i(u_1, \dots, u_k) = u_i,$$

and we may take as basis for $L = \mathrm{Hom}_{\mathbb{Z}}(M, \mathbb{Z})$ the linear forms $\{\ell_i\}_{i=1}^k$ defined by

$$\ell_i(u_1, \dots, u_k) = d_i^{-1}u_i.$$

It is straightforward to check that for $i = 1, \dots, k$,

$$E_{\ell_i}(f_\sigma)(x_1, \dots, x_k) = E_{\ell'_i}(g_{\sigma'})(x_1^{d_1}, \dots, x_k^{d_k}).$$

This implies the proposition. □

Lemma 4.4. *Put $[\mathbb{Z}^n \cap \mathbb{R}\langle J\rangle : M_J] = p^a$ and let $M \subseteq \mathbb{Z}^n \cap \mathbb{R}\langle J\rangle$ be a lattice containing $\mathbb{Z}\langle J\rangle$. Then $M \subseteq M_J$ if and only if $p^a \mid [\mathbb{Z}^n \cap \mathbb{R}\langle J\rangle : M]$.*

Proof. Suppose that $p^a \mid [\mathbb{Z}^n \cap \mathbb{R}\langle J\rangle : M]$. Then $[M : \mathbb{Z}\langle J\rangle] = e'$ with $(e', p) = 1$. In particular, $e'm \in \mathbb{Z}\langle J\rangle$ for all $m \in M$, so $M \subseteq M_J$. The other direction of the assertion is clear. □

There are restrictions on the lattices with respect to which f can be nondegenerate.

Proposition 4.5. *Let M be a lattice, $\mathbb{Z}\langle J\rangle \subseteq M \subseteq \mathbb{Z}^n \cap \mathbb{R}\langle J\rangle$.*

(a) *If f is nondegenerate relative to $(\Delta(f), M)$, then $M \subseteq M_J$.*

(b) *Suppose $M \subseteq M_J$. Then f is nondegenerate relative to $(\Delta(f), M)$ if and only if f is nondegenerate relative to $(\Delta(f), M_J)$.*

Proof. We may assume without loss of generality that $\mathbb{Z}\langle J\rangle$ is a subgroup of $\mathbb{Z}^n$ of rank n. For if $\operatorname{rank}(\mathbb{Z}\langle J\rangle) = k < n$, then by (4.1) we may take f to be a Laurent polynomial in $x_1, \ldots, x_k$, in which case $\mathbb{Z}\langle J\rangle$ is a subgroup of $\mathbb{Z}^n \cap \mathbb{R}\langle J\rangle (= \mathbb{Z}^k)$ of rank k.

We suppose M is not contained in M_J and prove that f must be degenerate relative to $(\Delta(f), M)$. By (4.2) and Lemma 4.4, we have $p^a \nmid [\mathbb{Z}\langle C\rangle : \mathbb{Z}\langle J\rangle]$. But $p^a \mid [\mathbb{Z}^n : \mathbb{Z}\langle J\rangle]$, so $p \mid [\mathbb{Z}^n : \mathbb{Z}\langle C\rangle]$. Arguing as in the proof of Equation (4.1) then shows that there exists a Laurent polynomial

$$h = \sum_{e\in E} c_e x^e \in \mathbb{F}_q[x_1^{\pm 1}, \ldots, x_n^{\pm 1}]$$

such that

$$g(x_1, \ldots, x_n) = h(x_1, \ldots, x_{n-1}, x_n^p). \tag{4.6}$$

To show f is degenerate relative to $(\Delta(f), M)$, it suffices by Proposition 4.3 to show that any Laurent polynomial g of the form (4.6) is degenerate relative to $(\Delta(g), \mathbb{Z}^n)$. The weight function w of Section 2 defines an increasing filtration on the ring $\mathbb{F}_q[x^u \mid u \in C(g) \cap \mathbb{Z}^n]$: level i/N of the filtration is spanned by the monomials of weight $\leq i/N$. If g were nondegenerate, then $\{x_i \partial g/\partial x_i\}_{i=1}^n$ would be a regular sequence in the associated graded ring and would generate a proper ideal of codimension $n!\,V(g)$ (by [Kouchnirenko 1976], see also [AS 1989, Section 2]). But, by Equation (4.6), $x_n \partial g/\partial x_n = 0$, and hence cannot be part of such a regular sequence. This contradiction establishes part (a) of Proposition 4.5.

Now suppose that $M \subseteq M_J$. Choose a basis $\{\boldsymbol{e}^{(i)}\}_{i=1}^n$ for M_J and integers $d_1, \ldots, d_n$ such that $\{d_i \mathbf{e}^{(i)}\}_{i=1}^n$ is a basis for M. By Lemma 4.4, $p \nmid d_1 \cdots d_n$. Let $\{\ell_i\}_{i=1}^n$ be the basis for $\operatorname{Hom}_{\mathbb{Z}}(M_J, \mathbb{Z})$ defined by

$$\ell_i(\boldsymbol{e}^{(j)}) = \delta_{ij} \quad \text{(Kronecker's delta)}.$$

Then $\{d_i^{-1}\ell_i\}_{i=1}^n$ is a basis for $\operatorname{Hom}_{\mathbb{Z}}(M, \mathbb{Z})$. And since $(d_i, p) = 1$ for all i, the $\{E_{\ell_i}(f_\sigma)\}_{i=1}^n$ have no common zero in $(\bar{\mathbb{F}}_q^\times)^n$ if and only if the same is true of the $\{E_{d_i^{-1}\ell_i}(f_\sigma)\}_{i=1}^n$. This establishes part (b) of Proposition 4.5. □

By Proposition 4.5(a), we must have $M \subseteq M_J$ if we hope to have f nondegenerate relative to $(\Delta(f), M)$. On the other hand, we must have $M_J \subseteq M$ in order for the trace formula (Theorem 2.5) to hold for M. Thus the only practical choice for M is to take $M = M_J$. Recall from Section 2 that if $g(t)$ is a power series with constant term 1, then $g(t)^\delta = g(t)/g(qt)$.

Theorem 4.7. *Suppose that $\mathbb{Z}\langle J\rangle$ has rank k and that f is nondegenerate relative to $(\Delta(f), M_J)$. Then $L(\mathbb{T}^n, f; t)^{(-1)^{n-1}} = P(t)^{\delta^{n-k}}$, where $P(t)$ is a polynomial of degree $k!\, V_{M_J}(f)$ and $V_{M_J}(f)$ denotes the volume of $\Delta(f)$ relative to Lebesgue measure on $\mathbb{R}\langle J\rangle$ normalized so that a fundamental domain for M_J has volume* 1.

Proof. One repeats the arguments of [AS 1989] with the modifications introduced for Theorem 2.5: replace $L(b)$ and B' by $L_0(b)$ and B'_0 and use $\tilde{\alpha}$ in place of α. We recall some of these details as they are needed in the proof of Theorem 4.17.

Let

$$\Omega^\bullet : 0 \to \Omega^0 \to \cdots \to \Omega^n \to 0$$

be the cohomological Koszul complex on B'_0 defined by the differential operators $\{\hat{D}_i\}_{i=1}^n$ constructed in [AS 1989, Section 2]. The endomorphism $\tilde{\alpha}$ of B'_0 constructed in Section 2 can be extended to an endomorphism $\tilde{\alpha}_\bullet$ of the complex $\Omega^\bullet$ by noting that $\Omega^i = (B'_0)^{\binom{n}{i}}$ and then defining $\tilde{\alpha}_i : \Omega^i \to \Omega^i$ to be

$$(q^{n-i}\tilde{\alpha})^{\binom{n}{i}} : (B'_0)^{\binom{n}{i}} \to (B'_0)^{\binom{n}{i}}. \tag{4.8}$$

Theorem 2.5 is equivalent to the assertion that

$$L(\mathbb{T}^n, f; t) = \prod_{i=0}^{n} \det(I - t\tilde{\alpha}_i \mid \Omega^i)^{(-1)^{i+1}},$$

which implies that

$$L(\mathbb{T}^n, f; t) = \prod_{i=0}^{n} \det(I - t\tilde{\alpha}_i \mid H^i(\Omega^\bullet))^{(-1)^{i+1}}. \tag{4.9}$$

Put $R = \mathbb{F}_q[x^u \mid u \in M_0(f)]$. The ring R has an increasing filtration defined by the weight function w of Section 2: $F_{i/N}R$ is the subspace spanned by $\{x^u \mid w(u) \le i/N\}$. Let $\bar{R} = \bigoplus_{i=0}^\infty \bar{R}_{i/N}$ be the associated graded ring, that is, $\bar{R}_{i/N} = F_{i/N}R/F_{(i-1)/N}$. Now suppose that f is nondegenerate relative to $(\Delta(f), M_J)$, let $\{\ell_i\}_{i=1}^k$ be a basis for $L = \mathrm{Hom}_{\mathbb{Z}}(M_J, \mathbb{Z})$, and let $\overline{E_{\ell_i}(f)} \in \bar{R}_1$ be the image in the associated graded ring of $E_{\ell_i}(f) \in F_1R$. The nondegeneracy hypothesis implies by the arguments in [Kouchnirenko 1976] that $\{\overline{E_{\ell_i}(f)}\}_{i=1}^k$ is a regular sequence in $\bar{R}$, that is, the (cohomological) Koszul complex on $\bar{R}$ defined by $\{\overline{E_{\ell_i}(f)}\}_{i=1}^k$ has vanishing cohomology except in top dimension. Furthermore, also by the methods in [Kouchnirenko 1976], one can show that the single nonvanishing cohomology group has dimension $k!\, V_{M_J}(f)$.

Since $M_J \subseteq \mathbb{Z}^n$, we may express the elements of L as linear forms in n variables. Write

$$\ell_i(u_1, \dots, u_n) = \sum_{j=1}^{n} a_{ij} u_j, \quad a_{ij} \in p^{-a}\mathbb{Z}.$$

Put $\hat{D}_{\ell_i} = \sum_{j=1}^{n} a_{ij} \hat{D}_j$ and let $\Omega_\ell^\bullet$ be the cohomological Koszul complex on B_0' defined by $\{\hat{D}_{\ell_i}\}_{i=1}^{k}$. The Frobenius action $\tilde{\alpha}_i : \Omega_\ell^i \to \Omega_\ell^i$ is defined to be

$$(q^{k-i}\tilde{\alpha})^{\binom{k}{i}} : (B_0')^{\binom{k}{i}} \to (B_0')^{\binom{k}{i}}.$$

The "reduction mod p" [AS 1989, Lemma 2.10] of $\Omega_\ell^\bullet$ is the Koszul complex on $\bar{R}$ defined by $\{\overline{E_{\ell_i}(f)}\}_{i=1}^{k}$. Monsky's cohomological lifting theorem [Monsky 1970, Theorem 8.5; AS 1989, Theorem A.1] then implies that the cohomology of $\Omega_\ell^\bullet$ vanishes except in top dimension and that $H^k(\Omega_\ell^\bullet)$ has dimension $k!\,V_{M_J}(f)$. But since $\{\hat{D}_{\ell_i}\}_{i=1}^{k}$ are linear combinations of $\{\hat{D}_i\}_{i=1}^{n}$ and vice versa, it follows that (as Frobenius modules)

$$H^i(\Omega^\bullet) \cong (H^k(\Omega_\ell^\bullet))^{\binom{n-k}{n-i}},$$

where it is understood that the right-hand side vanishes if $i < k$. In particular we have $H^n(\Omega^\bullet) \cong H^k(\Omega_\ell^\bullet)$, hence

$$\det(I - t\tilde{\alpha}_i \mid H^i(\Omega^\bullet)) = \det(I - q^{n-i}t\tilde{\alpha}_n \mid H^n(\Omega^\bullet))^{\binom{n-k}{n-i}}.$$

From Equation (4.9) we then get

$$L(\mathbb{T}^n, f; t) = \prod_{i=k}^{n} \det(I - q^{n-i}t\tilde{\alpha}_n \mid H^n(\Omega^\bullet))^{(-1)^{i+1}\binom{n-k}{n-i}}. \tag{4.10}$$

If we put

$$P(t) = \det(I - t\tilde{\alpha}_n \mid H^n(\Omega^\bullet)) \; (= \det(I - t\tilde{\alpha}_k \mid H^k(\Omega_\ell^\bullet))),$$

then $P(t)$ is a polynomial of degree $k!\,V_{M_J}(f)$ and (4.10) implies that

$$L(\mathbb{T}^n, f; t)^{(-1)^{n-1}} = P(t)^{\delta^{n-k}}.$$

This completes the proof of Theorem 4.7. □

Assume the hypotheses of Theorem 4.7. The quotient ring

$$\bar{R}/(\overline{E_{\ell_1}(f)}, \dots, \overline{E_{\ell_k}(f)})$$

is a graded ring of dimension $k!\,V_{M_J}(f)$ over $\mathbb{F}_q$. Put

$$a_i = \dim_{\mathbb{F}_q}(\bar{R}/(\overline{E_{\ell_1}(f)}, \dots, \overline{E_{\ell_k}(f)}))_{i/N}.$$

One can show that $a_i = 0$ for $i > kN$. By either repeating the argument of [AS 1989] or replacing the polynomial f by the polynomial $g(x_1^{e_1}, \dots, x_n^{e_n})$ constructed in the introduction and applying [AS 1989, Theorem 3.10], one obtains the following generalization of part of [AS 1989, Theorem 3.10].

Theorem 4.11. *Under the hypotheses of Theorem 4.7, the Newton polygon of the polynomial $P(t)$ relative to the valuation* ord_q *lies on or above the Newton polygon relative to* ord_q *of the polynomial* $\prod_{i=0}^{kN}(1-q^{i/N}t)^{a_i}$.

Remark. We recall the combinatorial description of the a_i. Let $W_0(i)$ be the counting function of Equation (3.1) and form the generating series

$$H(t) = \sum_{i=0}^{\infty} W_0(i) t^{i/N}.$$

Then

$$H(t) = \frac{\sum_{i=0}^{kN} a_i t^{i/N}}{(1-t)^k}.$$

Remark. The lower bound of Theorem 4.11 is generically sharp if, for some integer D depending on $\Delta(f)$, $p \equiv 1 \pmod{D}$ [Wan 1993].

We generalize Theorem 4.7 to the affine case. (The corresponding generalization of Theorem 4.11 is somewhat more involved so we postpone that to a future article.) Let

$$f = \sum_{j \in J} a_j x^j \in \mathbb{F}_q[x_1^{\pm 1}, \dots, x_k^{\pm 1}, x_{k+1}, \dots, x_n].$$

For each subset $A \subseteq \{k+1, \dots, n\}$, let f_A be the polynomial obtained from f by setting $x_i = 0$ for all $i \in A$. Then

$$f_A = \sum_{j \in J_A} a_j x^j \in \mathbb{F}[x_1^{\pm 1}, \dots, x_k^{\pm 1}, \{x_i\}_{i \notin A}], \tag{4.12}$$

where $J_A = \{j = (j_1, \dots, j_n) \in J \mid j_i = 0 \text{ for } i \in A\}$. We call f *convenient* if for each such A one has

$$\dim \Delta(f_A) = \dim \Delta(f) - |A|.$$

Suppose f is convenient and nondegenerate relative to $(\Delta(f), M_J)$. The hypothesis that f be convenient guarantees that f_A is also convenient, and the hypothesis that f be nondegenerate relative to $(\Delta(f), M_J)$ implies that f_A is nondegenerate relative to $(\Delta(f_A), M_J \cap \mathbb{R}\langle J_A \rangle)$. By Proposition 4.5(a), we must then have $M_J \cap \mathbb{R}\langle J_A \rangle \subseteq M_{J_A}$. The reverse inclusion is clear, so

$$M_{J_A} = M_J \cap \mathbb{R}\langle J_A \rangle, \tag{4.13}$$

and we conclude that f_A is nondegenerate relative to $(\Delta(f_A), M_{J_A})$. Applying Theorem 4.7, we get that

$$L(\mathbb{T}^{n-|A|}, f_A; t)^{(-1)^{n-|A|-1}} = P_A(t)^{\delta^{n-\dim \Delta(f)}}, \tag{4.14}$$

where $P_A(t)$ is a polynomial of degree

$$\deg P_A(t) = (\dim \Delta(f_A))!\, V_{M_{J_A}}(f_A). \tag{4.15}$$

The standard toric decomposition of affine space gives

$$S_m(\mathbb{T}^k \times \mathbb{A}^{n-k}, f) = \sum_{A \subseteq \{k+1,\dots,n\}} S_m(\mathbb{T}^{n-|A|}, f_A),$$

hence

$$L(\mathbb{T}^k \times \mathbb{A}^{n-k}, f; t)^{(-1)^{n-1}} = \prod_{A \subseteq \{k+1,\dots,n\}} (L(\mathbb{T}^{n-|A|}, f_A; t)^{(-1)^{n-|A|-1}})^{(-1)^{|A|}}. \tag{4.16}$$

Put

$$\nu(f) = \sum_{A \subseteq \{k+1,\dots,n\}} (-1)^{|A|} (\dim \Delta(f_A))!\, V_{M_{J_A}}(f_A).$$

Theorem 4.17. *If $f \in \mathbb{F}_q[x_1^{\pm 1}, \dots, x_k^{\pm 1}, x_{k+1}, \dots, x_n]$ is nondegenerate relative to $(\Delta(f), M_J)$ and convenient, then*

$$L(\mathbb{T}^k \times \mathbb{A}^{n-k}, f; t)^{(-1)^{n-1}} = Q(t)^{\delta^{n-\dim \Delta(f)}}, \tag{4.18}$$

where $Q(t)$ is a polynomial of degree $\nu(f)$.

Proof. It follows from Equations (4.14) and (4.16) that Equation (4.18) holds with

$$Q(t) = \prod_{A \subseteq \{k+1,\dots,n\}} P_A(t)^{(-1)^{|A|}}, \tag{4.19}$$

a rational function of degree $\nu(f)$ by Equation (4.15). It remains only to show that $Q(t)$ is a polynomial.

In the proof of Theorem 4.7, we constructed a complex $\Omega^\bullet$ satisfying

$$H^i(\Omega^\bullet) \cong (H^n(\Omega^\bullet))^{\binom{n-\dim \Delta(f)}{n-i}} \tag{4.20}$$

and $L(\mathbb{T}^n, f; t)^{(-1)^{n-1}} = P(t)^{\delta^{n-\dim \Delta(f)}}$, where

$$P(t) = \det(I - t\tilde{\alpha}_n \mid H^n(\Omega^\bullet)). \tag{4.21}$$

Since f is nondegenerate and convenient, each of the polynomials f_A satisfies the hypotheses of that theorem, so analogous assertions are true. Let

$$\Omega_A^\bullet : 0 \to \Omega_A^0 \to \cdots \to \Omega_A^{n-|A|} \to 0$$

be the corresponding cohomological Koszul complex with differential operators $\{\hat{D}_i^A\}_{i\notin A}$ and Frobenius operators $\{\tilde{\alpha}_i^A\}_{i=0}^{n-|A|}$. We have

$$H^i(\Omega_A^\bullet) = (H^{n-|A|}(\Omega_A^\bullet))^{\binom{n-\dim\Delta(f)}{n-|A|-i}} \tag{4.22}$$

and $L(\mathbb{T}^{n-|A|}, f_A; t)^{(-1)^{n-|A|-1}} = P_A(t)^{n-\dim\Delta(f)}$, where

$$P_A(t) = \det(I - t\tilde{\alpha}^A_{n-|A|} \mid H^{n-|A|}(\Omega_A^\bullet)). \tag{4.23}$$

There is an exact sequence of complexes [Libgober and Sperber 1995, Equation (4.1)]:

$$\Omega^\bullet \to \bigoplus_{|A|=1} \Omega_A^\bullet[-1] \to \bigoplus_{|A|=2} \Omega_A^\bullet[-2] \to \cdots \to \Omega^\bullet_{\{k+1,\dots,n\}}[-n+k] \to 0.$$

Let $\bar{\Omega}^\bullet = \ker(\Omega^\bullet \to \bigoplus_{|A|=1}\Omega_A^\bullet[-1])$, so that there is an exact sequence

$$0 \to \bar{\Omega}^\bullet \to \Omega^\bullet \to \bigoplus_{|A|=1} \Omega_A^\bullet[-1] \to \cdots \to \Omega^\bullet_{\{k+1,\dots,n\}}[-n+k] \to 0. \tag{4.24}$$

Equations (4.20), (4.22), (4.24), and induction on $n-k$ show that

$$H^i(\bar{\Omega}^\bullet) \cong (H^n(\bar{\Omega}^\bullet))^{\binom{n-\dim\Delta(f)}{n-i}}. \tag{4.25}$$

Equation (4.24) implies that

$$\prod_{i=0}^{n} \det(I - t\tilde{\alpha}_i \mid H^i(\bar{\Omega}^\bullet))^{(-1)^{i+1}} = \prod_{A\subseteq\{k+1,\dots,n\}} \biggl(\prod_{i=0}^{n-|A|} \det(I - t\tilde{\alpha}_i^A \mid H^i(\Omega_A^\bullet))^{(-1)^{i+|A|+1}}\biggr)^{(-1)^{|A|}}. \tag{4.26}$$

The inner product on the right-hand side of (4.26) equals $L(\mathbb{T}^{n-|A|}, f_A, t)^{(-1)^{|A|}}$, hence by (4.16) the right-hand side equals $L(\mathbb{T}^k\times\mathbb{A}^{n-k}, f; t)$. By (4.25) the left-hand side equals

$$\prod_{i=0}^{n-\dim\Delta(f)} \det(I - tq^i\tilde{\alpha}_n \mid H^n(\bar{\Omega}^\bullet))^{(-1)^{n-1}\binom{n-\dim\Delta(f)}{i}}.$$

We thus have

$$L(\mathbb{T}^k\times\mathbb{A}^{n-k}, f; t)^{(-1)^{n-1}} = \det(I - t\tilde{\alpha}_n \mid H^n(\bar{\Omega}^\bullet))^{\delta^{n-\dim\Delta(f)}}.$$

Comparison with Equation (4.18) then shows that

$$Q(t) = \det(I - t\tilde{\alpha}_n \mid H^n(\bar{\Omega}^\bullet)),$$

hence $Q(t)$ is a polynomial. □

We explain how to compute the archimedean absolute values of the roots of the polynomial $Q(t)$ under the hypothesis of Theorem 4.17. Take $M = M_J$ and let g be the Laurent polynomial associated to f by Equation (4.1). As noted in the proof of Proposition 4.3, the linear transformation $u_i \mapsto d_i^{-1} u_i$, $i = 1, \dots, k$, identifies the faces σ of $\Delta(f)$ with the faces σ' of $\Delta(g)$. In particular, the face $\Delta(f_A)$ of $\Delta(f)$ will correspond to some face σ'_A of $\Delta(g)$. Let g_A denote the sum of those terms of g whose exponents lie on the face σ'_A (so that $\Delta(g_A) = \sigma'_A$). By the Remark preceding Proposition 4.3, we have

$$L(\mathbb{T}^{n-|A|}, g_A; t)^{(-1)^{n-|A|-1}} = L(\mathbb{T}^{n-|A|}, f_A; t)^{(-1)^{n-|A|-1}} = P_A(t)^{\delta^{n-\dim \Delta(f)}}. \quad (4.27)$$

The nondegeneracy of f_A relative to $(\Delta(f_A), M_{J_A})$ implies the nondegeneracy of g_A relative to $(\Delta(g_A), \mathbb{Z}^{\dim \Delta(g_A)})$. We can thus apply the results of [AS 1990b] and [Denef and Loeser 1991] to g_A to compute the number of roots of $P_A(t)$ of a given archimedean weight. By Equation (4.19) and the fact that $Q(t)$ is a polynomial, we then get the number of roots of $Q(t)$ of a given archimedean weight.

For applications in the next section, we calculate the number of reciprocal roots of largest possible archimedean absolute value $q^{(\dim \Delta(f))/2}$ of $Q(t)$. For $A \neq \varnothing$, all reciprocal roots of $P_A(t)$ have absolute value $< q^{(\dim \Delta(f))/2}$, so this is just the number of reciprocal roots of $P_\varnothing(t)$ of absolute value $q^{(\dim \Delta(f))/2}$. By Equation (4.27), this can be obtained by applying [AS 1990b, Theorem 1.10] to g: the number $w_{\dim \Delta(f)}$ of reciprocal roots of highest weight is

$$w_{\dim \Delta(f)} = \sum_{(0,\dots,0) \subseteq \sigma' \subseteq \Delta(g)} (-1)^{\dim \Delta(g) - \dim \sigma'} (\dim \sigma')!\, V_{\mathbb{Z}^{\dim \sigma'}}(\sigma'). \quad (4.28)$$

Since $\Delta(g)$ is obtained from $\Delta(f)$ by an explicit linear transformation, we can express this in terms of invariants of $\Delta(f)$:

$$w_{\dim \Delta(f)} = \sum_{(0,\dots,0) \subseteq \sigma \subseteq \Delta(f)} (-1)^{\dim \Delta(f) - \dim \sigma} (\dim \sigma)!\, V_{M_{J_\sigma}}(\sigma), \quad (4.29)$$

where $J_\sigma = J \cap \sigma$.

We note an important special case of this formula. If every face of $\Delta(f)$ that contains the origin is of the form $\Delta(f_A)$ for some $A \subseteq \{k+1, \dots, n\}$, the right-hand side of Equation (4.29) is just $\nu(f)$. This gives the following result.

Corollary 4.30. *Under the hypothesis of Theorem 4.17, if every face of $\Delta(f)$ that contains the origin is of the form $\Delta(f_A)$ for some $A \subseteq \{k+1, \dots, n\}$, then all reciprocal roots of $Q(t)$ have archimedean absolute value $q^{(\dim \Delta(f))/2}$.*

As a special case of Corollary 4.30, we note the following result.

Corollary 4.31. *If $f \in \mathbb{F}_q[x_1, \dots, x_n]$ is nondegenerate relative to $(\Delta(f), M_J)$ and convenient, then $L(\mathbb{A}^n, f; t)^{(-1)^{n-1}}$ is a polynomial of degree $\nu(f)$ all of whose reciprocal roots have absolute value $q^{n/2}$.*

5. Examples

We explain how Theorem 4.17 implies a generalization of the result of Katz quoted in the introduction.

Proposition 5.1. *Let $f \in \mathbb{F}_q[x_1, \dots, x_n]$ have degree $d = p^k e$, $(e, p) = 1$, and suppose that every monomial appearing in f has degree divisible by p^k. If $f^{(d)}$, the homogeneous part of f of degree d, defines a smooth hypersurface in $\mathbb{P}^{n-1}$, then $L(\mathbb{A}^n, f; t)^{(-1)^{n-1}}$ is a polynomial of degree*

$$\nu(f) = \frac{1}{p^k}((d-1)^n + (-1)^n (p^k - 1)), \tag{5.2}$$

all of whose reciprocal roots have absolute value $q^{n/2}$.

Proof. Let $\boldsymbol{e}^{(1)}, \dots, \boldsymbol{e}^{(n)}$ denote the standard basis for $\mathbb{R}^n$. Over any sufficiently large extension field of $\mathbb{F}_q$, we can make a coordinate change on $\mathbb{A}^n$ so that f is convenient and for any $A \subseteq \{1, \dots, n\}$, the intersection of $f^{(d)} = 0$ with the coordinate hyperplanes $\{x_i = 0\}_{i \in A}$ is smooth. In particular, the equations $f_A^{(d)} = 0$ define smooth hypersurfaces in $\mathbb{P}^{n-|A|-1}$. The Newton polyhedron $\Delta(f)$ is then the simplex in $\mathbb{R}^n$ with vertices at the origin and the points $\{d\boldsymbol{e}^{(i)}\}_{i=1}^n$. The faces of $\Delta(f)$ not containing the origin are the convex hulls of the sets $\{d\boldsymbol{e}^{(i)}\}_{i \in A}$. It will be simpler to index these faces by their complements: let σ_A denote the face which is the convex hull of $\{d\boldsymbol{e}^{(i)}\}_{i \notin A}$.

Write $f = \sum_{j \in J} a_j x^j$, J a finite subset of $\mathbb{N}^n$. Let $M \subseteq \mathbb{Z}^n$ be the subgroup

$$M = \{(u_1, \dots, u_n) \in \mathbb{Z}^n \mid u_1 + \cdots + u_n \text{ is divisible by } p^k\}.$$

Since all monomials in f have degree divisible by p^k, it follows that $\mathbb{Z}\langle J \rangle \subseteq M$. In fact, $M_J \subseteq M$. To see this, let $(u_1, \dots, u_n) \in M_J$. By definition, there exists an integer c prime to p such that $c(u_1, \dots, u_n) \in \mathbb{Z}\langle J \rangle$. This implies that $c \sum_{i=1}^n u_i$ is divisible by p^k. But since $(c, p) = 1$, one has $\sum_{i=1}^n u_i$ divisible by p^k, therefore $(u_1, \dots, u_n) \in M$.

We claim that f is nondegenerate relative to $(\Delta(f), M)$. As basis for M we take the elements

$$(p^k, 0, \dots, 0) \cup \{(-1, 0, \dots, 0, 1, 0, \dots, 0)\}_{i=2}^n,$$

where the 1 occurs in the i-th position, and as basis for $L = \mathrm{Hom}_{\mathbb{Z}}(M, \mathbb{Z})$ we take the "dual basis", namely, the linear forms

$$\ell_1(u_1, \ldots, u_n) = p^{-k}(u_1 + \cdots + u_n)$$

and

$$\ell_i(u_1, \ldots, u_n) = u_i,$$

for $i = 2, \ldots, n$. Let $A \subseteq \{1, \ldots, n\}$ and let σ_A be the face of $\Delta(f)$ defined above. Note that

$$f_{\sigma_A} := \sum_{j \in J \cap \sigma_A} a_j x^j = f_A^{(d)}.$$

We must thus check that $\{E_{\ell_i}(f_A^{(d)})\}_{i=1}^n$ have no common zero in $(\bar{\mathbb{F}}_q^\times)^n$. But

$$E_{\ell_1}(f_A^{(d)}) = e^{-1} f_A^{(d)}$$

and

$$E_{\ell_i}(f_A^{(d)}) = x_i \frac{\partial f_A^{(d)}}{\partial x_i}$$

for $i = 2, \ldots, n$, so we must show that the system

$$f_A^{(d)} = x_2 \frac{\partial f_A^{(d)}}{\partial x_2} = \cdots = x_n \frac{\partial f_A^{(d)}}{\partial x_n} = 0 \tag{5.3}$$

has no solution in $(\bar{\mathbb{F}}_q^\times)^n$. Since $p \mid d$, the Euler relation implies that any common zero of $\{x_i \partial f_A^{(d)}/\partial x_i\}_{i=2}^n$ is also a zero of $x_1 \partial f_A^{(d)}/\partial x_1$, thus the system (5.3) is equivalent to the system

$$f_A^{(d)} = x_1 \frac{\partial f_A^{(d)}}{\partial x_1} = \cdots = x_n \frac{\partial f_A^{(d)}}{\partial x_n} = 0. \tag{5.4}$$

Furthermore, x_i does not appear in f_A if $i \in A$, hence the solutions of (5.4) in $(\bar{\mathbb{F}}_q^\times)^n$ are exactly the solutions of the set

$$\{f_A^{(d)}\} \cup \{\partial f_A^{(d)}/\partial x_i\}_{i \notin A} \tag{5.5}$$

in $(\bar{\mathbb{F}}_q^\times)^n$. However, the equation $f_A^{(d)} = 0$ defines a smooth hypersurface in $\mathbb{P}^{n-|A|-1}$, so any common zero of the set (5.5) must have $x_i = 0$ for all $i \notin A$. In particular, (5.5) has no common zero in $(\bar{\mathbb{F}}_q^\times)^n$. This implies that (5.4) has no solution in $(\bar{\mathbb{F}}_q^\times)^n$, proving the nondegeneracy of f relative to $(\Delta(f), M)$.

We can now compute $\nu(f)$. By Proposition 4.5(a) we have $M = M_J$, so

$$[\mathbb{Z}^{n-|A|} : M_{J_A}] = p^k \quad \text{for all } A \neq \{1, \ldots, n\}$$

and

$$(n-|A|)!\,V(f_A)/[\mathbb{Z}^{n-|A|}:M_{J_A}]=\begin{cases} d^{n-|A|}/p^k & \text{if } A\neq\{1,\dots,n\},\\ 1 & \text{if } A=\{1,\dots,n\}.\end{cases}$$

Then clearly

$$\nu(f)=\frac{1}{p^k}((d-1)^n+(-1)^n(p^k-1))$$

and the assertions of Proposition 5.1 follow from Theorem 4.17. Finally, note that if $L(\mathbb{A}^n, f; t)^{(-1)^{n-1}}$ is a polynomial of degree (5.2) over all sufficiently large extension fields of $\mathbb{F}_q$, then the same is true over $\mathbb{F}_q$ itself. The assertion about the absolute value of the roots follows immediately from Corollary 4.31. □

Remark. There are many results in the literature that, like Proposition 5.1, assert that $L(\mathbb{A}^n, f; t)^{(-1)^{n-1}}$ is a polynomial if $f^{(d)}$ defines a smooth hypersurface and some additional condition is satisfied (see [Deligne 1974, Théorème 8.4; AS 2000, Theorem 1.11 and the following remark; Katz 2005, Theorem 3.6.5; AS 2009, Theorem 3.1]). One might ask if any additional condition is really necessary. Consider the three-variable polynomial

$$f=(z^p-z)+x^{p-1}y+y^{p-1}z.$$

The homogeneous part of degree p is smooth but f has the same L-function as

$$g=x^{p-1}y+y^{p-1}z.$$

Since $\sum_{z\in\mathbb{F}_q}\Psi(y^{p-1}z)=0$ if $y\neq 0$, one calculates that $\sum_{x,y,z\in\mathbb{F}_q}\Psi(g(x,y,z))=q^2$. This gives $L(\mathbb{A}^3, f; t)=(1-q^2t)^{-1}$, showing that smoothness of $f^{(d)}$ alone is not sufficient to guarantee that $L(\mathbb{A}^n, f; t)^{(-1)^{n-1}}$ will be a polynomial.

We apply Theorem 4.17 to compute the zeta functions of some possibly singular hypersurfaces. Let $f\in\mathbb{F}_q[x_1,\dots,x_n]$ be a homogeneous polynomial and let $X\subseteq\mathbb{P}^{n-1}$ be the hypersurface $f=0$. Write the zeta function $Z(X/\mathbb{F}_q, t)$ of X in the form

$$Z(X/\mathbb{F}_q,t)=\frac{R(t)^{(-1)^{n-1}}}{(1-t)(1-qt)\dots(1-q^{n-2}t)},\tag{5.6}$$

where $R(t)$ is a rational function. The exponential sum associated to the polynomial $yf\in\mathbb{F}_q[x_1,\dots,x_n,y^{\pm1}]$ can be used to count points on the projective hypersurface X. The precise relation is given in [AS 1989, Equation (6.14)]:

$$L(\mathbb{A}^n\times\mathbb{T}, yf; t)^{(-1)^n}=R(qt)^{\delta}.\tag{5.7}$$

Proposition 5.8. *Suppose that $yf\in\mathbb{F}_q[x_1,\dots,x_n,y^{\pm1}]$ is nondegenerate relative to $(\Delta(yf), M_J)$ and convenient. Then $R(t)$ is a polynomial of degree $\nu(yf)$, all of whose reciprocal roots have absolute value $q^{(n-2)/2}$.*

Proof. The assertion about the degree of $R(t)$ follows immediately by applying Theorem 4.17 to Equation (5.7). The assertion about the absolute values of the roots of $R(t)$ follows immediately from Corollary 4.30. $\square$

As an illustration of Proposition 5.8, consider the projective hypersurface $X \subseteq \mathbb{P}^{n-1}$ over $\mathbb{F}_q$ defined by the homogeneous equation

$$f(x_1,\dots,x_n) = x_1^n + \dots + x_n^n + \lambda x_1 \dots x_n = 0,$$

where $\lambda \in \mathbb{F}_q$. If $p \nmid n$, this hypersurface is smooth for all but finitely many values of λ. If $p \mid n$, it is a singular hypersurface for all nonzero λ (except in the cases $p = n = 2$ and $p = n = 3$). We describe the zeta function when $p \mid n$.

Corollary 5.9. *Suppose that $n = p^k e$, where $k \geq 1$ and $(p, e) = 1$, and $\lambda \neq 0$. Then $R(t)$ is a polynomial of degree*

$$\deg R(t) = (p^k - 1)e^{n-1} + e^{-1}((e-1)^n + (-1)^n(e-1)), \tag{5.10}$$

all of whose reciprocal roots have absolute value $q^{(n-2)/2}$.

Remark. Note that the second summand on the right-hand side of Equation (5.10) is the dimension of the primitive part of middle-dimensional cohomology of a smooth hypersurface of degree e. When $\lambda = 0$, the hypersurface X_0 is smooth of degree e. (It is defined by the equation $x_1^e + \dots + x_n^e = 0$.)

Proof of Corollary 5.9. The proof is a direct application of Proposition 5.8. We sketch the details. It is straightforward to check that yf is convenient: $\Delta(yf)$ is the n-simplex in $\mathbb{R}^{n+1}$ with vertices at the origin and the points

$$(n, 0, \dots, 0, 1), (0, n, 0, \dots, 0, 1), \dots, (0, \dots, 0, n, 1),$$

and for each subset $A \subseteq \{1, \dots, n\}$, one has $\dim \Delta(yf_A) = n - |A|$. We have

$$J = \{(n, 0, \dots, 0, 1),\ (0, n, 0, \dots, 0, 1),\ \dots,\ (0, \dots, 0, n, 1),\ (1, \dots, 1, 1)\} \subseteq \mathbb{Z}^{n+1},$$

thus $\mathbb{R}\langle J\rangle$ is the hyperplane in $\mathbb{R}^{n+1}$ with equation $u_1 + \dots + u_n = nv$ and the lattice $\mathbb{Z}^{n+1} \cap \mathbb{R}\langle J\rangle$ has basis

$$B = \{(1, -1, 0, \dots, 0),\ (0, 1, -1, 0, \dots, 0),\ \dots,\ (0, \dots, 0, 1, -1, 0),\ (0, \dots, 0, n, 1)\}.$$

It follows that $n!\, V_n(yf) = n^{n-1}$. Similarly, we have

$$(n - |A|)!\, V_{n-|A|}(yf_A) = \begin{cases} n^{n-1-|A|} & \text{if } |A| \leq n-1, \\ 1 & \text{if } |A| = n. \end{cases}$$

Let the first $n-1$ vectors in B be denoted $\boldsymbol{a}_i$, $i = 1, \dots, n-1$. The lattice $\mathbb{Z}\langle J\rangle$ has basis

$$n\boldsymbol{a}_1, \dots, n\boldsymbol{a}_{n-2}, (n-1, -1, \dots, -1, 0), (1, \dots, 1, 1),$$

from which it follows that M_J has basis

$$p^k\boldsymbol{a}_1,\dots,p^k\boldsymbol{a}_{n-2},(n-1,-1,\dots,-1,0),(1,\dots,1,1). \tag{5.11}$$

One then checks that

$$[\mathbb{Z}^{n+1}\cap\mathbb{R}\langle J\rangle : M_J]=(p^k)^{n-2}.$$

If $|A|\geq 1$, then J_A consists of vectors $(0,\dots,0,n,0,\dots,0,1)$ for which the n occurs in the i-th entry for $i\notin A$ (the vector $(1,\dots,1,1)$ does not appear), and the calculation is easier. One gets

$$[\mathbb{Z}^{n+1}\cap\mathbb{R}\langle J_A\rangle : M_{J_A}]=\begin{cases}(p^k)^{n-2} & \text{if } A=\varnothing,\\ (p^k)^{n-1-|A|} & \text{if } 1\leq|A|\leq n-1,\\ 1 & \text{if } A=\{1,\dots,n\}.\end{cases}$$

We then have

$$\frac{(n-|A|)!\,V_{n-|A|}(yf_A)}{[\mathbb{Z}^{n+1}\cap\mathbb{R}\langle J_A\rangle : M_{J_A}]}=\begin{cases}p^ke^{n-1} & \text{if } A=\varnothing,\\ e^{n-1-|A|} & \text{if } 1\leq|A|\leq n-1,\\ 1 & \text{if } A=\{1,\dots,n\}.\end{cases}$$

It is now straightforward to check that $\nu(yf)$ equals the expression on the right-hand side of (5.10).

It remains to check that yf is nondegenerate relative to $(\Delta(yf),M_J)$. The dual basis of the basis (5.11) for M_J is the set of linear forms

$$\begin{aligned}\ell_i(u_1,\dots,u_n,v)&=\sum_{j=1}^{i}\frac{1}{p^k}u_j+\frac{n-i}{p^k}u_n-ev,\quad i=1,\dots,n-2,\\ \ell_{n-1}(u_1,\dots,u_n,v)&=-u_n+v,\\ \ell_n(u_1,\dots,u_n,v)&=v.\end{aligned}$$

The polynomials $(yf)_\sigma$ for faces σ of $\Delta(yf)$ that do not contain the origin are exactly the polynomials yf_A for $A\subset\{1,\dots,n\}$, $|A|<n$. If $A=\varnothing$, we have

$$E_{\ell_n}(yf)-E_{\ell_{n-1}}(yf)=\lambda yx_1\dots x_n,$$

which has no zero in $(\bar{\mathbb{F}}_q^\times)^{n+1}$. So suppose that $1\leq|A|\leq n-1$. Then

$$yf_A=\sum_{i\notin A}yx_i^n.$$

Suppose first that $n\notin A$. If $1\in A$, then

$$E_{\ell_1}(yf_A)+eE_{\ell_n}(yf_A)=-eyx_n^n,$$

and if $i \in A$ for some i, $2 \le i \le n-2$, then

$$E_{\ell_i}(yf_A) - E_{\ell_{i-1}}(yf_A) = -eyx_n^n.$$

Neither of these monomials vanishes on $(\bar{\mathbb{F}}_q^\times)^{n+1}$. If $i \notin A$ for all $i = 1, \dots, n-2$, then $A = \{n-1\}$. In this case we have

$$\begin{aligned} E_{\ell_1}(yf_A) + eE_{\ell_n}(yf_A) &= ey(x_1^n - x_n^n), \\ E_{\ell_i}(yf_A) - E_{\ell_{i-1}}(yf_A) &= ey(x_i^n - x_n^n) \quad \text{for } i = 2, \dots, n-2, \\ E_{\ell_n}(yf_A) &= y(x_1^n + \dots + x_{n-2}^n + x_n^n). \end{aligned}$$

If the first $n-2$ expressions vanish, then $yx_1^n = \cdots = yx_{n-2}^n = yx_n^n$. The vanishing of the last expression is then equivalent to $(n-1)yx_n^n = 0$, which is impossible in $(\bar{\mathbb{F}}_q^\times)^{n+1}$.

Now suppose that $n \in A$. If $1 \notin A$, then

$$E_{\ell_1}(yf_A) + eE_{\ell_n}(yf_A) = eyx_1^n,$$

and if $i \notin A$ for some i, $2 \le i \le n-2$, then

$$E_{\ell_i}(yf_A) - E_{\ell_{i-1}}(yf_A) = eyx_i^n.$$

Neither of these monomials vanishes on $(\bar{\mathbb{F}}_q^\times)^{n+1}$. If $i \in A$ for $i = 1, \dots, n-2$, then A contains all indices except $i = n-1$ and $E_{\ell_n}(yf_A) = yx_{n-1}^n$, which does not vanish on $(\bar{\mathbb{F}}_q^\times)^{n+1}$.

Thus yf satisfies the hypotheses of Proposition 5.8. ☐

Acknowledgements

We thank the referee for helpful comments and, in particular, a simpler proof of Proposition 4.5(a).

References

[AS 1987a] A. Adolphson and S. Sperber, "Newton polyhedra and the degree of the L-function associated to an exponential sum", *Invent. Math.* **88**:3 (1987), 555–569. MR 89d:11064 Zbl 0623.12012

[AS 1987b] A. Adolphson and S. Sperber, "p-adic estimates for exponential sums and the theorem of Chevalley–Warning", *Ann. Sci. École Norm. Sup.* (4) **20**:4 (1987), 545–556. MR 89d:11112 Zbl 0654.12011

[AS 1989] A. Adolphson and S. Sperber, "Exponential sums and Newton polyhedra: cohomology and estimates", *Ann. of Math.* (2) **130**:2 (1989), 367–406. MR 91e:11094 Zbl 0723.14017

[AS 1990a] A. Adolphson and S. Sperber, "p-adic estimates for exponential sums", pp. 11–22 in *p-adic analysis* (Trento, 1989), edited by F. Baldassarri et al., Lecture Notes in Math. **1454**, Springer, Berlin, 1990. MR 92d:11086 Zbl 0727.11056

[AS 1990b] A. Adolphson and S. Sperber, "Exponential sums on $(\mathbf{G}_m)^n$", *Invent. Math.* **101**:1 (1990), 63–79. MR 92c:11083 Zbl 0764.11037

[AS 2000] A. Adolphson and S. Sperber, "Exponential sums on $\mathbf{A}^n$. III", *Manuscripta Math.* **102**:4 (2000), 429–446. MR 2002g:11124 Zbl 0987.11076

[AS 2009] A. Adolphson and S. Sperber, "Exponential sums on $\mathbf{A}^n$, IV", *Int. J. Number Theory* **5**:5 (2009), 747–764. Zbl 05603968

[Ax 1964] J. Ax, "Zeroes of polynomials over finite fields", *Amer. J. Math.* **86** (1964), 255–261. MR 28 #3986 Zbl 0121.02003

[Deligne 1974] P. Deligne, "La conjecture de Weil, I", *Inst. Hautes Études Sci. Publ. Math.* **43** (1974), 273–307. MR 49 #5013 Zbl 0287.14001

[Denef and Loeser 1991] J. Denef and F. Loeser, "Weights of exponential sums, intersection cohomology, and Newton polyhedra", *Invent. Math.* **106**:2 (1991), 275–294. MR 93a:14019 Zbl 0763.14025

[Dwork 1962] B. Dwork, "On the zeta function of a hypersurface", *Inst. Hautes Études Sci. Publ. Math.* **12** (1962), 5–68. MR 28 #3039 Zbl 0173.48601

[Katz 2005] N. M. Katz, *Moments, monodromy, and perversity: a Diophantine perspective*, Annals of Mathematics Studies **159**, Princeton University Press, Princeton, NJ, 2005. MR 2006j:14020 Zbl 1079.14025

[Katz 2007] N. Katz, "Another look at the Dwork family", preprint, 2007, Available at http://www.math.princeton.edu/~nmk/dworkfam64.pdf.

[Kouchnirenko 1976] A. G. Kouchnirenko, "Polyèdres de Newton et nombres de Milnor", *Invent. Math.* **32**:1 (1976), 1–31. MR 54 #7454 Zbl 0328.32007

[Libgober and Sperber 1995] A. Libgober and S. Sperber, "On the zeta function of monodromy of a polynomial map", *Compositio Math.* **95**:3 (1995), 287–307. MR 96b:14022 Zbl 0968.14006

[Monsky 1970] P. Monsky, *p-adic analysis and zeta functions*, Lectures in Mathematics, Department of Mathematics, Kyoto University **4**, Kinokuniya Book-Store Co. Ltd., Tokyo, 1970. MR 44 #215 Zbl 0256.14009

[Moreno et al. 2004] O. Moreno, K. W. Shum, F. N. Castro, and P. V. Kumar, "Tight bounds for Chevalley–Warning–Ax–Katz type estimates, with improved applications", *Proc. London Math. Soc.* (3) **88**:3 (2004), 545–564. MR 2005g:11114 Zbl 1102.11032

[Rojas-León 2006] A. Rojas-León, "Purity of exponential sums on $\mathbb{A}^n$", *Compos. Math.* **142**:2 (2006), 295–306. MR 2007e:11091 Zbl 05033723

[Rojas-León and Wan 2007] A. Rojas-León and D. Wan, "Moment zeta functions for toric Calabi–Yau hypersurfaces", *Commun. Number Theory Phys.* **1**:3 (2007), 539–578. MR 2008m:11126 Zbl 05608723

[Serre 1962] Jean-Pierre. Serre, "Endomorphismes complètement continus des espaces de Banach p-adiques", *Inst. Hautes Études Sci. Publ. Math.* **12** (1962), 69–85. MR 26 #1733 Zbl 0104.33601

[Wan 1993] D. Q. Wan, "Newton polygons of zeta functions and L functions", *Ann. of Math.* (2) **137**:2 (1993), 249–293. MR 94f:11074 Zbl 0799.11058

Communicated by Hendrik W. Lenstra
Received 2008-08-13 Revised 2009-06-15 Accepted 2009-08-02

adolphs@math.okstate.edu Department of Mathematics, Oklahoma State University, Stillwater, OK 74078, United States

sperber@math.umn.edu School of Mathematics, University of Minnesota, Minneapolis, MN 55455, United States

ALGEBRA AND NUMBER THEORY 3:8(2009)

F-adjunction

Karl Schwede

In this paper we study singularities defined by the action of Frobenius in characteristic $p > 0$. We prove results analogous to inversion of adjunction along a center of log canonicity. For example, we show that if X is a Gorenstein normal variety then to every normal center of sharp F-purity $W \subseteq X$ such that X is F-pure at the generic point of W, there exists a canonically defined $\mathbb{Q}$-divisor Δ_W on W satisfying $(K_X)|_W \sim_{\mathbb{Q}} K_W + \Delta_W$. Furthermore, the singularities of X near W are "the same" as the singularities of (W, Δ_W). As an application, we show that there are finitely many subschemes of a quasiprojective variety that are compatibly split by a given Frobenius splitting. We also reinterpret Fedder's criterion in this context, which has some surprising implications.

1. Introduction

Suppose that X is a variety and Y is an effective integral Weil divisor on X such that $n(K_X + Y)$ is Cartier. If the singularities of X are mild (for example, if X is Cohen–Macaulay and normal) one has a restriction theorem $\omega_X(Y)/\omega_X = \omega_Y$. However $\mathcal{O}_X(n(K_X+Y))|_Y$ is not necessarily equal to nK_Y; there is an additional residue of $\mathcal{O}_X(n(K_X+Y))|_Y$ which (when divided by n) is called "the different", see [Kawamata et al. 1987, Lemma 5-1-9] and [Kollár et al. 1992, Chapter 16]. Even when Y is an arbitrary subvariety (that is, not a divisor) similar phenomena have been observed; see, for example, Kawamata [1997b; 1998; 2008] and [Ein and Mustaţă 2009]. In this paper we explore a related phenomenon in positive characteristic which we call *F-adjunction*, or Frobenius adjunction. In particular, we prove results very similar to the parts of what was known as the adjunction conjecture of Kawamata and Shokurov [Ambro 1999], which relates the singularities of X near a center of log canonicity $W \subseteq X$ to the singularities of W.

Suppose that R is a Gorenstein (or a sufficiently nice log-$\mathbb{Q}$-Gorenstein) normal F-finite ring. Then to every center of sharp F-purity $Q \in \operatorname{Spec} R$ (centers of sharp F-purity are characteristic p analogs of centers of log canonicity) such that R_Q

MSC2000: primary 14B05; secondary 13A35.

Keywords: F-pure, F-split, test ideal, log canonical, center of log canonicity, subadjunction, adjunction conjecture, different.

The author was partially supported by a National Science Foundation postdoctoral fellowship.

is F-pure and R/Q is normal we show that there exists a canonically defined $\mathbb{Q}$-divisor $\Delta_{R/Q}$ on Spec R/Q such that the singularities of R near Q are "the same" as the singularities of $(R/Q, \Delta_{R/Q})$.

A center of sharp F-purity is a characteristic $p > 0$ analog of a center of log canonicity; see for example [Kawamata 1997a, Definition 1.3] and [Schwede 2008a]. Technically speaking, a possibly nonclosed point $Q \in \operatorname{Spec} R$ is a *center of sharp F-purity* if, for every R-linear map $\phi : R^{1/p^e} \to R$, we have $\phi(Q^{1/p^e}) \subseteq Q$. In particular, if Spec R is F-split, then Spec R/Q is compatibly split with every Frobenius splitting of Spec R. Unfortunately, there may be infinitely many different maps that one needs to check to determine whether Q is a center of sharp F-purity. However, when R is Gorenstein and sufficiently local, there exists a "generating" map $\psi : R^{1/p} \to R$ such that Q is a center of sharp F-purity if and only if $\psi(Q^{1/p}) \subseteq Q$ for this single map ψ, see Proposition 4.1. A similar result also holds when R is $\mathbb{Q}$-Gorenstein with index not divisible by $p > 0$. It is the existence of this "generating map" that we use to prove our results.

We will now briefly outline the construction of $\Delta_{R/Q}$ on R/Q. On any scheme $X = \operatorname{Spec} R$ such that R is a normal local ring of characteristic $p > 0$, there is a bijection of sets

$$\left\{ \begin{matrix} \text{Effective } \mathbb{Q}\text{-divisors } \Delta \text{ such} \\ \text{that } (p^e - 1)(K_X + \Delta) \text{ is Cartier} \end{matrix} \right\} \longleftrightarrow \left\{ \begin{matrix} \text{Nonzero elements of} \\ \operatorname{Hom}_{\mathcal{O}_X}(F^e_* \mathcal{O}_X, \mathcal{O}_X) \end{matrix} \right\} \Big/ \sim,$$

where the equivalence relation on the right identifies two maps ϕ and ψ if there is a unit u such that $\phi(u \times \_) = \psi(\_)$; see Theorem 3.13. Statements related to this correspondence are well known and have appeared in several previous contexts, see [Hara and Watanabe 2002, Theorem 3.1, Proof 2] and [Mehta and Ramanathan 1985]. However, we do not think it has been explicitly described in the context of $\mathbb{Q}$-divisors and singularities defined by Frobenius.

With this bijection in mind, assume $(p^e - 1)K_X$ is Cartier, then the divisor 0 on $X = \operatorname{Spec} R$ determines a map $\phi \in \operatorname{Hom}_{\mathcal{O}_X}(F^e_* \mathcal{O}_X, \mathcal{O}_X)$. Setting $W = \operatorname{Spec} R/Q$, the map ϕ can be restricted to a map $\phi_Q \in \operatorname{Hom}_{\mathcal{O}_W}(F^e_* \mathcal{O}_W, \mathcal{O}_W)$ precisely because W is a center of sharp F-purity (the map is ϕ_Q is nonzero because R_Q is F-pure). But then ϕ_Q corresponds to a divisor $\Delta_{R/Q}$ on $W = \operatorname{Spec} R/Q$.

Once we have constructed $\Delta_{R/Q}$, we can relate the singularities of X and W. Roughly speaking, we can do this because the F-singularities of R (respectively, the F-singularities of R/Q) can all be defined by the images of certain

$$\phi \in \operatorname{Hom}_{\mathcal{O}_X}(F^e_* \mathcal{O}_X, \mathcal{O}_X)$$

(respectively $\phi_Q \in \operatorname{Hom}_{\mathcal{O}_W}(F^e_* \mathcal{O}_W, \mathcal{O}_W)$). Some of these results are summarized now.

Main Theorem (Theorem 5.2, Corollary 6.9, Remark 9.5). *Suppose that X is an integral separated normal F-finite noetherian scheme essentially of finite type*[1] *over an F-finite field of characteristic $p > 0$. Further suppose that Δ is an effective $\mathbb{Q}$-divisor on X such that $K_X + \Delta$ is $\mathbb{Q}$-Cartier with index not divisible by p. Let $W \subseteq X$ be an closed subscheme that satisfies the following properties*:

(a) *W is integral and normal.*

(b) *(X, Δ) is sharply F-pure at the generic point of W.*

(c) *The ideal sheaf of W is locally a center of sharp F-purity for (X, Δ).*

Then there exists a canonically determined effective divisor Δ_W on W satisfying the following properties:

(i) $(K_W + \Delta_W) \sim_{\mathbb{Q}} (K_X + \Delta)|_W$.

(ii) *Furthermore, if $(p^e - 1)(K_X + \Delta)$ is Cartier then $(p^e - 1)(K_W + \Delta_W)$ is Cartier and $(p^e - 1)\Delta_W$ is integral.*

(iii) *For any real number $t > 0$ and any ideal sheaf $\mathfrak{a}$ on X which is does not vanish on W, we have that $(X, \Delta, \mathfrak{a}^t)$ is sharply F-pure near W if and only if $(W, \Delta_W, \overline{\mathfrak{a}}^t)$ is sharply F-pure.*

(iv) *W is minimal among centers of sharp F-purity for (X, Δ), with respect to containment of topological spaces (in other words, the ideal sheaf of W is of maximal height as a center of sharp F-purity), if and only if (W, Δ_W) is strongly F-regular.*

(v) *There is a natural bijection between the centers of sharp F-purity of (W, Δ_W), and the centers of sharp F-purity of (X, Δ) which are properly contained in W as topological spaces.*

(vi) *There is a naturally defined ideal sheaf $\tau_b(X, \not\subseteq W; \Delta, \mathfrak{a}^t)$, which philosophically corresponds to an analog of an adjoint ideal in arbitrary codimension, such that*

$$\tau_b(X, \not\subseteq W; \Delta, \mathfrak{a}^t)|_W = \tau_b(W; \Delta_W, \overline{\mathfrak{a}}^t) = \text{“the big test ideal of } (R, \Delta, \overline{\mathfrak{a}}^t)\text{”}.$$

Here $\mathfrak{a}$ and $t > 0$ are as in (iii).

When the center W is not a normal scheme, some of these results can still be lifted to the normalization of W, see Proposition 8.2. Also see the concluding remarks to this paper. Part (vi) should be viewed as an ultimate generalization of the F-restriction theorems for test ideals found in Takagi [2007; 2008], also compare with [Hara and Watanabe 2002, Theorem 4.9, Remark 4.10].

[1]The essentially finite type hypothesis can be removed if one is willing to work on a sufficiently small affine chart or if X is the spectrum of a local ring.

The construction of Δ_W is local and does not require X to be projective. In particular, the statement of Theorem 5.2 is ring theoretic and may be more familiar to commutative algebraists. However the Δ_W constructed is canonical. In particular, the Δ_W glue together to give us the result in the global setting, see Remark 9.5.

When we combine this theory with the work of Fedder [1983], we obtain the following.

Theorem A (Theorem 5.5). *Suppose that S is a regular F-finite ring such that $F^e_* S$ is a free S module (for example, if S is local) and that $R = S/I$ is a quotient that is a normal domain. Further suppose that Δ_R is an effective $\mathbb{Q}$-divisor on* $\operatorname{Spec} R$ *such that* $\operatorname{Hom}_R(F^e_* R((p^e-1)\Delta), R)$ *is a rank one free $F^e_* R$-module (for example, if R is local and $(p^e-1)(K_R+\Delta)$ is Cartier). Then there exists an effective $\mathbb{Q}$-divisor Δ_S on* $\operatorname{Spec} S$ *such that*:

(a) *$(p^e-1)(K_S+\Delta_S)$ is Cartier.*

(b) *I is (Δ_S, F)-compatible and (S, Δ_S) is sharply F-pure at the minimal associated primes of I (that is, the generic points of* $\operatorname{Spec} S/I$).

(c) *Δ_S induces Δ_R as in the Main Theorem.*

We do not know of any similar result proved in characteristic 0 (except when R is a complete intersection [Ein et al. 2003]). The Δ_S in Theorem 5.5 is not canonically determined and therefore we do not see how to globalize this statement.

We also prove the following result.

Theorem B (Corollary 4.3, Remark 9.5). *Suppose that X is a normal variety of finite type over an F-finite field k. Suppose that $\phi : F^e_* \mathcal{O}_X \to \mathcal{O}_X$ is a (global) splitting of Frobenius. Then there exists an effective divisor Δ on X (determined uniquely by ϕ) such that*

(1) $K_X + \Delta \sim_{\mathbb{Q}} 0$;

(2) *(X, Δ) is sharply F-pure*;

(3) *The irreducible subvarieties compatibly split by ϕ coincide exactly with the centers of sharp F-purity of (X, Δ).*

Since centers of sharp F-purity are closely related to centers of log canonicity, the previous result should be viewed as a link between compatibly split subvarieties and centers of log canonicity (of log Calabi–Yau pairs).

Finally, also using these ideas, we prove that there are only finitely many centers of sharp F-purity for a sharply F-pure triple $(R, \Delta, \mathfrak{a}_\bullet)$ (the case when R is a local ring was done in [Schwede 2008a] using the techniques of [Enescu and Hochster 2008] or [Sharp 2007]). Here $\mathfrak{a}_\bullet$ is a graded system of ideals [Hara 2005; Schwede 2008a].

Theorem C (Theorem 5.8). *If* $(R, \Delta, \mathfrak{a}_\bullet)$ *is sharply* F*-pure, then there are at most finitely many centers of sharp* F*-purity.*

This also implies that if X is noetherian (although not necessarily affine) and (X, Δ) is locally sharply F-pure, then there are at most finitely many centers of sharp F-purity. This is the analog of the statement that *if* (X, Δ) *is log canonical, there exist at most finitely many centers of log canonicity*. Another implication of this is that for a globally F-split variety, there are at most finitely many subschemes compatibly split with any given splitting, see Corollary 5.10. In the case of a local ring, similar results have been obtained in [Enescu and Hochster 2008; Sharp 2007]; see also [Schwede 2008a, Corollary 5.2]. Finally, essentially the same result has been independently obtained by Mehta and Kumar [2009].

We close with a comparison of $\Delta_{R/Q}$ with related constructions which have been considered in characteristic zero (in particular, the aforementioned "different"). We then consider what happens if we normalize R/Q (in case R/Q is not normal). We conclude with several further remarks and questions. In particular see Remark 9.5 where a global version of the ideas of this paper are briefly discussed.

2. Preliminaries and notation

Throughout this paper, all schemes and rings are noetherian, excellent, reduced and of characteristic $p > 0$. We also assume that all rings R (and schemes X) have locally normalized dualizing complexes, $\omega_R^\bullet$ (respectively $\omega_X^\bullet$), see [Hartshorne 1966]. In fact, little is lost if one only considers rings that are of essentially finite type over a perfect field. Since we are primarily concerned with the affine or local setting, we will freely switch between the notation corresponding to a ring R and the associated scheme $X = \operatorname{Spec} R$. If $X = \operatorname{Spec} R$ and R is reduced, then we will use $k(X) = k(R)$ to denote the total field of fractions of R. If D is a divisor on $X = \operatorname{Spec} R$, we will mix notation and use $R(D)$ to denote the global sections of $\mathcal{O}_X(D)$. Furthermore, we will often use $F^e_* M$ to denote an R-module M viewed as an R-module via the e-iterated Frobenius, that is $r.x = r^{p^e} x$ (informally, this is just restriction of scalars). In particular, when R is reduced $F^e_* R$ is just another notation for R^{1/p^e}. The reason for this notation is that if $F^e : X \to X$ is the e-iterated Frobenius, then $F^e_* \mathcal{O}_X$ is just the sheaf associated to R^{1/p^e}.

We briefly review some properties of Weil divisors on normal schemes, compare with [Hartshorne 1977, Chapter II, Section 6; 1994] and [Bourbaki 1998, Chapter 7]. Recall that on a normal scheme X, a *Weil divisor* is finite formal sum of reduced and irreducible subschemes of codimension 1, and a *prime divisor* is a single irreducible subscheme of codimension 1. So if $X = \operatorname{Spec} R$, the Weil divisors carry the same information as formal sums of height one prime ideals. A $\mathbb{Q}$-*divisor* is an element of {group of Weil divisors} $\otimes_{\mathbb{Z}} \mathbb{Q}$; it can also be viewed as a finite formal sum $\sum a_i D_i$ where the $a_i \in \mathbb{Q}$ and the D_i are prime divisors. See [Kollár and Mori 1998] for basic facts about $\mathbb{Q}$-divisors from this point of view. A $\mathbb{Q}$-divisor for which all the a_i are integers is called an *integral divisor* (in other words, an integral divisor is a $\mathbb{Q}$-divisor that is also a Weil divisor). A $\mathbb{Q}$-divisor is called *$\mathbb{Q}$-Cartier* if there exists an integer $m > 0$ such that mD is an integral Cartier divisor. A $\mathbb{Q}$-divisor is called *m-Cartier* if mD is an integral Cartier divisor. A divisor (respectively a $\mathbb{Q}$-divisor) $D = \sum a_i D_i$ is called *effective* if each of the a_i are nonnegative integers (respectively, nonnegative rational numbers).

Since X is normal, for each prime divisor D on X, there is an associated discrete valuation v_D at the generic point of $D \subset X$. Then, for any nondegenerate element $f \in k(X)$ (an element is *nondegenerate* if it is nonzero on each generic point of $X = \operatorname{Spec} R$), there is a divisor $\operatorname{div} f$ which is defined as $\operatorname{div} f = \sum_{D \subset X} v_D(f) D$. Recall that associated to any divisor D on $X = \operatorname{Spec} R$ there is a coherent sheaf $\mathcal{O}_X(D)$ whose global sections we will denote by $R(D)$. Recall that the sheaf $R(D)$ is reflexive with respect to $\operatorname{Hom}_R(\_, R)$.

For the convenience of the reader, we record some useful properties of reflexive sheaves that we will use without comment.

Proposition 2.1 [Hartshorne 1977; 1994, Proposition 1.11, Theorem 1.12]. *Suppose that R is a normal domain and suppose that M and N are finitely generated torsion-free R-modules. Then*:

(1) *M is reflexive* (*that is*, *the natural map* $M \to \operatorname{Hom}_R(\operatorname{Hom}_R(M, R), R) = (M^\vee)^\vee$ *is an isomorphism*) *if and only if M is $S2$.*

(2) $\operatorname{Hom}_R(M, R) = M^\vee$ *is reflexive.*

(3) *If R is of characteristic p and F-finite* (*see Definition 2.6*), *then M is reflexive if and only if $F^e_* M$ is reflexive.*

(4) *If N is reflexive*, *then* $\operatorname{Hom}(M, N)$ *is also reflexive.*

(5) *Suppose M is reflexive*, *that* $X = \operatorname{Spec} R$ *and $Z \subset X$ is a closed subset of codimension 2. Set U to be $X \setminus Z$ and let $i : U \to X$ be the inclusion. Then* $i_*(M|_U) \cong M$.

(6) *With notation as in* (5), *the restriction map to U induces an equivalence of categories from reflexive coherent sheaves on X to reflexive coherent sheaves on U.*

Proposition 2.2 [Hartshorne 1994, Proposition 2.9; 2007, Remark 2.9]. *Suppose that X is a normal scheme and D is a divisor on X. Then*, *there is a one-to-one correspondence between effective divisors linearly equivalent to D and nondegenerate sections $s \in \Gamma(X, \mathcal{O}_X(D))$ modulo multiplication by units in $H^0(X, \mathcal{O}_X)$.*[2]

[2] A section is called *nondegenerate* if it is nonzero at the generic point of every irreducible component of X.

Definition 2.3. If X is equidimensional, then we set ω_X to be $h^{-\dim X}(\omega_X^{\bullet})$ and call it the *canonical module* of X. If, in addition, X is normal, then ω_X is a rank 1 reflexive sheaf and so it corresponds to an integral divisor class. A divisor D such that $\mathcal{O}_X(D) \cong \omega_X$ is called a *canonical divisor of* X and is denoted by K_X.

Remark 2.4. If X is not normal but instead Gorenstein in codimension 1 (G1) and S2, then one can still view ω_X as a divisor class (technically as an almost Cartier divisor / Weil divisorial subsheaf; see [Hartshorne 1994; Kollár et al. 1992]). Most of the results of this paper generalize to pairs (X, Δ) where X is G1 and S2 and Δ is an element from {almost Cartier divisors} $\otimes \mathbb{Q}$. However, there are several technical complications which we feel obscure the main points of this paper and so we will not work in this generality. In particular, one can have two different almost Cartier divisors / Weil divisorial subsheaves D and E such that $2D = 2E$ [Kollár et al. 1992, page 172]. Because of this, for a $\mathbb{Q}$-Weil divisorial subsheaf D, $\mathcal{O}_X(D)$ is not well defined. There are ways around this issue, although statements like Theorem 3.11(e,f) and the definition of sharply F-pure pairs would need to be amended. Another option is to do something similar to what is suggested in Remark 9.1.

Definition 2.5. A *pair* (X, Δ) is the combined information of a normal scheme X and an effective $\mathbb{Q}$-divisor Δ. A *triple* $(X, \Delta, \mathfrak{a}^t)$ is the combined information of a pair (X, Δ), an ideal sheaf $\mathfrak{a} \subseteq \mathcal{O}_X$ which on every chart $U = \operatorname{Spec} R$ satisfies $\mathfrak{a}|_U \cap R^{\circ} \neq \varnothing$, and a positive real number $t > 0$. If $X = \operatorname{Spec} R$, then we will sometimes write (R, Δ) instead of (X, Δ).

Now we define F-singularities, singularities defined by the action of Frobenius. These are classes of singularities associated with tight closure theory [Hochster and Huneke 1990], which are good analogs of singularities from the minimal model program [Kollár and Mori 1998].

Definition 2.6. We say that a ring R of positive characteristic $p > 0$ is *F-finite* if $F_* R = R^{1/p}$ is finite as an R-module.

Throughout the rest of this paper, *all* rings will be assumed to be F-finite. This is not too restrictive an assumption since any ring essentially of finite type over a perfect field is F-finite, see [Fedder 1983, Lemma 1.4].

Definition 2.7 [Hochster and Roberts 1976; Hochster and Huneke 1989; Hara and Watanabe 2002; Schwede 2008b]. Suppose that $(R, \mathfrak{m})$ is a local ring. We say that a triple $(R, \Delta, \mathfrak{a}^t)$ is *sharply F-pure* if there exists an integer $e > 0$, an element $a \in \mathfrak{a}^{\lceil t(p^e-1)\rceil}$ and a map $\phi \in \operatorname{Hom}_R(F^e_* R(\lceil (p^e-1)\Delta\rceil), R)$ such that $\phi(F^e_*(aR)) = R$. Here $F^e_*(aR) \subseteq F^e_* R(\lceil (p^e-1)\Delta\rceil)$. If $\Delta = 0$ and $\mathfrak{a} = R$, then we call the sharply F-pure triple $(R, \Delta, \mathfrak{a}^t)$ (or simply the ring R) *F-pure*.

Again, assuming R is local, a triple $(R, \Delta, \mathfrak{a}^t)$ is called *strongly F-regular* if for every $c \in R^\circ$ there is an integer $e > 0$, an element $a \in \mathfrak{a}^{\lceil t(p^e-1)\rceil}$, and a map $\phi \in \mathrm{Hom}_R(F^e_* R(\lceil (p^e-1)\Delta\rceil), R)$ such that $\phi(F^e_*(caR)) = R$.

If X is any scheme (for example $X = \operatorname{Spec} R$ where R is a nonlocal ring), then a triple $(X, \Delta, \mathfrak{a}^t)$ is called *sharply F-pure* (respectively, *strongly F-regular*) if for every closed point[3] $x \in X$, the localized triple $(\mathcal{O}_{X,x}, \Delta|_{\operatorname{Spec} \mathcal{O}_{X,x}}, \mathfrak{a}^t_x)$ is sharply F-pure (respectively, strongly F-regular).

Remark 2.8. In the case where R is a nonlocal ring, these definitions of strong F-regularity and sharp F-purity are slightly more general than the ones given in [Takagi 2004a; Takagi and Watanabe 2004; Schwede 2008a; 2008b]. Previously, a triple $(R, \Delta, \mathfrak{a}^t)$ (with R-not necessarily local) was called strongly F-regular (respectively sharply F-pure) if it satisfied the "local ring" version of the condition stated above. In the case that $\mathfrak{a} = R$ (or more generally, if $\mathfrak{a}$ is principal) then the various notions coincide (regardless of the Δ). The problem is that it is not clear whether a triple $(R, \Delta, \mathfrak{a}^t)$ is strongly F-regular (respectively sharply F-pure) if and only if it is strongly F-regular (respectively sharply F-pure) after localizing at every maximal ideal.

Remark 2.9. Suppose that R is local and that $(R, \Delta, \mathfrak{a}^t)$ is sharply F-pure and that e is as in the above definition, then for every integer $n > 0$ there exists a $\phi_n \in \mathrm{Hom}_R(F^{ne}_* R(\lceil (p^{ne}-1)\Delta\rceil), R)$ such that $1 \in \phi_n(F^{ne}_* \mathfrak{a}^{\lceil t(p^{ne}-1)\rceil})$. This follows from the same argument as in [Schwede 2008a, Lemma 2.8; 2008b, Proposition 3.3].

Remark 2.10. Sharply F-pure singularities are a characteristic $p > 0$ analog of log canonical singularities [Hara and Watanabe 2002; Schwede 2008b]. Strongly F-regular singularities are a characteristic $p > 0$ analog of Kawamata log terminal singularities [Hara and Watanabe 2002]. There are also good analogs of purely log terminal singularities that we will not discuss here, see [Takagi 2008].

Definition 2.11 [Hochster and Huneke 1990; Hara and Takagi 2004; Schwede 2008a; 2008b]. Suppose that $(R, \Delta, \mathfrak{a}^t)$ is a triple. An element $c \in R^\circ$ is called a big sharp test element for $(R, \Delta, \mathfrak{a}^t)$ if for all modules $N \subseteq M$ and all $z \in N_M^{*\Delta, \mathfrak{a}^t}$, one has that $c\mathfrak{a}^{\lceil t(p^e-1)\rceil} z^{p^e} \subseteq N_M^{[p^e]\Delta}$ for all $e \geq 0$.

For the definition of tight closure with respect to such a triple (and an explanation of the notation above), see [Schwede 2008a, Definition 2.14]. Also compare with [Hara and Yoshida 2003; Takagi 2004b; 2008.]

If R is reduced and F-finite, then there always exists a big sharp test element for any triple $(R, \Delta, \mathfrak{a}^t)$.

[3]If the condition holds at the closed points, then it also holds at the nonclosed points.

Definition 2.12 [Hochster and Huneke 1990; Lyubeznik and Smith 2001; Hara and Takagi 2004; Hochster 2007]. The *big test ideal* of a triple $(R, \Delta, \mathfrak{a}^t)$, denoted $\tau_b(R; \Delta, \mathfrak{a}^t)$, is defined as follows: Set $E = \oplus_{\mathfrak{m}\in \mathfrak{m}-\operatorname{Spec} R} E_{R/\mathfrak{m}}$, where $E_{R/\mathfrak{m}}$ is the injective hull of $R/\mathfrak{m}$. Then

$$\tau_b(R; \Delta, \mathfrak{a}^t) := \operatorname{Ann}_R 0_E^{*\Delta\mathfrak{a}^t} = \bigcap_{\mathfrak{m}} \operatorname{Ann}_R 0_{E_{R/\mathfrak{m}}}^{*\Delta,\mathfrak{a}^t}.$$

Remark 2.13. Big test ideals are characteristic $p > 0$ analogs of multiplier ideals, [Smith 2000; Hara 2001; Takagi 2004b; Hara and Yoshida 2003].

Remark 2.14. In [Schwede 2008a], we defined the big test ideal $\tau_b(R; \Delta, \mathfrak{a}^t)$ in a somewhat different way, essentially using the criterion for the big test ideal found in [Hara and Takagi 2004, Lemma 2.1]. While we will not state that definition here, we note that the big test ideal of [Schwede 2008a] was an ideal J of R which, when localized at any $\mathfrak{m}$, coincided with $\operatorname{Ann}_{R_\mathfrak{m}} 0_{E_{R/\mathfrak{m}}}^{*\Delta\mathfrak{a}^t}$. We now explain why such a J agrees with $\tau_b(R; \Delta, \mathfrak{a}^t)$. Note that this J is contained in each $\operatorname{Ann}_R 0_{E_{R/\mathfrak{m}}}^{*\Delta\mathfrak{a}^t}$, and so $J \subseteq \operatorname{Ann}_R 0_E^{*\Delta\mathfrak{a}^t}$. Conversely, we see that $\tau_b(R; \Delta, \mathfrak{a}^t)R_\mathfrak{m} \subseteq \operatorname{Ann}_{R_\mathfrak{m}} 0_{E_{R/\mathfrak{m}}}^{*\Delta\mathfrak{a}^t} \subseteq J_\mathfrak{m}$, which completes the proof.

Definition 2.15 [Schwede 2008a]. An ideal $I \subseteq R$ is said to be *F-compatible with respect to* $(R, \Delta, \mathfrak{a}^t)$ or equivalently *uniformly* $(\Delta, \mathfrak{a}^t, F)$*-compatible* or simply *F-compatible* if the context is clear, if for every $e > 0$, every $a \in \mathfrak{a}^{\lceil t(p^e-1)\rceil}$ and every map $\phi \in \operatorname{Hom}_R(F^e_* R(\lceil t(p^e-1)\Delta\rceil), R)$, we have $\phi(F^e_* aI) \subseteq I$. A *prime* ideal Q which is F-compatible with respect to $(R, \Delta, \mathfrak{a}^t)$ is called a *center of sharp F-purity for* $(R; \Delta, \mathfrak{a}^t)$, or simply a *center of F-purity* if the context is clear. We will also often abuse notation and call the subscheme $W := \operatorname{Spec} R/Q \subseteq \operatorname{Spec} R =: X$ a *center of F-purity* as well.

Remark 2.16. Centers of sharp F-purity are characteristic $p > 0$ analogs of centers of log canonicity. In particular, any center of log canonicity reduced from characteristic 0 to characteristic $p \gg 0$ is a center of sharp F-purity [Schwede 2008a, Theorem 6.7].

Lemma 2.17 [Schwede 2008a]. *Consider a triple* $(R, \Delta, \mathfrak{a}^t)$ (*recall all rings are assumed F-finite*). *The following properties of F-compatible ideals are satisfied.*

1. *Any (ideal-theoretic) intersection of F-compatible ideals is F-compatible.*
2. *Any (ideal-theoretic) sum of F-compatible ideals is F-compatible.*
3. *The radical of an F-compatible ideal is F-compatible.*
4. *The big test ideal $\tau_b(R; \mathfrak{a}^t, \Delta)$ is the unique smallest F-compatible ideal that has nontrivial intersection with R°.*
5. *The minimal primes of a radical F-compatible ideal are also F-compatible.*
6. *A pair (R, Δ) is strongly F-regular if and only if it has no centers of sharp F-purity besides the minimal primes of R.*

A version of Lemma 2.17(6) is true also for triples $(R, \Delta, \mathfrak{a}^t)$. Although in that case, one must use the "new" strong F-regularity condition, see Remark 2.8. In particular, [Schwede 2008a, Corollary 4.6] is probably not correct as stated. It should say: "$(R, \Delta, \mathfrak{a}_\bullet)$ is strongly F-regular after localizing at every maximal ideal of R if and only if $(R, \Delta, \mathfrak{a}_\bullet)$ has no centers of sharp F-purity besides the minimal primes of R." Thus the original statement of [Schwede 2008a, Corollary 4.6] is correct if one uses the definition of strong F-regularity from Definition 2.7. We believe this is the only instance of the issue described in Remark 2.8 causing a misstatement in that paper (although several results can be strengthened if one uses the "new" definition).

3. Relation between Frobenius and boundary divisors

In this section we'll describe a correspondence between maps $\phi : F^e_* \mathcal{O}_X \to \mathcal{O}_X$ and $\mathbb{Q}$-divisors Δ such that $K_X + \Delta$ is $\mathbb{Q}$-Cartier (with index not divisible by $p > 0$). Statements closely related to this correspondence have appeared in several previous contexts (see [Hara and Watanabe 2002, Theorem 3.1, Proof 2] and [Mehta and Ramanathan 1985]) and were known to experts. However, we do not think the correspondence has been explicitly written from a $\mathbb{Q}$-divisor perspective. As before, in this section we are assuming that X is the spectrum of a normal F-finite ring R with a locally normalized dualizing complex $\omega_R^\bullet$.

Roughly speaking, the correspondence goes like this. Suppose R is a local ring and set $X = \operatorname{Spec} R$:

- Given a $\phi \in \operatorname{Hom}_R(F^e_* R, R)$, this is the same as
- choosing a map (of $F^e_* R$-modules) $F^e_* R \to \operatorname{Hom}_R(F^e_* R, R)$ sending 1 to ϕ, which is the same as
- an effective Weil divisor D such that $\mathcal{O}_X(D) \cong \mathcal{O}_X((1 - p^e)K_X)$ (note that $F^e_* \mathcal{O}_X((1 - p^e)K_X) \cong \operatorname{Hom}_R(F^e_* R, R)$), which is the same as
- an effective $\mathbb{Q}$-divisor Δ where we set $\Delta = (1/(p^e - 1))D$.

The expert reader might wonder why we divide by $p^e - 1$ in the final step (and thus produce a $\mathbb{Q}$-divisor). It turns out that for the purposes of F-singularities, composing ϕ with itself (that is, $\phi \circ F^e_* \phi$) is harmless, see Section 4 below. Thus by dividing by $p^e - 1$ we are normalizing our divisor with respect to composition; see Theorem 3.11(e).

In order to make this correspondence precise and in order to be able to use it, we first need the following observations about maps $F^e_* \mathcal{O}_X \to \mathcal{O}_X$ (which of themselves are of independent interest). Lemma 3.1 is well known to experts; see [Fedder 1983; Mehta and Ramanathan 1985; Mehta and Srinivas 1991; Hara and

Watanabe 2002, Lemma 3.4]. However, the proof is short, so we include it for the convenience of the reader.

Lemma 3.1. *Suppose that (X, Δ) is a pair such that $(p^e-1)(K_X+\Delta)$ is a Cartier divisor. Then $\mathscr{H}om_{\mathcal{O}_X}(F^e_*\mathcal{O}_X((p^e-1)\Delta), \mathcal{O}_X)$ is an invertible sheaf when viewed as an $F^e_*\mathcal{O}_X$-module.*

Proof. It is enough to verify this locally, so we may assume that X is the spectrum of a local ring. Then observe that

$$\begin{aligned}\mathscr{H}om_{\mathcal{O}_X}(F^e_*\mathcal{O}_X((p^e-1)\Delta), \mathcal{O}_X) &\cong \mathscr{H}om_{\mathcal{O}_X}(F^e_*\mathcal{O}_X((p^e-1)\Delta + p^e K_X), \omega_X)\\ &\cong F^e_*\mathscr{H}om_{\mathcal{O}_X}(\mathcal{O}_X((p^e-1)\Delta + p^e K_X), \omega_X)\\ &\cong F^e_*\mathcal{O}_X((1-p^e)(K_X+\Delta))\\ &\cong F^e_*\mathcal{O}_X. \qquad \square\end{aligned}$$

Remark 3.2. We will often view $\mathscr{H}om_{\mathcal{O}_X}(F^e_*\mathcal{O}_X((p^e-1)\Delta), \mathcal{O}_X)$ as an $F^e_*\mathcal{O}_X$-submodule of $\mathscr{H}om_{\mathcal{O}_X}(F^e_*\mathcal{O}_X, \mathcal{O}_X)$.

Remark 3.3. For an arbitrary normal (nonlocal) F-finite scheme X, we do not know if one always has

$$\mathscr{H}om_{\mathcal{O}_X}(F^e_*\mathcal{O}_X((p^e-1)\Delta), \mathcal{O}_X) \cong \mathcal{O}_X((1-p^e)(K_X+\Delta)). \tag{3.3.1}$$

In the nonlocal case, if one is following the proof of Lemma 3.1, one should write

$$\begin{aligned}&\mathscr{H}om_{\mathcal{O}_X}(F^e_*\mathcal{O}_X((p^e-1)\Delta + p^e K_X), \omega_X)\\ &\qquad\cong F^e_*\mathscr{H}om_{\mathcal{O}_X}(\mathcal{O}_X((p^e-1)\Delta + p^e K_X), (F^e)^!\omega_X).\end{aligned}$$

The module $(F^e)^!\omega_X = \mathrm{Hom}_{\mathcal{O}_X}(F^e_*\mathcal{O}_X, \omega_X)$ is a canonical module on X, but these are only unique up to tensoring with an invertible sheaf. In the local case, tensoring with an invertible sheaf does nothing (and so ω_X is unique up to isomorphism — multiplication by a unit). Likewise, if X is of essentially finite type over an F-finite field, it is easy to see that $(F^e)^!\omega_X$ can be identified with ω_X (again, noncanonically, but up to multiplication by a unit of $H^0(X, \mathcal{O}_X)$). Of course, by passing to a sufficiently small affine chart, we can always assume that Equation (3.3.1) is satisfied. In fact, it may be that Equation (3.3.1) always holds.

The previous result also implies the following when interpreted using Fedder's criterion [Fedder 1983].

Corollary 3.4. *Suppose that $(R, \mathfrak{m})$ is a quasiGorenstein normal local ring (respectively, a $\mathbb{Q}$-Gorenstein local ring whose index is a factor of p^d-1). Further suppose that we can write $R = S/I$ where S is an F-finite regular local ring. Then for each $e > 0$ (respectively for each $e = nd$, $n > 0$) there exists an element $f_e \in R$ so that $(I^{[p^e]} : I) = I^{[p^e]} + (f_e)$.*

Proof. Simply note that $F^e_*(I^{[p^e]} : I)\operatorname{Hom}_S(F^e_* S, S)/F^e_* I^{[p^e]} \cong \operatorname{Hom}_R(F^e_* R, R)$ by [Fedder 1983, Lemma 1.6]. The quasiGorenstein or $\mathbb{Q}$-Gorenstein assumption implies that the right side of the equation is a free rank-one $F^e_* R$-module. □

Remark 3.5. If one fixes a generator T of $\operatorname{Hom}_S(F^e_* S, S)$, one can then view the element f_e as an S-module map $F^e_* S \to S$ that sends $F^e_* I$ into I.

Observation 3.6. Suppose (in the situation of Lemma 3.1) that X is the spectrum of a local ring, that $\Delta = 0$, and that $\mathcal{O}_X((p^e - 1)K_X)$ is a free rank-one $F^e_* \mathcal{O}_X$-module. Therefore, $\mathscr{H}om_{\mathcal{O}_X}(F^e_* \mathcal{O}_X, \mathcal{O}_X)$ has a generator T. If one composes T with its pushforward $F^e_* T : F^{2e}_* \mathcal{O}_X \to F^e_* \mathcal{O}_X$, one obtains a map

$$T_{2e} = T \circ F^e_* T : F^{2e}_* \mathcal{O}_X \to \mathcal{O}_X. \tag{3.6.1}$$

One can then ask whether that composition is a generator of the rank-one locally free $F^{2e}_* R$-module $\mathscr{H}om_{\mathcal{O}_X}(F^{2e}_* \mathcal{O}_X, \mathcal{O}_X)$? What can be said in the case that $\Delta \neq 0$? It turns out that the composition is indeed a generator (and in the case when $\Delta \neq 0$ as well). One can prove this using local duality, however it is no more difficult (and certainly more satisfying) to prove it directly. First however, let us compute a specific example.

Example 3.7. Consider the case when $X = \operatorname{Spec} \mathbb{F}_p[x_1, \dots, x_n] = \operatorname{Spec} R$ and choose T_e to be the generator of $\mathscr{H}om_R(F^e_* R, R)$ of the form

$$T_e(x_1^{l_1} x_2^{l_2} \dots x_n^{l_n}) = \begin{cases} 1, & \text{if } l_1 = l_2 = \dots = l_n = p^e - 1, \\ 0, & \text{whenever } l_i \leq p^e - 1 \text{ for all } i \text{ and } l_i < p^e - 1 \text{ for some } i. \end{cases}$$

Now consider $T_e \circ F^e_* T_e$, we claim it is equal to T_{2e}. Consider a monomial $m = x_1^{l_1} x_2^{l_2} \dots x_n^{l_n}$ such that $l_i \leq p^{2e} - 1$. We can write

$$m = (x_1^{k_1})^{p^e} (x_1^{j_1})(x_2^{k_2})^{p^e} (x_2^{j_2}) \dots (x_n^{k_n})^{p^e} (x_n^{j_n}),$$

where $k_i, j_i < p^e$ are integers. This implies that

$$T_e(F^e_* T_e(m)) = T_e(x_1^{k_1} \dots x_n^{k_n} T_e(x_1^{j_1} \dots x_n^{j_n})).$$

The claim is then easily verified since $p^e(p^e - 1) + (p^e - 1) = (p^{2e} - 1)$.

Remark 3.8. In the context of Example 3.7, it follows that $T_e(F^e_* I) = I^{[1/p^e]}$, where $I^{[1/p^e]}$ is the smallest ideal J such that $I \subseteq J^{[p^e]}$ [Blickle et al. 2008]. This was well known to experts.

In fact, Example 3.7 above is a special case of the following lemma (that is known to experts) which uses Hom-⊗ adjointness. For example, it is closely related to [Kunz 1986, Appendix F.17(a)].

Lemma 3.9. *Suppose that $R \to S$ is a finite map of rings such that $\mathrm{Hom}_R(S, R)$ is isomorphic to S as an S-module. Further suppose that M is a finite S-module.*

Then the natural map

$$\mathrm{Hom}_S(M, S) \times \mathrm{Hom}_R(S, R) \to \mathrm{Hom}_R(M, R) \tag{3.9.1}$$

induced by composition is surjective.

Proof. First, set α to be a generator (as an S-module) of $\mathrm{Hom}_R(S, R)$. Suppose we are given $f \in \mathrm{Hom}_R(M, R) \cong \mathrm{Hom}_R(M \otimes_S S, R)$. We wish to write it as a composition.

Using adjointness, this f induces an element $\Phi(f) \in \mathrm{Hom}_S(M, \mathrm{Hom}_R(S, R))$. Just as with the usual Hom-tensor adjointness, we define $\Phi(f)$ by the following rule:

$$(\Phi(f)(t))(s) = f(t \otimes s) = f(st) \text{ for } t \in M,\, s \in S.$$

Therefore, since $\mathrm{Hom}_R(S, R)$ is generated by α, for each f and $t \in M$ as above, we associate a unique element $a_{f,t} \in S$ with the property that $(\Phi(f)(t))(\_) = \alpha(a_{f,t}\,\_)$.

Thus using the isomorphism $\mathrm{Hom}_R(S, R) \cong S$, induced by sending α to 1, we obtain a map $\Psi : \mathrm{Hom}_R(M, R) \to \mathrm{Hom}_S(M, S)$ given by $\Psi(f)(t) = a_{f,t}$.

We now consider $\alpha \circ (\Psi(f))$. However,

$$\alpha(\Psi(f)(t)) = \alpha(a_{f,t}) = (\Phi(f)(t))(1) = f(t).$$

Thus $f = \alpha \circ (\Psi(f))$ and we see that the map (3.9.1) is surjective, as desired. $\square$

We need a certain variant of this in the context of pairs.

Corollary 3.10. *Suppose that (X, Δ) is a pair and that $K_X + \Delta$ is $(p^e - 1)$-Cartier. Then for every $d > 0$ the natural map Ψ,*

$$\begin{aligned}
&\mathscr{H}om_{F^e_*\mathcal{O}_X}(F^{e+d}_*\mathcal{O}_X(\lceil (p^d-1)\Delta\rceil), F^e_*\mathcal{O}_X)\\
&\qquad \otimes_{F^e_*\mathcal{O}_X} \mathscr{H}om_{\mathcal{O}_X}(F^e_*\mathcal{O}_X((p^e-1)\Delta), \mathcal{O}_X)\\
&\cong \mathscr{H}om_{F^e_*\mathcal{O}_X}(F^{e+d}_*\mathcal{O}_X(\lceil (p^{e+d}-1)\Delta\rceil), F^e_*\mathcal{O}_X((p^e-1)\Delta))\\
&\qquad \otimes_{F^e_*\mathcal{O}_X} \mathscr{H}om_{\mathcal{O}_X}(F^e_*\mathcal{O}_X((p^e-1)\Delta), \mathcal{O}_X)\\
&\to \mathscr{H}om_{\mathcal{O}_X}(F^{e+d}_*\mathcal{O}_X(\lceil (p^{e+d}-1)\Delta\rceil), \mathcal{O}_X)
\end{aligned}$$

induced by composition, is an isomorphism.

In other words, locally, every map $\phi : F^{e+d}_*\mathcal{O}_X(\lceil (p^{e+d}-1)\Delta\rceil) \to \mathcal{O}_X$ factors through some scaling of the (local) $F^e_*\mathcal{O}_X$-generator of

$$\mathscr{H}om_{\mathcal{O}_X}(F^e_*\mathcal{O}_X((p^e-1)\Delta), \mathcal{O}_X).$$

Proof. Notice that the map Ψ we are considering is a map of rank-one reflexive (that is, rank-one S2) $F^{e+d}_* \mathcal{O}_X$ sheaves and thus it is injective (since it is not zero). So to show it is an isomorphism, it is sufficient to show it is surjective in codimension one. Therefore we may consider the statement at the generic point γ of a codimension 1 subvariety (locally, this is localizing at a height one prime). Since X is Gorenstein in codimension one, we see that

$$(\mathscr{H}om_{\mathcal{O}_X}(F^1_* \mathcal{O}_X, \mathcal{O}_X))_\gamma$$

is a free rank-one $F^1_* \mathcal{O}_X$-module. We fix a generator T_1 and set T_n to be the generator of

$$(\mathscr{H}om_{\mathcal{O}_X}(F^n_* \mathcal{O}_X, \mathcal{O}_X))_\gamma$$

obtained by composing T_1 with itself $n-1$ times just as in (3.6.1) (T_n is a generator by Lemma 3.9).

If Δ does not contain the point γ in its support, we are done by the previous lemma. On the other hand, if Δ contains γ in its support, then we may express Δ at the stalk of η locally as z^t (where t is a rational number with denominator a factor of $p^e - 1$). Then we notice that

$$\begin{aligned} T_e(z^{(p^e-1)t}(F^d_* T_d(z_i^{\lceil (p^d-1)t \rceil} \_))) &= T_e(F^d_* T_d(z^{\lceil p^d(p^e-1)t+(p^d-1)t \rceil} \_)) \\ &= T_e(F^d_* T_d(z^{\lceil (p^{d+e}-1)t \rceil} \_)) = T_{e+d}(z^{\lceil (p^{d+e}-1)t \rceil} \_). \end{aligned}$$

This proves the corollary, since for any $n > 0$, $T_n(z^{\lceil (p^n-1)t \rceil} \_)$ generates the image of the $F^n_* \mathcal{O}_{X,\gamma}$-module

$$(\mathscr{H}om_{\mathcal{O}_X}(F^n_* \mathcal{O}_X(\lceil (p^n-1)\Delta \rceil), \mathcal{O}_X))_\gamma \text{ inside } (\mathscr{H}om_{\mathcal{O}_X}(F^n_* \mathcal{O}_X, \mathcal{O}_X))_\gamma . \qquad \square$$

We are now ready to explicitly relate $\phi : F^e_* \mathcal{O}_X \to \mathcal{O}_X$ to a $\mathbb{Q}$-divisor Δ. As mentioned before, parts of this theorem were likely known to experts, but to my knowledge, it has not been written down in the language of $\mathbb{Q}$-divisors.

Theorem 3.11. *Suppose R is a normal F-finite ring. For every map $\phi : F^e_* R \to R$, there exists an effective $\mathbb{Q}$-divisor $\Delta = \Delta_\phi$ on $X = \operatorname{Spec} R$ such that*:

(a) $(p^e-1)\Delta$ *is an integral divisor.*

(b) $(p^e-1)(K_X+\Delta)$ *is a Cartier divisor and* $\operatorname{Hom}_R(F^e_* R((p^e-1)\Delta), R) \cong F^e_* R$.

(c) *The natural map* $F^e_* R \cong \operatorname{Hom}_R(F^e_* R((p^e-1)\Delta), R) \to \operatorname{Hom}_R(F^e_* R, R)$ *sends some $F^e_* R$-module generator of* $\operatorname{Hom}_R(F^e_* R((p^e-1)\Delta), R)$ *to* ϕ.

(d) *The map ϕ is surjective if and only if the pair (R, Δ) is sharply F-pure.*

(e) *The composition map*

$$\phi_{(n+1)e} = \phi \circ F^e_* \phi \circ F^{2e}_* \phi \circ \ldots \circ F^{ne}_* \phi$$

also determines the same divisor Δ.

(f) *Another map* $\phi' : F^{e'}_* R \to R$ *determines the same* $\mathbb{Q}$*-divisor* Δ *if and only if for some positive integers* n *and* n' *such that* $(n+1)e = (n'+1)e'$ *(equivalently, for every such pair of integers) there exists a unit* $u \in R$ *such that*

$$\phi \circ F^e_* \phi \circ F^{2e}_* \phi \circ \ldots \circ F^{ne}_* \phi(ux) = \phi' \circ F^{e'}_* \phi' \circ F^{2e'}_* \phi' \circ \ldots \circ F^{n'e'}_* \phi'(x).$$

for all $x \in R$*. In other words,* ϕ *and* ϕ' *determine the same divisor if and only if* ϕ *composed with itself* n *times is a unit multiple of* ϕ' *composed with itself* n' *times.*

Proof. A map $\phi : F^e_* R \to R$ uniquely determines the map of $F^e_* R$-modules

$$\Phi : F^e_* R \to \mathrm{Hom}_R(F^e_* R, R)$$

which sends 1 to ϕ. This can also be viewed as applying the functor $\mathrm{Hom}_R(\_\,, R)$ to ϕ and factoring the map

$$\begin{tikzcd} R \arrow[r, "\sim"] \arrow[dr, "F^e"'] & \mathrm{Hom}_R(R, R) \arrow[r, "\phi^\vee"] & \mathrm{Hom}_R(F^e_* R, R) \\ & F^e_* R \arrow[ur, dotted, "1 \mapsto \phi"'] & \end{tikzcd} \tag{3.11.1}$$

through $F^e_* R$. We know that $\mathrm{Hom}_R(F^e_* R, R) \cong F^e_* R((1-p^e)K_X + M)$ for some Cartier divisor M (in many cases M is zero; see Remark 3.3). Therefore, the map Φ determines an effective divisor D which is linearly equivalent to $(1-p^e)K_X + M$; see [Hartshorne 1977] and Proposition 2.2. Set

$$\Delta := \frac{1}{p^e - 1} D.$$

Clearly property (a) is satisfied. For the first part of (b), simply note that

$$(p^e - 1)(K_X + \Delta) = (p^e - 1)K_X + D \sim (p^e - 1)K_X + (1 - p^e)K_X + M = M.$$

For the second part of (b), observe that

$$\begin{aligned} \mathrm{Hom}_R(F^e_* R((p^e - 1)\Delta), R) &\cong F^e_* R((1 - p^e)K_X + M - (p^e - 1)\Delta) \\ &\cong F^e_* R((1 - p^e)K_X + M - D) \cong F^e_* R. \end{aligned}$$

Let us now prove (c). At height one primes γ, the map

$$\Phi : F^e_* R_\gamma \to \mathrm{Hom}_R(F^e_* R, R)_\gamma \simeq F^e_* R_\gamma$$

as above, is multiplication (as an $F^e_* R$-module) by a generator of D. But so is the map from (c), $\Psi : F^e_* R \cong \mathrm{Hom}_R(F^e_* R((p^e - 1)\Delta), R) \to \mathrm{Hom}_R(F^e_* R, R)$. All the modules involved are rank-1 reflexive $F^e_* \mathcal{O}_X$-modules and that the domains of Φ and Ψ are isomorphic. Therefore the maps Φ and Ψ induce the same divisors and so Φ and Ψ can be identified (for an appropriate choice of isomorphism $\mathrm{Hom}_R(F^e_* R((p^e-1)\Delta), R) \cong F^e_* R$). Part (c) then follows.

To prove (d), suppose first that ϕ is surjective, or equivalently that 1 is in ϕ's image. Then there exists an R-module map α so that the composition

$$R \xrightarrow{\alpha} F^e_* R \xrightarrow{\phi} R$$

is the identity. Apply $\mathrm{Hom}_R(\_\,, R)$ to the diagram (3.11.1). This gives a diagram:

$$\begin{tikzcd}
R & & \mathrm{Hom}_R(\mathrm{Hom}_R(F^e_* R, R), R) \arrow[ll, "\phi"'] \arrow[dl, dotted] \arrow[r, leftrightarrow, "\sim"] & F^e_* R \\
& \mathrm{Hom}_R(F^e_* R, R) \arrow[ul] \arrow[d, leftrightarrow, "\sim"] & & \\
& F^e_* R(D) & &
\end{tikzcd}$$

and so we can factor ϕ as $F^e_* R \to F^e_* R(D) \to R$. This proves that (R, Δ) is a sharply F-pure pair. Conversely, suppose that (R, Δ) is sharply F-pure, then a single (equivalently every) generator α of $\mathrm{Hom}_R(F^e_* R((p^e-1)\Delta), R)$ satisfies $\alpha(F^e_* R) = R$. But ϕ is such a generator so $\phi(F^e_* R) = R$.

We now prove (e). It is enough to check the statement at a height one prime γ. We know that $\mathrm{Hom}_R(F^e_* R, R)_\gamma$ is locally free of rank one with generator T_e. We then see that $\phi_\gamma(\_) = T_e(d\,\_)$ where d is a defining equation for D when localized at γ. Composing this with itself n times, we obtain the map

$$\phi_\gamma \circ F^e_* \phi_\gamma \circ F^{2e}_* \phi_\gamma \circ \ldots \circ F^{ne}_* \phi_\gamma\,(F^{(n+1)e}_* z) = T_{(n+1)e}(F^{(n+1)e}_* d^{p^{ne}+p^{(n-1)e}+\cdots+p^e+1} z).$$

But now we notice that $(1/(p^{(n+1)e}-1))(p^{ne}+p^{(n-1)e}+\cdots+p^e+1)D$ is equal to $(1/(p^e-1))D$.

Finally, we prove (f). First note that changing a map by precomposing with multiplication by a unit does not change the associated divisor. Therefore, if maps ϕ and ϕ' satisfy the condition on their compositions (as above), then they determine the same divisor by (e). Conversely, suppose that the maps ϕ and ϕ' have the same associated divisor, and choose n and n' as above. Without loss of generality, by replacing ϕ and ϕ' with their compositions, we may assume that $e = e'$, and we simply have two maps $\phi, \phi' \in \mathrm{Hom}_R(F^e_* R, R)$ that determine the same divisor. In particular, the maps

$$\begin{aligned} F^e_* R &\to \mathrm{Hom}_R(F^e_* R, R) \\ 1 &\mapsto \phi \end{aligned} \qquad \text{and} \qquad \begin{aligned} F^e_* R &\to \mathrm{Hom}_R(F^e_* R, R) \\ 1 &\mapsto \phi' \end{aligned}$$

induce the same embedding of $\mathrm{Hom}_R(F^e_* R, R)$ into the total field of fractions of $F^e_* R$. Therefore the two maps differ by multiplication by a unit as desired; see [Hartshorne 2007] or Proposition 2.2. $\square$

Remark 3.12. Condition (a) above is redundant in view of condition (b).

Theorem 3.13. *Suppose that R is normal and F-finite as above. For every effective $\mathbb{Q}$-divisor Δ satisfying conditions (a) and (b) from Theorem 3.11, there exists a map $\phi \in \mathrm{Hom}_R(F^e_* R, R)$ such that the divisor associated to ϕ is Δ.*

Proof. We set ϕ to be the image of 1 under the composition

$$i \circ q \circ F^e : R \to F^e_* R \cong \mathrm{Hom}_R(F^e_* R((p^e - 1)\Delta), R) \to \mathrm{Hom}_R(F^e_* R, R),$$

where q is the isomorphism given by hypothesis, and i the map induced by the inclusion $F^e_* R \subseteq F^e_* R((p^e - 1)\Delta)$. It is straightforward to verify that applying $\mathrm{Hom}_R(\_, R)$ to the above composition also explicitly constructs (and factors) ϕ because of the isomorphism $\mathrm{Hom}_R(\mathrm{Hom}_R(F^e_* R, R), R) \cong F^e_* R$.

Applying $\mathrm{Hom}_R(\_, R)$ to this factorization of ϕ, and using the construction from Theorem 3.11 gives us back Δ. $\square$

In summary, we have shown that for a reduced normal F-finite local ring R there is a bijection between the sets

$$\left\{ \begin{array}{c} \text{Effective } \mathbb{Q}\text{-divisors } \Delta \text{ such} \\ \text{that } (p^e - 1)(K_X + \Delta) \text{ is Cartier} \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{Nonzero elements of} \\ \mathrm{Hom}_{\mathcal{O}_X}(F^e_* \mathcal{O}_X, \mathcal{O}_X) \end{array} \right\} \Big/ \sim,$$

where the equivalence relation on the right identifies two maps ϕ and ψ if there is a unit $u \in R$ such that $\phi(u \times \_) = \psi(\_)$. Remark 9.5 discusses how to make sense of such a correspondence in the nonlocal case.

One can even extend this correspondence further. Recall that putting an $R\{F^e\}$-module structure on an R-module M is equivalent to specifying an additive map

$$\phi_e : M \to M$$

such that $\phi_e(rm) = r^{p^e} \phi_e(m)$; see [Lyubeznik and Smith 2001] for additional details. Such maps can also be identified with R-module maps $M \to F^e_* M$.

Proposition 3.14. *Suppose that $(R, \mathfrak{m})$ is a complete normal local F-finite ring with injective hull of the residue field E_R. Then there is a bijection between the set of $R\{F^e\}$-module structures on E_R and the set of elements of $\mathrm{Hom}_R(F^e_* R, R)$.*

Proof. Consider a map $\phi : F^e_* R \to R$ and apply $\mathrm{Hom}_R(\_, E_R)$. This gives us a map

$$E_R = \mathrm{Hom}_R(R, E_R) \to \mathrm{Hom}_R(F^e_* R, E_R) = E_{F^e_* R} = F^e_* E_R.$$

Applying $\mathrm{Hom}_R(\_, E_R)$ gives us back ϕ. Note that there are (noncanonical) choices here when we identify $F^e_* E_R$ with $\mathrm{Hom}_R(F^e_* R, E_R)$. However, these are merely up to multiplication by units and so we can fix such isomorphisms. □

Therefore, in the case of a complete local normal ring, we have the following correspondences.

$$\left\{ \begin{array}{c} \text{Effective } \mathbb{Q}\text{-divisors } \Delta \text{ such} \\ \text{that } (p^e-1)(K_X+\Delta) \text{ is Cartier} \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{Nontrivial cyclic } F^e_* R\text{-submodules} \\ \text{of } \mathrm{Hom}_R(F^e_* R, R) \end{array} \right\}$$
$$\longleftrightarrow \left\{ \begin{array}{c} \text{Nonzero elements of} \\ \mathrm{Hom}_R(F^e_* R, R) \end{array} \right\} \Big/ \sim \;\longleftrightarrow \left\{ \begin{array}{c} \text{Nonzero } R\{F^e\}\text{-module} \\ \text{structures on } E_R \end{array} \right\} \Big/ \sim$$

The first equivalence relation identified two maps if they agree up to precomposition with multiplication by a unit of $F^e_* R$ (as above). The second equivalence relation identified two maps if they agree up to postcomposition with multiplication by a unit of $F^e_* R$.

Corollary 3.15. *Suppose that S is a regular F-finite ring such that $F^e_* S$ is free as an S-module and that $R = S/I$ is a quotient that is normal. Further suppose that* $\mathrm{Hom}_R(F^e_* R, R) \cong F^e_* R$ *(in particular, R is $\mathbb{Q}$-Gorenstein with index not divisible by p). Write $(I^{[p^e]} : I) = I^{[p^e]} + (f_e)$ just as in Corollary 3.4. Then for all $n > 0$,*

$$(I^{[p^{ne}]} : I) = I^{[p^{ne}]} + (f_e^{1+p^e+\dots+p^{(n-1)e}}).$$

4. Application to centers of sharp F-purity

In [Schwede 2008a], we introduced a notion called centers of sharp F-purity (also known as F-compatible ideals), a positive characteristic analog of a center of log canonicity; see for example [Kawamata 1997a; 1998]. Our main goal in this section is to prove several finiteness theorems about centers of sharp F-purity.

Recall that an ideal I is called F-compatible with respect to (R, Δ) if for every $e > 0$ and every $\phi \in \mathrm{Hom}_R(F^e_* R(\lceil (p^e-1)\Delta \rceil), \Delta)$, we have $\phi(F^e_* I) \subseteq I$. One limitation of the definition of F-compatible ideals is that it seems to require checking infinitely many $e > 0$ (and infinitely many ϕ). However, for radical ideals I, assuming that $(p^e-1)K_X$ is Cartier, we will show that it is enough to check the condition only for that e.

Proposition 4.1. *Suppose that R is a normal F-finite ring. Further suppose that Δ is an effective $\mathbb{Q}$-divisor such that* $\mathrm{Hom}_R(F^e_* R((p^e-1)\Delta), R)$ *is free as an $F^e_* \mathcal{O}_X$-module. Then a radical ideal $I \subset R$ is F-compatible with respect to (R, Δ) if and only if $T_e(F^e_* I) \subseteq I$ where T_e is a $F^e_* R$-module generator of*

$$\mathrm{Hom}(F^e_* R((p^e-1)\Delta), R).$$

Proof. Since a radical ideal I is F-compatible if and only if its minimal associated primes are F-compatible, see Lemma 2.17(5), without loss of generality we may assume that I is prime. Furthermore, since F-compatible ideals behave well with respect to localization, see [Schwede 2008a, Lemma 3.7], we may also assume that R is local and that $I = \mathfrak{m}$ is maximal.

Suppose that $\phi : F^b_* R(\lceil (p^b - 1)\Delta \rceil) \to R$ satisfies the property that $\phi(F^b_* \mathfrak{m}) \nsubseteq \mathfrak{m}$, we will obtain a contradiction. Therefore, for some element $x \in \mathfrak{m}$, we have that $\phi(F^b_* x) = u$ where u is a unit in R. By scaling ϕ, we may assume that $u = 1$. Now choose integers n and m such that $nb = me$. Consider the function $\psi : F^{nb}_* R \to R$ defined by the rule

$$\psi(F^{nb}_* \_) = \phi\Big(x F^b_* \phi\big(x F^{2b}_* \phi(x \cdots F^{(n-1)b}_* \phi(F^{nb}_* \_) \cdots)\big)\Big).$$

Notice that $\psi(F^{nb}_* x) = 1$. On the other hand, $\mathrm{Hom}_R(F^{me}_* R((p^{me} - 1)\Delta), R)$ is generated by T composed with itself $m-1$ times. Notice that since T sends $\mathfrak{m}$ into $\mathfrak{m}$, so does its composition. Therefore, to obtain our contradiction we simply have to check that $\psi \in \mathrm{Hom}_R(F^{nb}_* R, R)$ is an element of $\mathrm{Hom}_R(F^{me}_* R((p^{me}-1)\Delta), R)$. But that is straightforward since it was constructed by composing ϕ with itself (using the fact that we round up, not down, so that $p^a \lceil (p^b - 1)\Delta \rceil + \lceil (p^a - 1)\Delta \rceil \geq \lceil (p^{a+b} - 1)\Delta \rceil$). $\square$

Remark 4.2. For a sharply F-pure pair (R, Δ), all F-compatible ideals are radical.

Corollary 4.3. *Suppose that $\phi : F^e_* R \to R$ is a Frobenius splitting and R is an F-finite normal ring. Then the centers of sharp F-purity for the pair (R, Δ_ϕ) coincide with the subschemes of $X = \mathrm{Spec}\, R$ compatibly split with ϕ.*

Remark 4.4. One might ask if an analog of Proposition 4.1 holds for nonradical ideals, and we do not know the answer in general. However, in [Schwede 2008a], it was shown that the nonfinitistic/big test ideal is the unique smallest F-compatible ideal that intersects nontrivially with R°. Using the additional structure of the big test ideal, we are able to prove an analogous result (in fact, the proof is very similar to a special case of [Takagi 2008, Proposition 3.5(3)]).

Definition 4.5. Suppose that $\phi_e \in \mathrm{Hom}_R(F^e_* R, R)$ is a map. For every integer $n \geq 0$, we define $\phi_{ne} \in \mathrm{Hom}_R(F^{ne}_* R, R)$ to be the map obtained by composing ϕ_e with itself $n-1$ times, just as in Theorem 3.11(e). We set ϕ_0 to be the identity map in $\mathrm{Hom}_R(R, R)$.

Our next goal is to characterize the big test ideal using this machinery. First however, we need two lemmas.

Lemma 4.6. *Suppose that $\mathfrak{a}$ is an ideal generated by l elements and that m and k are integers. Then*:

$$(\mathfrak{a}^m)^{[p^k]} \supseteq \mathfrak{a}^{p^k m + l(p^k - 1)}.$$

Proof. Let $f_1, \dots, f_l$ be a set of generators for $\mathfrak{a}$. Then $\mathfrak{a}^{p^k m + l(p^k-1)}$ is generated by the elements of the form
$$f_1^{b_1} \dots f_l^{b_l},$$
where $\sum_{i=1}^{l} b_i = p^k m + l(p^k - 1)$. We will show that each such element is contained in $(\mathfrak{a}^m)^{[p^k]}$. Write each $b_i = q_i p^k + r_i$ where $0 \leq r_i < p^k$. Thus we have
$$f_1^{b_1} \cdots f_l^{b_l} = (f_1^{q_1} \cdots f_l^{q_l})^{p^k} (f_1^{r_1} \cdots f_l^{r_l}).$$
Note $\sum_{i=1}^{l} r_i \leq l(p^k - 1)$. Therefore,
$$p^k m + l(p^k - 1) = \sum_{i=1}^{l} b_i = \left(p^k \sum_{i=1}^{l} q_i\right) + \left(\sum_{i=1}^{l} r_i\right) \leq \left(p^k \sum_{i=1}^{l} q_i\right) + l(p^k - 1),$$
which implies that $p^k m \leq p^k \sum_{i=1}^{l} q_i$, in particular, $m \leq \sum_{i=1}^{l} q_i$. Therefore,
$$(f_1^{q_1} \cdots f_l^{q_l})^{p^k} \in (\mathfrak{a}^m)^{[p^k]}$$
and so $f_1^{b_1} \cdots f_l^{b_l} \in (\mathfrak{a}^m)^{[p^k]}$ as desired. $\square$

Lemma 4.7. *Suppose that $\mathfrak{a}$ is an ideal of R which can be generated by l elements and such that $\mathfrak{a} \cap R^\circ \neq \varnothing$. Fix an $e > 0$. Then there exists an element $c' \in R^\circ$ such that*
$$c' \mathfrak{a}^{\lceil t(p^{ne+k}-1) \rceil} \subseteq (\mathfrak{a}^{\lceil t(p^{ne}-1) \rceil})^{[p^k]}$$
for all $n > 0$ and all $k < e$.

Proof. First note that we have
$$(\mathfrak{a}^{\lceil t(p^{ne}-1) \rceil})^{[p^k]} \supseteq \mathfrak{a}^{p^k \lceil t(p^{ne}-1) \rceil + l(p^k-1)} \supseteq \mathfrak{a}^{p^k \lceil t(p^{ne}-1) \rceil + lp^e}.$$
The first containment holds by Lemma 4.6 above. Thus it is sufficient to find a c' such that $c' \mathfrak{a}^{\lceil t(p^{ne+k}-1) \rceil} \subseteq \mathfrak{a}^{p^k \lceil t(p^{ne}-1) \rceil + lp^e}$. Choose $c' \in \mathfrak{a}^{(l+1)p^e} \cap R^\circ$. We need to show that
$$(l+1)p^e + \lceil t(p^{ne+k} - 1) \rceil \geq p^k \lceil t(p^{ne} - 1) \rceil + lp^e.$$
However,
$$\begin{aligned} p^k \lceil t(p^{ne} - 1) \rceil + lp^e &\leq p^k \lfloor t(p^{ne} - 1) \rfloor + p^e + lp^e \\ &\leq \lfloor p^k t(p^{ne} - 1) \rfloor + (l+1)p^e \leq \lceil t(p^{ne+k} - 1) \rceil + (l+1)p^e \end{aligned}$$
as desired. $\square$

Proposition 4.8. *Suppose that R is a normal F-finite ring, that Δ is an effective $\mathbb{Q}$-divisor such that $(p^e - 1)\Delta$ is integral, and that $\operatorname{Hom}_R(F_*^e R((p^e - 1)\Delta), R)$ is of rank one and free as an $F_*^e R$-module with generator T_e (viewed as an element of*

$\mathrm{Hom}_R(F^e_* R, R))$. *Set T_{ne} to be the map obtained by composing T_e with itself $n-1$ times. Then we have the following*:

(i) *The big test ideal $\tau_b(R; \Delta)$ is the unique smallest ideal J whose intersection with R° is nontrivial and which satisfies $T_e(F^e_* J) \subseteq J$.*

(ii) *Furthermore, if $\mathfrak{a}$ is an ideal such that $\mathfrak{a} \cap R^\circ \neq \varnothing$ and $t > 0$ is a real number, then the big test ideal $\tau_b(R; \Delta, \mathfrak{a}^t)$ is the unique smallest ideal J whose intersection with R° is nontrivial and which satisfies $T_{ne}(F^{ne}_* \mathfrak{a}^{\lceil t(p^e-1)\rceil} J) \subseteq J$ for all integers $n > 0$.*

Proof. For (i), note that the big test ideal $\tau_b(R; \Delta)$ satisfies $T_e(F^e_* \tau_b(R; \Delta)) \subseteq \tau_b(R; \Delta)$, thanks to [Schwede 2008a, Proposition 6.1]. Thus we simply have to show it is the smallest such ideal. Likewise for (ii), $\tau_b(R; \Delta, \mathfrak{a}^t)$ satisfies the condition $T_{ne}(F^{ne}_* \mathfrak{a}^{\lceil t(p^e-1)\rceil} \tau_b(R; \Delta, \mathfrak{a}^t)) \subseteq \tau_b(R; \Delta, \mathfrak{a}^t)$ for all integers $n > 0$, so we must show that it is the smallest such ideal.

We now claim that statement is local in order to assume that $R = (R, \mathfrak{m})$ is a local ring. We outline the proof of this claim in case (i) since case (ii) is essentially the same. Suppose that J is an ideal which satisfies both $J \cap R^\circ \neq \varnothing$ and $T_e(F^e_* J) \subseteq J$. Then $J + \tau_b(R; \Delta)$ also satisfies both conditions. Note that J does not contain $\tau_b(R; \Delta)$ if and only if we have the strict containment $J + \tau_b(R; \Delta) \supsetneq J$. But in such a case, we can localize at a maximal ideal where the same strict containment holds. Thus we have reduced to the local case. Therefore, from this point forward, we assume that R is a local ring with maximal ideal $\mathfrak{m}$.

Suppose that J is an ideal such that $T_e(F^e_* J) \subseteq J$ — alternatively, such that $T_{ne}(F^{ne}_* \mathfrak{a}^{\lceil t(p^{ne}-1)\rceil} J \subseteq J$, for all $n > 0$ — and such that $J \cap R^\circ \neq \varnothing$. In case (i), notice also that $T_{ne}(F^{ne}_* J) \subseteq J$ for all positive integers n (and thus $\phi(F^{ne}_* J) \subseteq J$ for all $\phi \in \mathrm{Hom}_R(F^{ne}_* R((p^{ne-1} - 1)\Delta), R)$ since T_{ne} is also a generator by Corollary 3.10).

In the setting of (i), fix $d \in J \cap R^\circ$. By applying Matlis duality, we see that the composition

$$E_{R/J} \to E_R \to E_R \otimes_R F^{ne}_* R \to E_R \otimes_R F^{ne}_* R((p^{ne} - 1)\Delta) \xrightarrow{F^{ne}_*(\times d)} E_R \otimes_R F^{ne}_* R((p^{ne} - 1)\Delta)$$

is zero for every integer $n > 0$. Likewise, in the setting of (ii), for each $d \in J \cap R^\circ$ and each $a \in \mathfrak{a}^{\lceil t(p^{ne}-1)\rceil}$, we have that the composition

$$E_{R/J} \to E_R \to E_R \otimes_R F^{ne}_* R \to E_R \otimes_R F^{ne}_* R((p^{ne} - 1)\Delta) \xrightarrow{F^{ne}_*(\times da)} E_R \otimes_R F^{ne}_* R((p^{ne} - 1)\Delta)$$

is zero for every integer $n > 0$.

We now want to show that $E_{R/J} \subset 0^{*\Delta}_{E_R}$ (respectively $E_{R/J} \subset 0^{*\Delta,\mathfrak{a}^t}_{E_R}$) because $\mathrm{Ann}_R(0^{*\Delta}_{E_R}) = \tau_b(R;\Delta)$ (respectively $\mathrm{Ann}_R(0^{*\Delta,\mathfrak{a}^t}_{E_R}) = \tau_b(R;\Delta,\mathfrak{a}^t)$). Therefore, choose $z \in E_{R/J}$. By assumption $dz^{p^{ne}} = 0 \in E_R \otimes_R F^{ne}_* R((p^{ne}-1)\Delta)$ for all $n > 0$ (respectively, $d\mathfrak{a}^{\lceil t(p^{ne}-1)\rceil} z^{p^{ne}} = 0 \in E_R \otimes_R F^{ne}_* R((p^{ne}-1)\Delta)$ for all $n > 0$). We need to verify a similar statement for powers of p that are not multiples of e, and so now the proof becomes quite similar to [Hochster and Huneke 1990, Lemma 8.16].

In the setting of (i), we claim that $F^{ne}_* R((p^{ne}-1)\Delta)$ naturally maps to

$$F^{k+ne}_* R(\lceil (p^{k+ne}-1)\Delta\rceil)$$

for any $k > 0$ via the k-iterated action of Frobenius. To see this explicitly, apply $\mathrm{Hom}_R(R(-\lceil (p^{ne}-1)\Delta\rceil), \_)$ to the map $R \to F^k_* R(\lceil (p^k-1)\Delta\rceil)$. Tensoring with E_R then gives us a map

$$\begin{aligned} F^{ne}_* R((p^{ne}-1)\Delta) \otimes_R E_R &\to F^{k+ne}_* R(\lceil (p^{k+ne}-1)\Delta\rceil) \otimes_R E_R \\ dz^{p^n e} = d \otimes z &\mapsto d^{p^k} \otimes z = d^{p^k} z^{p^{k+ne}} \end{aligned}$$

which factors the map $E_R \to F^{k+ne}_* R(\lceil (p^{k+ne}-1)\Delta\rceil) \otimes_R E_R$. Hence, $d^{p^k} z^{p^{ne+k}}$ vanishes for all $k, n > 0$.

Choose $c = d^{p^{e-1}}$ and choose $j > 0$ arbitrary. Write $j = ne + k$ where $k < e$. Then

$$cz^{p^j} = d^{p^{e-1}} z^{p^{ne+k}} = d^{p^{e-1}-p^k} d^{p^k} z^{p^{ne+k}} = d^{p^{e-1}-p^k} 0 = 0,$$

as desired. Therefore, $E_{R/J} \subset 0^{*\Delta}_{E_R}$ so that

$$J = \mathrm{Ann}_R(E_{R/J}) \supseteq \mathrm{Ann}_R(0^{*\Delta}_{E_R}) = \tau_b(R;\Delta),$$

which proves (i).

In case (ii), using a similar argument, we still have $d^{p^k} (\mathfrak{a}^{\lceil t(p^{ne}-1)\rceil})^{[p^k]} z^{p^{ne+k}} = 0$ for all $k, n > 0$. By Lemma 4.7, there exists a $c' \in R^\circ$ such that

$$c' \mathfrak{a}^{\lceil t(p^{ne+k}-1)\rceil} \subseteq (\mathfrak{a}^{\lceil t(p^{ne}-1)\rceil})^{[p^k]}$$

for all $n > 0$ and all $k < e$.

Set $c = c' d^{p^{e-1}}$, choose $j > 0$ arbitrary and write $j = ne + k$ where $k < e$. Then

$$\begin{aligned} c\mathfrak{a}^{\lceil t(p^j-1)\rceil} z^{p^j} &= d^{p^{e-1}} c' \mathfrak{a}^{\lceil t(p^{ne+k}-1)\rceil} z^{p^{ne+k}} \\ &\subseteq d^{p^{e-1}-p^k} d^{p^k} (\mathfrak{a}^{\lceil t(p^{ne}-1)\rceil})^{[p^k]} z^{p^{ne+k}} \\ &= d^{p^{e-1}-p^k} 0 = 0, \end{aligned}$$

as desired. $\square$

5. *F*-adjunction

In this section, we reinterpret the following observation using the language from the previous sections.

Observation 5.1. Suppose that $(R, \mathfrak{m})$ is an F-finite local ring and

$$\phi \in \operatorname{Hom}_R(F^e_* R, R).$$

Further suppose that I is a proper ideal of R such that $\phi(F^e_* I) \subseteq I$. Then there is a diagram

$$\begin{tikzcd} F^e_* R \arrow[r, "\phi"] \arrow[d, "F^e_* \alpha"'] & R \arrow[d, "\alpha"] \\ F^e_*(R/I) \arrow[r, "\phi_I"] & R/I \end{tikzcd}$$

where the vertical arrows are the natural quotients.

- Because R is local, ϕ is surjective if and only if ϕ_I is surjective.

When we apply the correspondence between effective $\mathbb{Q}$-divisors and

$$\phi \in \operatorname{Hom}_R(F^e_* R, R),$$

we obtain the following result.

Theorem 5.2. *Suppose that R is a reduced F-finite normal ring and that (R, Δ). Assume also that $(p^e-1)\Delta$ is an integral divisor such that we have an isomorphism* $\operatorname{Hom}_R(F^e_* R((p^e-1)\Delta), R) \cong F^e_* R$ *of $F^e_* R$-modules. Further suppose that $I \subset R$ is F-compatible with respect to (R, Δ) and that R/I is normal. Finally suppose that (R, Δ) is sharply F-pure at the generic points of* $\operatorname{Spec} R/I$ (*that is*, *after localizing at the minimal primes of I*). *Then there exists a canonically determined effective $\mathbb{Q}$-divisor $\Delta_{R/I}$ on* $\operatorname{Spec} R/I$ *satisfying the following properties*:

(i) $(p^e - 1)(K_{R/I} + \Delta_{R/I})$ *is an integral Cartier divisor*

(ii) $\operatorname{Hom}_{R/I}\big(F^e_*((R/I)((p^e-1)\Delta_{R/I})), R/I\big) \cong F^e_*(R/I)$ *as $F^e_*(R/I)$-modules.*

(iii) (R, Δ) *is sharply F-pure near* $\operatorname{Spec} R/I$ *if and only if $(R/I, \Delta_{R/I})$ is sharply F-pure.*

(iv) *For any ideal $\mathfrak{a} \subseteq R$ which is not contained in any minimal prime of I and any real number $t > 0$, we have that $(R, \Delta, \mathfrak{a}^t)$ is sharply F-pure near* $\operatorname{Spec} R/I$ *if and only if $(R/I, \Delta_{R/I}, \overline{\mathfrak{a}}^t)$ is sharply F-pure.*

(v) *I is maximal with respect to containment among F-compatible ideals for the pair (R, Δ) (in other words, I is a minimal center of sharp F-purity), if and only if $(R/I, \Delta_{R/I})$ is a strongly F-regular pair and R/I is a domain.*[4]

[4] In fact, if we assume that I is maximal among F-compatible ideals, then it follows that R/I is a normal domain and so the assumption that R/I is normal is unnecessary.

(vi) *There exists a natural bijection between the centers of sharp F-purity of $(R/I, \Delta_{R/I})$ and the centers of sharp F-purity of (R, Δ) which contain I.*

Remark 5.3. Roughly speaking, properties (iii), (iv), (v) and (vi) imply that the singularities of $(R/I, \Delta_{R/I})$ are very closely related to the singularities of (R, Δ) near I. Compare with [Kawamata 1998; 2007; 2008; Ein and Mustaţă 2009; Ambro 1999; Ein et al. 2003].

Proof. Given Δ as above, associate a $\phi \in \operatorname{Hom}_R(F^e_* R, R)$ as in Theorem 3.13. Just as in Observation 5.1, we associate a $\phi_I \in \operatorname{Hom}_{R/I}(F^e_*(R/I), R/I)$, to which we associate a divisor $\Delta_{R/I}$. By construction (and using Theorem 3.11) we see that the existence and that properties (i) and (ii) are obvious. For the rest of the properties, it is harmless to assume that R is local. Notice that the map ϕ_I is not the zero map on any irreducible component of $\operatorname{Spec} R/I$ because (R, Δ) is sharply F-pure at the minimal primes of I. To show that $\Delta_{R/I}$ is canonically determined, note that if one chooses a different $\phi : F^e_* R \to R$ associated to Δ, the associated map ϕ_I will differ from the original choice by multiplication by a unit, and so $\Delta_{R/I}$ will not change. Likewise, if one chooses a different $e > 0$, then using Theorem 3.11(e,f), we obtain the same $\Delta_{R/I}$ yet again.

In terms of (iii), this simply follows from Observation 5.1. Notice now that (iv) is a generalization of (iii). Condition (iv) follows by an argument similar to the one in Observation 5.1 since we simply consider a diagram

$$\begin{tikzcd}
F^d_* R \arrow[r, "F^d_*(\times a)"] \arrow[d, "F^d_*\alpha"] & F^d_* R \arrow[r, "\phi^n"] \arrow[d, "F^d_*\alpha"] & R \arrow[d, "\alpha"] \\
F^d_* R/I \arrow[r, "F^d_*(\times \overline{a})"] & F^d_* R/I \arrow[r, "\phi^n_I"] & R/I
\end{tikzcd}$$

for each $d = ne$ instead and various $a \in \mathfrak{a}^{\lceil t(p^d-1)\rceil}$. In the diagram above, ϕ^n is the composition of ϕ with itself $n-1$ times as before. Now again, the map obtained by composing the bottom row is surjective if an only if the map obtained from composing the top row is surjective.

Condition (v) will follow from (vi) since a pair is strongly F-regular if and only if it has no centers of sharp F-purity. Therefore, we conclude by proving (vi). Suppose that $P \in \operatorname{Spec} R$ contains I, and corresponds to $\overline{P} \in \operatorname{Spec} R/I$. We will show that P is a center of sharp F-purity of (R, Δ) if and only if $\overline{P}$ is a center of sharp F-purity for $(R/I, \Delta_{R/I})$. First suppose that P is a center of sharp F-purity for (R, Δ). This is equivalent to the condition that $\phi(F^e_* P) \subseteq P$. This implies that $\phi_I(F^e_* \overline{P}) \subseteq \overline{P})$. The converse direction reverses this and is essentially the same as the argument given in the proof of [Schwede 2008a, Proposition 7.5]. $\square$

Remark 5.4. I do not know if one can somehow generalize the "centers of sharp F-purity" of condition (vi) to all F-compatible ideals. It is not hard to see that one does obtain a bijection between radical F-compatible ideals since they are intersections of centers of sharp F-purity. Section 6 is concerned with proving an analog of (vi) for the big test ideal.

Using the ideas of Fedder's criterion, we also obtain the following result.

Theorem 5.5. *Suppose that S is a regular F-finite ring such that $F^e_* S$ is a free S module* (*for example*, *if S is local*) *and that $R = S/I$ is a quotient that is a normal domain. Further suppose that Δ_R is an effective $\mathbb{Q}$-divisor on* $\operatorname{Spec} R$ *such that* $\operatorname{Hom}_R(F^e_* R((p^e-1)\Delta), R)$ *is a rank one free $F^e_* R$-module* (*for example*, *if R is local and $(p^e-1)(K_R+\Delta)$ is Cartier*). *Then there exists an effective $\mathbb{Q}$-divisor Δ_S on* $\operatorname{Spec} S$ *such that*:

(a) $(p^e-1)(K_S+\Delta_S)$ *is Cartier.*

(b) *I is (Δ_S, F)-compatible and (S, Δ_S) is sharply F-pure at the minimal associated primes of I* (*that is*, *at the generic points of* $\operatorname{Spec} S/I$).

(b) *Δ_S induces Δ_R as in Theorem 5.2.*

Proof. The key point is that every map $F^e_* R \to R$ is obtained by restricting a map $F^e_* S \to S$ to R, see [Fedder 1983, Lemma 1.6]. Note that condition (b) follows immediately since the map $F^e_* R \to R$ we are concerned with is nonzero. □

Remark 5.6. The Δ_S constructed in the above theorem is in no way canonically chosen.

Remark 5.7. I do not know of anything like a characteristic zero analog of this except in the case that $X \subseteq Y$ is a complete intersection [Ein and Mustaţă 2004]; also compare with [Kawakita 2008; Ein and Mustaţă 2009].

We now show that for an F-pure pair, there are at most finitely many centers of sharp F-purity (equivalently there are at most finitely many (Δ, F)-compatible ideals). We give a proof that is written using the language of divisors. However the same proof may be given without this language (this was done in a preprint of this paper). This result was proved for local rings in [Schwede 2008a, Corollary 5.2], using the method of [Enescu and Hochster 2008] or a modification of the method of [Sharp 2007]. Finally, essentially the same result has also been obtained independently in [Metha and Kumar 2009].

Theorem 5.8. *If $(R, \Delta, \mathfrak{a}_\bullet)$ is sharply F-pure, then there are at most finitely many centers of sharp F-purity.*

Proof. We may prove this on a finite affine cover of Spec R. Thus, we may assume[5] there exists a map $\phi : F^e_* R(\lceil (p^e-1)\Delta \rceil) \to R$ that sends some element $a \in F^e_* \mathfrak{a}_{p^e-1}$ to 1. Note, every center of sharp F-purity $Q \in \operatorname{Spec} R$ for $(R, \Delta, \mathfrak{a}_\bullet)$ satisfies $\phi(F^e_* a Q) \subseteq Q$. Our goal is to show that there are finitely many prime ideals Q such that $\phi(F^e_* a Q) \subseteq Q$.

First note that we can replace $\phi(\_)$ by $\phi(a \times \_)$ and so ignore the term a. For a contradiction, assume there are infinitely many such prime ideals Q such that $\phi(F^e_* Q) \subseteq Q$. We choose a collection $\mathfrak{Q}$ of infinitely many primes ideals Q satisfying:

(i) $\phi(F^e_* Q) \subseteq Q$.

(ii) All $Q \in \mathfrak{Q}$ have the same height.

(iii) The closure of the set $\mathfrak{Q}$ in the Zariski topology is an irreducible (possibly nonproper) closed subset W of Spec R. We set P to be the generic point of that subset W (in other words, $P = \cap_{Q \in \mathfrak{Q}} Q$).

Using the pigeonhole principle, it is not difficult to see that a set $\mathfrak{Q}$ satisfying conditions (i), (ii) and (iii) exists.

We make two observations about the prime ideal P:

- P must have smaller height than the elements of $\mathfrak{Q}$.
- P satisfies $\phi(F^e_* P) \subseteq P$ since P is the intersection of the elements of $\mathfrak{Q}$.

By restricting to an open affine set of Spec R containing P, we may assume that R/P is normal (the elements of $\mathfrak{Q}$ will still form a dense subset of Spec R/P). Therefore, ϕ induces a divisor Δ_P on Spec R/P as in Theorem 5.2. The set of elements in $\mathfrak{Q}$ restrict to centers of sharp F-purity for $(R/P, \Delta_P)$ by Theorem 5.2(vi). As noted above, $\{Q/P \mid Q \in \mathfrak{Q}\}$ is dense in Spec R/P and simultaneously $\{Q/P \mid Q \in \mathfrak{Q}\}$ is contained in the nonstrongly F-regular locus of $(R/P, \Delta_P)$, which is closed and proper. This is a contradiction. □

Remark 5.9. If one wishes to assume that R is not necessarily normal and that $\Delta = 0$, or even that Δ is some sort of appropriate generalization of a $\mathbb{Q}$-divisor (see for example [Hartshorne 2007] or [Kollár et al. 1992, Chapter 16]), the proof goes through without change.

Corollary 5.10. *Suppose that X is a noetherian F-finite Frobenius split scheme with splitting $\phi : F^e_* \mathcal{O}_X \to \mathcal{O}_X$, then there exists at most finitely many ϕ-compatibly split subschemes.*

Proof. Use a finite affine cover of X. On each open affine subset, there are finitely many compatibly split subschemes by the above argument. □

[5] This happens after localizing each point, so it happens in a neighborhood of each point, so we may use such neighborhoods to cover Spec R

6. Comments on adjoint-like test ideals and restriction theorems

Based on the work of Takagi, it is natural to hope that there is a restriction theorem of (generalized) adjoint-like test ideals, similar to the ones in [Takagi 2007; 2008]. Using the results of the previous section, we can accomplish this.

Definition 6.1. Suppose that R is F-finite normal ring and that $(R, \Delta, \mathfrak{a}^t)$ is a triple. Further suppose that :

(a) $Q \in \operatorname{Spec} R$ is a center of sharp F-purity for (R, Δ).

(b) $\mathfrak{a} \cap (R \setminus Q) \neq \varnothing$.

(c) $(R_Q, \Delta|_{\operatorname{Spec} R_Q})$ is sharply F-pure.

(d) R/Q is normal.

(e) There exists an integer e_0 such that $\operatorname{Hom}_R(F_*^{e_0} R((p^{e_0} - 1)\Delta), R)$ is free as an $F_*^{e_0} R$-module.

(f) The integer e_0 is the smallest positive integer satisfying condition (e).

Fix a map $\phi_{e_0} = \phi : F_*^{e_0} R \to R$ corresponding to Δ. We define the *big test ideal of* $(R, \Delta, \mathfrak{a}^t)$ *outside of* Q, denoted $\tau_b(R; \not\subseteq Q; \Delta, \mathfrak{a}^t)$ (if it exists), to be the smallest ideal J satisfying the following two conditions:

- J is not contained in Q (that is, $J \cap (R \setminus Q) \neq \varnothing$).
- $\phi_{ne_0}(F_*^{ne_0} \mathfrak{a}^{\lceil t(p^{ne_0} - 1) \rceil} J) \subseteq J$ for all $n \geq 0$ where ϕ_{ne_0} is as in Definition 4.5.

Remark 6.2. With regard to Definition 6.1(b), using the fact that

$$\mathfrak{a} \cap (R \setminus Q) \neq \varnothing,$$

we see that Q is a center of sharp F-purity for (R, Δ) if and only if it is a center of sharp F-purity for $(R, \Delta, \mathfrak{a}^t)$. Likewise, the localized pair $(R_Q, \Delta|_{\operatorname{Spec} R_Q})$ is sharply F-pure if and only if the localized triple

$$(R_Q, \Delta|_{\operatorname{Spec} R_Q}, (\mathfrak{a} R_Q)^t)$$

is sharply F-pure since $\mathfrak{a} R_Q = R_Q$.

Remark 6.3. It is unnecessary to choose e_0 to be the *smallest* integer satisfying condition (e). If one uses any integer e_0 satisfying condition (e), then one obtains the same $\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t)$. We will not verify this here as the proof is rather involved and is essentially the same argument as in Proposition 4.8.

Remark 6.4. It is also interesting to study the smallest ideal J which properly contains Q and such that $\phi_{ne_0}(F_*^{ne_0} \mathfrak{a}^{\lceil t(p^{ne_0} - 1) \rceil} J) \subseteq J$ for all $n \geq 0$ (again, if it exists). For future reference, we will denote that ideal by $\tau_b(R, \supseteq Q; \Delta, \mathfrak{a}^t)$.

Remark 6.5. If $\mathfrak{a} = R$, then $\tau_b(R, \not\subseteq Q; \Delta) = \tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t)$ is the unique smallest ideal not contained in Q such that $\phi_{e_0}(F^{e_0}_* J) \subseteq J$. Likewise, if $\mathfrak{a} = R$, $\tau_b(R, \supseteq Q; \Delta) = \tau_b(R, \supseteq Q; \Delta, \mathfrak{a}^t)$ is the smallest ideal properly containing Q such that $\phi_{e_0}(F^{e_0}_* J) \subseteq J$.

Remark 6.6. It is probably interesting to look at nonprime radical ideals Q which are F-compatible with respect to (R, Δ). Set $R^{\circ Q}$ to be the set of elements not contained in any minimal prime of Q. In that case, one should probably consider ideals J minimal with respect to the conditions that $J \cap R^{\circ Q} \neq \varnothing$ and $\phi(F^{e_0}_* J) \subseteq J$. If one takes Q to be the zero ideal of R, then $\tau_b(R, \not\subseteq Q; \Delta)$ is just the usual big test ideal, see Proposition 4.8. However, in this paper, we will not work in this generality.

Remark 6.7. Suppose that the ideals $\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t)$ and $\tau_b(R, \supseteq Q; \Delta, \mathfrak{a}^t)$ exist. Notice that $\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t) \subseteq \tau_b(R, \supseteq Q; \Delta, \mathfrak{a}^t)$. Furthermore, we claim that

$$\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t) + Q = \tau_b(R, \supseteq Q; \Delta, \mathfrak{a}^t). \tag{6.7.1}$$

The containment $\supseteq$ follows from the definition of $\tau_b(R, \supseteq Q; \Delta, \mathfrak{a}^t)$ because $\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t) + Q$ satisfies

$$\phi_{ne_0}(F^{ne_0}_* \mathfrak{a}^{\lceil t(p^{ne_0}-1)\rceil}(\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t) + Q)) \subseteq \tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t) + Q \tag{6.7.2}$$

since both Q and $\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t)$ satisfy the condition of Equation (6.7.2). But then since both $\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t)$ and Q are contained in $\tau_b(R, \supseteq Q; \Delta, \mathfrak{a}^t)$, we are done.

We can now prove that $\tau_b(R, \supseteq Q; \Delta, \mathfrak{a}^t)$ exists.

Proposition 6.8. *Suppose that $(R, \Delta, \mathfrak{a}^t)$ and $Q \in \operatorname{Spec} R$ are as in Definition 6.1. Further suppose that $\alpha : R \to R/Q$ is the natural surjection. Suppose that $\Delta_{R/Q}$ is the $\mathbb{Q}$-divisor on $\operatorname{Spec} R/Q$ corresponding to Δ as in Theorem 5.2. Then $\tau_b(R, \supseteq Q; \Delta, \mathfrak{a}^t)$ exists and is equal to $\alpha^{-1}(\tau_b(R/Q; \Delta_{R/Q}, \overline{\mathfrak{a}}^t))$. In particular*

$$\tau_b(R, \supseteq Q; \Delta, \mathfrak{a}^t)/Q = \tau_b(R, \supseteq Q; \Delta, \mathfrak{a}^t)|_{R/Q} = \tau_b(R/Q; \Delta_{R/Q}, \overline{\mathfrak{a}}^t).$$

Proof. As noted before, it is easy to see that if J contains Q and

$$\phi_{ne_0}(F^e_* \mathfrak{a}^{\lceil t(p^{ne_0}-1)\rceil} J) \subseteq J,$$

then $\phi_{ne_0, Q}(F^e_* \overline{\mathfrak{a}}^{\lceil t(p^{ne_0}-1)\rceil}(J/Q)) \subseteq J/Q$. Conversely, if we have an ideal $J \supseteq Q$ such that $\phi_{ne_0, Q}(F^e_* \overline{\mathfrak{a}}^{\lceil t(p^{ne_0}-1)\rceil}(J/Q)) \subseteq J/Q$ then

$$\phi_{ne_0}(F^e_* \mathfrak{a}^{\lceil t(p^{ne_0}-1)\rceil} J) \subseteq J + Q = J.$$

But ideals of R containing Q are in bijection with ideals of R/Q. This completes the proof. $\square$

Once we have verified that $\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t)$ exists, Proposition 6.8 will immediately imply the following restriction theorem.

Corollary 6.9. *Suppose that* $(R, \Delta, \mathfrak{a}^t)$ *and* $Q \in \operatorname{Spec} R$ *are as in Definition 6.1. Further suppose that* $\Delta_{R/Q}$ *is the* $\mathbb{Q}$*-divisor on* R/Q *corresponding to* Δ *as in Theorem 5.2. Then*

$$\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t)|_{R/Q} = (\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t) + Q)|_{R/Q} = \tau_b(R/Q; \Delta_{R/Q}, \overline{\mathfrak{a}}^t).$$

Proof. Apply Proposition 6.8 and Equation (6.7.1). The result will follow once we know that $\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t)$ exists. □

The rest of the section will be devoted to proving that the ideal $\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t)$ exists.

Remark 6.10. One way to do this is by working out a version of tight closure theory using $c \in R \setminus Q$ instead of $c \in R^\circ$. However, we will use a more direct approach.

We begin with several lemmas which are essentially the same as those used in the proof the existence of test elements. The main technical result of the section is Proposition 6.14, which combines the following three lemmas.

Lemma 6.11. *Suppose that* (R, Δ) *is a sharply* F*-pure pair,* $(p^e - 1)(K_R + \Delta)$ *is integral, and that* $\operatorname{Hom}_R(F^e_* R((p^e - 1)\Delta), R)$ *is free as an* $F^e_* R$*-module with generator* ϕ_e *(by restriction, we also view* ϕ_e *as an element of* $\operatorname{Hom}_R(F^e_* R, R)$*). Further suppose that* $d \in R$ *is an element not contained in any center of* F*-purity for* (R, Δ)*.*

Then:

(i) $1 \in \phi_{n_0 e}(F^{n_0 e}_*(dR))$ *for some* $n_0 > 0$.

(ii) *There exists* $n_0 > 0$ *such that* $1 \in \phi_{ne}(F^{ne}_*(dR))$ *for all* $n \geq n_0$.

Proof. We begin by proving (i). First we claim that the statement is local. Another way to phrase the conclusion of the lemma is that $\phi_{ne}(F^{ne}_*(dR)) = R$. However, $\phi_{ne}(F^{ne}_*(dR)) = R$ (for a fixed n) if and only if it is true after localizing at each maximal ideal. Conversely, if $(\phi_{n_i e})_{\mathfrak{m}_i}(F^{n_i e}_* dR_{\mathfrak{m}_i}) = R_{\mathfrak{m}_i}$ after localizing at some maximal ideal $\mathfrak{m}_i$ for some n_i, then it holds in a neighborhood of $\mathfrak{m}_i$ for the same n_i. Cover Spec R by a finite number of such neighborhoods and choose a sufficiently large n that works on all neighborhoods.[6] Therefore we may assume that $R = (R, \mathfrak{m})$ is local. Note that this is essentially the same as the usual proof that strong F-regularity localizes.

[6]Note that if $1 \in \phi_{ne}(F^{ne}_*(dR))$ then $1 \in \phi_{ne}(F^{ne}_* R)$. By composition, this implies that $1 \in \phi_{mne}(F^{mne}_*(dR))$ for all integers $m > 0$.

Choose a minimal center Q of sharp F-purity for (R, Δ) and mod out by Q. It follows that $(R/Q, \Delta_{R/Q})$ is strongly F-regular and also that $\overline{d} \neq 0 \in R/Q$.

In particular, for some $n > 0$, we have $\overline{\phi}_{ne}(F^{ne}_* \overline{d} R/Q) = R/Q$. Therefore, we can find an element $\overline{b} \in R/Q$ such that $\overline{\phi}_{ne}(F^{ne}_* \overline{db}) = 1 \in R/Q$. By choosing an arbitrary $b \in R$ such that the coset $b + Q = \overline{b}$, we see that $\phi_{ne}(F^{ne}_* db) = 1 + x$ for some $x \in Q$. Since R is local, $Q \subseteq \mathfrak{m}$ and $1+x$ is a unit, we have $1 \in \phi_{ne}(F^{ne}_*(dR))$ as desired.

We now prove (ii). Let n_0 be the integer from part (i). It follows that $1 \in \phi_{n_0 e}(F^{n_0 e}_* R)$ so there exists an element $f \in R$ such that $1 = \phi_{n_0 e}(F^{n_0 e}_* f)$. In particular, the map

$$\begin{aligned} F^{n_0 e} : R &\to F^{n_0 e}_* R((p^{n_0 e} - 1)\Delta) \\ 1 &\mapsto F^{n_0 e}_* 1 \end{aligned}$$

splits. This implies that $F^e : R \to F^e_* R((p^e - 1)\Delta)$ also splits. But then $1 \in \phi_e(F^e_* R)$ since ϕ_e was chosen as a generator of $\mathrm{Hom}_R(F^e_* R((p^e - 1)\Delta), R)$. Therefore we see that,

$$1 \in \phi_e(F^e_* R) = \phi_e(F^e_* \phi_{ne}(F^{ne}_*(dR))) = \phi_{(n+1)e}(F^{(n+1)e}_*(dR)).$$

Repeatedly applying ϕ_e will then complete the proof of (ii). □

Lemma 6.12. *Suppose that $(R, \Delta, \mathfrak{a}^t)$ is a triple and $Q \in \mathrm{Spec}\, R$ is a center of F-purity satisfying the conditions from Definition 6.1. Then there exists an element $c \in R \setminus Q$ that satisfies the following condition*:

For all $d \in R \setminus Q$ and for all sufficiently large $n > 0$, there exists an integer $m' > 0$ (which depends on both n and d) such that $c^{m'} \in \phi_{ne_0}(F^{ne_0}_* d\mathfrak{a}^{\lceil t(p^{ne}-1)\rceil})$.

Proof. Choose $c \in \mathfrak{a} \cap (R \setminus Q)$ so that

(a) $(R_c, \Delta|_{\mathrm{Spec}\, R_c})$ is sharply F-pure.

(b) There are no centers of sharp F-purity for $(R_c, \Delta|_{\mathrm{Spec}\, R_c})$ which contain QR_c (as an ideal).

(c) All centers of sharp F-purity for $(R_c, \Delta|_{\mathrm{Spec}\, R_c})$ are contained in QR_c (as ideals).

In particular, $d/1 \in R_c$ is not contained in any centers of sharp F-purity for $(R_c, \Delta|_{\mathrm{Spec}\, R_c})$. Note conditions (b) and (c) above may be summarized by saying that QR_c is the unique maximal height (as an ideal) center of sharp F-purity.

Therefore, by Lemma 6.11, we know that for all $n \gg 0$, $1 \in (\phi_{ne_0})_c(F^{ne_0}_*(dR_c))$. This implies that $c^{m'} \in \phi_{ne_0}(F^{ne_0}_* d\mathfrak{a}^{\lceil t(p^{ne_0}-1)\rceil})$ for some m'. □

Lemma 6.13. *Suppose that for some $e > 0$, we have a map $\gamma_e : F^e_* R \to R$ such that $b \in \gamma_e(F^e_* \mathfrak{a}^{\lceil t(p^e-1)\rceil})$. Then for all $n > 0$, $b^2 \in \gamma_{ne}(F^{ne}_* \mathfrak{a}^{\lceil t(p^{ne}-1)\rceil})$. Here γ_{ne} is the map obtained by composing γ with itself $n-1$ times, as in Definition 4.5.*

Proof. We proceed by induction. The case $n = 1$ was given by hypothesis. Now suppose the result holds for n (that is, $b^2 \in \gamma_{ne}(F^{ne}_* \mathfrak{a}^{\lceil t(p^{ne}-1)\rceil})$). However,

$$\begin{aligned} b^2 \in b\gamma_e(F^e_* \mathfrak{a}^{\lceil t(p^e-1)\rceil}) &= \gamma_e\big(F^e_* \mathfrak{a}^{\lceil t(p^e-1)\rceil} b^{p^e}\big) \subseteq \gamma_e\big(F^e_* \mathfrak{a}^{\lceil t(p^e-1)\rceil} b^2\big) \\ &\subseteq \gamma_e\big(F^e_* \mathfrak{a}^{\lceil t(p^e-1)\rceil} \gamma_{ne}(F^{ne}_* \mathfrak{a}^{\lceil t(p^{ne}-1)\rceil})\big) \\ &= \gamma_e\big(F^e_* \gamma_{ne}(F^{ne}_* (\mathfrak{a}^{\lceil t(p^e-1)\rceil})^{[p^{ne}]} \mathfrak{a}^{\lceil t(p^{ne}-1)\rceil})\big) \\ &\subseteq \gamma_{(n+1)e}\big(F^{(n+1)e}_* \mathfrak{a}^{\lceil t(p^{(n+1)e}-1)\rceil}\big), \end{aligned}$$

as desired. □

We now come to the main technical result of the section.

Proposition 6.14. *Assume the notation and conventions from Definition 6.1. There is an element $b \in R \setminus Q$ such that for every $d \in R \setminus Q$, there exists an integer $n_d > 0$ such that $b \in \phi_{n_d e_0}(F^{n_d e_0}_* d\mathfrak{a}^{\lceil t(p^{n_d e_0}-1)\rceil})$. Note that b does not depend on d.*

Proof. Fix $c \in R \setminus Q$ satisfying Lemma 6.12. Then there exist integers $n_1, m_1 > 0$ such that $c^{m_1} \in \phi_{n_1 e_0}(F^{n_1 e_0}_*(1)\mathfrak{a}^{\lceil t(p^{n_1 e_0}-1)\rceil})$. An application of Lemma 6.13 then implies that $c^{2m_1} \in \phi_{nn_1 e_0}(F^{nn_1 e_0}_*(1)\mathfrak{a}^{\lceil t(p^{nn_1 e_0}-1)\rceil})$ for all $n > 0$. We will show that $c^{3m_1} = b$ works.

Likewise, by Lemma 6.12, for some $n' > 0$ there exists m_d such that

$$c^{m_d} \in \phi_{n' e_0}\big(F^{n' e_0}_*(d)\mathfrak{a}^{\lceil t(p^{n' e_0}-1)\rceil}\big).$$

If $m_d < 3m_1$, we are done (set $n_d = n'$). Otherwise, choose $n > 0$ such that $m_1 p^{nn_1 e_0} \geq m_d$. Then,

$$\begin{aligned} c^{3m_1} &= c^{m_1} c^{2m_1} \in c^{m_1} \phi_{nn_1 e_0}\big(F^{nn_1 e_0}_* \mathfrak{a}^{\lceil t(p^{nn_1 e_0}-1)\rceil}\big) \\ &= \phi_{nn_1 e_0}\big(F^{nn_1 e_0}_* \mathfrak{a}^{\lceil t(p^{nn_1 e_0}-1)\rceil} c^{m_1 p^{nn_1 e_0}}\big) \subseteq \phi_{nn_1 e_0}\big(F^{nn_1 e_0}_* \mathfrak{a}^{\lceil t(p^{nn_1 e_0}-1)\rceil} c^{m_d}\big) \\ &\subseteq \phi_{nn_1 e_0}\big(F^{nn_1 e_0}_* \mathfrak{a}^{\lceil t(p^{nn_1 e_0}-1)\rceil} \phi_{n' e_0}(F^{n' e_0}_*(d)\mathfrak{a}^{\lceil t(p^{n' e_0}-1)\rceil})\big) \\ &= \phi_{nn_1 e_0}\big(F^{nn_1 e_0}_* \phi_{n' e_0}(F^{n' e_0}_*(d)(\mathfrak{a}^{\lceil t(p^{nn_1 e_0}-1)\rceil})^{[p^{n' e_0}]} \mathfrak{a}^{\lceil t(p^{n' e_0}-1)\rceil})\big) \\ &\subseteq \phi_{(nn_1+n')e_0}\big(F^{(nn_1+n')e_0}_*(d)\mathfrak{a}^{\lceil t(p^{(nn_1+n')e_0}-1)\rceil}\big). \end{aligned}$$

Thus we can choose $n_d = nn_1 + n'$, which completes the proof. □

Remark 6.15. The b from the previous proposition can be used as a big sharp test element for the variant of tight closure mentioned in Remark 6.10. In fact, to prove the existence of big sharp test elements, one still has to prove Proposition 6.14 or something closely related to it.

Definition 6.16 [Hara and Takagi 2004] . Fix b as in Proposition 6.14. Then we define the ideal $\widetilde{\tau}(R; b, \Delta, \mathfrak{a}^t)$ as follows:

$$\widetilde{\tau}(R; b, \Delta, \mathfrak{a}^t) := \sum_{n \geq 0} \phi_{ne_0}(F_*^{ne_0} b \mathfrak{a}^{\lceil t(p^{ne_0}-1)\rceil}).$$

Note that the sum stabilizes as a finite sum since R is noetherian.

We make several observations about this ideal (and then we will show it is equal to $\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t)$).

Lemma 6.17. *With notation as above, we have the following two results:*

(i) $b \in \widetilde{\tau}(R; b, \Delta, \mathfrak{a}^t)$. *In particular*, $\widetilde{\tau}(R; b, \Delta, \mathfrak{a}^t) \cap (R \setminus Q) \neq \varnothing$.

(ii) *For all* $n' \geq 0$, $\phi_{n'e_0}(F_*^{n'e_0} \mathfrak{a}^{\lceil t(p^{n'e_0}-1)\rceil} \widetilde{\tau}(R; b, \Delta, \mathfrak{a}^t)) \subseteq \widetilde{\tau}(R; b, \Delta, \mathfrak{a}^t)$.

Proof. For (i), simply set $d = b$ and apply Proposition 6.14. For (ii), notice we have the inclusion

$$\begin{aligned}
&\phi_{n'e_0}\big(F_*^{n'e_0} \mathfrak{a}^{\lceil t(p^{n'e_0}-1)\rceil} \widetilde{\tau}(R; b, \Delta, \mathfrak{a}^t)\big) \\
&\qquad = \phi_{n'e_0}\Big(F_*^{n'e_0} \mathfrak{a}^{\lceil t(p^{n'e_0}-1)\rceil} \sum_{n \geq 0} \phi_{ne_0}(F_*^{ne_0} b \mathfrak{a}^{\lceil t(p^{ne_0}-1)\rceil})\Big) \\
&\qquad \subseteq \phi_{n'e_0}\Big(F_*^{n'e_0} \sum_{n \geq 0} \phi_{ne_0}(F_*^{ne_0} b \mathfrak{a}^{\lceil t(p^{(n+n')e_0}-1)\rceil})\Big) \\
&\qquad = \sum_{n \geq n'} \phi_{ne_0}(F_*^{ne_0} b \mathfrak{a}^{\lceil t(p^{ne_0}-1)\rceil}) \subseteq \widetilde{\tau}(R; b, \Delta, \mathfrak{a}^t). \qquad \square
\end{aligned}$$

Theorem 6.18. *For* $b \in (R \setminus Q)$ *as in Proposition 6.14, the ideal* $\widetilde{\tau}(R; b, \Delta, \mathfrak{a}^t)$ *is the unique smallest ideal* J *that satisfies*

$$J \cap (R \setminus Q) \neq \varnothing \qquad \textit{and} \qquad \phi_{ne_0}(F_*^{ne_0} \mathfrak{a}^{\lceil t(p^{ne_0}-1)\rceil} J) \subseteq J \quad \textit{for all } n \geq 0.$$

Therefore $\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t) = \widetilde{\tau}(R; b, \Delta, \mathfrak{a}^t)$.

Proof. The previous lemma proves that $\widetilde{\tau}(R; b, \Delta, \mathfrak{a}^t)$ satisfies the two conditions. Suppose that J is any other ideal that also satisfies the two conditions in Theorem 6.18. Choose $d \in J \cap (R \setminus Q)$. By hypothesis,

$$\sum_{n \geq 0} \phi_{ne_0}(F_*^{ne_0} d \mathfrak{a}^{\lceil t(p^{ne_0}-1)\rceil}) \subseteq \sum_{n \geq 0} \phi_{ne_0}(F_*^{ne_0} \mathfrak{a}^{\lceil t(p^{ne_0}-1)\rceil} J) \subseteq J,$$

and so by Proposition 6.14, we see that $b \in J$. But then

$$\begin{aligned}
\widetilde{\tau}(R; b, \Delta, \mathfrak{a}^t) &= \sum_{n \geq 0} \phi_{ne_0}(F_*^{ne_0} b \mathfrak{a}^{\lceil t(p^{ne_0}-1)\rceil}) \\
&\subseteq \sum_{n > 0} \phi_{ne_0}(F_*^{ne_0} \mathfrak{a}^{\lceil t(p^{ne_0}-1)\rceil} J) \subseteq J. \qquad \square
\end{aligned}$$

Remark 6.19. Theorem 6.18 implies that $\widetilde{\tau}(R; b, \Delta, \mathfrak{a}^t)$ is also independent of the choice of b (as long as b is chosen via Proposition 6.14).

Remark 6.20. If b is as in Proposition 6.14, then for any multiplicative set T, it follows that $b/1$ satisfies Proposition 6.14 for the localized triple

$$(T^{-1}R, \Delta|_{\operatorname{Spec} T^{-1}R}, (T^{-1}\mathfrak{a})^t).$$

Therefore the formation of $\tau_b(R, \not\subseteq Q; \Delta, \mathfrak{a}^t) = \widetilde{\tau}(R; b, \Delta, \mathfrak{a}^t)$ commutes with localization. In particular, we can define $\tau_b(X, \not\subseteq W; \Delta, \mathfrak{a}^t)$ on a scheme X with center of F-purity W which locally satisfies the conditions of Definition 6.1.

7. Comments on codimension one centers of F-purity

Suppose that $(X = \operatorname{Spec} R, \Delta + D)$ is a pair, $D \subseteq X$ is an integral normal reduced and irreducible divisor, and Δ and D have no common components. Assume that $K_X + \Delta + D$ is $\mathbb{Q}$-Cartier with index not divisible by $p > 0$. Since X is normal, $(X, \Delta + D)$ is F-pure at the generic point of D and D is also a center of F-purity for the pair $(X, \Delta + D)$. In characteristic zero, there is the notion of the "different"; see [Kollár et al. 1992]. If Q is a defining ideal of D, the different is an effective divisor that plays a role similar to the divisor $\Delta_{R/Q}$ from Theorem 5.2.

We will show that the different and $\Delta_{R/Q}$ agree under the hypothesis that D is Cartier in codimension 2. Roughly speaking, this is the case where the different is uninteresting (it is also the case discussed in [Kollár and Mori 1998]). We will then give two applications of the methods used to prove this result. We expect that the different and $\Delta_{R/Q}$ coincide in general although we do not have a proof, see Remark 7.6.

First we need the following lemma. This lemma is implicit in the work we have done previously, but we provide an explicit proof for completeness. Lemma 7.1 is also closely related to the fact that the set of Frobenius actions on $H^d_{\mathfrak{m}}(R)$ is generated by the natural Frobenius action $F : H^{\dim R}_{\mathfrak{m}}(R) \to H^{\dim R}_{\mathfrak{m}}(R)$ [Lyubeznik and Smith 2001].

Lemma 7.1. *Suppose that R is an F-finite Gorenstein local ring. By dualizing the natural map $G : R \to F^e_* R$ (apply $\operatorname{Hom}_R(\_, \omega_R)$), we construct the map*

$$\Psi : F^e_* \omega_R \to \omega_R.$$

By fixing any isomorphism of ω_R with R (which we can do since R is Gorenstein), we obtain a map which we also call Ψ,

$$\Psi : F^e_* R \to R.$$

This map Ψ is an $F^e_* R$-module generator of $\operatorname{Hom}_R(F^e_* R, R)$. In particular, if R is normal, then Ψ corresponds to the divisor 0 via Theorem 3.11.

Proof. First note that the choices we made in the set-up of the lemma are all unique up to multiplication by a unit (note there is also the choice of isomorphism between $(F^e)^! \omega_R$ with $F^e_* \omega_R$ as in Remark 3.3). Therefore, these choices are irrelevant in terms of proving that Ψ is an $F^e_* R$-module generator. Suppose that ϕ is an arbitrary $F^e_* R$-module generator of $\mathrm{Hom}_R(F^e_* R, R)$; we can write $\Psi(\_) = \phi(d \cdot \_)$ for some $d \in F^e_* R$. Using the same isomorphisms we selected before, we can view ϕ as a map $F^e_* \omega_R \to \omega_R$. By duality for a finite morphism, we obtain $\phi^\vee : R \to F^e_* R$. Note now that $G(\_) = d \cdot \phi^\vee(\_)$. But G sends 1 to 1 which implies that d is a unit and completes the proof. □

We now need the following (useful) surjectivity. A similar argument (involving local duality) was used in the characteristic $p > 0$ inversion of adjunction result of [Hara and Watanabe 2002, Theorem 4.9].

Proposition 7.2. *Using the notation above, further suppose that D is Cartier in codimension 2 and that $(p^e - 1)(K_X + D + \Delta)$ is Cartier. Then the natural map of $F^e_* \mathcal{O}_X$-modules*:

$$\Phi : \mathrm{Hom}_{\mathcal{O}_X}(F^e_* \mathcal{O}_X((p^e-1)(D+\Delta)), \mathcal{O}_X) \to \mathrm{Hom}_{\mathcal{O}_D}(F^e_* \mathcal{O}_D((p^e-1)\Delta|_D), \mathcal{O}_D).$$

induced by restriction is surjective.

Proof. The statement is local so we may assume that $X = \mathrm{Spec}\, R$ where R is the spectrum of a local ring. Furthermore, because we are working locally, the domain of Φ is isomorphic to $F^e_* R$. Thus the image of Φ is cyclic as an $F^e_* \mathcal{O}_D$-module which implies that the image of Φ is a reflexive $F^e_* \mathcal{O}_D$-module. Therefore, it is sufficient to prove that Φ is surjective at the codimension one points of D (which correspond to codimension two points of X). We now assume that $X = \mathrm{Spec}\, R$ is the spectrum of a two-dimensional normal local ring and that D is a Cartier divisor defined by a local equation $(f = 0)$. Since D is normal and one-dimensional, D is Gorenstein, and so X is also Gorenstein. In particular, $(p^e - 1)\Delta$ is Cartier. This also explains how we can restrict $(p^e - 1)\Delta$ to D: perform the restriction at codimension 1 points of D, and then take the corresponding divisor.

Consider the following diagram of short exact sequences:

$$\begin{tikzcd}
0 \arrow[r] & R \arrow[r, "\times f"] \arrow[d, "1 \mapsto f^{p^e-1}"'] & R \arrow[r] \arrow[d, "1 \mapsto 1"] & R/f \arrow[r] \arrow[d, "1 \mapsto 1"] & 0 \\
0 \arrow[r] & F^e_* R \arrow[r, "F^e_* \times f"] & F^e_* R \arrow[r] & F^e_*(R/f) \arrow[r] & 0
\end{tikzcd}$$

Apply the functor $\mathrm{Hom}_R(\_, \omega_R)$ and note that we obtain the following diagram of short exact sequences.

$$\begin{tikzcd}
0 \arrow[r] & \omega_R \arrow[r, "\times f"] & \omega_R \arrow[r] & \omega_{R/f} \cong \mathrm{Ext}^1_R(R/f, \omega_R) \arrow[r] & 0 \\
0 \arrow[r] & F^e_* \omega_R \arrow[r, "F^e_* \times f"] \arrow[u, "\alpha"] & F^e_* \omega_R \arrow[r] \arrow[u, "\beta"] & F^e_* \omega_{R/f} \cong \mathrm{Ext}^1_R(F^e_*(R/f), \omega_R) \arrow[r] \arrow[u, "\delta"] & 0
\end{tikzcd}$$

The sequences are exact on the right because R is Gorenstein and hence Cohen–Macaulay. By Lemma 7.1, we see that δ and α can be viewed as $F^e_* R$-module generators of the modules

$$\mathrm{Hom}_{R/f}(F^e_*(R/f), R/f) \cong \mathrm{Hom}_{R/f}(F^e_* \omega_{R/f}, \omega_{R/f}),$$
$$\mathrm{Hom}_R(F^e_* R, R) \cong \mathrm{Hom}_R(F^e_* \omega_R, \omega_R),$$

respectively. Furthermore, the map labeled β can be identified with

$$\alpha \circ (F^e_*(\times f^{p^e-1})).$$

But the diagram proves exactly that the map $\beta \in \mathrm{Hom}_R(F^e_* R, R)$ restricts to a generator of $\mathrm{Hom}_{R/f}(F^e_* \omega_{R/f}, \omega_{R/f})$ which is exactly what we wanted to prove in the case that $\Delta = 0$. When $\Delta \neq 0$, we can simply premultiply the α, β and δ with a local generator of the Cartier divisor $(p^e - 1)\Delta$. $\square$

Remark 7.3. Suppose that X is normal, $\Delta = 0$ and D is Gorenstein in codimension 1 and S2 (but D is not necessarily normal or irreducible), then the map Φ from Proposition 7.2 is still surjective. The proof is unchanged.

The previous example also gives us the following corollary. Compare with [Kollár and Shepherd-Barron 1988, Theorem 5.1; Karu 2000, Theorem 2.5; Fedder and Watanabe 1989, Proposition 2.13; Schwede 2007, Theorem 5.1].

Corollary 7.4. *Suppose that R is normal, local and $\mathbb{Q}$-Gorenstein with index not divisible by p and that $f \in R$ is a nonzero divisor such that the map Φ from Proposition 7.2 (where $D = \mathrm{div}(f)$ and $\Delta = 0$) is surjective.*[7]

If $R[f^{-1}]$ is strongly F-regular and R/f is F-pure then R is strongly F-regular. In particular, both R and R/f are Cohen–Macaulay.

Proof. Since the map

$$\Phi : \mathrm{Hom}_R(F^e_* R((p^e - 1)\,\mathrm{div}(f)), R) \to \mathrm{Hom}_{R/f}(F^e_*(R/f), R/f).$$

is surjective, a splitting $\overline{\phi} \in \mathrm{Hom}_{R/f}(F^e_*(R/f), R/f)$ has a preimage

$$\phi \in \mathrm{Hom}_R(F^e_* R, R).$$

[7]Note that Φ is surjective if R/f is normal, or more generally if R/f is S2 and Gorenstein in codimension 1.

It then follows (just as in Observation 5.1) that the map ϕ is also surjective. In particular, ϕ sends some multiple of f^{p^e-1} to 1. But then since $R[f^{-1}]$ is strongly F-regular, we see that R itself is strongly F-regular. $\square$

Corollary 7.5. *Suppose that S is an F-finite regular local ring and I is a prime ideal of S such that $R = S/I$ is normal and satisfies*

$$(I^{[p^e]} : I) = I^{[p^e]} + (g)$$

for some $g \in S$ (note that this implies that $(p^e - 1)K_R$ is Cartier). Further suppose that $f \in S$ is an element whose image in R is nonzero and such that $R/(fR)$ is normal (or $S2$ and Gorenstein in codimension 1). Then

$$\Big((I + (f))^{[p^e]} : (I + (f))\Big) = (I + (f))^{[p^e]} + (f^{p^e-1}g).$$

Proof. If $A = S/(I+f)$, it follows from Proposition 7.2 that $\mathrm{Hom}_A(F^e_* A, A)$ is free of rank 1 as an $F^e_* A$-module and furthermore that a generator of $\mathrm{Hom}_A(F^e_* A, A)$ is obtained by multiplying a generator of $\mathrm{Hom}_R(F^e_* R, R)$ by f^{p^e-1} and restricting. The result then follows from [Fedder 1983, Lemma 1.6]. $\square$

Remark 7.6. Suppose that D is a normal prime divisor on X a normal scheme. Further suppose that Δ is an effective $\mathbb{Q}$-divisor (without common components with D) such that $K_X + \Delta + D$ is $\mathbb{Q}$-Cartier. Then there exists a canonically determined effective $\mathbb{Q}$-divisor Δ_D on D with $(K_X + \Delta + D)|_D \sim_{\mathbb{Q}} K_D + \Delta_D$; see [Kollár et al. 1992, Chapter 16] for a description of the construction of the different (which can be performed in any characteristic). Furthermore, in characteristic zero, the singularities of $(X, D + \Delta)$ near D are closely related to the singularities of (D, Δ_D) [Kollár et al. 1992; Kawakita 2007]. We expect that the different coincides with the divisor $\Delta_{R/Q}$ we have constructed, but we do not have a proof (the problem might be quite easy if approached correctly). One should note that we believe that the divisor called the "different" in [Takagi 2008, Theorem 4.3] is $\Delta_{R/Q}$. Again, we suspect that $\Delta_{R/Q}$ coincides with the different in general.

8. Comments on normalizing centers of F-purity

In the characteristic zero setting, one obstruction to working with an arbitrary log canonical centers $W \subseteq X$ is the fact that W may not be normal. One way around this issue is to normalize the subscheme W (even if W is a divisor). Therefore, it is tempting to do the same in positive characteristic. Using Lemma 8.1, one can do something like this in characteristic zero. In particular, in Proposition 8.2 we do obtain canonically determined $\mathbb{Q}$-divisors on the normalization of any center of F-purity. However, a full analog of inversion of adjunction on log canonicity via normalizing centers of F-purity is impossible, as we will see in Example 8.4.

Lemma 8.1. *Suppose that R is a reduced F-finite ring and that $\phi \in \operatorname{Hom}_R(F^e_* R, R)$. Set R^N to be the normalization of R inside the total field of fractions. Then ϕ extends to a unique R^N-linear map $\phi^N : F^e_* R^N \to R^N$ that restricts back to ϕ.*

Proof. To construct ϕ^N, simply tensor ϕ with the total field of fractions $k(R)$ of R and then restrict the domain to $F^e_* R^N$. The fact that the image of ϕ^N is contained inside R^N follows from [Brion and Kumar 2005, Hint to Exercise 1.2.E(4)]; for a complete proof see [Schwede 2008a, Proposition 7.11]. The fact that this ϕ^N is unique follows from the fact that the natural map

$$\operatorname{Hom}_R(F^e_* R, R) \to \operatorname{Hom}_R(F^e_* R, R) \otimes_R k(R) \cong \operatorname{Hom}_{k(R)}(F^e_* k(R), k(R))$$

is injective. □

Proposition 8.2. *Suppose that $X = \operatorname{Spec} R$ and (X, Δ) is a pair and that*

$$\mathscr{H}om_{\mathcal{O}_X}(F^e_* \mathcal{O}_X((p^e - 1)\Delta), \mathcal{O}_X)$$

is free of rank 1 as an $F^e_* \mathcal{O}_X$-module. Further suppose that $\operatorname{Spec} R/I = W \subset X$ is a reduced closed subscheme such that (X, Δ) is sharply F-pure at the generic points of W and I is F-compatible with respect to (R, Δ). Set

$$\eta : (\operatorname{Spec} R/I)^N = W^N \to W$$

to be the normalization map and write $W^N = \coprod_{i=1}^m W_i^N$; the disjoint union of W^N into its irreducible components.

Then there exists a canonically determined $\mathbb{Q}$-divisor Δ_{W^N} on W^N satisfying the following properties:

(i) *If $\Delta_{W^N,i}$ is set to the portion of Δ_{W^N} on W_i^N, then $(p^e - 1)(K_{W_i^N} + \Delta_{W^N,i})$ is Cartier, and $\mathscr{H}om_{\mathcal{O}_{W_i^N}}(F^e_* \mathcal{O}_{W_i^N}((p^e - 1)\Delta_{W^N,i}), \mathcal{O}_{W_i^N}) \cong F^e_* \mathcal{O}_{W_i^N}$ as $F^e_* \mathcal{O}_{W_i^N}$-modules.*

(ii) *The conductor ideal of (R/I) in $(R/I)^N$ is F-compatible with respect to $((R/I)^N, \Delta_{W^N})$.*

(iii) *The big test ideal $\tau_b((R/I)^N; \Delta_{W^N})$ of $((R/I)^N, \Delta_{W^N})$ is contained in the conductor ideal of $R/I \subseteq (R/I)^N$.*

(iv) *If (X, Δ) is sharply F-pure, then (W^N, Δ_{W^N}) is also sharply F-pure.*

(v) *If $\overline{J}$ is an ideal of $(R/I)^N$ which is F-compatible with respect to (R, Δ_{W^N}), then the inverse image J of $\overline{J}$ in R is F-compatible with respect to (R, Δ). (In particular, $\tau_b(R, \nsubseteq I; \Delta)$, defined as suggested in Remark 6.6, is contained in the inverse image of $\tau_b((R/I)^N, \Delta_{W^N})$).*

Remark 8.3. Even though W^N is not necessarily equidimensional, it is easy to define K_{W^N} since we can work on each component individually.

Proof. We can associate to Δ a map $\phi : F^e_* \operatorname{Hom}_R(F^e_* R, R)$ (up to scaling by a unit). By assumption, this ϕ restricts to a map $\phi_I \in \operatorname{Hom}_{R/I}(F^e_*(R/I), R/I)$ which is nonzero at the generic point of each irreducible component of R/I. By Lemma 8.1, this map extends to a map $\phi_I^N \in \operatorname{Hom}_{\mathcal{O}_{W^N}}(F^e_* \mathcal{O}_{W^N}, \mathcal{O}_{W^N})$. Thus this map gives us our Δ_{W^N} by Theorem 3.11. Notice that the image of a unit under $R \to (R/I)^N$ is still a unit, so that Δ_{W^N} is uniquely determined.

At this point, statement (i) is obvious. Statement (ii) follows from [Schwede 2008a, Proposition 7.10] and statement (iii) follows from the fact that the big test ideal is the smallest ideal F-compatible ideal with respect to $((R/I)^N, \Delta_{W^N})$. For statement (iv), note that if ϕ is surjective, then so is ϕ_I. But then it is easy to see that ϕ_I^N is also surjective.

To prove (v), we first note that $\phi_I(F^e_*(\overline{J} \cap R/I)) \subseteq \overline{J} \cap R/I$. But then we see that the preimage of $\overline{J} \cap R/I$ in R is F-compatible with respect to (R, Δ). $\square$

One might hope that the converse to property (iv) of Proposition 8.2 above holds, but unfortunately, this is not the case. Of course, it is easy to see that if ϕ_I^N is actually a splitting (that is, if it sends 1 to 1), then so is ϕ_I and thus ϕ is surjective near I (which would imply that (R, Δ) is sharply F-pure near I). However, it can happen that ϕ_I^N is surjective (that is, it sends some x to 1) but ϕ_I is not (in particular, the element x is in $(R/I)^N$ but *not* in R/I). The following example illustrates this phenomenon.

Example 8.4. Suppose that $R = k[a, b, c]$ where $k = \mathbb{F}_2$, the field with two elements (any perfect field of characteristic two will work). Set $I = (ac^2 + b^2)$. Set $\Delta = \operatorname{div}(I)$. It is easy to see that I is F-compatible with respect to (R, Δ). Notice that we can write

$$R/I = k[a, b, c]/(ac^2 + b^2) \cong k[x^2, xy, y].$$

Therefore, the normalization of R/I is simply $k[x, y]$. We will exhibit a map $\phi_I : F_*(R/I) \to R/I$, restricted from a map $\phi : F_* R \to R$, that is not surjective, but that the extension ϕ_I^N to the normalization is surjective. Of course, R/I is not weakly normal and so it is not F-pure, which implies that no such ϕ_* can be surjective.

To construct ϕ, we simply take the following map which is associated to Δ. Explicitly, we take the map $\psi : F_* R \to R$ that sends abc to 1 (and all other lower-degree monomials to zero) and precompose with multiplication by $ac^2 + b^2$. That is,

$$\phi(\_) = \psi((ac^2 + b^2) \cdot \_).$$

We compute ϕ on the relevant monomials.

$$\begin{aligned} &\phi(1) = 0, && \phi(a) = 0, && \phi(b) = 0, && \phi(c) = 0, \\ &\phi(ab) = 0, && \phi(bc) = c, && \phi(ac) = 0, && \phi(abc) = b, \end{aligned}$$

Thus we see that ϕ (and therefore also ϕ_I) is not surjective when localized at the origin. Now we wish to consider the corresponding map on $k[x, y]$. First we retranslate ϕ in terms of the variables x and y.

$$\begin{aligned} &\phi_I^N(1) = 0, && \phi_I^N(y) = 0, && \phi_I^N(xy^2) = y, \\ &\phi_I^N(x^2) = 0, && \phi_I^N(x^3y) = 0, && \phi_I^N(x^3y^2) = xy, \\ &\phi_I^N(xy) = 0, && \phi_I^N(x^2) = 0, && \end{aligned}$$

Therefore, $y = \phi_I^N(xy^2) = y\phi_I^N(x)$, which implies that $\phi_I^N(x) = 1$.

Remark 8.5. Of course, in the above example, there were certain purely inseparable field extensions in the normalization. In particular, R/I was not weakly normal. It may be that without such pure-inseparability, when ϕ_I^N is surjective so is ϕ.

9. Further remarks and questions

We conclude with some remarks and speculation.

Remark 9.1. It is natural to try to generalize the results of this paper outside of the case when R is normal. One approach to this is to normalize R as we discussed in the previous section. However, as we saw, this approach has limitations. Another more direct approach might be, instead of working with pairs (R, Δ) such that $(p^e - 1)(K_R + \Delta)$ is Cartier, to consider pairs (R, N) where N is a free (or perhaps locally free) subsheaf of $\mathrm{Hom}_R(F_*^e R, R)$ for some e.

Perhaps yet a better formulation would be to consider first the graded noncommutative algebra $\oplus_e \mathrm{Hom}_R(F_*^e R, R)$ where the multiplication is defined by composition. That is, for $\phi \in \mathrm{Hom}_R(F_*^d R, R)$ and $\psi \in \mathrm{Hom}_R(F_*^e R, R)$ the product $\phi \cdot \psi$ is defined to be $\phi \circ F_*^d \psi \in \mathrm{Hom}_R(F_*^{e+d} R, R)$. Dually, one could consider the noncommutative ring $\mathscr{F}(E_R)$ of [Lyubeznik and Smith 2001]. Then perhaps a pair could be the combined data of the ring R and a graded subalgebra $A \subseteq \oplus_e \mathrm{Hom}_R(F_*^e R, R)$ such that A is generated as an algebra over $A_0 \cong R$ by a single element $\phi \in \mathrm{Hom}_R(F_*^e R, R)$ for some e. Two pairs (R, A) and (R, A') would be said to be equivalent, if there is an integer $n > 0$ such that $A_{ne} = A'_{ne}$ for all e (here A_{ne} is the ne'th graded piece of A).

Almost all of the results of this paper can be generalized to such a setting.

Remark 9.2. This theory can also be used to help identify subschemes of a quasi-projective variety X that are compatibly split with a given Frobenius splitting. In particular, suppose that $\phi : F_*^e \mathcal{O}_X \to \mathcal{O}_X$ is a Frobenius splitting. We can then associate a divisor Δ_ϕ to ϕ. Any center of log canonicity of the pair (X, Δ) is a center of sharp F-purity [Schwede 2008a] and thus the associated scheme is compatibly split with ϕ.

Question 9.3. Suppose that R is a normal $\mathbb{Q}$-Gorenstein ring of finite type over a field of characteristic zero and that $Q \in \operatorname{Spec} R$ is a center of log canonicity. Further suppose that R_Q is log canonical and that, when reduced to characteristic $p \gg 0$ (or perhaps to infinitely many $p > 0$), $(R_p)_{Q_p}$ is F-pure. Then for each $p \gg 0$, we can associate a (canonically defined) Δ_{Q_p} on R_p/Q_p. We then ask whether or not Δ_{Q_p} is reduced from some $\mathbb{Q}$-divisor Δ on R?

Question 9.4. Is there a characteristic zero analog of $\tau_b(R, \not\subseteq Q; \Delta)$? Takagi has considered similar questions [Takagi 2007, Conjecture 2.8]. One possible analog is something along the following lines: for a log resolution $\pi : \widetilde{X} \to X = \operatorname{Spec} R$ of (R, Δ), let $E = \sum E_i$ be the sum of divisors E_i of $\widetilde{X}$ (exceptional or not) such that $Q \in \pi(E_i)$ and such that the discrepancy of (R, Δ) along E_i is ≤ -1. Then consider the ideal

$$\pi_* \mathcal{O}_{\widetilde{X}}\Big(\big\lceil K_{\widetilde{X}} - \pi^*(K_X + \Delta) + \epsilon \sum E_i \big\rceil\Big) \quad \text{for } \epsilon > 0 \text{ sufficiently small.}$$

Is it possible that this coincides with $\tau_b(R, \not\subseteq Q; \Delta)$ for infinitely many $p > 0$? Also compare with [Fujino 2008].

Finally, we consider the nonlocal setting.

Remark 9.5. Suppose that (X, Δ) is a pair where X is a (possibly proper) variety of finite type over an F-finite field k. In particular, we know that $(F^e)^! \omega_X$ can itself be identified with ω_X; see Remark 3.3. Further suppose that $K_X + \Delta$ is $\mathbb{Q}$-Cartier with index not divisible by $p > 0$. Now suppose that $W \subset X$ is a normal closed variety defined by an ideal sheaf I_W which is locally F-compatible with respect to Δ. Then on a sufficiently fine affine cover U_i of X, we can associate $\mathbb{Q}$-divisors Δ_{W_i} on $W_i = U_i \cap W$. It is easy to see that these divisors agree on overlaps since they were canonically determined. Therefore, there is a $\mathbb{Q}$-divisor Δ_W on W determined by (X, Δ).

Furthermore, we claim that

$$(p^e - 1)(K_X + \Delta)|_W \sim (p^e - 1)(K_W + \Delta_W). \tag{9.5.1}$$

One way to see this is to work globally (in particular, partially globalize Theorems 3.11 and 3.13). More precisely, there is a bijection of sets

$$\left\{ \begin{matrix} \text{Effective } \mathbb{Q}\text{-divisors } \Delta \text{ on } X \text{ such} \\ \text{that } (p^e-1)(K_X+\Delta) \text{ is Cartier} \end{matrix} \right\} \longleftrightarrow \left\{ \begin{matrix} \text{Line bundles } \mathcal{L} \text{ and nonzero} \\ \text{elements of } \operatorname{Hom}_{\mathcal{O}_X}(F^e_* \mathcal{L}, \mathcal{O}_X) \end{matrix} \right\} \Big/ \sim$$

The equivalence relation on the right side identifies two maps $\phi_1 : F^e_* \mathcal{L}_1 \to \mathcal{O}_X$ and $\phi_2 : F^e_* \mathcal{L}_2 \to \mathcal{O}_X$ if there is an isomorphism $\gamma : \mathcal{L}_1 \to \mathcal{L}_2$ and a commutative diagram

$$\begin{CD} F^e_* \mathscr{L}_1 @>{\phi_1}>> \mathcal{O}_X \\ @V{F^e_* \gamma}VV @VV{\mathrm{id}}V \\ F^e_* \mathscr{L}_2 @>{\phi_2}>> \mathcal{O}_X \end{CD}$$

We sketch the correspondence for the convenience of the reader. Given Δ, set $\mathscr{L} = \mathcal{O}_X((1-p^e)(K_X+\Delta))$. Then observe that

$$\mathscr{H}om_{\mathcal{O}_X}(F^e_* \mathscr{L}, \mathcal{O}_X) \cong F^e_* \mathscr{H}om_{\mathcal{O}_X}(\mathscr{L}, \mathcal{O}_X((1-p^e)K_X)) \cong F^e_* \mathcal{O}_X((p^e-1)\Delta).$$

We can choose a global section of $\mathcal{O}_X((p^e-1)\Delta)$ corresponding to the effective integral divisor $(p^e-1)\Delta$ (up to multiplication by a unit). This section may be viewed as a map $\phi_\Delta : F^e_* \mathscr{L} \to \mathcal{O}_X$ by the above isomorphism. For the converse direction, given such a ϕ we obtain a global section of $F^e_* \mathscr{L}^{-1}((1-p^e)K_X)$. This corresponds to an effective divisor D. Set $\Delta_\phi = (1/(p^e-1))D$. Again, as mentioned before, this is simply the globalized version of Theorems 3.11 and 3.13.

Now, since I_W is locally F-compatible with respect to Δ, we have that

$$\phi_\Delta(F^e_* I_W \mathscr{L}) \subseteq I_W.$$

By restriction, we obtain a map $\phi_W : \mathscr{L}|_W \to \mathcal{O}_W$. It is then clear that

$$\mathcal{O}_X((p^e-1)(K_X+\Delta))|_W$$

is isomorphic to $\mathcal{O}_W((p^e-1)(K_W+\Delta_W))$ as desired.

Acknowledgments

The author would like to thank Florin Ambro, Manuel Blickle, Mel Hochster, Karen Smith, Shunsuke Takagi, and Wenliang Zhang for several valuable discussions. The author would also like to thank Shunsuke Takagi and the referee for carefully reading an earlier draft and providing several useful comments.

References

[Ambro 1999] F. Ambro, "The adjunction conjecture and its applications", preprint, 1999. arXiv 9903060

[Blickle et al. 2008] M. Blickle, M. Mustaţă, and K. E. Smith, "Discreteness and rationality of F-thresholds", *Michigan Math. J.* **57** (2008), 43–61. MR 2492440 Zbl 05604519

[Bourbaki 1998] *Commutative algebra, Chapters 1–7, Elements of Mathematics*, Springer-Verlag, Berlin, 1998. Translated from the French, reprint of the 1989 English translation. MR1727221 (2001g:13001).

[Brion and Kumar 2005] M. Brion and S. Kumar, *Frobenius splitting methods in geometry and representation theory*, Progress in Mathematics **231**, Birkhäuser, Boston, MA, 2005. MR 2005k:14104 Zbl 1072.14066

[Ein and Mustaţă 2004] L. Ein and M. Mustaţă, "Inversion of adjunction for local complete intersection varieties", *Amer. J. Math.* **126**:6 (2004), 1355–1365. MR 2005j:14020 Zbl 1087.14012

[Ein and Mustaţă 2009] L. Ein and M. Mustaţă, "Jet schemes and singularities", pp. 505–546 in *Algebraic geometry* (Seattle, 2005), vol. 2, edited by D. Abramovich et al., Proc. Sympos. Pure Math. **80**, Amer. Math. Soc., Providence, RI, 2009. MR 2483946 Zbl 05548159

[Ein et al. 2003] L. Ein, M. Mustaţă, and T. Yasuda, "Jet schemes, log discrepancies and inversion of adjunction", *Invent. Math.* **153**:3 (2003), 519–535. MR 2004f:14028 Zbl 1049.14008

[Enescu and Hochster 2008] F. Enescu and M. Hochster, "The Frobenius structure of local cohomology", *Algebra Number Theory* **2**:7 (2008), 721–754. MR 2009i:13009 Zbl 05529343

[Fedder 1983] R. Fedder, "F-purity and rational singularity", *Trans. Amer. Math. Soc.* **278**:2 (1983), 461–480. MR 84h:13031 Zbl 0519.13017

[Fedder and Watanabe 1989] R. Fedder and K. Watanabe, "A characterization of F-regularity in terms of F-purity", pp. 227–245 in *Commutative algebra* (Berkeley, CA, 1987), edited by M. Hochster et al., Math. Sci. Res. Inst. Publ. **15**, Springer, New York, 1989. MR 91k:13009 Zbl 0738.13004

[Fujino 2008] O. Fujino, "Theory of non-lc ideal sheaves–basic properties", preprint, 2008. arXiv 0801.2198

[Hara 2001] N. Hara, "Geometric interpretation of tight closure and test ideals", *Trans. Amer. Math. Soc.* **353**:5 (2001), 1885–1906. MR 2001m:13009 Zbl 0976.13003

[Hara 2005] N. Hara, "A characteristic p analog of multiplier ideals and applications", *Comm. Algebra* **33**:10 (2005), 3375–3388. MR 2006f:13006 Zbl 1090.13003

[Hara and Takagi 2004] N. Hara and S. Takagi, "On a generalization of test ideals", *Nagoya Math. J.* **175** (2004), 59–74. MR 2005g:13009 Zbl 1094.13004

[Hara and Watanabe 2002] N. Hara and K.-I. Watanabe, "F-regular and F-pure rings vs. log terminal and log canonical singularities", *J. Algebraic Geom.* **11**:2 (2002), 363–392. MR 2002k:13009 Zbl 1013.13004

[Hara and Yoshida 2003] N. Hara and K.-I. Yoshida, "A generalization of tight closure and multiplier ideals", *Trans. Amer. Math. Soc.* **355**:8 (2003), 3143–3174. MR 2004i:13003 Zbl 1028.13003

[Hartshorne 1966] R. Hartshorne, *Residues and duality*, Lecture Notes in Math. **20**, Springer, Berlin, 1966. MR 36 #5145 Zbl 0212.26101

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Math. **52**, Springer, New York, 1977. MR 57 #3116 Zbl 0367.14001

[Hartshorne 1994] R. Hartshorne, "Generalized divisors on Gorenstein schemes", *K-Theory* **8**:3 (1994), 287–339. MR 95k:14008 Zbl 0826.14005

[Hartshorne 2007] R. Hartshorne, "Generalized divisors and biliaison", *Illinois J. Math.* **51**:1 (2007), 83–98. MR 2008j:14010 Zbl 1133.14005

[Hochster 2007] M. Hochster, "Foundations of tight closure theory", lecture notes from a course taught on the university of michigan, 2007, See http://www.math.lsa.umich.edu/~hochster/711F07/711.html.

[Hochster and Huneke 1989] M. Hochster and C. Huneke, *Tight closure and strong F-regularity*, Mém. Soc. Math. France (N.S.) **38**, 1989. MR 91i:13025 Zbl 0699.13003

[Hochster and Huneke 1990] M. Hochster and C. Huneke, "Tight closure, invariant theory, and the Briançon–Skoda theorem", *J. Amer. Math. Soc.* **3**:1 (1990), 31–116. MR 91g:13010

[Hochster and Roberts 1976] M. Hochster and J. L. Roberts, "The purity of the Frobenius and local cohomology", *Advances in Math.* **21**:2 (1976), 117–172. MR 54 #5230 Zbl 0348.13007

[Karu 2000] K. Karu, "Minimal models and boundedness of stable varieties", *J. Algebraic Geom.* **9**:1 (2000), 93–109. MR 2001g:14059 Zbl 0980.14008

[Kawakita 2007] M. Kawakita, "Inversion of adjunction on log canonicity", *Invent. Math.* **167**:1 (2007), 129–133. MR 2008a:14025 Zbl 1114.14009

[Kawakita 2008] M. Kawakita, "On a comparison of minimal log discrepancies in terms of motivic integration", *J. Reine Angew. Math.* **620** (2008), 55–65. MR 2427975 Zbl 1151.14014

[Kawamata 1997a] Y. Kawamata, "On Fujita's freeness conjecture for 3-folds and 4-folds", *Math. Ann.* **308**:3 (1997), 491–505. MR 99c:14008 Zbl 0909.14001

[Kawamata 1997b] Y. Kawamata, "Subadjunction of log canonical divisors for a subvariety of codimension 2", pp. 79–88 in *Birational algebraic geometry* (Baltimore, MD, 1996), edited by Y. Kawamata and V. Shokurov, Contemp. Math. **207**, Amer. Math. Soc., Providence, RI, 1997. MR 99a:14024 Zbl 0901.14004

[Kawamata 1998] Y. Kawamata, "Subadjunction of log canonical divisors, II", *Amer. J. Math.* **120**:5 (1998), 893–899. MR 2000d:14020 Zbl 0919.14003

[Kawamata et al. 1987] Y. Kawamata, K. Matsuda, and K. Matsuki, "Introduction to the minimal model problem", pp. 283–360 in *Algebraic geometry* (Sendai, 1985), edited by T. Oda, Adv. Stud. Pure Math. **10**, North-Holland, Amsterdam, 1987. MR 89e:14015 Zbl 0672.14006

[Kollár and Mori 1998] J. Kollár and S. Mori, *Birational geometry of algebraic varieties*, Cambridge Tracts in Mathematics **134**, Cambridge University Press, Cambridge, 1998. MR 2000b:14018 Zbl 0926.14003

[Kollár and Shepherd-Barron 1988] J. Kollár and N. I. Shepherd-Barron, "Threefolds and deformations of surface singularities", *Invent. Math.* **91**:2 (1988), 299–338. MR 88m:14022 Zbl 0642.14008

[Kollár et al. 1992] J. Kollár et al., *Flips and abundance for algebraic threefolds: Papers from the second summer seminar on algebraic geometry held at the University of Utah* (Salt Lake City, Utah, 1991), Astérisque **211**, Société Mathématique de France, Paris, 1992. MR 94f:14013

[Kunz 1986] E. Kunz, *Kähler differentials*, Vieweg, Braunschweig, 1986. MR 88e:14025

[Lyubeznik and Smith 2001] G. Lyubeznik and K. E. Smith, "On the commutation of the test ideal with localization and completion", *Trans. Amer. Math. Soc.* **353**:8 (2001), 3149–3180. MR 2002f:13010 Zbl 0977.13002

[Mehta and Ramanathan 1985] V. B. Mehta and A. Ramanathan, "Frobenius splitting and cohomology vanishing for Schubert varieties", *Ann. of Math.* (2) **122**:1 (1985), 27–40. MR 86k:14038 Zbl 0601.14043

[Mehta and Srinivas 1991] V. B. Mehta and V. Srinivas, "Normal F-pure surface singularities", *J. Algebra* **143**:1 (1991), 130–143. MR 92j:14044 Zbl 0760.14012

[Metha and Kumar 2009] V. B. Metha and S. Kumar, "Finiteness of the number of compatibly-split subvarieties", preprint, 2009. to appear in *Int. Math. Res. Not.* arXiv 0901.2098

[Schwede 2007] K. Schwede, "A simple characterization of Du Bois singularities", *Compos. Math.* **143**:4 (2007), 813–828. MR 2008k:14034 Zbl 1125.14002

[Schwede 2008a] K. Schwede, "Centers of F-purity", preprint, 2008. to appear in *Mathematische Zeitschrift*. arXiv 0807.1654

[Schwede 2008b] K. Schwede, "Generalized test ideals, sharp F-purity, and sharp test elements", *Math. Res. Lett.* **15**:6 (2008), 1251–1261. MR 2470398 Zbl 05505952

[Sharp 2007] R. Y. Sharp, "Graded annihilators of modules over the Frobenius skew polynomial ring, and tight closure", *Trans. Amer. Math. Soc.* **359**:9 (2007), 4237–4258. Zbl 1130.13002

[Smith 2000] K. E. Smith, "The multiplier ideal is a universal test ideal", *Comm. Algebra* **28**:12 (2000), 5915–5929. MR 2002d:13008 Zbl 0979.13007

[Takagi 2004a] S. Takagi, "F-singularities of pairs and inversion of adjunction of arbitrary codimension", *Invent. Math.* **157**:1 (2004), 123–146. MR 2006g:14028 Zbl 1121.13008

[Takagi 2004b] S. Takagi, "An interpretation of multiplier ideals via tight closure", *J. Algebraic Geom.* **13**:2 (2004), 393–415. MR 2005c:13002 Zbl 1080.14004

[Takagi 2007] S. Takagi, "Adjoint ideals along closed subvarieties of higher codimension", preprint, 2007. To appear in *J. Reine Angew. Math.* arXiv 0711.2342

[Takagi 2008] S. Takagi, "A characteristic p analogue of plt singularities and adjoint ideals", *Math. Z.* **259**:2 (2008), 321–341. MR 2009b:13004 Zbl 1143.13007

[Takagi and Watanabe 2004] S. Takagi and K.-i. Watanabe, "On F-pure thresholds", *J. Algebra* **282**:1 (2004), 278–297. MR 2006a:13010 Zbl 1082.13004

Communicated by Craig Huneke
Received 2009-02-03 Revised 2009-09-14 Accepted 2009-10-13

kschwede@umich.edu *Department of Mathematics, University of Michigan, East Hall, 530 Church Street, Ann Arbor, MI 48109, United States*
http://www.umich.edu/~kschwede

ALGEBRA AND NUMBER THEORY 3:8(2009)

Log minimal models according to Shokurov

Caucher Birkar

Following Shokurov's ideas, we give a short proof of the following klt version of his result: termination of terminal log flips in dimension d implies that any klt pair of dimension d has a log minimal model or a Mori fibre space. Thus, in particular, any klt pair of dimension 4 has a log minimal model or a Mori fibre space.

1. Introduction

All the varieties in this paper are assumed to be over an algebraically closed field k of characteristic zero. We refer the reader to Section 2 for notation and terminology.

The following conjecture is perhaps the most important problem in birational geometry.

Conjecture 1.1 (Minimal model). Let $(X/Z, B)$ be a Kawamata log terminal (klt) pair. Then it has a log minimal model or a Mori fibre space.

The 2-dimensional case of this conjecture is considered to be classical. The 3-dimensional case was settled in the 80's and 90's by the efforts of many mathematicians, in particular Mori, Shokurov and Kawamata. The higher-dimensional case has seen considerable progress in recent years, thanks primarily to Shokurov's work on the existence of log flips, which paved the way for further progress. The conjecture is also settled for pairs of general type [Birkar et al. 2006], and inductive arguments have been proposed for pairs of nonnegative Kodaira dimension [Birkar 2007]. For a more detailed account of the known cases of this conjecture, see the introduction to [Birkar 2007].

Shokurov [2006] proved that the log minimal model program (LMMP) in dimension $d-1$ and termination of terminal log flips in dimension d imply Conjecture 1.1 in dimension d even for log canonical (lc) pairs. (In this paper, by termination of *terminal log flips in dimension* d, we will mean termination of any sequence $X_i \dashrightarrow X_{i+1}/Z_i$ of log flips/Z starting with a d-dimensional klt pair $(X/Z, B)$ which is terminal in codimension ≥ 2; see Section 2 for a more precise formulation.)

MSC2000: 14E30.

Keywords: minimal models, Mori fibre spaces.

Following Shokurov's method and using results of [Birkar et al. 2006], we give a short proof of:

Theorem 1.2. *Termination of terminal log flips in dimension d implies Conjecture 1.1 in dimension d; more precisely, for a klt pair $(X/Z, B)$ of dimension d one constructs a log minimal model or a Mori fibre space by running the LMMP/Z on $K_X + B$ with scaling of a suitable big/Z $\mathbb{R}$-divisor and proving that it terminates.*

As in [Shokurov 2006], one immediately derives the following:

Corollary 1.3. *Conjecture 1.1 holds in dimension* 4.

Note that when $(X/Z, B)$ is effective (for example if it is of nonnegative Kodaira dimension), log minimal models are constructed in [Birkar 2007], using different methods, in dimension ≤ 5.

2. Basics

Let k be an algebraically closed field of characteristic zero. For an $\mathbb{R}$-divisor D on a variety X over k, we use $D^\sim$ to denote the birational transform of D on a specified birational model of X.

Definition 2.1. A *pair* $(X/Z, B)$ consists of normal quasiprojective varieties X, Z over k, an $\mathbb{R}$-divisor B on X with coefficients in $[0, 1]$ such that $K_X + B$ is $\mathbb{R}$-Cartier, and a projective morphism $X \to Z$. $(X/Z, B)$ is called *log smooth* if X is smooth and Supp B has simple normal crossing singularities.

For a prime divisor D on some birational model of X with a nonempty centre on X, $a(D, X, B)$ denotes the log discrepancy. $(X/Z, B)$ is *terminal in codimension* ≥ 2 if $a(D, X, B) > 1$ whenever D is exceptional/X. Log flips preserve this condition but divisorial contractions may not.

Let $(X/Z, B)$ be a klt pair. By a *log flip*/Z we mean the flip of a $K_X + B$-negative extremal flipping contraction/Z. A *sequence of log flips/Z starting with* $(X/Z, B)$ is a sequence $X_i \dashrightarrow X_{i+1}/Z_i$ in which $X_i \to Z_i \leftarrow X_{i+1}$ is a $K_{X_i} + B_i$-flip/Z and B_i is the birational transform of B_1 on X_1, and $(X_1/Z, B_1) = (X/Z, B)$. By termination of terminal log flips in dimension d we mean termination of such a sequence in which $(X_1/Z, B_1)$ is a d-dimensional klt pair which is terminal in codimension ≥ 2. Now assume that G is an $\mathbb{R}$-Cartier divisor on X. A *sequence of G-flops/Z with respect to* $(X/Z, B)$ is a sequence $X_i \dashrightarrow X_{i+1}/Z_i$ in which $X_i \to Z_i \leftarrow X_{i+1}$ is a G_i-flip/Z such that $K_{X_i} + B_i \equiv 0/Z_i$ where G_i is the birational transform of G on $X = X_1$.

Definition 2.2 ([Birkar 2007, §2]). Let $(X/Z, B)$ be a klt pair, $(Y/Z, B_Y)$ a $\mathbb{Q}$-factorial klt pair, $\phi\colon X \dashrightarrow Y/Z$ a birational map such that ϕ^{-1} does not contract divisors, and B_Y be the birational transform of B (Note that since $X \to Z$ and

$Y \to Z$ are both projective, by the definition of a pair, X and Y have the same image on Z). Moreover, assume that

$$a(D, X, B) \leq a(D, Y, B_Y)$$

for any prime divisor D on birational models of X and assume that the strict inequality holds for any prime divisor D on X which is exceptional/Y.

We say that $(Y/Z, B_Y)$ is a *log minimal model* of $(X/Z, B)$ if $K_Y + B_Y$ is nef/Z. On the other hand, we say that $(Y/Z, B_Y)$ is a *Mori fibre space* of $(X/Z, B)$ if there is a $K_Y + B_Y$-negative extremal contraction $Y \to Y'/Z$ such that $\dim Y' < \dim Y$.

Typically, one obtains a log minimal model or a Mori fibre space by a finite sequence of divisorial contractions and log flips.

Remark 2.3. Let $(X/Z, B)$ be a klt pair and $W \to X$ a log resolution. Let $B_W = B^\sim + (1 - \epsilon) \sum E_i$ where $0 < \epsilon \ll 1$ and E_i are the exceptional/X divisors on W. Remember that $B^\sim$ is the birational transform of B. If $(Y/X, B_Y)$ is a log minimal model of $(W/X, B_W)$, which exists by [Birkar et al. 2006], then by the negativity lemma $Y \to X$ is a small $\mathbb{Q}$-factorialisation of X. To find a log minimal model or a Mori fibre space of $(X/Z, B)$, it is enough to find one for $(Y/Z, B_Y)$. So, one could assume that X is $\mathbb{Q}$-factorial by replacing it with Y.

We recall a variant of the LMMP with scaling which we use in this paper. Let $(X/Z, B+C)$ be a $\mathbb{Q}$-factorial klt pair such that $K_X + B + C$ is nef/Z and $B, C \geq 0$. By [Birkar 2007, Lemma 2.7], either $K_X + B$ is nef/Z or there is an extremal ray R/Z such that

$$(K_X + B) \cdot R < 0 \ \text{ and } \ (K_X + B + \lambda_1 C) \cdot R = 0,$$

where

$$\lambda_1 := \inf\{t \geq 0 \mid K_X + B + tC \text{ is nef}/Z\}$$

and $K_X + B + \lambda_1 C$ is nef/Z. Now assume that R defines a divisorial contraction or a log flip $X \dashrightarrow X'$. We can consider $(X'/Z, B' + \lambda_1 C')$, where $B' + \lambda_1 C'$ is the birational transform of $B + \lambda_1 C$ and continue the argument. That is, either $K_{X'} + B'$ is nef/Z or there is an extremal ray R'/Z such that $(K_{X'} + B') \cdot R' < 0$ and $(K_{X'} + B' + \lambda_2 C') \cdot R' = 0$, where

$$\lambda_2 := \inf\{t \geq 0 \mid K_{X'} + B' + tC' \text{ is nef}/Z\}$$

and $K_{X'} + B' + \lambda_2 C'$ is nef/Z. By continuing this process, we obtain a special kind of LMMP on $K_X + B$ which we refer to as the *LMMP with scaling of C*. If it terminates, then we obviously get a log minimal model or a Mori fibre space for $(X/Z, B)$. Note that the required log flips exist by [Birkar et al. 2006].

3. Extremal rays

In this section, for convenience of the reader, we give the proofs of some results about extremal rays [Shokurov 2006, Corollary 9, Addendum 4]. The norm $\|G\|$ of an $\mathbb{R}$-divisor G denotes the maximum of the absolute value of its coeffecients.

Let $X \to Z$ be a projective morphism of normal quasiprojective varieties. A curve Γ on X is called *extremal/Z* if it generates an extremal ray R/Z which defines a contraction $X \to S/Z$, and if for some ample$/Z$ divisor H we have $H \cdot \Gamma = \min\{H \cdot \Sigma\}$, where Σ ranges over curves generating R. If $(X/Z, B)$ is divisorial log terminal (dlt) and $(K_X + B) \cdot R < 0$, then by [Kawamata 1991] there is a curve Σ generating R such that $(K_X + B) \cdot \Sigma \geq -2 \dim X$. On the other hand, since Γ and Σ both generate R we have

$$\frac{(K_X + B) \cdot \Gamma}{H \cdot \Gamma} = \frac{(K_X + B) \cdot \Sigma}{H \cdot \Sigma},$$

hence

$$(K_X + B) \cdot \Gamma = (K_X + B) \cdot \Sigma \frac{H \cdot \Gamma}{H \cdot \Sigma} \geq -2 \dim X.$$

Remark 3.1. Let $(X/Z, B)$ be a $\mathbb{Q}$-factorial klt pair, F be a reduced divisor on X whose support contains that of B, and V be the $\mathbb{R}$-vector space of divisors generated by the components of F.

(i) By [Shokurov 1992, Property 1.3.2; 1996, First Main Theorem 6.2 and Remark 6.4], the sets

$$\mathcal{L} = \{\Delta \in V \mid (X/Z, \Delta) \text{ is lc}\} \quad \text{and} \quad \mathcal{N} = \{\Delta \in \mathcal{L} \mid K_X + \Delta \text{ is nef}/Z\}$$

are rational polytopes in V. Since $B \in \mathcal{L}$, there are rational boundaries $B^1, \ldots, B^r \in \mathcal{L}$ and nonnegative real numbers $a_1, \ldots, a_r$ such that $B = \sum a_j B^j$, $\sum a_j = 1$, and each $(X/Z, B^j)$ is klt. In particular, there is $m \in \mathbb{N}$ such that $m(K_X + B^j)$ are Cartier, and for any curve Γ on X the intersection number $(K_X + B) \cdot \Gamma$ can be written as $\sum a_j \frac{n_j}{m}$ for certain $n_1, \ldots, n_r \in \mathbb{Z}$. Moreover, if Γ is extremal$/Z$, then the n_j satisfy $n_j \geq -2m \dim X$.

(ii) If $K_X + B$ is nef$/Z$, then $B \in \mathcal{N}$ and so one can choose the B^j so that $K_X + B^j$ are nef$/Z$.

Lemma 3.2. *Let $(X/Z, B)$ be a $\mathbb{Q}$-factorial klt pair. There is a real number $\alpha > 0$ such that*:

(i) *If Γ is any extremal curve$/Z$ and if $(K_X + B) \cdot \Gamma > 0$, then $(K_X + B) \cdot \Gamma > \alpha$.*

(ii) *If $K_X + B$ is nef$/Z$, then for any $\mathbb{R}$-divisor G, any sequence $X_i \dashrightarrow X_{i+1}/Z_i$ of G-flops$/Z$ with respect to $(X/Z, B)$, and any extremal curve Γ on X_i, if $(K_{X_i} + B_i) \cdot \Gamma > 0$, then $(K_{X_i} + B_i) \cdot \Gamma > \alpha$ where B_i is the birational transform of B.*

Proof. (i) If B is a $\mathbb{Q}$-divisor, then the statement is trivially true. Let $B^1, \dots, B^r$, $a_1, \dots, a_r$, and m be as in Remark 3.1(i). Let Γ be an extremal curve$/Z$. Then, $(K_X + B) \cdot \Gamma = \sum a_j (K_X + B^j) \cdot \Gamma$ and since for each j we have $(K_X + B^j) \cdot \Gamma \geq -2 \dim X$, the existence of α is clear for (i).

(ii) By Remark 3.1(ii) we may also assume that $K_X + B^j$ are nef$/Z$. Then, the sequence $X_i \dashrightarrow X_{i+1}/Z_i$ is also a sequence of flops with respect to each $(X/Z, B^j)$. In particular, $(X_i/Z, B_i^j)$ is klt and $m(K_{X_i} + B_i^j)$ is Cartier for any j, i where B_i^j is the birational transform of B^j. The rest is as in (i). □

Proposition 3.3. *Let $(X/Z, B)$ be a $\mathbb{Q}$-factorial klt pair, F a reduced divisor on X whose support contains that of B, and $\mathcal{L}$ as in Remark 3.1. There is a rational polytope $\mathcal{H} \subset \mathcal{L}$ of klt boundaries and of maximal dimension containing an open neighborhood of B in $\mathcal{L}$ (with respect to the topology on $\mathcal{L}$ induced by the norm $\|\cdot\|$) such that*

(i) *if $\Delta \in \mathcal{H}$ and $(K_X + \Delta) \cdot R < 0$ for an extremal ray R/Z, then $(K_X + B) \cdot R \leq 0$, and*

(ii) *if $K_X + B$ is nef$/Z$, $\Delta \in \mathcal{H}$, we have a sequence $X_i \dashrightarrow X_{i+1}/Z_i$ of $K_X + \Delta$-flops$/Z$ with respect to $(X/Z, B)$, and $(K_{X_i} + \Delta_i) \cdot R < 0$ for an extremal ray R/Z on some X_i, then $(K_{X_i} + B_i) \cdot R = 0$, where Δ_i, B_i are the birational transforms of Δ, B respectively.*

Proof. (i) Let $\mathcal{M} \subset \mathcal{L}$ be a rational polytope of klt boundaries and of maximal dimension containing an open neighborhood of B in $\mathcal{L}$. If the statement is not true then there is an infinite sequence of $\Delta_l \in \mathcal{M}$ and extremal rays R_l/Z such that for each l we have

$$(K_X + \Delta_l) \cdot R_l < 0, \quad (K_X + B) \cdot R_l > 0,$$

and $\|\Delta_l - B\|$ converges to 0. Let Ω_l be the point on the boundary of $\mathcal{M}$ such that $\Omega_l - \Delta_l = b_l(\Delta_l - B)$ for some real number $b_l \geq 0$ and such that $\|\Omega_l - B\|$ is maximal. So, Ω_l is the most far away point in $\mathcal{M}$ which is on the line determined by B and Δ_l, in the direction of Δ_l. Since $\|\Delta_l - B\|$ converges to 0, b_l converges to $+\infty$.

By assumptions, $(X/Z, \Omega_l)$ is klt and if Γ_l is an extremal curve$/Z$ generating R_l, then

$$(\Omega_l - \Delta_l) \cdot \Gamma_l = (K_X + \Omega_l) \cdot \Gamma_l - (K_X + \Delta_l) \cdot \Gamma_l \geq -2 \dim X.$$

This is not possible because by Lemma 3.2,

$$(K_X + \Delta_l) \cdot \Gamma_l + (B - \Delta_l) \cdot \Gamma_l = (K_X + B) \cdot \Gamma_l > \alpha,$$

and by the same arguments $(B - \Delta_l) \cdot \Gamma_l$ approaches 0.

By definition, the sequence $X_i \dashrightarrow X_{i+1}/Z_i$ is a sequence of $K_X + \Delta$-flips which are numerically trivial with respect to $K_X + B$. Let $\mathcal{H}$ be as in (i). Assume that R is an extremal ray$/Z$ on X_i such that $(K_{X_i} + \Delta_i) \cdot R < 0$ but $(K_{X_i} + B_i) \cdot R > 0$. Let Γ be an extremal curve$/Z$ generating R. Let Ω be the point on the boundary of $\mathcal{H}$ which is chosen for Δ similarly as in (i). By assumptions, $(X_i/Z, \Delta_i)$ and $(X_i/Z, \Omega_i)$ are klt where Ω_i is the birational transform of Ω. So,

$$(\Omega_i - \Delta_i) \cdot \Gamma = (K_{X_i} + \Omega_i) \cdot \Gamma - (K_{X_i} + \Delta_i) \cdot \Gamma \geq -2 \dim X.$$

On the other hand, $(K_{X_i} + B_i) \cdot \Gamma > \alpha$ where α is as in Lemma 3.2. By construction, there is some $b \geq 0$ such that $b(\Delta_i - B_i) = \Omega_i - \Delta_i$. Therefore,

$$(K_{X_i} + \Delta_i) \cdot \Gamma = (K_{X_i} + B_i) \cdot \Gamma + (\Delta_i - B_i) \cdot \Gamma > \alpha - \frac{2 \dim X}{b}.$$

which is not possible if $b\alpha > 2 \dim X$. In other words, if Δ is close enough to B then the statement of (ii) also holds, that is, we only need to shrink $\mathcal{H}$ appropriately. □

4. Proof of the main results

Proof of Theorem 1.2. Let $(X/Z, B)$ be a klt pair of dimension d. By Remark 2.3, we can assume that X is $\mathbb{Q}$-factorial. Let $H \geq 0$ be an $\mathbb{R}$-divisor which is big$/Z$ so that $K_X + B + H$ is klt and nef$/Z$. Run the LMMP$/Z$ on $K_X + B$ with scaling of H. If the LMMP terminates, then we get a log minimal model or a Mori fibre space. Suppose that we get an infinite sequence $X_i \dashrightarrow X_{i+1}/Z_i$ of log flips$/Z$, where we may also assume that $(X_1/Z, B_1) = (X/Z, B)$.

Let λ_i be the threshold on X_i determined by the LMMP with scaling as explained in Section 2. So, $K_{X_i} + B_i + \lambda_i H_i$ is nef$/Z$,

$$(K_{X_i} + B_i) \cdot R_i < 0 \quad \text{and} \quad (K_{X_i} + B_i + \lambda_i H_i) \cdot R_i = 0,$$

where B_i and H_i are the birational transforms of B and H and, R_i is the extremal ray which defines the flipping contraction $X_i \to Z_i$. Obviously, $\lambda_i \geq \lambda_{i+1}$.

Put $\lambda = \lim_{i\to\infty} \lambda_i$. If the limit is attained, that is, $\lambda = \lambda_i$ for some i, then the sequence terminates by Corollary 1.4.2 of [Birkar et al. 2006]. So, we assume that the limit is not attained. Actually, if $\lambda > 0$, again [Birkar et al. 2006] implies that the sequence terminates. However, we do not need to use [Birkar et al. 2006] in this case. In fact, by replacing B_i with $B_i + \lambda H_i$, we can assume that $\lambda = 0$ hence $\lim_{i\to\infty} \lambda_i = 0$.

Put $\Lambda_i := B_i + \lambda_i H_i$. Since we are assuming that terminal log flips terminate, or, alternatively, by Corollary 1.4.3 of [Birkar et al. 2006], we can construct a terminal (in codimension ≥ 2) crepant model $(Y_i/Z, \Theta_i)$ of $(X_i/Z, \Lambda_i)$. A slight modification of the argument in Remark 2.3 would do this. Note that we can assume that all the Y_i are isomorphic to Y_1 in codimension one, perhaps after truncating the sequence. Let $\Delta_1 = \lim_{i\to\infty} \Theta_i^{\sim}$ on Y_1 and let Δ_i be its birational transform on Y_i. The limit is obtained componentwise.

Since H_i is big$/Z$ and $K_{X_i} + \Lambda_i$ is klt and nef$/Z$, $K_{X_i} + \Lambda_i$ and $K_{Y_i} + \Theta_i$ are semiample$/Z$ by the base point freeness theorem for $\mathbb{R}$-divisors. Thus, $K_{Y_i}+\Delta_i$ is a limit of movable$/Z$ divisors which in particular means that it is pseudo-effective$/Z$. Note that if $K_{Y_i}+\Delta_i$ is not pseudo-effective$/Z$, we get a contradiction by Corollary 1.3.2 of [Birkar et al. 2006].

Now run the LMMP$/Z$ on $K_{Y_1} + \Delta_1$. No divisor will be contracted again because $K_{Y_1} + \Delta_1$ is a limit of movable$/Z$ divisors. Since $K_{Y_1} + \Delta_1$ is terminal in codimension ≥ 2, by assumptions, the LMMP terminates with a log minimal model $(W/Z, \Delta)$. By construction, Δ on W is the birational transform of Δ_1 on Y_1, and $G_i := \Theta_i^{\sim} - \Delta$ on W satisfies $\lim_{i\to\infty} G_i = 0$.

By Proposition 3.3, for each G_i with $i \gg 0$, we can run the LMMP$/Z$ on $K_W + \Delta + G_i$ which will be a sequence of G_i-flops, that is, $K + \Delta$ would be numerically zero on all the extremal rays contracted in the process. No divisor will be contracted because $K_W+\Delta+G_i$ is movable$/Z$. The LMMP ends up with a log minimal model $(W_i/Z, \Omega_i)$. Here, Ω_i is the birational transform of $\Delta + G_i$ and so of Θ_i. Let S_i be the lc model of $(W_i/Z, \Omega_i)$ which is the same as the lc model of $(Y_i/Z, \Theta_i)$ and that of $(X_i/Z, \Lambda_i)$ because $K_{W_i} + \Omega_i$ and $K_{Y_i} + \Theta_i$ are nef$/Z$ with W_i and Y_i being isomorphic in codimension one, and $K_{Y_i} + \Theta_i$ is the pullback of $K_{X_i} + \Lambda_i$. Also note that since $K_{X_i} + B_i$ is pseudoeffective$/Z$, $K_{X_i} + \Lambda_i$ is big$/Z$; hence S_i is birational to X_i.

By construction, $K_{W_i} + \Delta^{\sim}$ is nef$/Z$ and it turns out that $K_{W_i} + \Delta^{\sim} \sim_{\mathbb{R}} 0/S_i$. Suppose that this is not the case. Then, $K_{W_i}+\Delta^{\sim}$ is not numerically zero$/S_i$ hence there is some curve C/S_i such that $(K_{W_i}+\Delta^{\sim}+G_i^{\sim})\cdot C = 0$ but $(K_{W_i}+\Delta^{\sim})\cdot C > 0$ which implies that $G_i^{\sim}\cdot C < 0$. Hence, there is a $K_{W_i} + \Delta^{\sim} + (1+\tau)G_i^{\sim}$-negative extremal ray R/S_i for any $\tau > 0$. This contradicts Proposition 3.3 because we must have

$$(K_{W_i} + \Delta^{\sim} + G_i^{\sim}) \cdot R = (K_{W_i} + \Delta^{\sim}) \cdot R = 0.$$

Therefore, $K_{W_i} + \Delta^{\sim} \sim_{\mathbb{R}} 0/S_i$. Now $K_{X_i} + \Lambda_i \sim_{\mathbb{R}} 0/Z_i$ implies that Z_i is over S_i and so $K_{Y_i} + \Delta_i \sim_{\mathbb{R}} 0/S_i$. On the other hand, $K_{X_i} + B_i$ is the pushdown of $K_{Y_i} + \Delta_i$; hence $K_{X_i} + B_i \sim_{\mathbb{R}} 0/S_i$. Thus, $K_{X_i} + B_i \sim_{\mathbb{R}} 0/Z_i$ and this contradicts the fact that $X_i \to Z_i$ is a $K_{X_i} + B_i$-flipping contraction. So, the sequence of flips terminates and this completes the proof. □

Proof of Corollary 1.3. Since terminal log flips terminate in dimension 4 by [Fujino 2004; Shokurov 2004] (see also [Alexeev et al. 2007]), the result follows from Theorem 1.2. □

Acknowledgements

I thank V. V. Shokurov and Y. Kawamata for their comments and suggestions, which improved this paper.

References

[Alexeev et al. 2007] V. Alexeev, C. Hacon, and Y. Kawamata, "Termination of (many) 4-dimensional log flips", *Invent. Math.* **168**:2 (2007), 433–448. MR 2008f:14028 Zbl 1118.14017

[Birkar 2007] C. Birkar, "On existence of log minimal models", preprint, 2007. arXiv 0706.1792

[Birkar et al. 2006] C. Birkar, P. Cascini, C. Hacon, and J. McKernan, "Existence of minimal models for varieties of log general type", preprint, 2006. arXiv math/0610203

[Fujino 2004] O. Fujino, "Termination of 4-fold canonical flips", *Publ. Res. Inst. Math. Sci.* **40**:1 (2004), 231–237. Addendum: **41**:1 (2005), 251–257. MR 2004k:14020 Zbl 1069.14017

[Kawamata 1991] Y. Kawamata, "On the length of an extremal rational curve", *Invent. Math.* **105**:3 (1991), 609–611. MR 92m:14026 Zbl 0751.14007

[Shokurov 1992] V. V. Shokurov, "Three-dimensional log flips", *Izv. Ross. Akad. Nauk Ser. Mat.* **56**:1 (1992), 105–203. In Russian; translated in *Russian Acad. Sci. Izv. Math.* **40**:1 (1993), 95–202. MR 93j:14012 Zbl 0785.14023

[Shokurov 1996] V. V. Shokurov, "3-fold log models", *Journal of Math. Sci.* **81**:3 (1996), 2667–2699. MR 97i:14015 Zbl 0873.14014

[Shokurov 2004] V. V. Shokurov, "Letters of a bi-rationalist, V: Minimal log discrepancies and termination of log flips", pp. 328–351 in *Algebr. Geom. Metody, Svyazi i Prilozh.*, Tr. Mat. Inst. Steklova **246**, 2004. In Russian; translated in *Proc. Stekl. Inst. Mathematics* **246**, (2004), 315–336. MR 2006b:14019 Zbl 1107.14012

[Shokurov 2006] V. V. Shokurov, "Letters of a bi-rationalist, VII: Ordered termination", preprint, 2006. arXiv math/0607822

Communicated by Yujiro Kawamata
Received 2009-03-12 Revised 2009-09-07 Accepted 2009-10-06

c.birkar@dpmms.cam.ac.uk *DPMMS, Centre for Mathematical Sciences, Cambridge University, Wilberforce Road, Cambridge, CB3 0WB, United Kingdom*

ALGEBRA AND NUMBER THEORY 3:8(2009)

Centers of graded fusion categories

Shlomo Gelaki, Deepak Naidu and Dmitri Nikshych

Let $\mathcal{C}$ be a fusion category faithfully graded by a finite group G and let $\mathcal{D}$ be the trivial component of this grading. The center $\mathcal{Z}(\mathcal{C})$ of $\mathcal{C}$ is shown to be canonically equivalent to a G-equivariantization of the relative center $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})$. We use this result to obtain a criterion for $\mathcal{C}$ to be group-theoretical and apply it to Tambara–Yamagami fusion categories. We also find several new series of modular categories by analyzing the centers of Tambara–Yamagami categories. Finally, we prove a general result about the existence of zeroes in S-matrices of weakly integral modular categories.

1. Introduction

Throughout this paper we work over an algebraically closed field k of characteristic 0. All categories considered in this paper are finite, abelian, semisimple, and k-linear. We freely use the language and basic theory of fusion categories, module categories over them, braided categories, and Frobenius–Perron dimensions [Bakalov and Kirillov 2001; Ostrik 2003; Etingof et al. 2005].

Let G be a finite group. A fusion category $\mathcal{C}$ is *G-graded* if there is a decomposition

$$\mathcal{C} = \bigoplus_{g \in G} \mathcal{C}_g$$

of $\mathcal{C}$ into a direct sum of full abelian subcategories such that the tensor product of $\mathcal{C}$ maps $\mathcal{C}_g \times \mathcal{C}_h$ to $\mathcal{C}_{gh}$, for all $g, h \in G$. A *G-extension* of a fusion category $\mathcal{D}$ is a G-graded fusion category $\mathcal{C}$ whose trivial component $\mathcal{C}_e$, where e is the identity of G, is equivalent to $\mathcal{D}$.

MSC2000: primary 16W30; secondary 18D10.
Keywords: fusion categories, braided categories, graded tensor categories.

Gradings and extensions play an important role in the study and classification of fusion categories. For example, *nilpotent* fusion categories (that is, those categories that can be obtained from the trivial category by a sequence of group extensions) were studied in [Gelaki and Nikshych 2008]. It was proved in [Etingof et al. 2005] that every fusion category of prime power dimension is nilpotent. Group-theoretical properties of such categories were studied in [Drinfeld et al. 2007]. Recently, fusion categories of dimension $p^n q^m$, where p, q are primes, were shown to be Morita equivalent to nilpotent categories [Etingof et al. 2009].

The main goal of this paper is to describe the center $\mathcal{Z}(\mathcal{C})$ of a G-graded fusion category $\mathcal{C}$ in terms of its trivial component $\mathcal{D}$ (Theorem 3.5) and apply this description to the study of structural properties of $\mathcal{C}$ and the construction of new examples of modular categories.

The organization of the paper is as follows. In Section 2 we recall some basic notions, results, and examples of fusion categories, notably the notions of the relative center of a bimodule category [Majid 1991], group action on a fusion category and crossed product [Tambara 2001], equivariantization and de-equivariantization theory [Arkhipov and Gaitsgory 2003; Bruguières 2000; Gaitsgory 2005; Kirillov 2002; Müger 2000; Drinfeld et al. 2009], and braided G-crossed fusion categories [Turaev 2000; 2008].

In Section 3 we study the center $\mathcal{Z}(\mathcal{C})$ of a G-graded fusion category $\mathcal{C}$. We show that if $\mathcal{D}$ is the trivial component of $\mathcal{C}$, then the relative center $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})$ has a canonical structure of a braided G-crossed category and there is an equivalence of braided fusion categories $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})^G \cong \mathcal{Z}(\mathcal{C})$ (Theorem 3.5). Thus, the structure of $\mathcal{Z}(\mathcal{C})$ can be understood in terms of a smaller and more transparent category $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})$. In particular, there is a canonical braided action (studied in detail in [Etingof et al. 2009]) of G on $\mathcal{Z}(\mathcal{D})$. In Corollary 3.10 we use this action to prove that $\mathcal{C}$ is group-theoretical if and only if $\mathcal{Z}(\mathcal{D})$ contains a G-stable Lagrangian subcategory. As an illustration, we describe the center of a crossed product fusion category $\mathcal{C} = \mathcal{D} \rtimes G$.

We apply the results from Section 4 to the study of Tambara–Yamagami categories [Tambara and Yamagami 1998]. We obtain a convenient description of the centers of such categories as equivariantizations and compute their modular data, that is, S- and T-matrices. This computation was previously done in [Izumi 2001] using different techniques. We establish a criterion for a Tambara–Yamagami category to be group-theoretical (Theorem 4.6). We also extend the construction of non-group-theoretical semisimple Hopf algebras from Tambara–Yamagami categories given in [Nikshych 2008].

In Section 5 we construct a series of new modular categories as factors of the centers of Tambara–Yamagami categories. One associates a pair of such categories $\mathcal{E}(q, \pm)$ with any nondegenerate quadratic form q on an abelian group A of odd order. The categories $\mathcal{E}(q, \pm)$ have dimension $4|A|$. They are group-theoretical if and only if A contains a Lagrangian subgroup with respect to q. We compute the S- and T-matrices of $\mathcal{E}(q, \pm)$ and write down several small examples explicitly.

Section 6 is independent from the rest of the paper and contains a general result about existence of zeroes in S-matrices of weakly integral modular categories (Theorem 6.1). This is a categorical analogue of a classical result of Burnside in character theory.

2. Preliminaries

2A. ***Dual fusion categories and Morita equivalence.*** Let $\mathcal{C}$ be a fusion category and let $\mathcal{M}$ be an indecomposable right $\mathcal{C}$-module category $\mathcal{M}$. The category $\mathcal{C}^*_{\mathcal{M}}$ of $\mathcal{C}$-module endofunctors of $\mathcal{M}$ is a fusion category, called the dual of $\mathcal{C}$ with respect to $\mathcal{M}$ [Etingof et al. 2005; Ostrik 2003].

Following [Müger 2003a], we say that two fusion categories $\mathcal{C}$ and $\mathcal{D}$ are *Morita equivalent* if $\mathcal{D}$ is equivalent to $\mathcal{C}^*_{\mathcal{M}}$, for some indecomposable right $\mathcal{C}$-module category $\mathcal{M}$. A fusion category is said to be *pointed* if all its simple objects are invertible (any such category is equivalent to the category Vec_G^{ω} of vector spaces graded by a finite group G with the associativity constraint given by a 3-cocycle $\omega \in Z^3(G,\, k^\times)$). A fusion category is called *group-theoretical* if it is Morita equivalent to a pointed fusion category. See [Ostrik 2003; Etingof et al. 2005; Nikshych 2008] for details of the theory of group-theoretical categories.

2B. ***The center of a bimodule category and the relative center of a fusion category.*** Let $\mathcal{C}$ be a fusion category with unit object $\mathbf{1}$ and associativity constraint $\alpha_{X,Y,Z} : (X \otimes Y) \otimes Z \xrightarrow{\sim} X \otimes (Y \otimes Z)$ and let $\mathcal{M}$ be a $\mathcal{C}$-bimodule category.

Definition 2.1. The *center* of $\mathcal{M}$ is the category $\mathcal{Z}_{\mathcal{C}}(\mathcal{M})$ of $\mathcal{C}$-bimodule functors from $\mathcal{C}$ to $\mathcal{M}$.

Explicitly, the objects of $\mathcal{Z}_{\mathcal{C}}(\mathcal{M})$ are pairs (M, γ), where M is an object of $\mathcal{M}$ and

$$\gamma = \{\gamma_X : X \otimes M \xrightarrow{\sim} M \otimes X\}_{X \in \mathcal{C}} \tag{1}$$

is a natural family of isomorphisms making the diagram

$$\begin{tikzcd}[column sep=small]
& X \otimes (M \otimes Y) \arrow[rr, "\alpha^{-1}_{X,M,Y}"] & & (X \otimes M) \otimes Y \arrow[dr, "\gamma_X"] & \\
X \otimes (Y \otimes M) \arrow[ur, "\gamma_Y"] \arrow[dr, "\alpha^{-1}_{X,Y,M}"'] & & & & (M \otimes X) \otimes Y, \\
& (X \otimes Y) \otimes M \arrow[rr, "\gamma_{X\otimes Y}"'] & & M \otimes (X \otimes Y) \arrow[ur, "\alpha^{-1}_{M,X,Y}"'] &
\end{tikzcd} \tag{2}$$

commutative, where the α's denote the associativity constraints in $\mathcal{M}$.

Indeed, a $\mathscr{C}$-bimodule functor $F : \mathscr{C} \to \mathcal{M}$ is completely determined by the pair $(F(\mathbf{1}),\ \{\gamma_X\}_{X\in\mathscr{C}})$, where $\gamma = \{\gamma_X\}_{X\in\mathscr{C}}$ is the collection of isomorphisms

$$\gamma_X : X \otimes F(\mathbf{1}) \xrightarrow{\sim} F(X) \xrightarrow{\sim} F(\mathbf{1}) \otimes X,$$

coming from the $\mathscr{C}$-bimodule structure on F.

We will call the natural family of isomorphisms (1) the *central structure* of an object $X \in \mathfrak{Z}_{\mathscr{C}}(\mathcal{M})$.

Remark 2.2. (i) The definition of the center of a bimodule category is parallel to that of the center of a bimodule over a ring.

(ii) We will often suppress the central structure while working with objects of $\mathfrak{Z}_{\mathscr{C}}(\mathcal{M})$ and refer to (M, γ) simply as M.

(iii) $\mathfrak{Z}_{\mathscr{C}}(\mathcal{M})$ is a semisimple abelian category. It has the obvious canonical structure of a $\mathfrak{Z}(\mathscr{C})$-module category, where $\mathfrak{Z}(\mathscr{C})$ is the center of $\mathscr{C}$ (see, for example, [Kassel 1995, Section XIII.4] for the definition of $\mathfrak{Z}(\mathscr{C})$).

Here is an important special case of this construction. Let $\mathscr{C}$ be a fusion category and let $\mathfrak{D} \subset \mathscr{C}$ be a fusion subcategory. Then $\mathscr{C}$ is a $\mathfrak{D}$-bimodule category. We will call $\mathfrak{Z}_{\mathfrak{D}}(\mathscr{C})$ the *relative center* of $\mathscr{C}$.

Remark 2.3. The aforementioned construction of the relative center is a special case of a more general construction considered in [Majid 1991, Definition 3.2 and Theorem 3.3].

It is easy to see that $\mathfrak{Z}_{\mathfrak{D}}(\mathscr{C})$ is a tensor category with tensor product defined as follows. If (X, γ) and (X', γ') are objects in $\mathfrak{Z}_{\mathfrak{D}}(\mathscr{C})$ then

$$(X, \gamma) \otimes (X', \gamma') := (X \otimes X', \tilde{\gamma}),$$

where $\tilde{\gamma}_V : V \otimes (X \otimes X') \xrightarrow{\sim} (X \otimes X') \otimes V,\ V \in \mathfrak{D}$, is defined by the diagram

$$\begin{tikzcd}
V \otimes (X \otimes X') \arrow[r, "\alpha^{-1}_{V,X,X'}"] \arrow[d, "\tilde{\gamma}_V"'] & (V \otimes X) \otimes X' \arrow[r, "\gamma_V"] & (X \otimes V) \otimes X' \arrow[d, "\alpha_{X,V,X'}"] \\
(X \otimes X') \otimes V & X \otimes (X' \otimes V) \arrow[l, "\alpha^{-1}_{X,X',V}"'] & X \otimes (V \otimes X'). \arrow[l, "\gamma'_V"']
\end{tikzcd} \tag{3}$$

The unit object of $\mathfrak{Z}_{\mathfrak{D}}(\mathscr{C})$ is $(\mathbf{1},\ \mathrm{id})$. The dual of (X, γ) is $(X^*,\ \overline{\gamma})$, where $\overline{\gamma}_V := (\gamma_{^*V})^*$.

Remark 2.4. Let $\mathscr{C}$ and $\mathfrak{D}$ be as above.

(i) $\mathfrak{Z}_{\mathfrak{D}}(\mathscr{C})$ is dual to the fusion category $\mathfrak{D} \boxtimes \mathscr{C}^{\mathrm{rev}}$ (where $\mathscr{C}^{\mathrm{rev}}$ is the fusion category obtained from $\mathscr{C}$ by reversing the tensor product and $\boxtimes$ is Deligne's tensor product of fusion categories) with respect to its module category $\mathscr{C}$, where $\mathcal{D}$ and $\mathcal{C}^{\text{rev}}$ act on $\mathcal{C}$ via the right and left multiplication respectively. In particular, $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})$ is a fusion category.

(ii) $\mathrm{FPdim}(\mathcal{Z}_{\mathcal{D}}(\mathcal{C})) = \mathrm{FPdim}(\mathcal{C})\,\mathrm{FPdim}(\mathcal{D})$, where FPdim denotes the Frobenius–Perron dimension of a category.

(iii) $\mathcal{Z}_{\mathcal{C}}(\mathcal{C})$ coincides with the center $\mathcal{Z}(\mathcal{C})$ of $\mathcal{C}$. This category has a canonical braiding given by

$$c_{(X,\gamma),\,(X',\gamma')} = \gamma_{X'} : (X,\gamma)\otimes(X',\gamma') \xrightarrow{\sim} (X',\gamma')\otimes(X,\gamma). \tag{4}$$

(iv) There is an obvious forgetful tensor functor:

$$\mathcal{Z}(\mathcal{C}) \mapsto \mathcal{Z}_{\mathcal{D}}(\mathcal{C}) : (X,\gamma) \mapsto (X,\gamma|_{\mathcal{D}}). \tag{5}$$

2C. ***Centralizers in braided fusion categories.*** Let $\mathcal{C}$ be a braided fusion category with braiding c. Two objects X and Y of $\mathcal{C}$ are said to *centralize* each other [Müger 2003b] if $c_{Y,X}c_{X,Y} = \mathrm{id}_{X\otimes Y}$.

For any fusion subcategory $\mathcal{D}\subseteq\mathcal{C}$ its *centralizer* $\mathcal{D}'$ is the full fusion subcategory of $\mathcal{C}$ consisting of all objects $X\in\mathcal{C}$ centralizing every object in $\mathcal{D}$. The category $\mathcal{C}$ is said to be *nondegenerate* if $\mathcal{C}' = \mathrm{Vec}$. In this case one has $\mathcal{D}'' = \mathcal{D}$ [Müger 2003b]. If $\mathcal{C}$ is a premodular category, that is, has a spherical structure, then it is nondegenerate if and only if it is modular.

A braided fusion category $\mathcal{E}$ is called *Tannakian* if it is equivalent to the representation category $\mathrm{Rep}(G)$ of a finite group G as a braided fusion category. Here $\mathrm{Rep}(G)$ is considered with its standard symmetric braiding. The group G is defined by $\mathcal{E}$ up to an isomorphism [Deligne 1990].

A fusion subcategory $\mathcal{L}$ of a braided fusion category is called *Lagrangian* if it is Tannakian and $\mathcal{L} = \mathcal{L}'$.

Theorem 2.5 [Drinfeld et al. 2007]. *A fusion category $\mathcal{C}$ is group-theoretical if and only if $\mathcal{Z}(\mathcal{C})$ contains a Lagrangian subcategory.*

2D. ***Group actions on fusion categories and equivariantization.*** Let G be a finite group, and let $\underline{G}$ denote the monoidal category whose objects are elements of G, whose morphisms are identities, and whose tensor product is given by multiplication in G. Recall that an action of G on a fusion category $\mathcal{C}$ is a monoidal functor $\underline{G}\to\mathrm{Aut}_{\otimes}(\mathcal{C}) : g\mapsto T_g$. For any $g,h\in G$, let

$$\gamma_{g,h} = T_g\circ T_h \simeq T_{gh}$$

be the isomorphism defining the monoidal structure on the functor $\underline{G}\to\mathrm{Aut}_{\otimes}(\mathcal{C})$.

Definition 2.6. A *G-equivariant object* in $\mathcal{C}$ is a pair $(X,\{u_g\}_{g\in G})$ consisting of an object X of $\mathcal{C}$ together with a collection of isomorphisms $u_g : T_g(X)\simeq X$, $g\in G$, such that the diagram

$$\begin{tikzcd} T_g(T_h(X)) \arrow[r, "T_g(u_h)"] \arrow[d, "\gamma_{g,h}(X)"'] & T_g(X) \arrow[d, "u_g"] \\ T_{gh}(X) \arrow[r, "u_{gh}"] & X \end{tikzcd}$$

commutes for all $g, h \in G$. One defines morphisms of equivariant objects to be morphisms in $\mathscr{C}$ commuting with $u_g,\ g \in G$.

Equivariant objects in $\mathscr{C}$ form a fusion category, called the *equivariantization* of $\mathscr{C}$ and denoted by $\mathscr{C}^G$ [Tambara 2001; Arkhipov and Gaitsgory 2003; Gaitsgory 2005]. One has $\mathrm{FPdim}(\mathscr{C}^G) = |G|\,\mathrm{FPdim}(\mathscr{C})$.

There is another fusion category that comes from an action of G on $\mathscr{C}$. It is the *crossed product* category $\mathscr{C} \rtimes G$ defined as follows [Tambara 2001; Nikshych 2008]. As an abelian category, $\mathscr{C} \rtimes G := \mathscr{C} \boxtimes \mathrm{Vec}_G$, where Vec_G denotes the fusion category of G-graded vector spaces. The tensor product in $\mathscr{C} \rtimes G$ is given by

$$(X \boxtimes g) \otimes (Y \boxtimes h) := (X \otimes T_g(Y)) \boxtimes gh, \qquad X, Y \in \mathscr{C}, \quad g, h \in G. \tag{6}$$

The unit object is $\mathbf{1} \boxtimes e$ and the associativity and unit constraints come from those of $\mathscr{C}$. Clearly, $\mathscr{C} \rtimes G$ is faithfully G-graded with the trivial component $\mathscr{C}$.

As explained in [Nikshych 2008], $\mathscr{C}$ is a right $\mathscr{C} \rtimes G$-module category via

$$Y \otimes (X \boxtimes g) := T_{g^{-1}}(Y \otimes X),$$

and the corresponding dual category $(\mathscr{C} \rtimes G)^*_{\mathscr{C}}$ is equivalent to $\mathscr{C}^G$. It follows from [Müger 2003a] that there is an equivalence of braided fusion categories

$$\mathscr{Z}(\mathscr{C} \rtimes G) \cong \mathscr{Z}(\mathscr{C}^G).$$

Let G be a finite group. For any conjugacy class K of G fix a representative $a_K \in K$. Let G_K denote the centralizer of a_K in G.

Proposition 2.7. *Let $\mathscr{C} = \bigoplus_{g \in G} \mathscr{C}_g$ be a G-graded fusion category with an action $g \mapsto T_g$ of G on $\mathscr{C}$ such that T_g carries $\mathscr{C}_h$ to $\mathscr{C}_{ghg^{-1}}$. Let $H := \{g \in G \mid \mathscr{C}_g \neq 0\}$. There is a bijection between the set of isomorphism classes of simple objects of $\mathscr{C}^G$ and pairs $(K,\ X)$, where $K \subset H$ is a conjugacy class of G and X is a simple G_K-equivariant object of $\mathscr{C}_{a_K}$.*

Proof. A simple G-equivariant object of $\mathscr{C}$ must be supported on a single conjugacy class K. Let $Y = \oplus_{g \in K} Y_g$ be such an object. Then Y_{a_K} is a simple G_K-equivariant object.

Conversely, given a G_K-equivariant object X in $\mathscr{C}_{a_K}$ let

$$Y = \bigoplus_h T_h(X),$$

where the summation is taken over the set of representatives of cosets of G_K in G. It is easy to see that Y acquires the structure of a simple G-equivariant object.

Clearly, the two constructions are inverses of each other. □

Remark 2.8. The Frobenius–Perron dimension of the simple object corresponding to a pair (K, X) in Proposition 2.7 is $|K|\,\mathrm{FPdim}(X)$.

2E. ***De-equivariantization of fusion categories.*** Let $\mathscr{C}$ be a fusion category. Let $\mathscr{E} = \mathrm{Rep}(G)$ be a Tannakian category along with a braided tensor functor $\mathscr{E} \to \mathscr{Z}(\mathscr{C})$ such that the composition $\mathscr{E} \to \mathscr{Z}(\mathscr{C}) \to \mathscr{C}$ (where the second arrow is the forgetful functor) is fully faithful. The following construction was introduced in [Bruguières 2000] and [Müger 2000]. Let $A := \mathrm{Fun}(G)$ be the algebra of functions on G. It is a commutative algebra in $\mathscr{E}$ and thus its image is a commutative algebra in $\mathscr{Z}(\mathscr{C})$. This fact allows us to view the category $\mathscr{C}_G$ of A-modules in $\mathscr{C}$ as a fusion category, called *de-equivariantization* of $\mathscr{C}$. There is a canonical surjective tensor functor

$$F : \mathscr{C} \to \mathscr{C}_G : X \mapsto A \otimes X. \tag{7}$$

It was explained in [Müger 2000; Drinfeld et al. 2009] that the group G acts on $\mathscr{C}_G$ by tensor autoequivalences (this action comes from the action of G on A by right translations). Furthermore, there is a bijection between subcategories of $\mathscr{C}$ containing the image of $\mathscr{E} = \mathrm{Rep}(G)$ and G-stable subcategories of $\mathscr{C}_G$. This bijection preserves Tannakian subcategories.

The procedures of equivariantization and de-equivariantization are inverses of each other: that is, there are canonical equivalences $(\mathscr{C}_G)^G \cong \mathscr{C}$ and $(\mathscr{C}^G)_G \cong \mathscr{C}$.

In particular, the construction above applies when $\mathscr{C}$ is a braided fusion category containing a Tannakian subcategory $\mathscr{E} = \mathrm{Rep}(G)$. In this case the braiding of $\mathscr{C}$ gives rise to an additional structure on the de-equivariantization functor (7). Namely, there is natural family of isomorphisms

$$X \otimes F(Y) \xrightarrow{\sim} F(Y) \otimes X, \qquad X \in \mathscr{C}_G,\ Y \in \mathscr{C}, \tag{8}$$

satisfying obvious compatibility conditions. In other words, F can be factored through a braided functor $\mathscr{C} \to \mathscr{Z}(\mathscr{C}_G)$, that is, F is a *central* functor.

If $\mathscr{E} \subset \mathscr{C}'$ then $\mathscr{C}_G$ is a braided fusion category with the braiding inherited from that of $\mathscr{C}$. If $\mathscr{E} = \mathscr{C}'$, the category $\mathscr{C}_G$ is nondegenerate. (In the presence of a spherical structure this category is called the *modularization* of $\mathscr{C}$ by $\mathscr{E}$ [Bruguières 2000; Müger 2000].)

Remark 2.9. The category $\mathscr{C}_G$ is not braided in general. However it does have an additional structure — it is a *braided G-crossed fusion category*. See next section (2F) for details.

2F. *Braided G-crossed categories.* Let G be a finite group. Kirillov [2002] and Müger [2004] found a description of all braided fusion categories $\mathcal{D}$ containing $\mathrm{Rep}(G)$. Namely, they showed that the datum of a braided fusion category $\mathcal{D}$ containing $\mathrm{Rep}(G)$ is equivalent to the datum of a braided G-crossed category $\mathcal{C}$; see Theorem 2.12. The notion of a braided G-crossed category is due to Turaev [2000; 2008] and is recalled below.

Definition 2.10. A *braided G-crossed fusion category* is a fusion category $\mathcal{C}$ equipped with (i) a (not necessarily faithful) grading $\mathcal{C} = \bigoplus_{g\in G} \mathcal{C}_g$, (ii) an action $g \mapsto T_g$ of G on $\mathcal{C}$ such that $T_g(\mathcal{C}_h) \subset \mathcal{C}_{ghg^{-1}}$, and (iii) a natural collection of isomorphisms

$$c_{X,Y} : X \otimes Y \simeq T_g(Y) \otimes X, \qquad X \in \mathcal{C}_g,\ g \in G \text{ and } Y \in \mathcal{C}, \tag{9}$$

called the *G-braiding*. These structures are required to satisfy certain compatibility conditions, which we now state. Let $\gamma_{g,h} : T_g T_h \xrightarrow{\sim} T_{gh}$ denote the tensor structure of the functor $g \mapsto T_g$ and μ_g the tensor structure of T_g.

(a) The diagram

$$\begin{tikzcd}
T_g(X)\otimes T_g(Y) \arrow[r, "c_{T_g(X),T_g(Y)}"] & T_{ghg^{-1}}(T_g(Y))\otimes T_g(X) \arrow[d, "(\gamma_{ghg^{-1},g})_Y\otimes \mathrm{id}_{T_g(X)}"] \\
T_g(X\otimes Y) \arrow[u, "(\mu_g)^{-1}_{X,Y}"] \arrow[d, "T_g(c_{X,Y})"'] & T_{gh}(Y)\otimes T_g(X) \\
T_g(T_h(Y)\otimes X) \arrow[r, "(\mu_g)^{-1}_{T_g(Y),X}"'] & T_g(T_h(Y))\otimes T_g(X) \arrow[u, "(\gamma_{g,h})_Y\otimes \mathrm{id}_{T_g(X)}"']
\end{tikzcd} \tag{10}$$

commutes for all $g, h \in G$ and objects $X \in \mathcal{C}_h,\ Y \in \mathcal{C}$.

(b) The diagram

$$\begin{tikzcd}
 & (X\otimes Y)\otimes Z \arrow[dl, "\alpha_{X,Y,Z}"'] \arrow[dr, "c_{X,Y}\otimes \mathrm{id}_Z"] & \\
X\otimes (Y\otimes Z) \arrow[d, "c_{X,Y\otimes Z}"'] & & (T_g(Y)\otimes X)\otimes Z \arrow[d, "\alpha_{T_g(Y),X,Z}"] \\
T_g(Y\otimes Z)\otimes X \arrow[d, "(\mu_g)^{-1}_{Y,Z}\otimes \mathrm{id}_X"'] & & T_g(Y)\otimes (X\otimes Z) \arrow[d, "\mathrm{id}_{T_g(Y)}\otimes c_{X,Z}"] \\
(T_g(Y)\otimes T_g(Z))\otimes X \arrow[rr, "\alpha_{T_g(Y),T_g(Z),X}"] & & T_g(Y)\otimes (T_g(Z)\otimes X)
\end{tikzcd} \tag{11}$$

commutes for all $g \in G$ and objects $X \in \mathcal{C}_g,\ Y, Z \in \mathcal{C}$.

(c) The diagram

$$\begin{tikzcd}[column sep=small]
 & X \otimes (Y \otimes Z) \arrow[dr, "\mathrm{id}_X \otimes c_{Y,Z}"] & \\
(X \otimes Y) \otimes Z \arrow[ur, "\alpha_{X,Y,Z}"] & & X \otimes (T_h(Z) \otimes Y) \arrow[d, "\alpha^{-1}_{X,T_h(Z),Y}"] \\
T_{gh}(Z) \otimes (X \otimes Y) \arrow[u, "c^{-1}_{X\otimes Y,Z}"] & & (X \otimes T_h(Z)) \otimes Y \arrow[d, "c_{X,T_h(Z)}\otimes \mathrm{id}_Y"] \\
T_g T_h(Z) \otimes (X \otimes Y) \arrow[u, "(\gamma_{g,h})_Z \otimes \mathrm{id}_{X\otimes Y}"] \arrow[rr, "\alpha^{-1}_{T_gT_h(Z),X,Y}"] & & (T_g T_h(Z) \otimes X) \otimes Y
\end{tikzcd} \tag{12}$$

commutes for all $g, h \in G$ and objects $X \in \mathscr{C}_g,\ Y \in \mathscr{C}_h,\ Z \in \mathscr{C}$.

Remark 2.11. The trivial component $\mathscr{C}_e$ of a braided G-crossed fusion category $\mathscr{C}$ is a braided fusion category with the action of G by braided autoequivalences. This can be seen by taking $X,\ Y \in \mathscr{C}_e$ in diagrams (10)–(12).

Theorem 2.12 ([Kirillov 2002; Müger 2004]). *The equivariantization and de-equivariantization constructions establish a bijection between the set of equivalence classes of G-crossed braided fusion categories and the set of equivalence classes of braided fusion categories containing* $\mathrm{Rep}(G)$ *as a symmetric fusion subcategory.*

We shall now sketch the proof of this theorem. An alternative approach is given in [Drinfeld et al. 2009].

Suppose $\mathscr{C}$ is a braided G-crossed fusion category. We define a braiding $\tilde{c}$ on its equivariantization $\mathscr{C}^G$ as follows.

Let $(X,\ \{u_g\}_{g\in G})$ and $(Y,\ \{v_g\}_{g\in G})$ be objects in $\mathscr{C}^G$. Let $X = \oplus_{g\in G} X_g$ be a decomposition of X with respect to the grading of $\mathscr{C}$. Define an isomorphism

$$\tilde{c}_{X,Y} : X \otimes Y = \bigoplus_{g\in G} X_g \otimes Y \xrightarrow{\oplus c_{X_g,Y}} \bigoplus_{g\in G} T_g(Y) \otimes X_g \xrightarrow{\oplus v_g \otimes \mathrm{id}_{X_g}} \bigoplus_{g\in G} Y \otimes X_g = Y \otimes X. \tag{13}$$

It follows from condition (a) of Definition 2.10 that $\tilde{c}_{X,Y}$ respects the equivariant structures, that is, it is an isomorphism in $\mathscr{C}^G$. Its naturality is clear. The fact that $\tilde{c}$ is a braiding on $\mathscr{C}^G$ (that is, the hexagon axioms) follows from the commutativity of diagrams (11) and (12). It is easy to check that $\tilde{c}$ restricts to the standard braiding on $\mathrm{Rep}(G) = \mathrm{Vec}^G \subset \mathscr{C}^G$. Hence, $\mathscr{C}^G$ contains a Tannakian subcategory $\mathrm{Rep}(G)$.

Conversely, let $\mathscr{C}$ be a braided fusion category with braiding c containing a Tannakian subcategory $\mathrm{Rep}(G)$. The restriction of the de-equivariantization functor F from (7) on $\mathrm{Rep}(G)$ is isomorphic to the fiber functor $\mathrm{Rep}(G) \to \mathrm{Vec}$. Hence for any object X in $\mathscr{C}_G$ and any object V in $\mathrm{Rep}(G)$ we have an automorphism of

$F(V)\otimes X$ defined as the composition

$$F(V)\otimes X \xrightarrow{\sim} X\otimes F(V) \xrightarrow{\sim} F(V)\otimes X, \tag{14}$$

where the first isomorphism comes from the fact that $F(V)\in \mathrm{Vec}$ and the second one is (8).

When X is simple we have an isomorphism $\mathrm{Aut}_{\mathscr{C}}(F(V)\otimes X)\cong \mathrm{Aut}_{\mathrm{Vec}}(F(V))$, hence we obtain a tensor automorphism i_X of $F|_{\mathrm{Rep}(G)}$. Since $\mathrm{Aut}_{\otimes}(F|_{\mathrm{Rep}(G)})\cong G$ we have an assignment $X\mapsto i_X\in G$. The hexagon axiom of braiding implies that this assignment is multiplicative, that is, that $i_Z = i_X i_Y$ for any simple object Z contained in $X\otimes Y$. Thus, it defines a G-grading on $\mathscr{C}$:

$$\mathscr{C}=\bigoplus_{g\in G}\mathscr{C}_g, \qquad \text{where } \mathbb{O}(\mathscr{C}_g)=\{X\in\mathbb{O}(\mathscr{C}) \mid i_X=g\}. \tag{15}$$

It is straightforward to check that $i_{T_g(X)} = ghg^{-1}$ whenever $i_X=h$.

Finally, to construct a G-crossed braiding on $\mathscr{C}$, observe that $\mathscr{C}$ and $\mathscr{C}^{\mathrm{rev}}$ are embedded into the crossed product category $\mathscr{C}\rtimes G=(\mathscr{C}^G)^*_{\mathscr{C}}$ as subcategories $\mathscr{C}_{\mathrm{left}}$ and $\mathscr{C}_{\mathrm{right}}$, consisting, respectively, of functors of left and right multiplications by objects of $\mathscr{C}$. Clearly, there is a natural family of isomorphisms

$$X\otimes Y \xrightarrow{\sim} Y\otimes X, \qquad \text{with } X\in\mathscr{C}_{\mathrm{left}} \text{ and } Y\in\mathscr{C}_{\mathrm{right}}, \tag{16}$$

satisfying obvious compatibility conditions. Note that $\mathscr{C}_{\mathrm{left}}$ is identified with the diagonal subcategory of $\mathscr{C}\rtimes G$ spanned by objects $X\boxtimes g$, $X\in\mathscr{C}_g$, $g\in G$, and $\mathscr{C}_{\mathrm{right}}$ is identified with the trivial component subcategory $\mathscr{C}\boxtimes e$. Using (6) we conclude that isomorphisms (16) give rise to a G-crossed braiding on $\mathscr{C}$.

One can check that the two constructions above (from braided fusion categories containing $\mathrm{Rep}(G)$ to braided G-crossed categories and vice versa) are inverses of each other; see [Kirillov 2002; Müger 2004; Drinfeld et al. 2009] for details.

Remark 2.13. Let $\mathscr{C}=\oplus_{g\in G}\mathscr{C}_g$ be a braided G-crossed fusion category. It was shown in [Drinfeld et al. 2009] that the braided category $\mathscr{C}^G$ is nondegenerate if and only if $\mathscr{C}_e$ is nondegenerate and the G-grading of $\mathscr{C}$ is faithful.

3. The center of a graded fusion category

Let G be a finite group and let $\mathscr{D}$ be a fusion category. Throughout this section $\mathscr{C}$ will denote a fusion category with a faithful G-grading, whose trivial component is $\mathscr{D}$; that is, $\mathscr{C}$ is a G-extension of $\mathscr{D}$:

$$\mathscr{C}=\bigoplus_{g\in G}\mathscr{C}_g, \qquad \mathscr{C}_e=\mathscr{D}. \tag{17}$$

In what follows we consider only *faithful* gradings: that is, those such that $\mathscr{C}_g \neq 0$ for all $g \in G$. An object of $\mathscr{C}$ contained in $\mathscr{C}_g$ will be called *homogeneous* of degree g.

Our goal is to describe the center $\mathscr{Z}(\mathscr{C})$ as an equivariantization of the relative center $\mathscr{Z}_{\mathscr{D}}(\mathscr{C})$ defined in Section 2B.

3A. ***The relative center $\mathscr{Z}_{\mathscr{D}}(\mathscr{C})$ as a braided G-crossed category.*** Let us define a canonical braided G-crossed category structure on $\mathscr{Z}_{\mathscr{D}}(\mathscr{C})$.

First of all, there is an obvious faithful G-grading on $\mathscr{Z}_{\mathscr{D}}(\mathscr{C})$:

$$\mathscr{Z}_{\mathscr{D}}(\mathscr{C}) = \bigoplus_{g\in G} \mathscr{Z}_{\mathscr{D}}(\mathscr{C}_g). \tag{18}$$

Indeed, it is clear that for every simple object X of $\mathscr{Z}_{\mathscr{D}}(\mathscr{C})$ the forgetful image of X in $\mathscr{C}$ must be homogeneous.

We now define the action of G on $\mathscr{Z}_{\mathscr{D}}(\mathscr{C})$. Take $g, h \in G$. Let $\mathrm{Fun}_{\mathscr{D}\boxtimes\mathscr{D}^{\mathrm{rev}}}(\mathscr{C}_g, \mathscr{C}_h)$ denote the category of $\mathscr{D}$-bimodule functors from $\mathscr{C}_g$ to $\mathscr{C}_h$. Clearly, it is a $\mathscr{Z}(\mathscr{D})$-bimodule category.

Proposition 3.1. *Let $g, h \in G$. The functors*

$$L_{g,h}:\mathscr{Z}_{\mathscr{D}}(\mathscr{C}_h) \xrightarrow{\sim} \mathrm{Fun}_{\mathscr{D}\boxtimes\mathscr{D}^{rev}}(\mathscr{C}_g, \mathscr{C}_{hg}) : Z \mapsto Z\otimes ?, \tag{19}$$

$$R_{g,h}:\mathscr{Z}_{\mathscr{D}}(\mathscr{C}_h) \xrightarrow{\sim} \mathrm{Fun}_{\mathscr{D}\boxtimes\mathscr{D}^{rev}}(\mathscr{C}_g, \mathscr{C}_{gh}) : Z \mapsto ?\otimes Z. \tag{20}$$

are equivalences of $\mathscr{Z}(\mathscr{D})$-bimodule categories.

Proof. We prove that (19) is an equivalence. Let $\mathrm{Fun}_{\mathscr{D}}(\mathscr{C}_g, \mathscr{C}_{hg})$ be the category of right $\mathscr{D}$-module functors from $\mathscr{C}_g$ to $\mathscr{C}_{hg}$. It suffices to prove that

$$M_{g,h} : \mathscr{C}_h \to \mathrm{Fun}_{\mathscr{D}}(\mathscr{C}_g, \mathscr{C}_{hg}) : X \mapsto X\otimes ? \tag{21}$$

is an equivalence. Indeed, $\mathscr{D}$-bimodule functor structures on $M_{g,h}(X)$ for $X \in \mathscr{C}_h$ are in bijection with central structures on X.

For every $g \in G$ choose a simple object $X_g \in \mathscr{C}_g$. Then $A_g := X_g \otimes X_g^*$ is an algebra in $\mathscr{D}$. It follows from [Ostrik 2003, Theorem 1] that the functor $Y \mapsto Y \otimes X_g^*$ is a left $\mathscr{C}$-module category equivalence between $\mathscr{C}$ and the category of right A_g-modules in $\mathscr{C}$. Since $Y \otimes X_g^*$ belongs to $\mathscr{D}$ if and only if Y is in $\mathscr{C}_g$ we see that the functor above restricts to a left $\mathscr{D}$-module category equivalence between $\mathscr{C}_g$ and the category of right A_g-modules in $\mathscr{D}$. There are also similar equivalences of right module categories.

It follows that for all $g, h \in G$ there is an equivalence

$$Y \mapsto X_g \otimes Y \otimes X_{hg}^* \tag{22}$$

between $\mathscr{C}$ and the category of $(A_g - A_{hg})$-bimodules in $\mathscr{C}$. The right-hand side of (22) belongs to $\mathscr{D}$ if and only if Y is in $\mathscr{C}_h$. Hence, (22) restricts to an equivalence between $\mathcal{C}_h$ and the category of $(A_g - A_{hg})$-bimodules in $\mathcal{D}$. The latter category is identified with the category of right $\mathcal{D}$-module functors between the categories of right A_g-modules and A_{hg}-modules in $\mathcal{D}$, that is, with $\mathrm{Fun}_{\mathcal{D}}(\mathcal{C}_g, \mathcal{C}_{hg})$. It is easy to see that upon this identification the restriction of equivalence (22) to $\mathcal{C}_h$ coincides with (21).

The proof of the equivalence (20) is completely similar. $\square$

We define tensor functors

$$T_{g,h} := L^{-1}_{g,ghg^{-1}} R_{g,h} : \mathcal{Z}_{\mathcal{D}}(\mathcal{C}_h) \to \mathcal{Z}_{\mathcal{D}}(\mathcal{C}_{ghg^{-1}}), \qquad g, h \in G, \tag{23}$$

and set

$$T_g := \bigoplus_{h \in G} T_{g,h} : \mathcal{Z}_{\mathcal{D}}(\mathcal{C}) \to \mathcal{Z}_{\mathcal{D}}(\mathcal{C}). \tag{24}$$

The definition of T_g along with Proposition 3.1 give rise to the following natural isomorphism of $\mathcal{D}$-bimodule functors from $\mathcal{C}_g$ to $\mathcal{C}$:

$$c_{-,Y} : ? \otimes Y \xrightarrow{\sim} T_g(Y) \otimes ?. \tag{25}$$

It translates to a natural family of isomorphisms

$$c_{X,Y} : X \otimes Y \xrightarrow{\sim} T_g(Y) \otimes X, \qquad X \in \mathcal{C}_g,\ Y \in \mathcal{Z}_{\mathcal{D}}(\mathcal{C}),\ g \in G, \tag{26}$$

satisfying natural compatibility conditions corresponding to the $\mathcal{D}$-bimodule structure on (25). Since the grading (18) is faithful, we have $T_g(\mathcal{Z}_{\mathcal{D}}(\mathcal{C}_h)) \subset \mathcal{Z}_{\mathcal{D}}(\mathcal{C}_{ghg^{-1}})$.

Take $X_1 \in \mathcal{C}_{g_1}$, $X_2 \in \mathcal{C}_{g_2}$ and set $X = X_1 \otimes X_2$ in (26). We obtain a natural isomorphism

$$T_{g_1} T_{g_2}(Y) \otimes X_1 \otimes X_2 \xrightarrow{\sim} T_{g_1 g_2}(Y) \otimes X_1 \otimes X_2. \tag{27}$$

Since every object $Z \in \mathcal{C}_{g_1 g_2}$ is contained in $X_1 \otimes X_2$ for some $X_1 \in \mathcal{C}_{g_1}$, $X_2 \in \mathcal{C}_{g_2}$, using naturality of (27) we obtain a natural isomorphism

$$T_{g_1} T_{g_2}(Y) \otimes Z \xrightarrow{\sim} T_{g_1 g_2}(Y) \otimes Z, \qquad Z \in \mathcal{C}_{g_1 g_2}, \tag{28}$$

of $\mathcal{D}$-bimodule functors $T_{g_1} T_{g_2}(Y) \otimes ?$ and $T_{g_1 g_2}(Y) \otimes ?$. By Proposition 3.1 this gives an isomorphism $T_{g_1} T_{g_2}(Y) \xrightarrow{\sim} T_{g_1 g_2}(Y)$, $Y \in \mathcal{Z}_{\mathcal{D}}(\mathcal{C})$, that is, an isomorphism of functors $T_{g_1} T_{g_2} \xrightarrow{\sim} T_{g_1 g_2}$. Thus, the assignment $g \mapsto T_g$ is an action of G on $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})$ by tensor autoequivalences.

Suppose that X is an object in $\mathcal{Z}(\mathcal{C}_g)$. Then both sides of (26) have structure of objects in $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})$ obtained by composing central structures of X and Y.

Lemma 3.2. *Isomorphisms (26) define a G-braiding on $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})$.*

Proof. That isomorphisms (26) are indeed morphisms in $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})$ follows from commutativity of the diagram

$$\begin{tikzcd}
X \otimes Y \otimes V \arrow[r, "\mathrm{id}_X \otimes \delta_V"] \arrow[d, "c_{X,Y} \otimes \mathrm{id}_V"'] & X \otimes V \otimes Y \arrow[r, "\gamma_V \otimes \mathrm{id}_Y"] \arrow[dl, "c_{X \otimes V, Y}"'] & V \otimes X \otimes Y \arrow[d, "\mathrm{id}_V \otimes c_{X,Y}"] \arrow[dl, "c_{V \otimes X, Y}"'] \\
T_g(Y) \otimes X \otimes V \arrow[r, "\mathrm{id}_{T_g(Y)} \otimes \gamma_V"'] & T_g(Y) \otimes V \otimes X \arrow[r, "T_g(\delta)_V \otimes \mathrm{id}_X"'] & V \otimes T_g(Y) \otimes X,
\end{tikzcd} \tag{29}$$

where $(X, \gamma) \in \mathcal{Z}_{\mathcal{D}}(\mathcal{C}_g)$, $(Y, \delta) \in \mathcal{Z}_{\mathcal{D}}(\mathcal{C})$, and $V \in \mathcal{D}$. Indeed, the parallelogram in the middle commutes by naturality of c, and the two triangular faces commute since the natural isomorphism (25) is an isomorphism of $\mathcal{D}$-bimodule functors.

It is straightforward to check that isomorphisms $c_{X,Y}$ satisfy the compatibility conditions of Definition 2.10. □

The constructions and arguments above prove the following theorem.

Theorem 3.3. *Let G be a finite group and let $\mathcal{C}$ be a fusion category with a faithful G-grading whose trivial component is $\mathcal{D}$. The relative center $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})$ has a canonical structure of a braided G-crossed category.*

Remark 3.4. In particular, to every G-extension of a fusion category $\mathcal{D}$ we assigned an action of G by braided autoequivalences of $\mathcal{Z}(\mathcal{D})$. This assignment is studied in detail in [Etingof et al. 2009].

3B. ***The center $\mathcal{Z}(\mathcal{C})$ as an equivariantization.*** As before, let G be a finite group and let $\mathcal{C}$ be a fusion category with a faithful G-grading (17). Let $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})$ be the braided G-crossed category constructed in Section 3A.

Theorem 3.5. *There is an equivalence of braided fusion categories*

$$\mathcal{Z}_{\mathcal{D}}(\mathcal{C})^G \xrightarrow{\sim} \mathcal{Z}(\mathcal{C}). \tag{30}$$

Proof. We see from (26) that a G-equivariant object in $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})$ has a structure of a central object in $\mathcal{C}$ defined as in (13). It follows from definitions that the corresponding tensor functor $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})^G \to \mathcal{Z}(\mathcal{C})$ is braided.

Conversely, given an object Y in $\mathcal{Z}(\mathcal{C})$, consider its forgetful image $\tilde{Y}$ in $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})$. Combining the central structure of Y with isomorphism (26) we obtain a family of isomorphisms

$$\tilde{Y} \otimes X \xrightarrow{\sim} T_g(\tilde{Y}) \otimes X, \qquad X \in \mathcal{C}_g,\ g \in G,$$

which gives rise to the isomorphism of $\mathcal{D}$-bimodule functors $\tilde{Y} \otimes ? \xrightarrow{\sim} T_g(\tilde{Y}) \otimes ? : \mathcal{C}_g \to \mathcal{C}$. By Proposition 3.1 we obtain a natural isomorphism $\tilde{Y} \xrightarrow{\sim} T_g(\tilde{Y})$ and, hence, a G-equivariant structure on $\tilde{Y}$. Thus, we have a tensor functor $\mathcal{Z}(\mathcal{C}) \to \mathcal{Z}_{\mathcal{D}}(\mathcal{C})^G$. It is clear that the two functors are quasiinverses of each other. □

We describe the Tannakian subcategory $\mathcal{E} \cong \operatorname{Rep}(G) \subset \mathcal{Z}(\mathcal{C})$ corresponding to equivalence (30). For any representation $\pi : G \to GL(V)$ of the grading group G, consider an object I_π in $\mathcal{Z}(\mathcal{C})$ where $I_\pi = V \otimes \mathbf{1}$ as an object of $\mathcal{C}$ with the permutation isomorphism

$$c_{I_\pi, X} := \pi(g) \otimes \mathrm{id}_X : I_\pi \otimes X \cong X \otimes I_\pi, \qquad \text{when } X \in \mathcal{C}_g. \tag{31}$$

Then $\mathcal{E}$ is the subcategory of $\mathcal{Z}(\mathcal{C})$ consisting of objects I_π, where π runs through all finite-dimensional representations of G.

Remark 3.6. Here is another description of the subcategory $\mathcal{E}$: it consists of all objects in $\mathcal{Z}(\mathcal{C})$ sent to Vec by the forgetful functor $\mathcal{Z}(\mathcal{C}) \to \mathcal{Z}_{\mathcal{D}}(\mathcal{C})$.

Corollary 3.7. *Let $\mathcal{C}$ be a faithfully G-graded fusion category with the trivial component $\mathcal{D}$. Let $\mathcal{E} = \operatorname{Rep}(G) \subset \mathcal{Z}(\mathcal{C})$ be the Tannakian subcategory constructed above. Then the de-equivariantization category $(\mathcal{E}')_G$ is braided tensor equivalent to $\mathcal{Z}(\mathcal{D})$.*

Proof. The statement follows from Theorem 3.5 since $(\mathcal{E}')_G$ is the trivial component of the grading of $\mathcal{Z}(\mathcal{C})_G = \mathcal{Z}_{\mathcal{D}}(\mathcal{C})$. □

Remark 3.8. The assignment above

$$\{G\text{-extensions of } \mathcal{D}\} \mapsto \{\text{braided } G\text{-crossed extensions of } \mathcal{Z}(\mathcal{D})\} \tag{32}$$

can be thought of as an analogue of the center construction for G-extensions.

Next, we describe simple objects of $\mathcal{Z}(\mathcal{C})$. For any conjugacy class K in G fix a representative $a_K \in K$. Let G_K denote the centralizer of a_K in G. Note that the action (24) of G on $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})$ restricts to the action of G_K on $\mathcal{Z}_{\mathcal{D}}(\mathcal{C}_{a_K})$.

Proposition 3.9. *There is a bijection between the set of isomorphism classes of simple objects of $\mathcal{Z}(\mathcal{C})$ and pairs (K, X), where K is a conjugacy class of G and X is a simple G_K-equivariant object of $\mathcal{Z}_{\mathcal{D}}(\mathcal{C}_{a_K})$.*

Proof. By Theorem 3.5 we have $\mathcal{Z}(\mathcal{C}) \simeq \mathcal{Z}_{\mathcal{D}}(\mathcal{C})^G$, so the stated parameterization is immediate from the description of simple objects of the equivariantization category given in Proposition 2.7. □

3C. ***A criterion for a graded fusion category to be group-theoretical.*** We have seen in Corollary 3.7 that $\mathcal{Z}(\mathcal{C})$ contains a Tannakian subcategory $\mathcal{E} = \operatorname{Rep}(G)$ such that the de-equivariantization $(\mathcal{E}')_G$ is braided equivalent to $\mathcal{Z}(\mathcal{D})$, where $\mathcal{D}$ is the trivial component of $\mathcal{C}$. Furthermore, by Remark 2.11, there is a canonical action of G on $\mathcal{Z}(\mathcal{D})$, by braided autoequivalences. By [Drinfeld et al. 2009], Tannakian subcategories of $\mathcal{Z}(\mathcal{C})$ containing $\mathcal{E}$ bijectively correspond to G-stable Tannakian subcategories of $(\mathcal{E}')_G \simeq \mathcal{Z}(\mathcal{D})$. Combining this observation with Theorem 2.5(ii) we obtain the following criterion.

Corollary 3.10. *A graded fusion category $\mathscr{C} = \bigoplus_{g\in G} \mathscr{C}_g$, $\mathscr{C}_e = \mathscr{D}$, is group-theoretical if and only if $\mathscr{Z}(\mathscr{D})$ contains a G-stable Lagrangian subcategory.*

Corollary 3.10 will be useful in Section 4D, where we characterize group-theoretical Tambara–Yamagami categories.

We can specialize Corollary 3.10 to equivariantization categories. Let G be a finite group acting on a fusion category $\mathscr{C}$. The equivariantization $\mathscr{C}^G$ is Morita equivalent to the crossed product category $\mathscr{C} \rtimes G$ (see Section 2D). Therefore, $\mathscr{Z}(\mathscr{C}^G) \cong \mathscr{Z}(\mathscr{C} \rtimes G)$. Clearly, the trivial component of $\mathscr{Z}(\mathscr{C} \rtimes G)_G$ is $\mathscr{Z}(\mathscr{C})$ and the canonical action of G on $\mathscr{Z}(\mathscr{C})$ is induced from the action of G on $\mathscr{C}$ in an obvious way.

Corollary 3.11. *The equivariantization $\mathscr{C}^G$ is group-theoretical if and only if there exists a G-stable Lagrangian subcategory of $\mathscr{Z}(\mathscr{C})$.*

Remark 3.12. Let G act on $\mathscr{C}$ as before. One can check (independently from the results of this section) that the G-set of Lagrangian subcategories of $\mathscr{Z}(\mathscr{C})$ is isomorphic to the G-set consisting of indecomposable $\mathscr{C}$-module categories $\mathscr{M}$ such that the dual category $\mathscr{C}^*_{\mathscr{M}}$ is pointed. This isomorphism is given by the map constructed in [Naidu and Nikshych 2008, Theorem 4.17]. Thus, the criterion in Corollary 3.11 is the same as [Nikshych 2008, Corollary 3.6].

3D. ***Example: the relative center of a crossed product category.*** Let G be a finite group and let $g \mapsto T_g$, $g \in G$, be an action of G on a fusion category $\mathscr{D}$. Let $\mathscr{C} := \mathscr{D} \rtimes G$ be the crossed product category defined in Section 2D. It has a natural grading

$$\mathscr{C} = \bigoplus_{g\in G} \mathscr{C}_g, \qquad \text{where } \mathscr{C}_g = \{Y \boxtimes g \mid Y \in \mathscr{D}\}.$$

We describe the braided G-crossed fusion category structure on the relative center

$$\mathscr{Z}_{\mathscr{D}}(\mathscr{C}) = \bigoplus_{g\in G} \mathscr{Z}_{\mathscr{D}}(\mathscr{C}_g).$$

By definition, the objects of $\mathscr{Z}_{\mathscr{D}}(\mathscr{C}_g)$ are pairs $(Y \boxtimes g, \gamma)$, where $Y \in \mathscr{D}$ and

$$\gamma = \{\gamma_X : X \otimes Y \xrightarrow{\sim} Y \otimes T_g(X)\}_{X\in\mathscr{D}} \tag{33}$$

is a natural family of isomorphisms satisfying natural compatibility conditions. Thus, $\mathscr{Z}_{\mathscr{D}}(\mathscr{C}_g)$ can be viewed as a "deformation" of $\mathscr{Z}(\mathscr{D})$ by means of T_g.

The action of G on $\mathscr{D}$ induces an action $h \mapsto \tilde{T}_h$ on $\mathscr{Z}_{\mathscr{D}}(\mathscr{C})$ defined as follows. Applying T_h, $h \in G$, to $\gamma_{T_{h^{-1}}(X)}$ in (33), we obtain an isomorphism

$$\tilde{\gamma}_X : X \otimes T_h(Y) \xrightarrow{\sim} T_h(Y) \otimes T_{hgh^{-1}}(X). \tag{34}$$

Set $\tilde{T}_h(Y \boxtimes g, \gamma) := (T_h(Y) \boxtimes hgh^{-1}, \tilde{\gamma})$. Thus, $\tilde{T}_h$ maps $\mathscr{Z}_{\mathscr{D}}(\mathscr{C}_g)$ to $\mathscr{Z}_{\mathscr{D}}(\mathscr{C}_{hgh^{-1}})$.

Finally, the G-braiding between objects $(X\boxtimes h)\in\mathcal{Z}_{\mathcal{D}}(\mathcal{C}_h)$ and $(Y\boxtimes g)\in\mathcal{Z}_{\mathcal{D}}(\mathcal{C}_g)$ comes from the isomorphism

$$\begin{aligned}(X\boxtimes h)\otimes(Y\boxtimes g)=(X\otimes T_h(Y))\boxtimes hg \xrightarrow{\tilde{\gamma}} (T_h(Y)\otimes T_{hgh^{-1}}(X))\boxtimes hg&\\ =(T_h(Y)\boxtimes hgh^{-1})\otimes(X\boxtimes h)&\\ =\tilde{T}_h(Y\boxtimes g)\otimes(X\boxtimes h).&\end{aligned}$$

By Theorem 3.5, the category $\mathcal{Z}(\mathcal{D}\rtimes G)\cong\mathcal{Z}(\mathcal{D}^G)$ is equivalent to the equivariantization of the braided G-crossed category above.

4. The centers of Tambara–Yamagami categories

Our goal in this section is to apply techniques developed in Section 3 to Tambara–Yamagami categories introduced in [Tambara and Yamagami 1998] (see Section 4A below for the definition). Namely, using the techniques in Section 3 we establish a criterion for a Tambara–Yamagami category to be group-theoretical. We then use this criterion together with Corollary 3.11 to produce a series of non-group-theoretical semisimple Hopf algebras. In this section we assume that our ground field k is the field of complex numbers $\mathbb{C}$. We begin by recalling the definition of a Tambara–Yamagami category.

4A. ***Definition of Tambara–Yamagami categories.*** Let $\mathbb{Z}_2=\langle\,\delta\mid\delta^2=1\,\rangle$ be the cyclic group of order 2.

Tambara and Yamagami [1998] completely classified all $\mathbb{Z}_2$-graded fusion categories in which all but one simple objects are invertible and the noninvertible simple object has nontrivial graded degree.

They showed that any such category $\mathcal{TY}(A,\chi,\tau)$ is determined, up to an equivalence, by a finite abelian group A, a nondegenerate symmetric bilinear form $\chi : A\times A\to k^{\times}$, and a square root $\tau\in k$ of $|A|^{-1}$. The category $\mathcal{TY}(A,\chi,\tau)$ is described as follows. It is a skeletal category (that is, such that any two isomorphic objects are equal) with simple objects $\{a\mid a\in A\}$ and m, and tensor product

$$a\otimes b=a+b,\quad a\otimes m=m,\quad m\otimes a=m,\quad m\otimes m=\bigoplus_{a\in A}a,$$

for all $a,b\in A$, and the unit object $0\in A$. The associativity constraints are given by

$$\alpha_{a,b,c}=\mathrm{id}_{a+b+c},\quad \alpha_{a,b,m}=\mathrm{id}_m,\quad \alpha_{a,m,b}=\chi(a,b)\ \mathrm{id}_m,\quad \alpha_{m,a,b}=\mathrm{id}_m,$$

$$\alpha_{a,m,m}=\bigoplus_{b\in A}\mathrm{id}_b,\quad \alpha_{m,a,m}=\bigoplus_{b\in A}\chi(a,b)\ \mathrm{id}_b,$$

$$\alpha_{m,m,a}=\bigoplus_{b\in A}\mathrm{id}_b,\quad \alpha_{m,m,m}=\bigoplus_{a,b\in A}\tau\,\chi(a,b)^{-1}\ \mathrm{id}_m\,.$$

The unit constraints are the identity maps. The category $\mathcal{TY}(A,\chi,\tau)$ is rigid with $a^*=-a$ and $m^*=m$ (with obvious evaluation and coevaluation maps).

Let $n:=|A|$. The dimensions of simple objects of $\mathcal{TY}(A,\chi,\tau)$ are $\mathrm{FPdim}(a)=1$, $a\in A$, and $\mathrm{FPdim}(m)=\sqrt{n}$. We have $\mathrm{FPdim}(\mathcal{TY}(A,\chi,\tau))=2n$.

The $\mathbb{Z}_2$-grading on $\mathcal{TY}(A,\chi,\tau)$ is

$$\mathcal{TY}(A,\chi,\tau)=\mathcal{TY}(A,\chi,\tau)_1\oplus\mathcal{TY}(A,\chi,\tau)_\delta,$$

where $\mathcal{TY}(A,\chi,\tau)_1$ is the full fusion subcategory generated by the invertible objects $a\in A$ and $\mathcal{TY}(A,\chi,\tau)_\delta$ is the full abelian subcategory generated by the object m.

Let $\mathcal{C}:=\mathcal{TY}(A,\chi,\tau)$ and $\mathcal{D}:=\mathcal{TY}(A,\chi,\tau)_1$.

4B. ***Braided $\mathbb{Z}_2$-crossed category $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})$.*** First, let us describe the simple objects of $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})=\mathcal{Z}(\mathcal{C}_1)\oplus\mathcal{Z}_{\mathcal{D}}(\mathcal{C}_\delta)$. Let $\widehat{A}:=\mathrm{Hom}(A,k^\times)$. Clearly, $\mathcal{Z}(\mathcal{C}_1)=\mathcal{Z}(\mathrm{Vec}_A)$, so its simple objects are parameterized by $(a,\phi)\in A\times\widehat{A}$. The object $X_{(a,\phi)}$ corresponding to such a pair is equal to a as an object of $\mathcal{C}$ and its central structure is given by

$$\phi(x)\,\mathrm{id}_{a+x}:x\otimes X_{(a,\phi)}\xrightarrow{\sim}X_{(a,\phi)}\otimes x. \tag{35}$$

Using Definition 2.1 we see that simple objects of $\mathcal{Z}_{\mathcal{D}}(\mathcal{C}_\delta)$ are parameterized by functions $\rho:A\to k^\times$ satisfying

$$\rho(a+b)=\chi(a,b)^{-1}\rho(a)\rho(b),\quad a,b\in A \tag{36}$$

(clearly, such functions form a torsor over $\widehat{A}$). The corresponding object Z_ρ is equal to m as an object of $\mathcal{C}$ and has the relative central structure

$$\rho(x)\,\mathrm{id}_m:x\otimes Z_\rho\xrightarrow{\sim}Z_\rho\otimes x,\quad x\in A. \tag{37}$$

Let $A\to\widehat{A}:a\mapsto\widehat{a}$ be the homomorphism defined by $\widehat{a}(x)=\chi(x,\,a)$. Similarly, let $\widehat{A}\to A:\phi\mapsto\widehat{\phi}$ be the homomorphism defined by $\phi(x)=\chi(x,\,\widehat{\phi})$ (recall that χ is nondegenerate). Clearly, these two maps are inverses of each other.

The fusion rules of $\mathcal{Z}_{\mathcal{D}}(\mathcal{C})$ are computed using formula (3) :

$$\begin{aligned}X_{(a,\phi)}\otimes X_{(b,\psi)}&=X_{(a+b,\phi+\psi)},\\ X_{(a,\phi)}\otimes Z_\rho&=Z_{\rho\phi(-\widehat{a})},\\ Z_\rho\otimes X_{(a,\phi)}&=Z_{\rho\phi(-\widehat{a})},\\ Z_{\rho'}\otimes Z_\rho&=\bigoplus_{a\in A}X_{(a,\widehat{a}\rho'/\overline{\rho})}.\end{aligned}$$

We have $X^*_{(a,\phi)}=X_{(-a,-\phi)}$ and $Z^*_\rho=Z_{\overline{\rho}}$, where $\overline{\rho}(x)=\rho(-x),\ x\in A$.

Using the construction given in Section 3A we see that the action of $\mathbb{Z}_2$ on $\mathfrak{X}_{\mathfrak{D}}(\mathscr{C})$ is given by

$$T_1 = \mathrm{id}_{\mathfrak{X}_{\mathfrak{D}}(\mathscr{C})}; \qquad T_\delta(X_{(a,\phi)}) = X_{(-\widehat{\phi},\, -\widehat{a})}, \quad T_\delta(Z_\rho) = Z_\rho. \tag{38}$$

The monoidal functor structure on $\mathbb{Z}_2 \to \mathrm{Aut}_\otimes(\mathfrak{X}_{\mathfrak{D}}(\mathscr{C}))$ is given by the natural isomorphism $\gamma := \gamma_{\delta,\delta} : T_\delta \circ T_\delta \xrightarrow{\sim} T_1$ defined by

$$\gamma_{X_{(a,\phi)}} = \phi(a)\,\mathrm{id}_{X_{(a,\phi)}}, \quad \gamma_{Z_\rho} = \Bigl(\tau \sum_{x\in A} \rho(x)^{-1}\Bigr)\,\mathrm{id}_{Z_\rho}\,.$$

The crossed braiding morphisms on $\mathfrak{X}_{\mathfrak{D}}(\mathscr{C})$ are given by

$$\begin{aligned}
c_{X_{(a,\phi)},X_{(b,\psi)}} &= \psi(a)\,\mathrm{id}_{a+b} : X_{(a,\phi)} \otimes X_{(b,\psi)} \xrightarrow{\sim} X_{(b,\psi)} \otimes X_{(a,\phi)},\\
c_{X_{(a,\phi)},Z_\rho} &= \rho(a)\,\mathrm{id}_m : X_{(a,\phi)} \otimes Z_\rho \xrightarrow{\sim} Z_\rho \otimes X_{(a,\phi)},\\
c_{Z_\rho,X_{(a,\phi)}} &= \mathrm{id}_m : Z_\rho \otimes X_{(a,\phi)} \xrightarrow{\sim} X_{(-\widehat{\phi},\,-\widehat{a})} \otimes Z_\rho,\\
c_{Z_{\rho'},Z_\rho} &= \oplus_{a\in A}\, \rho(-a)^{-1}\,\mathrm{id}_a : Z_{\rho'} \otimes Z_\rho \xrightarrow{\sim} Z_\rho \otimes Z_{\rho'}.
\end{aligned}$$

4C. ***The equivariantization category*** $\mathfrak{X}_{\mathfrak{D}}(\mathscr{C})^{\mathbb{Z}_2}$**.** A simple calculation of $\mathbb{Z}_2$-equivariant objects in $\mathfrak{X}_{\mathfrak{D}}(\mathscr{C})$ establishes the following.

Proposition 4.1. *The following is a complete list of simple objects of* $\mathfrak{X}_{\mathfrak{D}}(\mathscr{C})^{\mathbb{Z}_2} \cong \mathfrak{X}(\mathscr{T}\mathscr{Y}(A,\chi,\tau))$ *up to an isomorphism*:

(1) $2n$ *invertible objects parameterized by pairs* (a,ϵ), *where* $a \in A$ *and* $\epsilon^2 = \chi(a,a)^{-1}$. *The corresponding object* $X_{a,\epsilon}$ *is equal to* $X_{(a,-\widehat{a})}$ *as an object of* $\mathfrak{X}_{\mathfrak{D}}(\mathscr{C})$ *and has* $\mathbb{Z}_2$*-equivariant structure*

$$\epsilon\,\mathrm{id}_{X_{(a,-\widehat{a})}} : T_\delta(X_{(a,-\widehat{a})}) \xrightarrow{\sim} X_{(a,-\widehat{a})};$$

(2) $\frac{n(n-1)}{2}$ *two-dimensional objects parameterized by unordered pairs* $(a,\, b)$ *of distinct objects in* A. *The corresponding object* $Y_{a,b}$ *is equal to* $X_{(a,-\widehat{b})} \oplus X_{(b,-\widehat{a})}$ *as an object of* $\mathfrak{X}_{\mathfrak{D}}(\mathscr{C})$ *and has* $\mathbb{Z}_2$*-equivariant structure*

$$\bigl(\mathrm{id}_{X_{(a,-\widehat{b})}} \oplus \chi(a,b)^{-1}\,\mathrm{id}_{X_{(b,-\widehat{a})}}\bigr) : T_\delta(X_{(a,-\widehat{b})} \oplus X_{(b,-\widehat{a})}) \xrightarrow{\sim} X_{(a,-\widehat{b})} \oplus X_{(b,-\widehat{a})};$$

(3) $2n$ $\sqrt{n}$*-dimensional objects parameterized by pairs* $(\rho,\,\Delta)$, *where* $\rho : A \to k^\times$ *satisfies* (36) *and* $\Delta^2 = \tau \sum_{x\in A} \rho(x)^{-1}$. *The corresponding object* $Z_{\rho,\Delta}$ *is equal to* Z_ρ *as an object of* $\mathfrak{X}_{\mathfrak{D}}(\mathscr{C})$ *and has* $\mathbb{Z}_2$*-equivariant structure*

$$\Delta\,\mathrm{id}_{Z_\rho} : T_\delta(Z_\rho) \xrightarrow{\sim} Z_\rho.$$

Recall from [Etingof et al. 2005] that in a braided fusion category of an integer Frobenius–Perron dimension there is a canonical choice of a twist θ such that the categorical dimensions of objects coincide with their Frobenius–Perron dimensions. Namely, for any simple object X the scalar θ_X is defined in such a way that the composition

$$\mathbf{1} \xrightarrow{\mathrm{coev}_X} X \otimes X^* \xrightarrow{\theta_X c_{X,X^*}} X^* \otimes X \xrightarrow{\mathrm{ev}_X} \mathbf{1} \tag{39}$$

is equal to $\mathrm{FPdim}(X)\,\mathrm{id}_X$.

Let θ be the canonical twist on $\mathcal{Z}(\mathcal{C})$. Using the previous observation, explicit formulas from Section 4B, and Section 2F, we immediately obtain the following.

$$\theta_{X_{a,\epsilon}} = \chi(a,a)^{-1}, \quad \theta_{Y_{a,b}} = \chi(a,b)^{-1}, \quad \theta_{Z_{\rho,\Delta}} = \Delta.$$

Using the fusion rules of $\mathcal{Z}(\mathcal{C})$ (which may be computed using the explicit formulas in Section 4B), values of the twists above, and the well known formula

$$S_{X,Y} = \theta_X^{-1}\theta_Y^{-1} \sum_Z N_{X,Y}^Z \theta_Z d_Z, \tag{40}$$

we obtain the S- and T-matrices of $\mathcal{Z}(\mathcal{C})$:

$$\begin{aligned}
&S_{X_{a,\epsilon},X_{a',\epsilon'}} = \chi(a,a')^2, \quad S_{X_{a,\epsilon},Y_{b,c}} = 2\chi(a,b+c),\\
&S_{X_{a,\epsilon},Z_{\rho,\Delta}} = \epsilon\sqrt{n}\rho(a), \quad S_{Y_{a,b},Y_{c,d}} = 2\left(\chi(a,d)\chi(b,c) + \chi(a,c)\chi(b,d)\right),\\
&S_{Y_{a,b},Z_{\rho,\Delta}} = 0, \qquad S_{Z_{\rho,\Delta},Z_{\rho',\Delta'}} = \frac{1}{\Delta\Delta'}\sum_{a\in A}\chi(a,a)^2\rho(a)\rho'(a);\\
&T_{X_{a,\epsilon}} = \chi(a,a)^{-1}, \quad T_{Y_{a,b}} = \chi(a,b)^{-1}, \quad T_{Z_{\rho,\Delta}} = \Delta.
\end{aligned}$$

Proposition 4.2. *The maximal pointed subcategory of $\mathcal{Z}(\mathcal{C})$ is nondegenerate if and only if $|A|$ is odd.*

Proof. Let $a \in A$ be an element of order 2. Then $X_{a,\epsilon}$ centralizes every invertible object of $\mathcal{Z}(\mathcal{C})$. □

Remark 4.3. We note that simple objects and the S- and T-matrices of $\mathcal{Z}(\mathcal{C})$ were described in [Izumi 2001] using very different methods.

4D. ***A criterion for a Tambara–Yamagami category to be group-theoretical.*** The group $A\times\widehat{A}$ is equipped with a canonical nondegenerate quadratic form $q : A\times\widehat{A} \to k^\times$ given by

$$q((a,\phi)) := \phi(a), \quad (a,\phi) \in A\times\widehat{A}.$$

We will call a subgroup $B \subset A\times\widehat{A}$ *Lagrangian* if $q|_B = 1$ and $B = B^\perp$ with respect to the bilinear form defined by q. Lagrangian subgroups of $A\times\widehat{A}$ correspond to Lagrangian subcategories of $\mathcal{Z}(\mathrm{Vec}_A) \cong \mathrm{Vec}_{A\times\widehat{A}}$.

The braided tensor autoequivalence T_δ of $\mathcal{Z}(\mathrm{Vec}_A)$ defined in Section 4B determines an order 2 automorphism of $A\times\widehat{A}$, which we denote simply by δ:

$$\delta((a,\phi)) = (-\widehat{\phi}, -\widehat{a}), \quad (a,\phi) \in A\times\widehat{A}. \tag{41}$$

Definition 4.4. We will say that a subgroup $L \subset A$ is *Lagrangian (with respect to χ)* if $L = L^{\perp}$ with respect to the inner product on A given by χ. Equivalently, $|L|^2 = |A|$ and $\chi|_L = 1$.

Lemma 4.5. *Let A be an abelian 2-group such that $|A| = 2^{2n}$ and let χ be a nondegenerate symmetric bilinear form on A. Then A contains a Lagrangian subgroup.*

Proof. It suffices to show that A contains an isotropic element, that is, an element $x \in A$, $x \neq 0$, such that $\chi(x, x) = 1$. Then one can pass from A to $\langle x\rangle^{\perp}/\langle x\rangle$ and use induction.

Suppose that A is cyclic with a generator a. Then $2^{2n}a = 0$ and $\chi(a, a)$ is a (2^{2n})th root of unity, hence $\chi(2^n a, 2^n a) = \chi(a, a)^{2^{2n}} = 1$.

If A is not cyclic then it contains a subgroup $A_0 = \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$. Let x_1, x_2 be distinct nonzero elements of A_0. Suppose $\chi(x_i, x_i) \neq 1$, $i = 1, 2$. Then $\chi(x_i, x_i) = -1$ and $\chi(x_1 + x_2, x_1 + x_2) = 1$, as desired. □

Theorem 4.6. *Let $\mathscr{C} = \mathscr{T}\mathscr{Y}(A, \chi, \tau)$ be a Tambara–Yamagami fusion category. Then $\mathscr{C}$ is group-theoretical if and only if A contains a Lagrangian subgroup* (*with respect to χ*).

Proof. By Corollary 3.10, $\mathscr{C}$ is group-theoretical if and only if $\mathscr{Z}(\mathscr{D})$ contains a T_δ-stable Lagrangian subcategory. Equivalently, $\mathscr{C}$ is group-theoretical if and only if $A \times \widehat{A}$ contains a Lagrangian subgroup B stable under the action

$$(a, \phi) \mapsto (\widehat{\phi}, \widehat{a}). \tag{42}$$

This condition on B is the same as being stable under the action of δ from (41).

Let L be a Lagrangian (with respect to χ) subgroup of A and let $\widehat{L} := \{\widehat{a} \mid a \in L\}$. Then $L \times \widehat{L}$ is a Lagrangian subgroup of $A \times \widehat{A}$ stable under (42). Hence $\mathscr{C}$ is group-theoretical.

Conversely, suppose that $\mathscr{C}$ is group-theoretical. Let us write $A = A_{\text{even}} \oplus A_{\text{odd}}$, where A_{even} is the Sylow 2-subgroup of A and A_{odd} is the maximal odd order subgroup of A. Since $|A|$ must be a square, we conclude that $|A_{\text{even}}|$ is a square, and so A_{even} contains a Lagrangian subgroup with respect to $\chi|_{A_{\text{even}}}$ by Lemma 4.5.

So it remains to show that A_{odd} contains a Lagrangian subgroup with respect to $\chi|_{A_{\text{odd}}}$. For this end we may assume that $|A|$ is odd. Let $B \subset A \times \widehat{A}$ be a Lagrangian subgroup stable under (42). Then $B = B_+ \oplus B_-$, where

$$B_{\pm} := \{(a, \pm\widehat{a}) \mid (a, \pm\widehat{a}) \in B\}.$$

Let $L_{\pm} = B_{\pm} \cap (A \times \{1\})$. Then $|L_+||L_-| = |A|$, and $\chi|_{L_{\pm}} = 1$. Hence, $L_{\pm}$ are Lagrangian subgroups of A. □

Remark 4.7. It was observed in [Etingof et al. 2005, Remark 8.48] that for an odd prime p and elliptic bicharacter χ on $A = (\mathbb{Z}/p\mathbb{Z})^2$, the category $\mathscr{T}\mathscr{Y}((\mathbb{Z}/p\mathbb{Z})^2, \chi, \tau)$ is not group-theoretical. The criterion from Theorem 4.6 extends this observation.

4E. ***A series of non-group-theoretical semisimple Hopf algebras obtained from Tambara–Yamagami categories.*** Here we apply Corollary 3.11 to produce a series of non-group-theoretical fusion categories admitting fiber functors (that is, representation categories of non-group-theoretical semisimple Hopf algebras), generalizing examples constructed in [Nikshych 2008]. We refer the reader to [Montgomery 1993] as a reference on Hopf algebra theory.

Let A be a finite abelian group with a nondegenerate bilinear form χ. Let $\mathrm{Aut}(A, \chi)$ denote the group of automorphisms of A preserving χ.

The following proposition was proved in [Nikshych 2008, Proposition 2.10].

Proposition 4.8. *There is an action of* $\mathrm{Aut}(A, \chi)$ *on* $\mathcal{TY}(A, \chi, \tau)$ *given by* $g \mapsto T_g$, *where*

$$T_g(A) = g(a), \quad T_g(m) = m, \qquad a \in A,\ g \in \mathrm{Aut}(A, \chi),$$

with the tensor structure of T_g *given by identity morphisms.*

Corollary 4.9. *Let* G *be a subgroup of* $\mathrm{Aut}(A, \chi)$. *Then the fusion category* $\mathcal{TY}(A, \chi, \tau)^G$ *is group-theoretical if and only if there is a Lagrangian subgroup of* $(A,\ \chi)$ *stable under the action of* G.

Proof. Combine Corollary 3.11 and Theorem 4.6. □

We will say that a nondegenerate symmetric bilinear form $\chi : A \times A \to k^\times$ is *hyperbolic* if there are Lagrangian subgroups $L,\ L' \subset A$ such that $A = L \oplus L'$. In this case L' is isomorphic to the group $\widehat{L} = \mathrm{Hom}(L,\ k^\times)$ of characters of L and χ is identified with the canonical bilinear form on $L \oplus \widehat{L}$.

It was demonstrated in Tambara [2000] that when $n = |A|$ is odd the category $\mathcal{TY}(A, \chi, \tau)$ admits a fiber functor (that is, $\mathcal{TY}(A, \chi, \tau)$ is equivalent to the representation category of a semisimple Hopf algebra) if and only if τ^{-1} is a positive integer and χ is hyperbolic.

Corollary 4.10. *Let* p *be an odd prime, let* $L = (\mathbb{Z}/p\mathbb{Z})^N$, $N \geq 1$, *let* $A = L \oplus \widehat{L}$, *and let* $\chi : A \times A \to k^\times$ *be the canonical bilinear form defined by*

$$\chi((a,\ \phi),\ (b,\ \psi)) = \psi(a)\phi(b), \quad a, b \in A,\ \phi, \psi \in \widehat{A}.$$

Suppose that G *is a subgroup of* $\mathrm{Aut}(A,\ \chi)$ *not contained in any conjugate of* $\mathrm{Aut}(L) \subset \mathrm{Aut}(A,\ \chi)$. *Then the equivariantization category* $\mathcal{TY}(A,\ \chi,\ p^{-N})^G$ *is a non-group-theoretical fusion category equivalent to the representation category of a semisimple Hopf algebra of dimension* $2p^{2N}|G|$.

Proof. Note that $\mathrm{Aut}(A,\ \chi)$ acts transitively on the set of Lagrangian subgroups of $(A,\ \chi)$ and the stabilizer of L is $\mathrm{Aut}(L)$. Apply Corollary 4.9. □

Remark 4.11. The series of fusion categories in Corollary 4.10 extends the one constructed in [Nikshych 2008], where the case of $N = 1$ and $G = \mathbb{Z}/2\mathbb{Z}$ was considered.

5. Examples of modular categories arising from quadratic forms

As before, let $\mathscr{C} := \mathscr{T}\mathscr{Y}(A, \chi, \tau)$ be a Tambara–Yamagami category and let $\mathscr{D} := \mathscr{T}\mathscr{Y}(A, \chi, \tau)_1$ be the trivial component of $\mathbb{Z}_2$-grading of $\mathscr{T}\mathscr{Y}(A, \chi, \tau)$. In this section we assume that our ground field k is the field of complex numbers $\mathbb{C}$.

Suppose that the symmetric bicharacter $\chi : A \times A \to k^\times$ comes from a quadratic form on A, that is, there is a function $q : A \to k^\times$ such that

$$q(a+b) = q(a)q(b)\chi(a,b), \quad a, b \in A \quad \text{and} \quad q(-a) = q(a).$$

From the description obtained in Section 4B we observe that $\mathscr{Z}_{\mathscr{D}}(\mathscr{C})$ contains a fusion subcategory spanned by the simple objects $X_{(a,\widehat{a})}$, $a \in A$, and $Z_{q^{-1}}$. It is clear from the Tambara–Yamagami classification in Section 4A that this category is equivalent to $\mathscr{C}$.

Proposition 5.1. *Suppose that the symmetric bicharacter χ comes from a quadratic form on A. Then $\mathscr{C}$ admits a $\mathbb{Z}_2$-crossed braided category structure. The equivariantization $\mathscr{C}^{\mathbb{Z}_2}$ is nondegenerate if and only if $|A|$ is odd.*

Proof. Clearly, $\mathscr{C}$ inherits the $\mathbb{Z}_2$-crossed braided category structure from $\mathscr{Z}_{\mathscr{D}}(\mathscr{C})$. The nondegeneracy claim follows from Proposition 4.2 and Remark 2.13. □

Let us assume that $n := |A|$ is odd. Then χ corresponds to a unique quadratic form q. Let $\mathscr{E}(q, \pm) := \mathscr{C}^{\mathbb{Z}_2}$ be the modular category constructed in Proposition 5.1 (the $\pm$ corresponding to $\tau = \pm\frac{1}{\sqrt{n}}$, respectively). In what follows we describe the fusion rules and S- and T-matrices of $\mathscr{E}(q, \pm)$.

5A. *Fusion rules of* $\mathscr{E}$.

Clearly, $\mathscr{E}(q, \pm)$ is a fusion category of dimension $4n$. It has the following simple objects:

- two invertible objects, $\mathbf{1} = X_+$ and X_-;
- $\frac{n-1}{2}$ two-dimensional objects Y_a, $a \in A - \{0\}$ (with $Y_{-a} = Y_a$); and
- two $\sqrt{n}$-dimensional objects Z_l, $l \in \mathbb{Z}/2\mathbb{Z}$.

Here we simplify the notation used in Section 4C and define

$$X_\pm := X_{0,\pm 1}, \quad Y_a := Y_{a,-a}, \quad Z_l := Z_{q^{-1},\Delta_l},$$

where Δ_l, $l \in \mathbb{Z}/2\mathbb{Z}$, are distinct square roots of $\pm\frac{1}{\sqrt{n}} \sum_{a\in A} q(a)$.

The fusion rules of $\mathscr{E}(q,\pm)$ are given by

$$
\begin{aligned}
&X_-\otimes X_- = X_+, && X_\pm\otimes Y_a = Y_a, && X_+\otimes Z_l = Z_l,\\
&X_-\otimes Z_l = Z_{l+1}, && Y_a\otimes Y_b = Y_{a+b}\oplus Y_{a-b}, && Y_a\otimes Y_a = X_+\oplus X_-\oplus Y_{2a},\\
&Y_a\otimes Z_l = Z_0\oplus Z_1, && Z_l\otimes Z_l = X_+\oplus\left(\oplus Y_a\right), && Z_l\otimes Z_{l+1} = X_-\oplus\left(\oplus Y_a\right),
\end{aligned}
$$

where $a,\ b\in A$ ($a\neq b$) and $l\in\mathbb{Z}/2\mathbb{Z}$. All objects of $\mathscr{E}(q,\pm)$ are self-dual.

Remark 5.2. Note that the fusion rules of $\mathscr{E}(q,\pm)$ do not depend on the quadratic form q and the number τ. We show below that the S- and T-matrices of $\mathscr{E}(q,\pm)$ do depend on q and τ.

5B. *S- and T-matrices of $\mathscr{E}$.*

Lemma 5.3. *The Gauss sums corresponding to q and q^2 are equal up to a sign, that is,*

$$\frac{\sum_{a\in A} q(a)^2}{\sum_{a\in A} q(a)}\in\{\pm 1\}.$$

Proof. Consider the group $A\times A$ with a nondegenerate quadratic form $Q=q\times q$. The Gaussian sum for this form is

$$\tau(A\times A,Q)=\sum_{a,b\in A} q(a)q(b)=\tau(A,q)^2.$$

The restriction of Q on the diagonal subgroup $D:=\{(a,a)\mid a\in A\}$ is nondegenerate since $|A|$ is odd. The restriction of Q on the orthogonal complement $D^\perp=\{(a,-a)\mid a\in A\}$ is nondegenerate as well. By the multiplicativity of Gaussian sums we have

$$\tau(A\times A,Q)=\tau(D,Q)\tau(D^\perp,Q)=\Big(\sum_{a\in A} q(a)^2\Big)^2,$$

which implies the result. □

Using the formulas for the S- and T- matrices of $\mathscr{Z}(\mathscr{C})$ given in Section 4C we can write down the S- and T- matrices of $\mathscr{E}(q,\pm)$:

$$
\begin{aligned}
&S_{X_\pm,X_\pm}=1, \qquad S_{X_\mp,X_\pm}=1, \qquad S_{X_\pm,Y_a}=2, \quad S_{Y_a,Z_l}=0,\\
&S_{X_+,Z_l}=\sqrt{n}, \quad S_{X_-,Z_l}=-\sqrt{n}, \quad S_{Y_a,Y_b}=2\left(\frac{q(a+b)^2}{q(a)^2q(b)^2}+\frac{q(a)^2q(b)^2}{q(a+b)^2}\right),
\end{aligned}
$$

$$S_{Z_l,Z_l}=\begin{cases}\pm\sqrt{n} & \text{if the Gauss sums of } q \text{ and } q^2 \text{ coincide,}\\ \mp\sqrt{n} & \text{otherwise,}\end{cases}$$

$$S_{Z_l,Z_{l+1}}=\begin{cases}\mp\sqrt{n} & \text{if the Gauss sums of } q \text{ and } q^2 \text{ coincide,}\\ \pm\sqrt{n} & \text{otherwise.}\end{cases}$$

$$T_{X_\pm} = 1, \quad T_{Y_a} = q(a)^2, \quad T_{Z_l} = \Delta_l.$$

(Recall that $\Delta_l, l \in \mathbb{Z}/2\mathbb{Z}$, are distinct square roots of $\pm\frac{1}{\sqrt{n}} \sum_{a\in A} q(a)$.)

5C. ***Example with*** $A = \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$**.** Let p be an odd prime and let $A := \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. Let $\left(\frac{\cdot}{p}\right)$ denote the Legendre symbol modulo p, that is, $\left(\frac{a}{p}\right) = 1$ if $a \in (\mathbb{Z}/p\mathbb{Z})^\times$ is a square modulo p and -1 otherwise.

Let $a, b \in (\mathbb{Z}/p\mathbb{Z})^\times$ and $\xi := e^{2\pi i/p}$. Consider the following nondegenerate quadratic form q on A:

$$q(x_1, x_2) = \xi^{ax_1^2 - bx_2^2}.$$

It is hyperbolic if $\left(\frac{ab}{p}\right) = 1$ and elliptic if $\left(\frac{ab}{p}\right) = -1$.

Lemma 5.4. *For every $a, b \in A^\times$, we have*

$$\sum_{x\in\mathbb{Z}/p\mathbb{Z}} \xi^{ax^2} = \begin{cases} \left(\frac{a}{p}\right)\sqrt{p} & \text{if } p \equiv 1 \pmod 4, \\ \left(\frac{a}{p}\right) i\sqrt{p} & \text{if } p \equiv 3 \pmod 4, \end{cases}$$

and

$$\sum_{(x_1,x_2)\in\mathbb{Z}/p\mathbb{Z}\times\mathbb{Z}/p\mathbb{Z}} \xi^{ax_1^2 - bx_2^2} = \left(\frac{ab}{p}\right) p.$$

Proof. The first assertion is well known; see, for example, [Ireland and Rosen 1990]. The second assertion is an easy consequence of the first. □

Using Lemma 5.4 we can explicitly write the S-matrix of $\mathscr{E}(q, \pm)$:

$$\begin{aligned} S_{X_\pm,X_\pm} &= 1, S_{X_\mp,X_\pm} = 1, & S_{X_\pm,Y_{(x_1,x_2)}} &= 2, \\ S_{X_+,Z_l} &= p,\ S_{X_-,Z_l} = -p, & S_{Y_{(x_1,x_2)},Y_{(y_1,y_2)}} &= 4\,\mathrm{Re}(\xi^{4ax_1y_1 - 4bx_2y_2}), \\ S_{Y_{(x_1,x_2)},Z_l} &= 0, \ \ S_{Z_l,Z_l} = \pm p, & S_{Z_l,Z_{l+1}} &= \mp p, \end{aligned}$$

and its T-matrix:

$$T_{X_\pm} = 1, \quad T_{Y_{(x_1,x_2)}} = \xi^{2ax_1^2 - 2bx_2^2}, \quad T_{Z_l} = \Delta_l,$$

where $\Delta_l, l \in \mathbb{Z}/2\mathbb{Z}$, are distinct square roots of $\pm\left(\frac{ab}{p}\right)$.

The central charge of the modular category $\mathscr{E}(q, \pm)$ is

$$\zeta(\mathscr{E}(q, \pm)) = \left(\frac{ab}{p}\right).$$

Below we give the S- and T-matrices of the modular category $\mathscr{E}(q, \pm)$ for $p = 3$. Order simple objects of $\mathscr{E}(q, \pm)$ as follows: $\mathbf{1}, X_-, Y_{(0,1)}, Y_{(1,0)}, Y_{(1,1)}, Y_{(1,2)}, Z_+, Z_-$. There are four modular categories $\mathscr{E}(q, \pm)$ of dimension 36 corresponding to the choices of hyperbolic/elliptic q and $\tau = \pm\frac{1}{3}$.

(a) When q is hyperbolic we have

$$S=\begin{pmatrix}1&1&2&2&2&2&3&3\\1&1&2&2&2&2&-3&3\\2&2&-2&4&-2&-2&0&0\\2&2&4&-2&-2&-2&0&0\\2&2&-2&-2&4&-2&0&0\\2&2&-2&-2&-2&4&0&0\\3&-3&0&0&0&0&\pm3&\mp3\\3&-3&0&0&0&0&\mp3&\pm3\end{pmatrix},$$

$$T=\operatorname{diag}\{1,1,\xi^2,\xi,1,1,1,-1\}\quad\text{when } \tau=\tfrac{1}{3},$$
$$T=\operatorname{diag}\{1,1,\xi^2,\xi,1,1,i,-i\}\quad\text{when } \tau=-\tfrac{1}{3}.$$

Note that both the corresponding modular categories are group-theoretical with central charge 1; in fact the one with $\tau=\frac{1}{3}$ is equivalent to the representation category of the double $D(S_3)$ of the symmetric group S_3 and the one with $\tau=-\frac{1}{3}$ is equivalent to the twisted double of S_3.

(b) When q is elliptic we have

$$S=\begin{pmatrix}1&1&2&2&2&2&3&3\\1&1&2&2&2&2&-3&3\\2&2&-2&4&-2&-2&0&0\\2&2&4&-2&-2&-2&0&0\\2&2&-2&-2&-2&4&0&0\\2&2&-2&-2&4&-2&0&0\\3&-3&0&0&0&0&\pm3&\mp3\\3&-3&0&0&0&0&\mp3&\pm3\end{pmatrix},$$

$$T=\operatorname{diag}\{1,1,\xi,\xi,\xi^2,\xi^2,i,-i\}\quad\text{when } \tau=\frac{1}{3},$$
$$T=\operatorname{diag}\{1,1,\xi,\xi,\xi^2,\xi^2,1,-1\}\quad\text{when } \tau=-\frac{1}{3}.$$

Both the corresponding modular categories are not group-theoretical. They both have central charge -1 and so are not equivalent to centers of fusion categories. In particular, they are not equivalent to representation categories of any twisted group doubles.

5D. ***Example with $A=\mathbb{Z}/p\mathbb{Z}$.*** Let p be an odd prime and let $A:=\mathbb{Z}/p\mathbb{Z}$. Let $a\in(\mathbb{Z}/p\mathbb{Z})^\times$ and $\xi:=e^{2\pi i/p}$. Up to isomorphism there are two nondegenerate quadratic forms q on A:

$$q(x)=\xi^{ax^2},$$

one corresponding to $\left(\frac{a}{p}\right) = 1$ and another to $\left(\frac{a}{p}\right) = -1$.

Using Lemma 5.4 we can explicitly write the S-matrix of $\mathcal{E}(q, \pm)$:

$$\begin{aligned} S_{X_\pm, X_\pm} &= 1, & S_{X_\mp, X_\pm} &= 1, & S_{X_\pm, Y_x} &= 2, \\ S_{X_+, Z_l} &= \sqrt{p}, & S_{X_-, Z_l} &= -\sqrt{p}, & S_{Y_x, Y_y} &= 4\,\mathrm{Re}(\xi^{4axy}), \\ S_{Y_a, Z_l} &= 0, & S_{Z_l, Z_l} &= \pm\left(\frac{2}{p}\right)\sqrt{p}, & S_{Z_l, Z_{l+1}} &= \mp\left(\frac{2}{p}\right)\sqrt{p}. \end{aligned}$$

Further, we have

$$T_{X_\pm} = 1, \quad T_{Y_x} = \xi^{-2ax^2}, \quad T_{Z_l} = \Delta_l,$$

where

$$\Delta_l,\ l \in \mathbb{Z}/2\mathbb{Z}, \text{ are distinct square roots of } \begin{cases} \pm\left(\frac{a}{p}\right) & \text{if } p \equiv 1 \pmod 4, \\ \pm\left(\frac{a}{p}\right) i & \text{if } p \equiv 3 \pmod 4. \end{cases}$$

The central charge of the modular category $\mathcal{E}(q, \pm)$ is

$$\zeta(\mathcal{E}(q, \pm)) = \begin{cases} \left(\frac{2a}{p}\right) & \text{if } p \equiv 1 \pmod 4, \\ -\left(\frac{2a}{p}\right) i & \text{if } p \equiv 3 \pmod 4. \end{cases}$$

Below we give the S- and T-matrices of the modular category $\mathcal{E}(q, \pm)$ for $p = 3$ and 5. For $p = 3$ we order the simple objects as $\mathbf{1}, X_-, Y_1, Z_0, Z_1$ and for $p = 5$ we order them as $\mathbf{1}, X_-, Y_1, Y_2, Z_0, Z_1$. (In (c) and (d) below, $\xi = e^{2\pi i/5}$.)

(a) When $p = 3$ and $a = 1$ we have

$$S = \begin{pmatrix} 1 & 1 & 2 & \sqrt{3} & \sqrt{3} \\ 1 & 1 & 2 & -\sqrt{3} & -\sqrt{3} \\ 2 & 2 & -2 & 0 & 0 \\ \sqrt{3} & -\sqrt{3} & 0 & \mp\sqrt{3} & \pm\sqrt{3} \\ \sqrt{3} & -\sqrt{3} & 0 & \pm\sqrt{3} & \mp\sqrt{3} \end{pmatrix},$$

$$T = \mathrm{diag}\left\{1, 1, \frac{-1 + i\sqrt{3}}{2}, \frac{1+i}{\sqrt{2}}, \frac{-1-i}{\sqrt{2}}\right\} \quad \text{when } \tau = \frac{1}{\sqrt{3}},$$

$$T = \mathrm{diag}\left\{1, 1, \frac{-1 + i\sqrt{3}}{2}, \frac{1-i}{\sqrt{2}}, \frac{-1+i}{\sqrt{2}}\right\} \quad \text{when } \tau = -\frac{1}{\sqrt{3}}.$$

The central charge of both the corresponding modular categories is i.

(b) When $p=3$ and $a=2$ we have

$$S= \text{ the } S\text{-matrix in (a)},$$

$$T=\operatorname{diag}\left\{1,\,1,\,\frac{-1-i\sqrt{3}}{2},\,\frac{1-i}{\sqrt{2}},\,\frac{-1+i}{\sqrt{2}}\right\}\quad \text{when } \tau=\frac{1}{\sqrt{3}},$$

$$T=\operatorname{diag}\left\{1,\,1,\,\frac{-1-i\sqrt{3}}{2},\,\frac{1+i}{\sqrt{2}},\,\frac{-1-i}{\sqrt{2}}\right\}\quad \text{when } \tau=\frac{1}{\sqrt{3}}.$$

The central charge of both the corresponding modular categories is $-i$.

(c) When $p=5$ and $a=1$ we have

$$S=\begin{pmatrix} 1 & 1 & 2 & 2 & \sqrt{5} & \sqrt{5} \\ 1 & 1 & 2 & 2 & -\sqrt{5} & -\sqrt{5} \\ 2 & 2 & \sqrt{5}-1 & -\sqrt{5}-1 & 0 & 0 \\ 2 & 2 & -\sqrt{5}-1 & \sqrt{5}-1 & 0 & 0 \\ \sqrt{5} & -\sqrt{5} & 0 & 0 & \mp\sqrt{5} & \pm\sqrt{5} \\ \sqrt{5} & -\sqrt{5} & 0 & 0 & \pm\sqrt{5} & \mp\sqrt{5} \end{pmatrix},$$

$$T=\operatorname{diag}\left\{1,\,1,\,\xi^3,\,\xi^2,\,1,\,-1\right\}\quad \text{when } \tau=\frac{1}{\sqrt{5}},$$

$$T=\operatorname{diag}\left\{1,\,1,\,\xi^3,\,\xi^2,\,i,\,-i\right\}\quad \text{when } \tau=-\frac{1}{\sqrt{5}}.$$

The central charge of both the corresponding modular categories is -1.

(d) When $p=5$ and $a=2$ we have

$$S=\begin{pmatrix} 1 & 1 & 2 & 2 & \sqrt{5} & \sqrt{5} \\ 1 & 1 & 2 & 2 & -\sqrt{5} & -\sqrt{5} \\ 2 & 2 & -\sqrt{5}-1 & \sqrt{5}-1 & 0 & 0 \\ 2 & 2 & \sqrt{5}-1 & -\sqrt{5}-1 & 0 & 0 \\ \sqrt{5} & -\sqrt{5} & 0 & 0 & \mp\sqrt{5} & \pm\sqrt{5} \\ \sqrt{5} & -\sqrt{5} & 0 & 0 & \pm\sqrt{5} & \mp\sqrt{5} \end{pmatrix},$$

$$T=\operatorname{diag}\left\{1,\,1,\,\xi,\,\xi^4,\,i,\,-i\right\}\quad \text{when } \tau=\frac{1}{\sqrt{5}},$$

$$T=\operatorname{diag}\left\{1,\,1,\,\xi,\,\xi^4,\,1,\,-1\right\}\quad \text{when } \tau=-\frac{1}{\sqrt{5}}.$$

The central charge of both the corresponding modular categories is 1.

6. Appendix: Zeroes in S-matrices

There is a classical result of Burnside in character theory saying that if χ is an irreducible character of a finite group G and $\chi(1) > 1$, then $\chi(g) = 0$ for some $g \in G$; see [Berkovich and Zhmud' 1999, Chapter 21].

In this appendix we establish a categorical analogue of this result for weakly integral modular categories. Recall from [Etingof et al. 2008] that a fusion category $\mathscr{C}$ is called *weakly integral* if its Frobenius–Perron dimension is an integer. In this case the Frobenius–Perron dimension of every simple object of $\mathscr{C}$ is the square root of an integer [Etingof et al. 2005].

Let $\mathscr{C}$ be a weakly integral modular category with the S-matrix S. Let $\mathbb{O}(\mathscr{C})$ denote the set of all (representatives of isomorphism classes of) simple objects of $\mathscr{C}$. Given $X \in \mathbb{O}(\mathscr{C})$ define the sets

$$T_X = \{Y \in \mathbb{O}(\mathscr{C}) \mid S_{X,Y} = 0\}, \quad D_X = \mathbb{O}(\mathscr{C}) - (T_X \cup \{\mathbf{1}\}).$$

Clearly, we have a partition $\mathbb{O}(\mathscr{C}) = T_X \cup D_X \cup \{\mathbf{1}\}$. Let $\mathscr{T}_X$ and $\mathscr{D}_X$ be full abelian subcategories of $\mathscr{C}$ generated by T_X and D_X, respectively.

Let K be the field extension of $\mathbb{Q}$ generated by the entries of S. It is known [de Boer and Goeree 1991; Coste and Gannon 1994] that there is a root of unity ξ such that $K \subset \mathbb{Q}(\xi)$. In particular, the operation of taking the square of an absolute value of an element of S is well defined. Let $G := \mathrm{Gal}(K/\mathbb{Q})$. Every element $\sigma \in G$ comes from a permutation σ of $\mathbb{O}(\mathscr{C})$ such that $\sigma(S_{X,Y}) = S_{X,\sigma(Y)}$ for all $X, Y \in \mathbb{O}(\mathscr{C})$.

Let $\mathscr{C}$ be a weakly integral modular category. It was shown in [Etingof et al. 2005] that there is a canonical spherical structure on $\mathscr{C}$ such that categorical dimensions in $\mathscr{C}$ coincide with Frobenius–Perron dimensions. Let us fix this structure for the remainder of this section. For any $X \in \mathbb{O}(\mathscr{C})$ let d_X denote the dimension of X. For any full abelian subcategory $\mathscr{A}$ of $\mathscr{C}$ let $\dim \mathscr{A}$ denote the sum of squares of dimensions of simple objects of $\mathscr{A}$.

Theorem 6.1. *Let $\mathscr{C}$ be a weakly integral modular category with the S-matrix S. Then T_X is not empty for every noninvertible simple object X of $\mathscr{C}$. That is, every row (column) of S corresponding to a noninvertible simple object contains at least one zero entry.*

Proof. Note that the statement of Proposition does not depend on the choice of spherical structure.

We have $\sum_{Y \in \mathbb{O}(\mathscr{C})} |S_{X,Y}|^2 = \dim \mathscr{C}$; hence,

$$1 = \frac{\dim \mathscr{C}}{d_X^2} - \sum_{Y \in D_X} \left| \frac{S_{X,Y}}{d_X} \right|^2 = \frac{1 + \dim \mathscr{T}_X}{d_X^2} - \left(\sum_{Y \in D_X} \left| \frac{S_{X,Y}}{d_X} \right|^2 - \frac{\dim \mathscr{D}_X}{d_X^2} \right), \tag{43}$$

where d_X denotes the dimension of X. It suffices to check that

$$\frac{1}{\dim \mathcal{D}_X} \sum_{Y \in D_X} \left| \frac{S_{X,Y}}{d_X} \right|^2 \geq \frac{1}{d_X^2}, \tag{44}$$

since then (43) implies that $1 \leq (1 + \dim \mathcal{T}_X)/d_X^2$, whence

$$\dim \mathcal{T}_X \geq d_X^2 - 1. \tag{45}$$

But X is noninvertible so $d_X > 1$ and $\mathcal{T}_X \neq 0$.

Rewriting the left hand side of (44) as the sum of $\dim \mathcal{D}_X$ terms and using the inequality of arithmetic and geometric means we obtain

$$\begin{aligned} \frac{1}{\dim \mathcal{D}_X} \sum_{Y \in D_X} \left| \frac{S_{X,Y}}{d_X} \right|^2 &= \frac{1}{\dim \mathcal{D}_X} \sum_{Y \in D_X} d_Y^2 \left| \frac{S_{X,Y}}{d_X d_Y} \right|^2 \\ &\geq \frac{1}{d_X^2} \left(\prod_{Y \in D_X} \left| \frac{S_{X,Y}}{d_Y} \right|^{2d_Y^2} \right)^{1/\dim \mathcal{D}_X}. \end{aligned}$$

The set D_X is clearly stable under all automorphisms in the Galois group, and hence so is the product $\prod_{Y \in D_X} \left| S_{X,Y}/d_Y \right|^{2d_Y^2}$. Therefore, this product belongs to $\mathbb{Q}$. Its factors are squares of absolute values of characters of $K_0(\mathcal{C})$ on X and hence are algebraic integers. Since all factors are positive, the product is ≥ 1, which implies (44). □

For $X \in \mathcal{O}(\mathcal{C})$ define

$$U_X = \{Y \in \mathcal{O}(\mathcal{C}) \mid |S_{X,Y}| = d_Y\}.$$

Let $\mathcal{U}_X$ be the full abelian subcategory of $\mathcal{C}$ generated by U_X.

Proposition 6.2. *Let $\mathcal{C}$ be a weakly integral modular category and let X be a simple noninvertible object in $\mathcal{C}$. Then*

$$3 \dim \mathcal{T}_X + \dim \mathcal{U}_X > \dim \mathcal{C}. \tag{46}$$

Proof. We may assume $d_X \geq \sqrt{2}$.

We will use the following theorem of Siegel [1945] from number theory. Let $K/\mathbb{Q}$ be a finite Galois extension with the Galois group $G = \mathrm{Gal}(K/\mathbb{Q})$. Let α be a totally positive algebraic integer in K, $\alpha \neq 1$. Then

$$\frac{1}{|G|} \sum_{\sigma \in G} \sigma(\alpha) \geq \frac{3}{2}.$$

We apply this to the situation when K is the extension of $\mathbb{Q}$ generated by entries of S. We compute

$$\begin{aligned}\dim \mathscr{C} &= \sum_{Y\in\mathscr{C}} |S_{X,Y}|^2 = d_X^2 + \sum_{Y\in U_X} d_Y^2 + \sum_{Y\in\mathbb{O}(\mathscr{C})-(T_X\cup U_X\cup\{\mathbf{1}\})} |S_{X,Y}|^2 \\ &= d_X^2 + \dim \mathscr{U}_X + \sum_{Y\in\mathbb{O}(\mathscr{C})-(T_X\cup U_X\cup\{\mathbf{1}\})} d_Y^2 \left(\frac{1}{|G|} \sum_{\sigma\in G} \sigma\left(\frac{|S_{X,Y}|^2}{d_Y^2} \right) \right) \\ &\geq 2 + \dim \mathscr{U}_X + \tfrac{3}{2}(\dim \mathscr{C} - \dim \mathscr{T}_X - \dim \mathscr{U}_X - 1);\end{aligned}$$

therefore $3 \dim \mathscr{T}_X + \dim \mathscr{U}_X \geq \dim \mathscr{C} + 1 > \dim \mathscr{C}$, as required. $\square$

Remark 6.3. Our proofs of Theorem 6.1 and Proposition 6.2 imitate the corresponding proofs for group characters given in [Berkovich and Zhmud' 1999].

Acknowledgments

We are grateful to P. Etingof, M. Müger, and V. Ostrik for useful discussions. Part of this work was done while Gelaki was on sabbatical in the Departments of Mathematics at the University of New Hampshire and MIT; he is grateful for their warm hospitality. His research was partially supported by the Israel Science Foundation (grant No. 125/05). Nikshych's research of was supported by the NSA grant H98230-07-1-0081 and the NSF grant DMS-0800545.

References

[Arkhipov and Gaitsgory 2003] S. Arkhipov and D. Gaitsgory, "Another realization of the category of modules over the small quantum group", *Adv. Math.* **173**:1 (2003), 114–143. MR 2004e:17010 Zbl 1025.17004

[Bakalov and Kirillov 2001] B. Bakalov and A. Kirillov, Jr., *Lectures on tensor categories and modular functors*, University Lecture Series **21**, American Mathematical Society, Providence, RI, 2001. MR 2002d:18003 Zbl 0965.18002

[Berkovich and Zhmud' 1999] Y. G. Berkovich and E. M. Zhmud', *Characters of finite groups*, vol. 2, Translations of Mathematical Monographs **181**, American Mathematical Society, Providence, RI, 1999. MR 99j:20007 Zbl 0934.20009

[de Boer and Goeree 1991] J. de Boer and J. Goeree, "Markov traces and II_1 factors in conformal field theory", *Comm. Math. Phys.* **139**:2 (1991), 267–304. MR 93i:81211 Zbl 0760.57002

[Bruguières 2000] A. Bruguières, "Catégories prémodulaires, modularisations et invariants des variétés de dimension 3", *Math. Ann.* **316**:2 (2000), 215–236. MR 2001d:18009 Zbl 0943.18004

[Coste and Gannon 1994] A. Coste and T. Gannon, "Remarks on Galois symmetry in rational conformal field theories", *Phys. Lett. B* **323**:3-4 (1994), 316–321. MR 95h:81031

[Deligne 1990] P. Deligne, "Catégories tannakiennes", pp. 111–195 in *The Grothendieck Festschrift*, vol. 2, edited by P. Cartier et al., Progr. Math. **87**, Birkhäuser, Boston, MA, 1990. MR 92d:14002 Zbl 0727.14010

[Drinfeld et al. 2007] V. Drinfeld, S. Gelaki, D. Nikshych, and V. Ostrik, "Group-theoretical properties of nilpotent modular categories", preprint, 2007. arXiv 0704.0195

[Drinfeld et al. 2009] V. Drinfeld, S. Gelaki, D. Nikshych, and V. Ostrik, "On braided fusion categories I", preprint, 2009. arXiv 0906.0620

[Etingof et al. 2005] P. Etingof, D. Nikshych, and V. Ostrik, "On fusion categories", *Ann. of Math.* (2) **162**:2 (2005), 581–642. MR 2006m:16051 Zbl 1125.16025

[Etingof et al. 2008] P. Etingof, D. Nikshych, and V. Ostrik, "Weakly group-theoretical and solvable fusion categories", preprint, 2008. arXiv 0809.3031

[Etingof et al. 2009] P. Etingof, D. Nikshych, and V. Ostrik, "Fusion categories and homotopy theory", preprint, 2009. arXiv 0909.3140

[Gaitsgory 2005] D. Gaitsgory, "The notion of category over an algebraic stack", preprint, 2005. arXiv 0507192

[Gelaki and Nikshych 2008] S. Gelaki and D. Nikshych, "Nilpotent fusion categories", *Adv. Math.* **217**:3 (2008), 1053–1071. MR 2009b:18015 Zbl 1168.18004

[Ireland and Rosen 1990] K. Ireland and M. Rosen, *A classical introduction to modern number theory*, 2nd ed., Graduate Texts in Mathematics **84**, Springer, New York, 1990. MR 92e:11001 Zbl 0712.11001

[Izumi 2001] M. Izumi, "The structure of sectors associated with Longo–Rehren inclusions, II: Examples", *Rev. Math. Phys.* **13**:5 (2001), 603–674. MR 2002k:46161

[Kassel 1995] C. Kassel, *Quantum groups*, Graduate Texts in Mathematics **155**, Springer, New York, 1995. MR 96e:17041 Zbl 0808.17003

[Kirillov 2002] A. Kirillov, Jr., "Modular categories and orbifold models", *Comm. Math. Phys.* **229**:2 (2002), 309–335. MR 2003m:17024 Zbl 1073.17011

[Majid 1991] S. Majid, "Representations, duals and quantum doubles of monoidal categories", *Rend. Circ. Mat. Palermo* (2) *Suppl.* **26** (1991), 197–206. MR 93c:18008 Zbl 0762.18005

[Montgomery 1993] S. Montgomery, *Hopf algebras and their actions on rings*, CBMS Regional Conference Series in Mathematics **82**, Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1993. MR 94i:16019 Zbl 0793.16029

[Müger 2000] M. Müger, "Galois theory for braided tensor categories and the modular closure", *Adv. Math.* **150**:2 (2000), 151–201. MR 2001a:18008

[Müger 2003a] M. Müger, "From subfactors to categories and topology, I: Frobenius algebras in and Morita equivalence of tensor categories", *J. Pure Appl. Algebra* **180**:1-2 (2003), 81–157. MR 2004f:18013

[Müger 2003b] M. Müger, "On the structure of modular categories", *Proc. London Math. Soc.* (3) **87**:2 (2003), 291–308. MR 2004g:18009

[Müger 2004] M. Müger, "Galois extensions of braided tensor categories and braided crossed G-categories", *J. Algebra* **277**:1 (2004), 256–281. MR 2005b:18011

[Naidu and Nikshych 2008] D. Naidu and D. Nikshych, "Lagrangian subcategories and braided tensor equivalences of twisted quantum doubles of finite groups", *Comm. Math. Phys.* **279**:3 (2008), 845–872. MR 2009b:16092 Zbl 1139.16028

[Nikshych 2008] D. Nikshych, "Non-group-theoretical semisimple Hopf algebras from group actions on fusion categories", *Selecta Math.* (*N.S.*) **14**:1 (2008), 145–161. MR 2009k:16075 Zbl 05574000

[Ostrik 2003] V. Ostrik, "Module categories, weak Hopf algebras and modular invariants", *Transform. Groups* **8**:2 (2003), 177–206. MR 2004h:18006 Zbl 1044.18004

[Siegel 1945] C. L. Siegel, "The trace of totally positive and real algebraic integers", *Ann. of Math.* (2) **46** (1945), 302–312. MR 6,257a Zbl 0063.07009

[Tambara 2000] D. Tambara, "Representations of tensor categories with fusion rules of self-duality for abelian groups", *Israel J. Math.* **118** (2000), 29–60. MR 2001j:18015 Zbl 0969.18007

[Tambara 2001] D. Tambara, "Invariants and semi-direct products for finite group actions on tensor categories", *J. Math. Soc. Japan* **53**:2 (2001), 429–456. MR 2002e:18010 Zbl 0980.18003

[Tambara and Yamagami 1998] D. Tambara and S. Yamagami, "Tensor categories with fusion rules of self-duality for finite abelian groups", *J. Algebra* **209**:2 (1998), 692–707. MR 2000b:18013 Zbl 0923.46052

[Turaev 2000] V. Turaev, "Homotopy field theory in dimension 3 and crossed group-categories", preprint, 2000. arXiv 0005291

[Turaev 2008] V. Turaev, "Crossed group-categories", *Arab. J. Sci. Eng. Sect. C Theme Issues* **33**:2 (2008), 483–503. MR 2500054

Communicated by Susan Montgomery
Received 2009-05-21 Revised 2009-08-31 Accepted 2009-11-09

gelaki@math.technion.ac.il *Department of Mathematics, Technion-Israel Institute of Technology, 32000 Haifa, Israel*

dnaidu@math.tamu.edu *Department of Mathematics, Texas A&M University, College Station, TX 77843, United States*

nikshych@math.unh.edu *Department of Mathematics and Statistics, University of New Hampshire, Durham, NH 03824, United States*

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in ANT are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@mathscipub.org with details about how your graphics were generated.

White Space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory

Volume 3 No. 8 2009