Algebra & Number Theory

Volume 12
2018
No. 8

msp

Algebra & Number Theory

msp.org/ant

See inside back cover or msp.org/ant for submission instructions.

The subscription price for 2018 is US $340/year for the electronic version, and $535/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o 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 MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ALGEBRA AND NUMBER THEORY 12:8 (2018)
dx.doi.org/10.2140/ant.2018.12.1823

On the relative Galois module structure of rings of integers in tame extensions

Adebisi Agboola and Leon R. McCulloh

Let F be a number field with ring of integers O_F and let G be a finite group. We describe an approach to the study of the set of realisable classes in the locally free class group $\mathrm{Cl}(O_F G)$ of $O_F G$ that involves applying the work of McCulloh in the context of relative algebraic K theory. For a large class of soluble groups G, including all groups of odd order, we show (subject to certain mild conditions) that the set of realisable classes is a subgroup of $\mathrm{Cl}(O_F G)$. This may be viewed as being a partial analogue in the setting of Galois module theory of a classical theorem of Shafarevich on the inverse Galois problem for soluble groups.

MSC2010: primary 11R33; secondary 11R32, 11R65, 19F99.
Keywords: Galois module structure, realisable classes, rings of integers, inverse Galois problem, relative K-group.

Introduction

Suppose that F is a number field with ring of integers O_F, and let G be a finite group. If F_π/F is any tame Galois G-algebra extension of F, then a classical theorem of E. Noether implies that the ring of integers O_π of F_π is a locally free O_FG-module, and so determines a class (O_π) in the locally free class group $\mathrm{Cl}(O_FG)$ of O_FG. Hence, if we write $H^1_t(F, G)$ for the pointed set of isomorphism classes of tame G-extensions of F, then we obtain a map of pointed sets

$$\psi : H^1_t(F, G) \to \mathrm{Cl}(O_FG), \quad [\pi] \mapsto (O_\pi).$$

Even when G is abelian, so that $H^1_t(F, G)$ is actually a group, this map is almost never a group homomorphism. We say that an element $c \in \mathrm{Cl}(O_FG)$ is *realisable* if $c = (O_\pi)$ for some tame Galois G-algebra extension F_π/F, and we write $\mathcal{R}(O_FG)$ for the collection of realisable classes in $\mathrm{Cl}(O_FG)$. These classes are natural objects of study, and they have arisen in a number of different contexts in Galois module theory. The problem of describing $\mathcal{R}(O_FG)$ for a given G may be viewed as being a loose analogue of the inverse Galois problem in the setting of arithmetic Galois module theory.

When G is abelian, McCulloh [1987] has given a complete description of $\mathcal{R}(O_FG)$ by showing that it is equal to the kernel of a certain Stickelberger homomorphism on $\mathrm{Cl}(O_FG)$. In particular, he has shown that $\mathcal{R}(O_FG)$ is in fact a group. In subsequent unpublished work McCulloh [2011; 2012] showed that, for arbitrary G, the set $\mathcal{R}(O_FG)$ is always contained in the kernel of this Stickelberger homomorphism, and he raised the question of whether or not $\mathcal{R}(O_FG)$ is in fact always equal to this kernel. This question has inspired research by a number of authors, and we refer the reader to, e.g., [Byott and Sodaïgui 2005; Byott et al. 2006; Farhat and Sodaïgui 2015] and to the bibliographies of these papers for further information concerning previous work on this problem.

In this paper we shall describe a new approach to studying this topic that involves combining the methods introduced by McCulloh [1987; 2011] with techniques involving relative algebraic K-theory and categorical twisted forms introduced by D. Burns and Agboola [2006]. This enables us to both clarify certain aspects of the theory of realisable classes and to establish new results. Although our perspective is somewhat different, it should be stressed that many of the main ideas that we use are in fact already present in some form in [McCulloh 1987; 2011].

Let us now describe the contents of this paper in more detail. In Section 2 we recall some basic facts concerning principal homogeneous spaces, Galois algebras and resolvends; these play a key role in everything that follows. Next, we assemble a number of technical results explaining how resolvends may be used to compute discriminants of rings of integers in Galois G-extensions. We also discuss how certain Galois cohomology groups may be expressed in terms of resolvends in a manner that is very useful for calculations in class groups and K-groups. In Section 4 we explain how determinants of resolvends may be represented in terms of certain character maps, and we recall an approximation theorem of A. Siviero (which is in turn a variant of [McCulloh 1987, Theorem 2.14]).

We begin Section 5 by outlining the results we need about twisted forms and relative algebraic K-groups from [Agboola and Burns 2006]. Each tame G-extension F_π/F of F has an associated resolvend isomorphism

$$\boldsymbol{r}_G : F_\pi \otimes_F F^c \simeq F^c G$$

of F^cG-modules, and this may be used to construct a categorical twisted form which is represented by an element $[O_\pi, O_F G; \boldsymbol{r}_G]$ in a certain relative algebraic K-group $K_0(O_F G, F^c)$. The group $K_0(O_F G, F^c)$ admits a natural surjection onto the locally free class group $\mathrm{Cl}(O_F G)$, sending $[O_\pi, O_F G; \boldsymbol{r}_G]$ to (O_π), and so there is a map of pointed sets

$$\Psi : H^1_t(F, G) \to K_0(O_F G, F^c), \quad [\pi] \mapsto [O_\pi, O_F G; \boldsymbol{r}_G]$$

which is a refinement (more precisely, a lifting) of the map ψ above.

Crucial to our approach is the fact that each of the constructions that we have just described admits a local variant. Let v be any place of F, and write $H^1_t(F_v, G)$ for the pointed set of isomorphism classes of tame G-extensions of F_v. Then there is a localisation homomorphism

$$\lambda_v : K_0(O_F G, F^c) \to K_0(O_{F_v} G, F^c_v)$$

as well as a map of pointed sets

$$\Psi_v : H^1_t(F_v, G) \to K_0(O_{F_v} G, F^c_v), \quad [\pi_v] \mapsto [O_{\pi_v}, O_{F_v} G; \boldsymbol{r}_G].$$

The following result reflects the fact that $[O_\pi, O_F G; \boldsymbol{r}_G]$ is a much finer structure invariant than (O_π) (see Proposition 13.1 below):

Proposition A. *The kernel of Ψ is finite.*

Let G' denote the derived subgroup of G. We may identify $H^1(F, G')$ with a subset of $H^1(F, G)$ via the exact sequence $0 \to G' \to G \to G^{\mathrm{ab}} \to 0$. Proposition A is proved by showing that $\mathrm{Ker}(\Psi)$ is a subset of the pointed set $H^1_{\mathrm{fnr}}(F, G')$ of isomorphism classes of G'-Galois F-algebras that are unramified at all finite places of F; this last set is finite because there are only finitely many unramified extensions of F of bounded degree. If G is abelian, the map Ψ is injective (see Proposition 14.3). In many cases one can show that $\mathrm{Ker}(\Psi) = H^1_{\mathrm{fnr}}(F, G')$, but we do not know whether this equality always holds.

Write $K\mathcal{R}(O_F G)$ for the image of Ψ, i.e., for the collection of realisable classes of $K_0(O_F G, F^c)$. The central conjecture of this paper gives a precise description of $K\mathcal{R}(O_F G)$ in terms of a local-global principle for the relative algebraic K-group $K_0(O_F G, F^c)$. This may be described as follows.

For each place v of F, let $H^1_{\mathrm{nr}}(F_v, G)$ denote the subset $H^1_t(F_v, G)$ consisting of isomorphism classes of unramified G-extensions of F_v. We define a pointed set of ideles $J(H^1_t(F, G))$ of $H^1_t(F, G)$ to be the restricted direct product over all places v of the sets $H^1_t(F_v, G)$ with respect to the subsets $H^1_{\mathrm{nr}}(F_v, G)$ (see Definition 6.2). The natural maps $H^1_t(F, G) \to H^1_t(F_v, G)$ for each v induce a map $H^1_t(F, G) \to J(H^1_t(F, G))$. We also define a group of ideles $J(K_0(O_F G, F^c))$ of $K_0(O_F G, F^c)$ to be the restricted direct product over all places of F of the groups $K_0(O_{F_v} G, F^c_v)$ with respect to the subgroups $K_0(O_{F_v}G, O_{F_v^c})$ (see Definition 5.8). We show that the maps λ_v above induce an injective localisation map

$$\lambda : K_0(O_F G, F^c) \to J(K_0(O_F G, F^c))$$

(see Proposition 5.9), and that the maps Ψ_v induce an idelic version

$$\Psi^{\mathrm{id}} : J(H_t^1(F, G)) \to J(K_0(O_F G, F^c))$$

of the map Ψ (see Definition 6.2). We conjecture that $K\mathcal{R}(O_F G)$ has the following description (see Conjecture 6.5 below):

Conjecture B. $K\mathcal{R}(O_F G) = \lambda^{-1}(\mathrm{Im}(\Psi^{\mathrm{id}}))$.

In other words, our conjecture predicts that an element x lies in the image of Ψ if and only if $\lambda_v(x)$ lies in the image of Ψ_v for every place v of F. We remark that it follows directly from the definitions that

$$K\mathcal{R}(O_F G) \subseteq \lambda^{-1}(\mathrm{Im}(\Psi^{\mathrm{id}})).$$

We point out that, in contrast to $\mathcal{R}(O_F G)$, it is not difficult to show that if G is nontrivial, then $K\mathcal{R}(O_F G)$ is never a subgroup of $K_0(O_F G, F^c)$ (cf. [Agboola and Burns 1998, Remark 2.10(iii); 2006, Remarks 6.13(i)].) Nevertheless, by applying the methods of [McCulloh 1987; 2011] in the present context, we show that Conjecture B implies both an affirmative answer to McCulloh's question concerning $\mathcal{R}(O_F G)$ as well as a positive solution to the inverse Galois problem for G over F (see Theorems 6.6, 6.7 and 13.6 below):

Theorem C. *If Conjecture B holds, then $\mathcal{R}(O_F G)$ is a subgroup of $\mathrm{Cl}(O_F G)$. Furthermore, if $c \in \mathcal{R}(O_F G)$, then there exist infinitely many $[\pi] \in H_t^1(F, G)$ such that F_π is a field and $(O_\pi) = c$. The extensions F_π/F may be chosen to have ramification disjoint from any finite set S of places of F. In particular, the inverse Galois problem for G admits a positive solution over F.*

In order to orient the reader, we shall now briefly indicate the main ideas involved in the proof of Theorem C.

We begin by observing that the long exact sequence of relative algebraic K-theory yields a sequence

$$K_1(F^c G) \xrightarrow{\partial^1} K_0(O_F G, F^c) \xrightarrow{\partial^0} \mathrm{Cl}(O_F G) \to 0.$$

Hence, in order to show that $\mathcal{R}(O_F G) = \mathrm{Im}(\psi)$ is a subgroup of $\mathrm{Cl}(O_F G)$, it suffices to show that $\partial^1(K_1(F^c G)) \cdot \mathrm{Im}(\Psi)$ is a subgroup of $K_0(O_F G, F^c)$.

To do this, we first show that it suffices to prove that

$$\lambda(\partial^1(K_1(F^c G))) \cdot \mathrm{Im}(\Psi^{\mathrm{id}})$$

is a subgroup of $J(K_0(O_F G, F^c))$. Once this is done, it is not hard to show that $\partial^1(K_1(F^c G)) \cdot \mathrm{Im}(\Psi)$ is equal to the kernel of the homomorphism

$$K_0(O_F G, F^c) \xrightarrow{\lambda} J(K_0(O_F G, F^c)) \to \frac{J(K_0(O_F G, F^c))}{\lambda[\partial^1(K_1(F^c G))] \cdot \mathrm{Im}(\Psi^{\mathrm{id}})},$$

and so is indeed a subgroup of $K_0(O_F G, F^c)$ (see Theorem 6.7 below). The crux of the proof of the first part of Theorem C therefore consists of showing that $\lambda(\partial^1(K_1(F^c G))) \cdot \mathrm{Im}(\Psi^{\mathrm{id}})$ is a subgroup of $K_0(O_F G, F^c)$.

This is accomplished as follows. Write $G(-1)$ for the group G (viewed as a set) endowed with an action of Ω_F via the inverse cyclotomic character. Although in general this is only an action on G as a set (rather than via automorphisms of G), the induced action on conjugacy classes of G does induce an action on the centre $Z(F^c[G])$ of the group ring $F^c G$. We write $Z(F^c[G(-1)])$ to denote $Z(F^c[G])$ endowed with this action. We set

$$\Lambda(FG) := Z(F^c[G(-1)])^{\Omega_F},$$

and we write $\Lambda(O_F G)$ for the (unique) O_F-maximal order in $\Lambda(FG)$. For each place v of F, we define $\Lambda(F_v G)$ and $\Lambda(O_{F_v} G)$ in an analogous manner. We write $J(\Lambda(FG))$ for the restricted direct product over all places of F of the groups $\Lambda(F_v G)^\times$ with respect to the subgroups $\Lambda(O_{F_v} G)^\times$.

Let $\mathrm{Irr}(G)$ denote the set of irreducible characters of G. Motivated by an analysis of normal integral basis generators of tame local extensions, we define a Stickelberger pairing

$$\langle -, - \rangle_G : \mathrm{Irr}(G) \times G \to \mathbb{Q}.$$

(Loosely speaking, this may be viewed as being a monodromy-type pairing that encodes ramification data associated to tame extensions of local fields in a uniform manner (cf. Definition 10.6 below).) We then use this pairing to construct a K-theoretic transpose Stickelberger homomorphism

$$K\Theta^t : J(\Lambda(FG)) \to J(K_0(O_F G, F^c)).$$

The homomorphism $K\Theta^t$ is closely related to the map Ψ^{id} in the following way. We show that even though the map Ψ_v is just a map of pointed sets, the image $\Psi_v(H^1_{\mathrm{nr}}(F_v, G))$ of the restriction of Ψ_v to $H^1_{\mathrm{nr}}(F_v, G)$ is in fact a subgroup of $K_0(O_{F_v} G, F_v^c)$ for each v. Using an approximation theorem for $J(\Lambda(FG))$, we show further that, for a suitable choice of auxiliary ideal $\mathfrak{a}$ of O_F, the homomorphism $K\Theta^t$ may be used to construct a homomorphism

$$\Theta^t_{\mathfrak{a}} : \mathrm{Cl}'^{+}_{\mathfrak{a}}(\Lambda(O_F G)) \to \frac{J(K_0(O_F G, F^c))}{\lambda[\partial^1(K_1(F^c G))] \cdot \prod_v \Psi_v(H^1_{\mathrm{nr}}(F_v, G))},$$

where $\mathrm{Cl}'^{+}_{\mathfrak{a}}(\Lambda(O_F G))$ is a certain finite quotient of $J(\Lambda(FG))$. We prove that

$$\mathrm{Im}(\Theta^t_{\mathfrak{a}}) = \mathrm{Im}(\overline{\Psi^{\mathrm{id}}}),$$

where $\overline{\Psi^{\mathrm{id}}}$ denotes the composition of Ψ^{id} with the obvious quotient map

$$J(K_0(O_F G, F^c)) \to \frac{J(K_0(O_F G, F^c))}{\lambda[\partial^1(K_1(F^c G))] \cdot \prod_v \Psi_v(H^1_{\mathrm{nr}}(F_v, G))}.$$

We then show that this in turn implies that

$$\lambda(\partial^1(K_1(F^c G))) \cdot \mathrm{Im}(K\Theta^t) = \lambda(\partial^1(K_1(F^c G))) \cdot \mathrm{Im}(\Psi^{\mathrm{id}}). \tag{0-1}$$

In particular, this proves that the right-hand side of (0-1) is a subgroup of $J(K_0(O_F G, F^c))$, as claimed. This completes our outline of the proof of the first part of Theorem C.

The strategy of the proof of the second part of Theorem C may be very roughly described as follows. Suppose that $x \in \lambda^{-1}(\mathrm{Im}(\Psi^{\mathrm{id}}))$. By using the map $K\Theta^t$ together with a suitable approximation theorem on $J(K_0(O_F G, F^c))$, we show that there are infinitely many $y \in \lambda^{-1}(\mathrm{Im}(\Psi^{\mathrm{id}}))$ such that (i) $\partial^0(y) = \partial^0(x)$, and (ii) each y corresponds via Conjecture B to an element $[\pi_y] \in H^1_t(F, G)$ which is ramified (away from S) in such a way that $\pi_y \in \mathrm{Hom}(\Omega_F, G)$ is forced to be surjective. This in turn implies that F_{π_y} is a field (rather than just a Galois algebra), and so the inverse Galois problem for G admits a positive solution over F.

Let us now turn to our results concerning the validity of Conjecture B.

When G is abelian, we obtain the following refinement of [McCulloh 1987, Theorem 6.7] (see Theorem 14.2 below):

Theorem D. *Conjecture B is true if G is abelian.*

By combining our methods with work of Neukirch, we are able to establish a variant of Conjecture B for a large class of soluble groups, including all groups of odd order (see Theorems 16.4 and 16.5 below). We thereby obtain the following result, which may be viewed as being a partial analogue of a classical theorem of Shafarevich [1954] on the inverse Galois problem for soluble groups in the context of arithmetic Galois module theory. (See Theorem 16.7 of the main text.)

Theorem E. *Suppose that G is of odd order and that $(|G|, h_F) = 1$, where h_F denotes the class number of F. Suppose also that F contains no nontrivial $|G|$-th roots of unity. Then $\mathcal{R}(O_F G)$ is a subgroup of $\mathrm{Cl}(O_F G)$. If $c \in \mathcal{R}(O_F G)$, then there exist infinitely many $[\pi] \in H^1_t(F, G)$ such that F_π is a field and $(O_\pi) = c$. The extensions F_π/F may be chosen to have ramification disjoint from any finite set S of places of F.*

While it is perhaps conceivable that it might be possible to remove the hypothesis $(|G|, h_F) = 1$ of Theorem E using methods similar to those of the present paper (although we do not as yet know how to do this), the same probably cannot be said of the condition concerning the number of roots of unity in F. This latter hypothesis is forced upon us because our proof makes crucial use of a lifting theorem of Neukirch (see Section 15) where such hypotheses are unavoidable (cf. the last paragraph of the introduction of [Neukirch 1979]). It would be interesting to determine whether or not the methods of [Shafarevich 1954] can be used to prove a result similar to Theorem E for all soluble groups.

The results and techniques introduced in this paper suggest a number of different avenues of further investigation. For example, our methods may also be applied in the context of the relative Galois module structure of the square root of the inverse different as studied by C. Tsang [2016; 2017], and it seems reasonable to expect that an analogue of Theorem E holds in this setting. Applying the methods of [Agboola 2012] to the study of counting and equidistribution problems involving cohomological classes in relative algebraic K-groups should lead to new results concerning similar problems for number fields, generalising certain aspects of e.g., [Wright 1989; Malle 2002]. Our techniques may also be applied in the setting of global function fields [Agboola and Burns 2001; 2006], and it would be of interest to further investigate the connection between the approach adopted here and that taken in e.g., [Chinburg 1994] (cf., for example, [Agboola and Burns 2006, §4]).

Here is an outline of the rest of this paper. In Section 7, we explain a hitherto unpublished result of McCulloh that describes how resolvends of normal integral bases of tamely ramified extensions of nonarchimedean local fields admit certain *Stickelberger factorisations* (see Definition 7.12); this is a nonabelian analogue of a version of Stickelberger's factorisation of abelian Gauss sums. A somewhat analogous (but much simpler) framework over $\mathbb{R}$ is described in Section 8.

In Section 9, we recall the definition and properties of the Stickelberger pairing. We also give a new character-theoretic description of this pairing (see Proposition 9.2) as well as an application of this description (see Corollary 9.4).

We construct a K-theoretic version of the transpose Stickelberger homomorphism in Section 10, and we also briefly describe an alternative approach to defining the Stickelberger pairing and establishing its basic properties. In Section 11 we construct transpose Stickelberger homomorphisms $\Theta^t_{\mathfrak{a}}$ on modified narrow ray class groups $\mathrm{Cl}'^{+}_{\mathfrak{a}}(\Lambda(O_F G))$. These are used in Section 12 to prove Theorem 6.6, thereby completing the proof of the first part of Theorem C.

In Section 13 we prove Proposition A, and we explain how a weaker form of Conjecture B implies that every realisable class in $\mathrm{Cl}(O_F G)$ may be realised (in infinitely many ways) by rings of integers of tame field (and not merely Galois algebra) G-extensions of F. This proves the second part of Theorem C.

We give a proof of Theorem D in Section 14. In Section 15, we describe work of Neukirch on the solution to an embedding problem that is required for the proof of Theorem E. This proof is completed in Section 16 via showing that a suitable variant of Conjecture B holds for a large class of soluble groups (see Definition 16.1 and Theorems 16.3 and 16.4).

1. Notation and conventions

For any field L, we write L^c for an algebraic closure of L, and we set

$$\Omega_L := \mathrm{Gal}(L^c/L).$$

If L is a number field or a nonarchimedean local field (by which we shall always mean a finite extension of $\mathbb{Q}_p$ for some prime p), then O_L denotes the ring of integers of L. If L is an archimedean local field, then we adopt the usual convention of setting $O_L = L$.

Throughout this paper, F will denote a number field. For each place v of F, we fix an embedding $F^c \to F^c_v$, and we view Ω_{F_v} as being a subgroup of Ω_F via this choice of embedding. We write I_v for the inertia subgroup of Ω_{F_v} when v is finite.

The symbol G will always denote a finite group upon which Ω_F acts trivially. If H is any finite group, we write $\mathrm{Irr}(H)$ for the set of irreducible F^c-valued characters of H and R_H for the corresponding ring of virtual characters. We write $\mathbf{1}_H$ (or simply $\mathbf{1}$ if there is no danger of confusion) for the trivial character in R_H. If $h \in H$, then we write $c(h)$ for the conjugacy class of h in H and $\mathcal{C}(H)$ for the set of conjugacy classes of H. We denote the derived subgroup of H by H'.

If L is a number field or a local field, and Γ is any group upon which Ω_L acts continuously, we identify Γ-torsors over L (as well as their associated algebras, which are Hopf–Galois extensions associated to $A_\Gamma := (L^c\Gamma)^{\Omega_L}$) with elements of the set $Z^1(\Omega_L, \Gamma)$ of Γ-valued continuous 1-cocycles of Ω_L (see [Serre 1997, I.5.2] and Section 2 below). If $\pi \in Z^1(\Omega_L, \Gamma)$, then we write L_π/L for the corresponding Hopf–Galois extension of L, and O_π for the integral closure of O_L in L_π. (Thus $O_\pi = L_\pi$ if L is an archimedean local field.) Each such L_π is a principal homogeneous space (p.h.s.) of the Hopf algebra $\operatorname{Map}_{\Omega_L}(\Gamma, L^c)$ of Ω_L-equivariant maps from Γ to L^c. It may be shown that if $\pi_1, \pi_2 \in Z^1(\Omega_L, \Gamma)$, then $L_{\pi_1} \simeq L_{\pi_2}$ if and only if π_1 and π_2 differ by a coboundary. The set of isomorphism classes of Γ-torsors over L may be identified with the pointed cohomology set $H^1(L, \Gamma) := H^1(\Omega_L, \Gamma)$. We write $[\pi] \in H^1(L, \Gamma)$ for the class of L_π in $H^1(L, \Gamma)$. If L is a number field or a nonarchimedean local field we write $H^1_t(L, \Gamma)$ for the subset of $H^1(L, \Gamma)$ consisting of those $[\pi] \in H^1(L, \Gamma)$ for which L_π/L is at most tamely ramified. If L is an archimedean local field, we set $H^1_t(L, G) = H^1(L, G)$. We denote the subset of $H^1_t(L, \Gamma)$ consisting of those $[\pi] \in H^1_t(L, \Gamma)$ for which L_π/L is unramified at all (including infinite) places of L by $H^1_{\mathrm{nr}}(L, \Gamma)$. (So, with this convention, if L is an archimedean local field, we have $H^1_{\mathrm{nr}}(L, \Gamma) = 0$.) If L is a number field, we write $H^1_{\mathrm{fnr}}(F, \Gamma)$ for the subset of $H^1_t(F, \Gamma)$ consisting of those $[\pi] \in H^1_t(F, \Gamma)$ for which L_π/L is unramified at all finite places of L.

If A is any algebra, we write $Z(A)$ for the centre of A. If A is semisimple, we write

$$\mathrm{nrd} : A^\times \to Z(A)^\times, \quad \mathrm{nrd} : K_1(A) \to Z(A)^\times$$

for the reduced norm maps on $A^\times$ and $K_1(A)$ respectively [Fröhlich 1983, Chapter II, §1]. If A is an R-algebra for some ring R, and $R \to R_1$ is an extension of R, we write $A_{R_1} := A \otimes_R R_1$ to denote extension of scalars from R to R_1.

If S_1 and S_2 are sets, we sometimes use the notation $S_1 \xrightarrow{\text{epi}} S_2$ to denote a surjective map from S_1 to S_2.

2. Principal homogeneous spaces and resolvends

In this section we shall describe some basic facts concerning principal homogeneous spaces and resolvends.

Throughout this section, the symbol L denotes either a number field or a local field.

Principal homogeneous spaces. [McCulloh 1987, §1; Byott 1998, §1]. Let Γ be any finite group upon which Ω_L acts continuously on the left, and write $Z^1(\Omega_L, \Gamma)$ for the set of Γ-valued continuous Ω_L 1-cocycles. If $\pi \in Z^1(\Omega_L, \Gamma)$, then we write ${}^\pi\Gamma$ for the set Γ endowed with the following modified action of Ω_L: if

$$\Gamma \to {}^\pi\Gamma, \quad \gamma \mapsto \bar{\gamma}$$

is the identity map on the underlying sets, then

$$\bar{\gamma}^\omega = \overline{\pi(\omega) \cdot \gamma^\omega}$$

for each $\gamma \in \Gamma$ and $\omega \in \Omega_L$. The group Γ acts on ${}^{\pi}\Gamma$ via right multiplication.

We define an associated L-algebra L_π by

$$L_\pi := \mathrm{Map}_{\Omega_L}({}^{\pi}\Gamma, L^c);$$

this is the algebra of L^c-valued functions on ${}^{\pi}\Gamma$ that are fixed under the action of Ω_L. The Hopf algebra

$$A = A_L := (L^c\Gamma)^{\Omega_L}$$

acts on L_π via the rule

$$(\alpha \cdot a)(\gamma) = \sum_{g \in \Gamma} \alpha_g \cdot a(\gamma \cdot g)$$

for all $\gamma \in \Gamma$ and $\alpha = \sum_{g\in\Gamma} \alpha_g \cdot g \in A$. The algebra L_π is a principal homogeneous space (p.h.s. for short) of the Hopf algebra

$$B := \mathrm{Map}_{\Omega_L}(\Gamma, L^c). \tag{2-1}$$

It may be shown that every p.h.s. of B is isomorphic to an algebra of the form L_π for some π, and so every such p.h.s. may be viewed as being a subset of the L^c-algebra $\mathrm{Map}(\Gamma, L^c)$. It is easy to check that

$$L_\pi \otimes_L L^c = L^c\Gamma \cdot \ell_\Gamma,$$

where $\ell_\Gamma \in \mathrm{Map}(\Gamma, L^c)$ is defined by

$$\ell_\Gamma(\gamma) = \begin{cases} 1 & \text{if } \gamma = 1, \\ 0 & \text{otherwise.} \end{cases}$$

This implies that L_π is a free, rank one A-module.

The Wedderburn decomposition of L_π may be described as follows. For any $\bar{\gamma} \in {}^{\pi}\Gamma$, write $\mathrm{Stab}(\bar{\gamma})$ for the stabiliser of $\bar{\gamma}$ in Ω_L, and set

$$L(\bar{\gamma}) := (L^c)^{\mathrm{Stab}(\bar{\gamma})}.$$

Then

$$L_\pi \simeq \prod_{\Omega_L \backslash {}^{\pi}\Gamma} L(\bar{\gamma}),$$

where $\Omega_L \backslash {}^{\pi}\Gamma$ denotes the set of Ω_L-orbits of ${}^{\pi}\Gamma$, and the product is taken over a set of orbit representatives. In general, the field $L(\bar{\gamma})$ is not normal over L. However, if Ω_L acts trivially on Γ, then $Z^1(\Omega_L, \Gamma) = \mathrm{Hom}(\Omega_L, \Gamma)$, and for each $\bar{\gamma} \in^{\pi} \Gamma$, we have

$$L(\bar{\gamma}) = (L^c)^{\mathrm{Ker}(\pi)} =: L^\pi, \tag{2-2}$$

with $\mathrm{Gal}(L^\pi/L) \simeq \pi(\Omega_L)$. In this case, we have that

$$L_\pi \simeq \prod_{\Gamma/\pi(\Omega_L)} L^\pi, \tag{2-3}$$

and this isomorphism depends only upon the choice of a transversal of $\pi(\Omega_L)$ in Γ.

Remark 2.1. For most of this paper we shall only need to consider the case in which Ω_L acts trivially on Γ; in this situation $A = L\Gamma$, and L_π is a Γ-Galois L-algebra. A notable exception to this will occur in Section 7, when we take L to be a nonarchimedean local field, and we construct a canonical subextension of a tame extension L_π/L (see Definitions 7.4 and 7.6). This canonical subextension is complementary to the maximal unramified subextension of L_π/L, and is not usually a Galois algebra extension of L. It is however, a p.h.s. of a Hopf algebra of the form (2-1) associated to a certain group Γ equipped (as a set) with a nontrivial Ω_L-action.

Resolvends. [McCulloh 1987, §1; Byott 1998, §2]. Since every p.h.s. of B may be viewed as being a subset of $\mathrm{Map}(\Gamma, L^c)$, it is natural to consider the Fourier transforms of elements of $\mathrm{Map}(\Gamma, L^c)$. These arise via the *resolvend map*

$$\boldsymbol{r}_\Gamma : \mathrm{Map}(\Gamma, L^c) \to L^c\Gamma, \quad a \mapsto \sum_{s\in\Gamma} a(s)s^{-1}.$$

The map $\boldsymbol{r}_\Gamma$ is an isomorphism of left $L^c\Gamma$-modules, but not of algebras, because it does not preserve multiplication. It is easy to show that for any $a \in \mathrm{Map}(\Gamma, L^c)$, we have that $a \in L_\pi$ if and only if $\boldsymbol{r}_\Gamma(a)^\omega = \boldsymbol{r}_\Gamma(a)\cdot\pi(\omega)$ for all $\omega \in \Omega_L$. It may also be shown that an element $a \in L_\pi$ generates L_π as an A-module if and only if $\boldsymbol{r}_\Gamma(a) \in (L^c\Gamma)^\times$. Two elements $a_1, a_2 \in \mathrm{Map}(\Gamma, L^c)$ with $\boldsymbol{r}_\Gamma(a_1), \boldsymbol{r}_\Gamma(a_2) \in (L^c\Gamma)^\times$ generate the same p.h.s. as an A-module if and only if $\boldsymbol{r}_\Gamma(a_1) = b\cdot\boldsymbol{r}_\Gamma(a_2)$ for some $b \in A^\times$. If a is any generator of L_π as an A-module, then a Γ-valued Ω_L 1-cocycle that represents the class $[\pi]$ of π in the pointed cohomology set $H^1(L, \Gamma)$ is given by

$$\omega \mapsto \boldsymbol{r}_\Gamma(a)^{-1}\cdot\boldsymbol{r}_\Gamma(a)^\omega.$$

We define pointed sets (where in each case the distinguished element is afforded by $1 \in A^\times_{L^c} = (L^c\Gamma)^\times$)

$$H(A) := \{\alpha \in A^\times_{L^c} : \alpha^{-1}\cdot\alpha^\omega \in \Gamma,\ \forall\omega \in \Omega_L\} \quad\text{and}\quad \mathcal{H}(A) := H(A)/\Gamma = \{\alpha\cdot\Gamma : \alpha \in H(A)\},$$

and we write $r_\Gamma(a) \in \mathcal{H}(A)$ for the image in $\mathcal{H}(A)$ of $\boldsymbol{r}_\Gamma(a) \in H(A)$. The element $r_\Gamma(a)$ is referred to as the *reduced resolvend* of a. If $\mathfrak{A}$ is any O_L-order in A, then we define $H(\mathfrak{A})$ and $\mathcal{H}(\mathfrak{A})$ in a similar manner. Hence we have

$$H(\mathfrak{A}) = \mathfrak{A}_{O_{L^c}} \cap H(A) \quad\text{and}\quad \mathcal{H}(\mathfrak{A}) = H(\mathfrak{A})/\Gamma.$$

Write L^t for the maximal, tamely ramified extension of L. We set

$$H_t(A) := \{\alpha \in H(A) : \alpha^\omega = \alpha,\ \forall\omega \in \Omega_{L^t}\} \quad\text{and}\quad \mathcal{H}_t(A) := H_t(A)/\Gamma = \{\alpha\cdot\Gamma : \alpha \in H_t(A)\},$$

and we define $H_t(\mathfrak{A})$ and $\mathcal{H}_t(\mathfrak{A})$ analogously for any O_L-order $\mathfrak{A}$ in A.

We shall now give a characterisation of the set $H(A)$ that avoids any explicit mention of Galois action. This is a nonabelian version of a description of $H(A)$ in terms of primitive elements of quotients of groups of units in Hopf algebras in the abelian case [Agboola and Burns 2006, Theorem 6.4].

In order to do this, we first note that there are Ω_L-equivariant homomorphisms of algebras

$$\Delta, i_1, i_2 : A_{L^c} \to A_{L^c} \otimes_{L^c} A_{L^c}$$

induced by the maps

$$\Delta(\gamma) = \gamma \otimes \gamma, \quad i_1(\gamma) = \gamma \otimes 1, \quad i_2(\gamma) = 1 \otimes \gamma$$

for $\gamma \in \Gamma$.

We define a map of pointed sets

$$\mathcal{P} : A_{L^c}^{\times} \to (A_{L^c} \otimes_{L^c} A_{L^c})^{\times}, \quad x \mapsto \Delta(x) \cdot [i_1(x) \cdot i_2(x)]^{-1}.$$

It is easy to verify that

$$\mathcal{P}(x_1 \cdot x_2) = \Delta(x_1) \cdot \mathcal{P}(x_2) \cdot [i_1(x_1) \cdot i_2(x_1)]^{-1}.$$

As $\mathcal{P}(\gamma) = 1$ for each $\gamma \in \Gamma$, it follows that $\mathcal{P}$ induces a map of pointed sets (which we denote by the same symbol)

$$\mathcal{P} : A_{L^c}^{\times}/\Gamma \to (A_{L^c} \otimes_{L^c} A_{L^c})^{\times}.$$

Theorem 2.2. *Let $x \in A_{L^c}^{\times}$. Then $x \in H(A)$ if and only if $\mathcal{P}(x) \in (A \otimes_L A)^{\times}$.*

Proof. Suppose that $x \in H(A)$. Then if $\omega \in \Omega_L$, we have

$$x^{\omega} = x \cdot \gamma_{\omega}$$

for some $\gamma_{\omega} \in \Gamma$. Hence

$$\begin{aligned}
[\Delta(x)(i_1(x)i_2(x))^{-1}]^{\omega} &= \Delta(x)(\gamma_\omega \otimes \gamma_\omega)[i_1(x)(\gamma_\omega \otimes 1)i_2(x)(1 \otimes \gamma_\omega)]^{-1} \\
&= \Delta(x)(\gamma_\omega \otimes \gamma_\omega)(1 \otimes \gamma_\omega)^{-1} i_2(x)^{-1}(\gamma_\omega \otimes 1)^{-1} i_1(x)^{-1} \\
&= \Delta(x)(\gamma_\omega \otimes \gamma_\omega)(1 \otimes \gamma_\omega)^{-1}(\gamma_\omega \otimes 1)^{-1} i_2(x)^{-1} i_1(x)^{-1} \\
&= \Delta(x)[i_1(x)i_2(x)]^{-1}.
\end{aligned}$$

This shows that

$$\mathcal{P}(x) \in [(A_{L^c} \otimes_{L^c} A_{L^c})^{\times}]^{\Omega_L} = (A \otimes_L A)^{\times}.$$

Suppose conversely that $\mathcal{P}(x) \in (A \otimes_L A)^{\times}$, and that $x^{\omega} = x \cdot u_{\omega}$ for each $\omega \in \Omega_L$. We wish to show that $u_\omega \in \Gamma$. As the maps Δ, i_1, and i_2 are Ω_L-equivariant, we have that

$$\Delta(x)^{\omega} = \Delta(x) \cdot \Delta(u_\omega), \quad i_1(x)^{\omega} = i_1(x) \cdot i_1(u_\omega), \quad i_2(x)^{\omega} = i_2(x) \cdot i_2(u_\omega),$$

and a straightforward computation shows that

$$\mathcal{P}(x)^{\omega} = \Delta(x) \cdot \mathcal{P}(u_\omega) \cdot [i_1(x) \cdot i_2(x)]^{-1}.$$

As $\mathcal{P}(x) = \mathcal{P}(x)^{\omega}$, this implies that $\mathcal{P}(u_\omega) = 1$, i.e., that

$$\Delta(u_\omega) = i_1(u_\omega) \cdot i_2(u_\omega).$$

It now follows that $u_\omega \in \Gamma$ via an argument identical to that given in [Agboola and Burns 2006, Theorem 6.4]. □

Let F be a number field. Our next result shows that the pointed set $H(A_F)$ of resolvends satisfies a Hasse principle.

Proposition 2.3. *Let F be a number field, and suppose that $x \in (F^c\Gamma)^\times$. Then $x \in H(A_F)$ if and only if $\mathrm{loc}_v(x) \in H(A_{F_v})$ for every finite place v of F.*

Proof. We first observe that the map $\mathcal{P}$ commutes with localisation, i.e., for each finite place v of F, we have

$$\mathrm{loc}_v(\mathcal{P}(x)) = \mathcal{P}(\mathrm{loc}_v(x)) \tag{2-4}$$

for all $x \in (F^c\Gamma)^\times$. Hence we have

$$\begin{aligned} x \in H(A_F) &\iff \mathcal{P}(x) \in (A_F \otimes_F A_F)^\times && \text{(from Theorem 2.2)}\\ &\iff \mathrm{loc}_v(\mathcal{P}(x)) \in (A_{F_v} \otimes_{F_v} A_{F_v})^\times && \text{for each finite } v\\ &\iff \mathcal{P}(\mathrm{loc}_v(x)) \in (A_{F_v} \otimes_{F_v} A_{F_v})^\times && \text{for each finite } v \text{ (from (2-4))}\\ &\iff \mathrm{loc}_v(x) \in H(A_{F_v}) && \text{for each finite } v \text{ (from Theorem 2.2).} \end{aligned}$$ □

Remark 2.4. It is also possible to give a proof of Proposition 2.3 directly from the definition of $H(A_F)$. The standard such proof that was known to the authors is valid only for abelian groups Γ; we are grateful to an anonymous referee for explaining how this proof may be modified so as to hold for arbitrary finite groups.

Suppose that $x \in A_{F^c}^\times$ is such that, for each finite place v of F, we have $\mathrm{loc}_v(x) \in H(A_{F_v})$. We wish to show that $x \in H(A_F)$.

Let E/F be any finite Galois extension such that Ω_E fixes x. Then the action of Ω_F on x factors through the action of the finite group $D := \mathrm{Gal}(E/F)$. Hence, to prove the desired result, it suffices to show that for any $\delta \in D$, we have $x^\delta = x \cdot \gamma_\delta$, with $\gamma_\delta \in \Gamma$.

Let $\mathcal{G}_F$ denote the subgroup of Ω_F generated by the subgroups Ω_{F_v} as v runs over the finite places of F. As each element of Ω_F is conjugate to an element of Ω_{F_v} for some v, it follows via the Chebotarev density theorem that the image $\overline{\mathcal{G}}_F$ of $\mathcal{G}_F$ in D has nontrivial intersection with every conjugacy class of D. A lemma of Jordan now implies that $\overline{\mathcal{G}}_F$ must be equal to the whole of D [Serre 2003, p. 435, Theorem 4']. The result we seek now follows at once.

3. Resolvends and cohomology

Recall that F is a number field and G is a finite group upon which Ω_F acts trivially. In this section, we explain, following [McCulloh 1987, §2], how resolvends may be used to compute discriminants of rings of integers of G-Galois extensions of F, and to describe certain Galois cohomology groups.

For each $[\pi] \in H^1(F, G)$, the standard trace map

$$\mathrm{Tr} : \mathrm{Map}(G, F^c) \to F^c$$

induces a trace map

$$\mathrm{Tr} : F_\pi \to F$$

via restriction. This in turn yields an associated, nondegenerate bilinear form $(a, b) \mapsto \mathrm{Tr}(ab)$ on F_π. If M is any full O_F-lattice in F_π, then we set

$$M^* := \{b \in F_\pi \mid \mathrm{Tr}(b \cdot M) \subseteq O_F\} \quad \text{and} \quad \mathrm{disc}(O_\pi/O_F) := [O_\pi^* : O_\pi]_{O_F},$$

where the symbol $[- : -]_{O_F}$ denotes the O_F-module index. We see from the isomorphism (2-3) that we have

$$\mathrm{disc}(O_\pi/O_F) = \mathrm{disc}(O_{F^\pi}/O_F)^{[G:\pi(\Omega_F)]},$$

where $\mathrm{disc}(O_{F^\pi}/O_F)$ denotes the usual discriminant of the number field F^π over F, and so it follows that

$$\mathrm{disc}(O_\pi/O_F) = O_F$$

if and only if F_π/F is unramified at all finite places of F.

Definition 3.1. We write $[-1]$ for the maps induced on $\mathrm{Map}(G, F^c)$ and F^cG by the map $g \mapsto g^{-1}$ on G.

Lemma 3.2. *Suppose that $a, b \in F_\pi$ for some $[\pi] \in H^1(F, G)$. Then*

$$\boldsymbol{r}_G(a) \cdot \boldsymbol{r}_G(b)^{[-1]} = \sum_{s \in G} \mathrm{Tr}(a^s b) \cdot s^{-1} \in FG.$$

Proof. This may be verified via a straightforward calculation (see, e.g., [McCulloh 1983, (1.6)], and note that the calculation given there is valid for an arbitrary finite group G). □

Corollary 3.3. *Suppose that $F_\pi = FG \cdot a$. Then we have*:

(i) $\boldsymbol{r}_G(a)^{-1} = \boldsymbol{r}_G(b)^{[-1]}$, *where $b \in F_\pi$ satisfies* $\mathrm{Tr}(a^s b^t) = \delta_{s,t}$.

(ii) $(O_F G \cdot a)^* = O_F G \cdot b$.

(iii) $[(O_F G \cdot a)^* : O_F G \cdot a]_{O_F} = [O_F G : O_F G \cdot \boldsymbol{r}_G(a) \cdot \boldsymbol{r}_G(a)^{[-1]}]_{O_F}$.

(iv) $\boldsymbol{r}_G(a) \in (O_{F^c} G)^\times$ *if and only if* $O_\pi = O_F G \cdot a$ *and* $\mathrm{disc}(O_\pi/O_F) = O_F$.

Analogous results hold if F is replaced by F_v for any finite place v of F.

Proof. Exactly as in [McCulloh 1987, 2.10 and 2.11]. □

Lemma 3.4. *Suppose that L is either a number field or a local field. Then*

(i) $H^1(L, (L^cG)^\times) = 1$,

(ii) $H^1(L, Z(L^cG)^\times) = 1$.

Proof. For each $\chi \in \mathrm{Irr}(G)$, write $d(\chi)$ for the degree of χ, and $M_{d(\chi)}(L^c)$ for the algebra of $d(\chi) \times d(\chi)$-matrices over L^c. Then the Wedderburn isomorphism of algebras

$$L^c G \simeq \bigoplus_{\chi \in \mathrm{Irr}(G)} M_{d(\chi)}(L^c)$$

yields isomorphisms of groups

$$(L^c G)^\times \simeq \bigoplus_{\chi \in \mathrm{Irr}(G)} \mathrm{GL}_{d(\chi)}(L^c), \quad Z(L^c G)^\times \simeq \bigoplus_{\chi \in \mathrm{Irr}(G)} (L^c)^\times.$$

Let $\chi_1, \dots, \chi_m \in \mathrm{Irr}(G)$ be a set of representatives of $\Omega_L \backslash \mathrm{Irr}(G)$. Write $\mathrm{Stab}(\chi_i)$ for the stabiliser of χ_i in Ω_L, and set $L[\chi_i] := (L^c)^{\mathrm{Stab}(\chi_i)}$. There are isomorphisms of Ω_L-modules

$$(L^c G)^\times \simeq \bigoplus_{i=1}^m \mathrm{Ind}_{\Omega_{L[\chi_i]}}^{\Omega_L}(\mathrm{GL}_{d(\chi_i)}(L^c)), \quad Z(L^c G)^\times \simeq \bigoplus_{i=1}^m \mathrm{Ind}_{\Omega_{L[\chi_i]}}^{\Omega_L}(L^c)^\times.$$

We have

$$H^1(L, (L^c G)^\times) \simeq H^1(L, \bigoplus_{i=1}^m \mathrm{Ind}_{\Omega_{L[\chi_i]}}^{\Omega_L} \mathrm{GL}_{d(\chi_i)}(L^c)) \simeq \bigoplus_{i=1}^m H^1(L[\chi_i], \mathrm{GL}_{d(\chi_i)}(L^c)) = 1,$$

where the second isomorphism follows via Shapiro's lemma and the final equality is a standard consequence of Hilbert's Theorem 90. This proves (i). The proof of (ii) is very similar. □

Recall that two pointed sets S_1 and S_2 are said to be *isomorphic* if there is a bijection of sets

$$f : S_1 \to S_2$$

with $f(x_1) = f(x_2)$, where x_i is the distinguished element of S_i, $(i = 1, 2)$.

A sequence

$$\cdots \to S_{i-1} \xrightarrow{f_i} S_i \xrightarrow{f_{i+1}} S_{i+1} \to \cdots$$

of pointed sets is said to be *exact* if there is an equality of sets

$$\mathrm{Im}(f_i) = f_{i+1}^{-1}(x_{i+1}),$$

where x_{i+1} is the distinguished element of S_{i+1}.

Theorem 3.5. (1) *There is an exact sequence of pointed sets*

$$1 \to G \to (FG)^\times \to \mathcal{H}(FG) \to H^1(F, G) \to 1. \tag{3-1}$$

(2) *For each finite place v of F, recall that $H^1_{\mathrm{nr}}(F_v, G)$ denotes the subset of $H^1(F_v, G)$ consisting of those $[\pi_v] \in H^1(F_v, G)$ for which the associated G-Galois extension F_{π_v}/F_v is unramified. Then there is an exact sequence of pointed sets*

$$1 \to G \to (O_{F_v} G)^\times \to \mathcal{H}(O_{F_v} G) \to H^1_{\mathrm{nr}}(F_v, G) \to 1. \tag{3-2}$$

(3) *There are exact sequences of pointed sets*

$$1 \to G \to (FG)^\times \to \mathcal{H}_t(FG) \to H^1_t(F, G) \to 1, \tag{3-3}$$

and

$$1 \to G \to (F_v G)^\times \to \mathcal{H}_t(F_v G) \to H^1_t(F_v, G) \to 1 \tag{3-4}$$

for each place v of F.

Proof. When G is abelian, parts (a) and (b) are proved in [McCulloh 1987, p. 268 and p. 273] by considering the Ω_F and Ω_{F_v}-cohomology of the exact sequences of abelian groups

$$1 \to G \to (F^c G)^\times \to (F^c G)^\times / G \to 1 \tag{3-5}$$

and

$$1 \to G \to (O_{F_v^c} G)^\times \to (O_{F_v^c} G)^\times / G \to 1$$

respectively. If G is nonabelian, and these exact sequences are viewed as exact sequences of pointed sets instead, then a similar proof of part (a) also holds, as is pointed out in [McCulloh 1987, p. 268]: taking Ω_F-cohomology of the exact sequence (3-5) of pointed sets yields an exact sequence

$$1 \to G \to (FG)^\times \to \mathcal{H}(FG) \to H^1(F, G) \to H^1(F, (F^c G)^\times), \tag{3-6}$$

and since $H^1(F, (F^c G)^\times) = 1$ (see Lemma 3.4(i)), (3-1) immediately follows.

Alternatively, we could also argue directly (as is done in [McCulloh 1987]) that the map $\mathcal{H}(FG) \to H^1(F, G)$ in (3-6) is surjective. Let us briefly describe the argument given in [McCulloh 1987]. Suppose that $[\pi] \in H^1(F, G)$, and let $a \in F_\pi$ be a normal basis generator of F_π / F. Set $\alpha = \boldsymbol{r}_G(a)$; then the coset $\alpha \cdot G \in \mathcal{H}(FG)$ lies in the preimage of $[\pi]$, and so it follows that (3-6) is indeed surjective on the right, as claimed.

Part (b) follows from Corollary 3.3(iv) (cf. the proof of (2.12) on [McCulloh 1987, p. 273]).

The proof of (c) is very similar to that of (a). Let F^t and F^t_v denote the maximal tamely ramified extensions of F and F_v respectively, and set $\Omega^t_F := \mathrm{Gal}(F^t/F)$, $\Omega^t_{F_v} := \mathrm{Gal}(F^t_v/F_v)$. Then (c) follows via considering the Ω^t_F and $\Omega^t_{F_v}$-cohomology of the exact sequences of pointed sets

$$1 \to G \to (F^t G)^\times \to (F^t G)^\times / G \to 1$$

and

$$1 \to G \to (F^t_v G)^\times \to (F^t_v G)^\times / G \to 1$$

respectively, using the direct argument given in [McCulloh 1987, p. 268] that we have described above. □

Suppose that L is a number field or a local field. Recall that $Z(LG)$ denotes the centre of LG. Before stating our next result, we note that the reduced norm map

$$\mathrm{nrd} : (LG)^\times \to Z(LG)^\times$$

induces an injection $G^{\mathrm{ab}} \to Z(LG)^{\times}$. (More explicitly, if we identify $Z(L^cG)^{\times}$ with $\prod_{\chi \in \mathrm{Irr}(G)} (L^c)^{\times}$ via the Wedderburn decomposition of L^cG (see the proof of Lemma 3.4), then the injection $G^{\mathrm{ab}} \to Z(L^G)^{\times}$ is induced by the map $G \to Z(L^cG)^{\times}$ given by $g \mapsto [(\det(\chi))(g)]_{\chi}$, where $\det(\chi)$ is the abelian character of G defined below in Definition 4.3. See also (4-5).) In what follows, we shall identify G^{ab} with its image in $Z(LG)^{\times}$ under this map. We set

$$\begin{aligned} H(Z(LG)) &:= \{\alpha \in Z(L^cG)^{\times} : \alpha^{-1} \cdot \alpha^{\omega} \in G^{\mathrm{ab}},\ \forall \omega \in \Omega_L\}, \\ \mathcal{H}(Z(LG)) &:= H(Z(LG))/G^{\mathrm{ab}} = \{\alpha \cdot G^{\mathrm{ab}} : \alpha \in H(Z(LG))\}. \end{aligned}$$

We define $H(Z(\mathfrak{A}))$ and $\mathcal{H}(Z(\mathfrak{A}))$ analogously for any O_L-order $\mathfrak{A}$ in LG.

Proposition 3.6. *Let L be a number field or a local field. Then there is an exact sequence of abelian groups*:

$$1 \to G^{\mathrm{ab}} \to Z(LG)^{\times} \to \mathcal{H}(Z(LG)) \to H^1(L, G^{\mathrm{ab}}) \to 1. \tag{3-7}$$

Proof. This follows at once from taking Ω_L cohomology of the exact sequence of abelian groups

$$1 \to G^{\mathrm{ab}} \to Z(L^cG)^{\times} \to Z(L^cG)^{\times}/G^{\mathrm{ab}} \to 1,$$

arising from the injection $G^{\mathrm{ab}} \to Z(L^cG)^{\times}$ induced by the reduced norm map $\mathrm{nrd} : (LG)^{\times} \to Z(LG)^{\times}$ as described above, and noting that $H^1(\Omega_L, Z(L^cG)^{\times}) = 1$, via Lemma 3.4(ii). □

It is easy to see that the group $(LG)^{\times}$ acts on the pointed set $\mathcal{H}(LG)$ by left multiplication. Write $(LG)^{\times} \backslash \mathcal{H}(LG)$ for the quotient set afforded by this action. It follows from Theorem 3.5 and Proposition 3.6 that there are isomorphisms

$$H^1(L, G) \xrightarrow{\sim} (LG)^{\times} \backslash \mathcal{H}(LG) \quad \text{and} \quad H^1(L, G^{\mathrm{ab}}) \xrightarrow{\sim} Z(LG)^{\times} \backslash \mathcal{H}(Z(LG))$$

of pointed sets and abelian groups respectively, and that the following diagram commutes:

$$\begin{tikzcd} H^1(L, G) \arrow[r, "\sim"] \arrow[d] & (LG)^{\times} \backslash \mathcal{H}(LG) \arrow[d, "\mathrm{nrd}"] \\ H^1(L, G^{\mathrm{ab}}) \arrow[r, "\sim"] & Z(LG)^{\times} \backslash \mathcal{H}(Z(LG)). \end{tikzcd} \tag{3-8}$$

(Here the left-hand vertical arrow is induced by the quotient map $G \to G^{\mathrm{ab}}$, while the right-hand vertical arrow is induced by the reduced norm map $\mathrm{nrd} : (L^cG)^{\times} \to Z(L^cG)^{\times}$.)

We shall need the following result in Section 6.

Proposition 3.7. *Let F be a number field. For each finite place v of F, the image of the map*

$$\mathrm{nrd} : (O_{F_v}G)^{\times} \backslash \mathcal{H}(O_{F_v}G) \to Z(O_{F_v}G)^{\times} \backslash \mathcal{H}(Z(O_{F_v}G))$$

of pointed sets is in fact a group.

Proof. Just as in the case of (3-8), we see from the exact sequences (3-2) and (3-7) that there is a commutative diagram

$$\begin{tikzcd}
H^1_{\mathrm{nr}}(F_v, G) \arrow[r, "\sim"] \arrow[d] & (O_{F_v}G)^\times \backslash \mathcal{H}(O_{F_v}G) \arrow[d, "\mathrm{nrd}"] \\
H^1_{\mathrm{nr}}(F_v, G^{\mathrm{ab}}) \arrow[r] \arrow[d, "\cap"] & Z(O_{F_v}G)^\times \backslash \mathcal{H}(Z(O_{F_v}G)) \arrow[d, "\cap"] \\
H^1(F_v, G^{\mathrm{ab}}) \arrow[r, "\sim"] & Z(F_vG)^\times \backslash \mathcal{H}(Z(F_vG)).
\end{tikzcd} \tag{3-9}$$

The middle horizontal arrow of this commutative diagram is therefore injective, and its image is a subgroup of $Z(O_{F_v}G)^\times \backslash \mathcal{H}(Z(O_{F_v}G))$. Hence, to prove the desired result, it suffices to show that the map $H^1_{\mathrm{nr}}(F_v, G) \to H^1_{\mathrm{nr}}(F_v, G^{\mathrm{ab}})$ is surjective. This is in turn an immediate consequence of the fact that the Galois group $\mathrm{Gal}(F_v^{\mathrm{nr}}/F_v)$ is profinite free on a single generator. □

4. Determinants and character maps

In this section we shall describe how determinants of resolvends may be represented in terms of certain character maps.

Let L be a number field or a local field.

Suppose that Γ is any finite group upon which the absolute Galois group Ω_L of L acts (possibly trivially). Then Ω_L also acts on the ring R_Γ of virtual characters of Γ according to the following rule: if $\chi \in \mathrm{Irr}(\Gamma)$ and $\omega \in \Omega_L$, then, for each $\gamma \in \Gamma$, we have $\chi^\omega(\gamma) = \omega(\chi(\omega^{-1}(\gamma)))$.

We begin by recalling some well-known facts and definitions concerning determinant maps (see, e.g., [Fröhlich 1983, Chapter II; 1984, Chapter I]).

Definition 4.1. For each element a of $\mathrm{GL}_n(L^cG)$, we define an element

$$\mathrm{Det}(a) \in \mathrm{Hom}(R_G, (L^c)^\times) \simeq Z(L^cG)^\times \tag{4-1}$$

in the following way: if T is any representation of G over L^c with character ϕ, then we set

$$\mathrm{Det}(a)(\phi) := \det(T(a)).$$

It may be shown that this definition depends only upon the character ϕ, and not upon the choice of representation T. The map

$$\mathrm{Det} : \mathrm{GL}_n(L^cG) \to \mathrm{Hom}(R_G, (L^c)^\times)$$

is Ω_L-equivariant, and so induces a map

$$\mathrm{Det} : \mathrm{GL}_n(LG) \to \mathrm{Hom}_{\Omega_L}(R_G, (L^c)^\times).$$

Remark 4.2. The map Det in (4-1) above is essentially the same as the reduced norm map. Let

$$\mathrm{nrd} : (L^cG)^\times \to Z(L^cG)^\times \tag{4-2}$$

denote the reduced norm. Then (4-2) induces an isomorphism

$$\mathrm{nrd}: K_1(L^cG) \xrightarrow{\sim} Z(L^cG)^\times \simeq \mathrm{Hom}(R_G, (L^c)^\times) \tag{4-3}$$

(see, e.g., [Curtis and Reiner 1987, Theorem 45.3]). Suppose now that ϕ is any L^c-valued character of G and let $a \in (L^cG)^\times$. Then we have that

$$\mathrm{Det}(a)(\phi) = \mathrm{nrd}(a)(\phi)$$

(see [Fröhlich 1984, Chapter I, Proposition 2.7]).

Definition 4.3. Suppose that $\chi \in \mathrm{Irr}(G)$. We define an abelian character $\det(\chi)$ of G as follows. Let T be any representation of G over L^c affording χ. For each element $g \in G$, we set

$$(\det(\chi))(g) = \mathrm{Det}(T(g)).$$

Then $\det(\chi)$ is independent of the choice of T, and may be viewed as being a character of G^{ab}. We extend det to a homomorphism $R_G \to (G^{\mathrm{ab}})^\wedge$, where $(G^{\mathrm{ab}})^\wedge$ denotes the group of characters of G^{ab}, by defining

$$\det\Biggl(\sum_{\chi\in\mathrm{Irr}(G)} a_\chi \chi\Biggr) = \prod_{\chi\in\mathrm{Irr}(G)} (\det(\chi))^{a_\chi},$$

and we set

$$A_G := \mathrm{Ker}(\det).$$

Hence we have an exact sequence of groups

$$0 \to A_G \to R_G \xrightarrow{\det} (G^{\mathrm{ab}})^\wedge \to 0. \tag{4-4}$$

Applying the functor $\mathrm{Hom}(-, (L^c)^\times)$ to (4-4), we obtain an exact sequence

$$0 \to G^{\mathrm{ab}} \to \mathrm{Hom}(R_G, (L^c)^\times) \xrightarrow{\mathrm{rag}} \mathrm{Hom}(A_G, (L^c)^\times) \to 0, \tag{4-5}$$

which is surjective on the right because $(L^c)^\times$ is divisible. It follows that there are Ω_L-equivariant isomorphisms

$$\mathrm{Hom}(A_G, (L^c)^\times) \simeq \mathrm{Hom}(R_G, (L^c)^\times)/G^{\mathrm{ab}} \simeq Z(L^cG)^\times/G^{\mathrm{ab}}. \tag{4-6}$$

In what follows, we shall sometimes identify $\mathrm{Hom}(A_G, (L^c)^\times)$ with $Z(L^cG)^\times/G^{\mathrm{ab}}$ via (4-6) without explicit mention.

Taking Ω_L-cohomology of (4-5) yields an exact sequence

$$0 \to G^{\mathrm{ab}} \to \mathrm{Hom}_{\Omega_L}(R_G, (L^c)^\times) \xrightarrow{\mathrm{rag}} \mathrm{Hom}_{\Omega_L}(A_G, (L^c)^\times) \to H^1(L, G^{\mathrm{ab}}) \to 1, \tag{4-7}$$

which is surjective on the right via Lemma 3.4(ii).

Definition 4.4. Let R_G^s denote the (additive) subgroup of R_G generated by the symplectic characters of G. Thus, R_G^s is generated by the irreducible symplectic characters of G, together with elements of the form $\chi + \overline{\chi}$, where $\chi \in R_G$ and $\overline{\chi}$ denotes the complex conjugate of χ. All virtual characters lying in R_G^s are real-valued.

If F is a number field, and v is a real place of F, we write

$$\mathrm{Hom}^{+}_{\Omega_{F_v}}(R_G, (F_v^c)^\times)$$

for those elements $f \in \mathrm{Hom}_{\Omega_{F_v}}(R_G, (F_v^c)^\times)$ for which $f(\eta) > 0$ for all $\eta \in R_G^s$. Note that if $f \in \mathrm{Hom}_{\Omega_{F_v}}(R_G, (F_v^c)^\times)$ and $\chi \in R_G$, then we automatically have

$$f(\chi + \overline{\chi}) = f(\chi) \cdot \overline{f(\chi)} > 0.$$

Hence in fact $f \in \mathrm{Hom}^{+}_{\Omega_{F_v}}(R_G, (F_v^c)^\times)$ if and only if f is positive on all irreducible, symplectic characters of G. In particular, if G has no nontrivial irreducible symplectic characters (e.g., if $|G|$ is odd), then we have

$$\mathrm{Hom}^{+}_{\Omega_{F_v}}(R_G, (F_v^c)^\times) = \mathrm{Hom}_{\Omega_{F_v}}(R_G, (F_v^c)^\times).$$

We write $Z(F_vG)^\times_+$ for the image of $\mathrm{Hom}^{+}_{\Omega_{F_v}}(R_G, (F_v^c)^\times)$ in $Z(F_vG)^\times$ under the isomorphism

$$\mathrm{Hom}_{\Omega_{F_v}}(R_G, (F_v^c)^\times) \xrightarrow{\sim} Z(F_vG)^\times.$$

Proposition 4.5. *Let F be a number field. For each place v of F, we write*

$$\mathrm{Det} : (F_v^cG)^\times \to \mathrm{Hom}(R_G, (F_v^c)^\times) \simeq Z(F_v^cG)^\times \tag{4-8}$$

for the determinant homomorphism afforded by Definition 4.1.

(1) *If v is real, then* (4-8) *induces an isomorphism*

$$\mathrm{Det}((F_vG)^\times) \simeq \mathrm{Hom}^{+}_{\Omega_{F_v}}(R_G, (F_v^c)^\times) \simeq Z(F_vG)^\times_+. \tag{4-9}$$

(2) *If v is finite or complex, then the map* (4-8) *induces isomorphisms*

$$\mathrm{Det}((F_vG)^\times) \simeq \mathrm{Hom}_{\Omega_{F_v}}(R_G, (F_v^c)^\times) \simeq Z(F_vG)^\times, \tag{4-10}$$

$$\mathrm{Det}(\mathcal{H}(F_vG)) \simeq \mathrm{Hom}_{\Omega_{F_v}}(A_G, (F_v^c)^\times). \tag{4-11}$$

(3) *If v is finite of residue characteristic coprime to $|G|$, so $O_{F_v}G$ is an O_{F_v}-maximal order in F_vG, then* (4-8) *induces isomorphisms*

$$\mathrm{Det}((O_{F_v}G)^\times) \simeq \mathrm{Hom}_{\Omega_{F_v}}(R_G, (O_{F_v^c})^\times) \simeq Z(O_{F_v}G)^\times, \tag{4-12}$$

$$\mathrm{Det}(\mathcal{H}(O_{F_v}G)) \simeq \mathrm{Hom}_{\Omega_{F_v}}(A_G, (O_{F_v^c})^\times). \tag{4-13}$$

Proof. The isomorphisms (4-9), (4-10) and (4-12) are standard and are explained in e.g., [Fröhlich 1983, Chapter II, §1].

Suppose that v is either finite or complex. Theorem 3.5(a) and (4-10) yield the commutative diagram

$$\begin{tikzcd}
G \arrow[r, "\subseteq"] \arrow[d] & (F_vG)^\times \arrow[r] \arrow[d, "\mathrm{Det}"] & \mathcal{H}(F_vG) \arrow[r, "\mathrm{epi}"] \arrow[d, "\mathrm{Det}"] & H^1(F_v, G) \arrow[d, "\mathrm{epi}"] \\
G^{\mathrm{ab}} \arrow[r, "\subseteq"] \arrow[d, equal] & \mathrm{Det}((F_vG)^\times) \arrow[r] \arrow[d, "\sim"] & \mathrm{Det}(\mathcal{H}(F_vG)) \arrow[r, "\mathrm{epi}"] \arrow[d] & H^1(F_v, G^{\mathrm{ab}}) \arrow[d, equal] \\
G^{\mathrm{ab}} \arrow[r, "\subseteq"] & \mathrm{Hom}_{\Omega_{F_v}}(R_G, (F_v^c)^\times) \arrow[r] & \mathrm{Hom}_{\Omega_{F_v}}(A_G, (F_v^c)^\times) \arrow[r, "\mathrm{epi}"] & H^1(F_v, G^{\mathrm{ab}}),
\end{tikzcd} \tag{4-14}$$

and this implies that the map

$$\mathrm{Det}(\mathcal{H}(F_vG)) \to \mathrm{Hom}_{\Omega_{F_v}}(A_G, (F_v^c)^\times)$$

is an isomorphism, which proves (4-11).

Suppose now that v is finite of residue characteristic coprime to $|G|$. In order to establish (4-13), we first observe that applying the functor $\mathrm{Hom}(-, (O_{F_v^c})^\times)$ to the exact sequence (4-4) yields a sequence

$$0 \to G^{\mathrm{ab}} \to \mathrm{Hom}(R_G, (O_{F_v^c})^\times) \to \mathrm{Hom}(A_G, (O_{F_v^c})^\times) \to 1 \tag{4-15}$$

which is surjective on the right because $(O_{F_v^c})^\times$ is divisible. Taking Ω_{F_v}-cohomology of (4-15) yields

$$\begin{aligned} 0 \to G^{\mathrm{ab}} \to \mathrm{Hom}_{\Omega_{F_v}}(R_G, (O_{F_v^c})^\times) \to \mathrm{Hom}_{\Omega_{F_v}}(A_G, (O_{F_v^c})^\times) \to \\ \to H^1(F_v, G^{\mathrm{ab}}) \xrightarrow{f} H^1(F_v, \mathrm{Hom}(R_G, (O_{F_v^c})^\times)). \end{aligned} \tag{4-16}$$

Now since v does not divide the order of G, $Z(O_{F_v}G)$ is an O_{F_v}-maximal order in (the split algebra) $Z(F_vG)$ and

$$Z(O_{F_v^c}G)^\times \simeq \mathrm{Hom}(R_G, (O_{F_v^c})^\times)$$

(see (4-12)). Suppose that $\pi \in \mathrm{Ker}(f)$. Then there exists $u \in Z(O_{F_v^c}G)^\times$ such that $u^\omega \cdot u^{-1} = \pi(\omega)$ for all $\omega \in \Omega_{F_v}$. This implies that $u^{|G^{\mathrm{ab}}|} \in Z(O_{F_v}G)^\times$. As $v \nmid |G^{\mathrm{ab}}|$ and $Z(O_{F_v}G)$ is a maximal order, it follows that $u \in Z(O_{F_v^{\mathrm{nr}}}G)^\times$, and so $\pi \in H^1_{\mathrm{nr}}(F_v, G^{\mathrm{ab}})$. Hence there is an exact sequence

$$0 \to G^{\mathrm{ab}} \to \mathrm{Hom}_{\Omega_{F_v}}(R_G, (O_{F_v^c})^\times) \to \mathrm{Hom}_{\Omega_{F_v}}(A_G, (O_{F_v^c})^\times) \to H^1_{\mathrm{nr}}(F_v, G^{\mathrm{ab}}). \tag{4-17}$$

We recall also (see the proof of Proposition 3.7) that the natural map $H^1_{\mathrm{nr}}(F_v, G) \to H^1_{\mathrm{nr}}(F_v, G^{\mathrm{ab}})$ is surjective because the group $\mathrm{Gal}(F_v^{\mathrm{nr}}/F_v)$ is profinite free on a single generator. Theorem 3.5(b) together with (4-12) and (4-17) now yield the following commutative diagram:

$$\begin{tikzcd}
G \arrow[r, "\subseteq"] \arrow[d] & (O_{F_v}G)^{\times} \arrow[r] \arrow[d, "\mathrm{Det}"] & \mathcal{H}(O_{F_v}G) \arrow[r, "\mathrm{epi}"] \arrow[d, "\mathrm{Det}"] & H^1_{\mathrm{nr}}(F_v, G) \arrow[d, "\mathrm{epi}"] \\
G^{\mathrm{ab}} \arrow[r, "\subseteq"] \arrow[d, equal] & \mathrm{Det}((O_{F_v}G)^{\times}) \arrow[r] \arrow[d, "\sim"] & \mathrm{Det}(\mathcal{H}(O_{F_v}G)) \arrow[r, "\mathrm{epi}"] \arrow[d] & H^1_{\mathrm{nr}}(F_v, G^{\mathrm{ab}}) \arrow[d, equal] \\
G^{\mathrm{ab}} \arrow[r, "\subseteq"] & \mathrm{Hom}_{\Omega_{F_v}}(R_G, (O_{F_v^c})^{\times}) \arrow[r] & \mathrm{Hom}_{\Omega_{F_v}}(A_G, (O_{F_v^c})^{\times}) \arrow[r] & H^1_{\mathrm{nr}}(F_v, G^{\mathrm{ab}}).
\end{tikzcd} \tag{4-18}$$

It follows from (4-18) that the third row of this diagram is surjective on the right. Since $\mathrm{Det}(\mathcal{H}(O_{F_v}G))$ is a subgroup of $\mathrm{Hom}_{\Omega_{F_v}}(A_G, (O_{F_v^c})^{\times})$, we see that the map

$$\mathrm{Det}(\mathcal{H}(O_{F_v}G)) \to \mathrm{Hom}_{\Omega_{F_v}}(A_G, (O_{F_v^c})^{\times})$$

is an isomorphism. This establishes (4-13). □

If on the other hand v is finite and $v \mid |G|$, so $O_{F_v}G$ is not an O_{F_v}-maximal order in F_vG, then we have

$$\mathrm{Det}(\mathcal{H}(O_{F_v}G)) \subseteq \mathrm{Hom}_{\Omega_{F_v}}(A_G, (O^c_{F_v})^{\times}),$$

but this inclusion is not in general an equality. If $\mathfrak{a}$ is any integral ideal of O_F, set

$$U_{\mathfrak{a}}(O_{F_v^c}) := (1 + \mathfrak{a}O_{F_v^c}) \cap (O_{F_v^c})^{\times},$$

and write $U_{\alpha}(O_{F_v^c})$ instead of $U_{\mathfrak{a}}(O_{F_v^c})$ when $\mathfrak{a} = \alpha O_F$. We shall need the following result of A. Siviero (which is a variant of [McCulloh 1987, Theorem 2.14]) in Section 11.

Proposition 4.6 (A. Siviero). *Let v be a finite place of F. Then if $N \in \mathbb{Z}_{>0}$ is divisible by a sufficiently large power of $|G|$, we have*

$$\mathrm{Hom}_{\Omega_{F_v}}(A_G, U_N(O_{F_v^c})) \subseteq \mathrm{Det}(\mathcal{H}(O_{F_v}G)) \subseteq \mathrm{Hom}_{\Omega_{F_v}}(A_G, (O_{F_v^c})^{\times}).$$

Proof. This is shown in [Siviero 2013, Theorem 5.1.10] when G is abelian, and the proof for arbitrary finite G is quite similar. As the reference is not widely accessible, we describe the argument.

If $v \nmid |G|$, then Proposition 4.5(iii) implies that we have

$$\mathrm{Hom}_{\Omega_{F_v}}(A_G, O^{\times}_{F_v^c}) = \mathrm{Det}(\mathcal{H}(O_{F_v}G)) = \mathrm{Hom}_{\Omega_{F_v}}(A_G, (O_{F_v^c})^{\times}),$$

and so it follows that the desired result holds in this case. We may therefore suppose that $v \mid |G|$.

We first observe that the group

$$\frac{\mathrm{Hom}_{\Omega_{F_v}}(A_G, (O_{F_v^c})^{\times})}{\mathrm{Det}((O_{F_v}G)^{\times}/G)}$$

is annihilated by $|G^{\mathrm{ab}}|[\mathrm{Det}(\mathcal{M}_v^{\times}) : \mathrm{Det}(O_{F_v}G)^{\times}]$, where $\mathcal{M}_v$ denotes any O_{F_v}-maximal order in F_vG containing $O_{F_v}G$. Since A_G is finitely generated, it follows that $\mathrm{Det}((O_{F_v}G)^{\times}/G)$ is of finite index in

$\mathrm{Hom}_{\Omega_{F_v}}(A_G, (O_{F_v^c})^{\times})$, and so is an open subgroup of $\mathrm{Hom}_{\Omega_{F_v}}(A_G, (O_{F_v^c})^{\times})$. The result now follows from the fact that, because $v \mid |G|$, the collection of groups

$$\{\mathrm{Hom}_{\Omega_{F_v}}(A_G, U_{|G|^n}(O_{F_v^c})) \mid n \geq 0\}$$

is a fundamental system of neighbourhoods of the identity of $\mathrm{Hom}_{\Omega_{F_v}}(A_G, (O_{F_v^c})^{\times})$. □

Remark 4.7. When G is abelian, it follows from [McCulloh 1987, Theorem 2.14] that we may take $N = |G|^2$ in Proposition 4.6.

We shall also require the following related result in Section 15.

Proposition 4.8. *Let Γ be a finite group with an action of Ω_F. Suppose that $v \mid |\Gamma|$ is a finite place of F, and write $\mathfrak{p}_v$ for the maximal ideal of O_{F_v}. Then for all sufficiently large n, we have*

$$\mathrm{Hom}_{\Omega_{F_v}}(A_\Gamma, U_{\mathfrak{p}_v^n}(O_{F_v^c})) \subseteq \mathrm{rag}[\mathrm{Hom}_{\Omega_{F_v^c}}(R_\Gamma, (O_{F_v^c})^{\times})].$$

Proof. The proof of this is very similar to that of Proposition 4.6. We observe that

$$|\Gamma^{\mathrm{ab}}| \cdot \mathrm{Hom}_{\Omega_{F_v^c}}(A_\Gamma, (O_{F_v^c})^{\times}) \subseteq \mathrm{rag}[\mathrm{Hom}_{\Omega_{F_v^c}}(R_\Gamma, (O_{F_v^c})^{\times})],$$

which implies that $\mathrm{rag}[\mathrm{Hom}_{\Omega_{F_v^c}}(R_\Gamma, (O_{F_v^c})^{\times})]$ is an open subgroup of $\mathrm{Hom}_{\Omega_{F_v^c}}(A_\Gamma, (O_{F_v^c})^{\times})$ because A_Γ is finitely generated. The desired result now follows since the collection of groups $\{\mathrm{Hom}_{\Omega_{F_v}}(A_\Gamma, U_{\mathfrak{p}_v^n}(O_{F_v^c})) \mid n \geq 0\}$ is a fundamental system of neighbourhoods of the identity of $\mathrm{Hom}_{\Omega_{F_v^c}}(A_\Gamma, (O_{F_v^c})^{\times})$. □

5. Twisted forms and relative K-groups

Recall that G is a finite group upon which Ω_F acts trivially. In this section, we shall recall some basic facts concerning categorical twisted forms and relative algebraic K-groups. The reader may consult [Agboola and Burns 2006; Swan 1968, Chapter 15] for some of the details that we omit.

Twisted forms. Suppose that R is a Dedekind domain with field of fractions L of characteristic zero. (For notational convenience, we shall sometimes also allow ourselves to take $R = L$.) Let $\mathfrak{A}$ be any R-algebra which is finitely generated as an R-module and which satisfies $\mathfrak{A} \otimes_R L \simeq LG$.

Definition 5.1. Let Λ be any extension of R, and write $\mathcal{P}(\mathfrak{A})$ and $\mathcal{P}(\mathfrak{A} \otimes_R \Lambda)$ for the categories of finitely generated, projective $\mathfrak{A}$ and $\mathfrak{A} \otimes_R \Lambda$-modules respectively. A *categorical* Λ*-twisted* $\mathfrak{A}$*-form* (or *twisted form* for short) is an element of the fibre product category $\mathcal{P}(\mathfrak{A}) \times_{\mathcal{P}(\mathfrak{A} \otimes_R \Lambda)} \mathcal{P}(\mathfrak{A})$, where the fibre product is taken with respect to the functor $\mathcal{P}(\mathfrak{A}) \to \mathcal{P}(\mathfrak{A} \otimes_R \Lambda)$ afforded by extension of scalars. In concrete terms therefore, a twisted form consists of a triple $(M, N; \xi)$, where M and N are finitely generated, projective $\mathfrak{A}$-modules, and

$$\xi : M \otimes_R \Lambda \xrightarrow{\sim} N \otimes_R \Lambda$$

is an isomorphism of $\mathfrak{A} \otimes_R \Lambda$-modules.

Example 5.2. If F_π/F is any G-extension and $\mathcal{L}_\pi \subseteq F_\pi$ is any nonzero projective $O_F G$-module, then $(\mathcal{L}_\pi, O_F G; \boldsymbol{r}_G)$ is a categorical F^c-twisted $O_F G$-form. In particular, if F_π/F is a tame G-extension, then $(O_\pi, O_F G; \boldsymbol{r}_G)$ is a categorical F^c-twisted $O_F G$-form. Similarly, if v is any place of F, then (still assuming F_π/F to be tame) $(O_{\pi,v}, O_{F_v} G; \boldsymbol{r}_G)$ is a categorical F_v^c-twisted $O_{F_v} G$-form. We shall mainly be concerned with twisted forms of these types in this paper.

We write $K_0(\mathfrak{A}, \Lambda)$ for the Grothendieck group associated to the fibre product category $\mathcal{P}(\mathfrak{A}) \times_{\mathcal{P}(\mathfrak{A} \otimes_R \Lambda)} \mathcal{P}(\mathfrak{A})$, and we write $[M, N; \xi]$ for the isomorphism class of the twisted form $(M, N; \xi)$ in $K_0(\mathfrak{A}, \Lambda)$. The group $K_0(\mathfrak{A}, \Lambda)$ is often called *the relative K-group with respect to the homomorphism* $\mathfrak{A} \to \Lambda$. Recall [Swan 1968, Theorem 15.5] that there is a long exact sequence of relative algebraic K-theory:

$$K_1(\mathfrak{A}) \to K_1(\mathfrak{A} \otimes_R \Lambda) \xrightarrow{\partial^1_{\mathfrak{A},\Lambda}} K_0(\mathfrak{A}, \Lambda) \xrightarrow{\partial^0_{\mathfrak{A},\Lambda}} K_0(\mathfrak{A}) \to K_0(\mathfrak{A} \otimes_R \Lambda). \tag{5-1}$$

The first and last arrows in this sequence are afforded by extension of scalars from R to Λ. The map $\partial^0_{\mathfrak{A},\Lambda}$ is defined by

$$\partial^0_{\mathfrak{A},\Lambda}([M, N; \lambda]) = [M] - [N].$$

The map $\partial^1_{\mathfrak{A},\Lambda}$ is defined by first recalling that the group $K_1(\mathfrak{A} \otimes_R \Lambda)$ is generated by pairs of the form (V, ϕ), where V is a finitely generated, free, $\mathfrak{A} \otimes_R \Lambda$-module, and $\phi : V \xrightarrow{\sim} V$ is an $\mathfrak{A} \otimes_R \Lambda$-isomorphism. If T is any projective $\mathfrak{A}$-submodule of V satisfying $T \otimes_{\mathfrak{A}} \Lambda \simeq V$, then we set

$$\partial^1_{\mathfrak{A},\Lambda}(V, \phi) = [T, T; \phi].$$

It may be shown that this definition is independent of the choice of T.

We shall often ease notation and write e.g., ∂^0 rather than $\partial^0_{\mathfrak{A},\Lambda}$ when no confusion is likely to result.

Idelic description and localisation. [Fröhlich 1983, Chapter II,§1]. Let us retain the notation established above, and suppose in addition that we now work over a number field F. The reduced norm map

$$\mathrm{nrd} : (FG)^\times \to Z(FG)^\times$$

induces isomorphisms

$$K_1(FG) \simeq \mathrm{nrd}(K_1(FG)) \simeq \mathrm{nrd}((FG)^\times) \simeq \mathrm{Det}((FG)^\times) \subseteq Z(FG)^\times \tag{5-2}$$

and

$$K_1(F_v G) \simeq \mathrm{nrd}(K_1(F_v G)) \simeq \mathrm{nrd}((F_v G)^\times) \simeq \mathrm{Det}((F_v G)^\times) \subseteq Z(F_v G)^\times \tag{5-3}$$

for each place v of F. In general the natural map $K_1(\mathfrak{A}_v) \to K_1(F_v G)$ is not injective, and so the reduced norm map

$$\mathrm{nrd} : K_1(\mathfrak{A}_v) \to Z(\mathfrak{A}_v)^\times$$

is not an isomorphism (although it is surjective if $\mathfrak{A}_v$ is an O_{F_v}-maximal order in $F_v G$). If we write $K_1(\mathfrak{A}_v)'$ for the image of $K_1(\mathfrak{A}_v)$ in $K_1(F_v G)$, then (5-3) induces isomorphisms

$$K_1(\mathfrak{A}_v)' \simeq \mathrm{nrd}(K_1(\mathfrak{A}_v)') \simeq \mathrm{nrd}((\mathfrak{A}_v)^\times) \simeq \mathrm{Det}(\mathfrak{A}_v^\times). \tag{5-4}$$

We shall make frequent use of the identifications (5-2), (5-3) and (5-4) (as well as those afforded by Proposition 4.5) in what follows, sometimes without explicit mention.

For each place v of F, we write

$$\mathrm{loc}_v : K_1(FG) \to K_1(F_v G)$$

for the obvious localisation map.

Definition 5.3. We define the group of ideles $J(K_1(FG))$ of $K_1(FG)$ to be the restricted direct product over all places v of F of the groups $\mathrm{Det}(F_v G)^\times \simeq K_1(F_v G)$ with respect to the subgroups $\mathrm{Det}(O_{F_v} G)^\times$. We define the group of finite ideles $J_f(K_1(FG))$ in a similar manner but with the restricted direct product taken over all finite places v of F.

If E is any extension of F, then the homomorphism

$$\mathrm{Det}(FG)^\times \to J(K_1(FG)) \times \mathrm{Det}(EG)^\times, \quad x \mapsto ((\mathrm{loc}_v(x))_v, x^{-1})$$

induces a homomorphism

$$\Delta_{\mathfrak{A},E} : \mathrm{Det}(FG)^\times \to \frac{J(K_1(FG))}{\prod_v \mathrm{Det}(\mathfrak{A}_v)^\times} \times \mathrm{Det}(EG)^\times.$$

Theorem 5.4. (a) *There is a natural isomorphism*

$$\mathrm{Cl}(\mathfrak{A}) \xrightarrow{\sim} \frac{J(K_1(FG))}{\mathrm{Det}(FG)^\times \prod_v \mathrm{Det}(\mathfrak{A}_v)^\times}.$$

(b) *There is a natural isomorphism*

$$h_{\mathfrak{A},E} : K_0(\mathfrak{A}, E) \xrightarrow{\sim} \mathrm{Coker}(\Delta_{\mathfrak{A},E}).$$

Proof. Part (a) is a well-known result of A. Fröhlich [1984, Chapter I]. Part (b) is proved in [Agboola and Burns 2006, Theorem 3.5]. □

Remark 5.5. If $[M, N; \xi] \in K_0(\mathfrak{A}, E)$ and M, N are locally free $\mathfrak{A}$-modules of rank one (which is the only case that we shall need in this paper), then $h_{\mathfrak{A},E}([M, N; \xi])$ may be described explicitly as follows.

For each place v of F, we choose $\mathfrak{A}_v$-bases m_v of M_v and n_v of N_v. We also choose an FG basis n_∞ of N_F, as well as an FG-module isomorphism $\theta : M_F \xrightarrow{\sim} N_F$. Then, for each v, we may write $n_v = \nu_v \cdot n_\infty$, with $\nu_v \in (F_v G)^\times$. As $\theta^{-1}(n_\infty)$ is an FG-basis of M_F, we may write $m_v = \mu_v \cdot \theta^{-1}(n_\infty)$, with $\mu_v \in (F_v G)^\times$. Finally, writing θ_E for the map $M_E \to N_E$ afforded by θ via extension of scalars from F to E, we have that $(\xi \circ \theta_E^{-1})(n_\infty) = \nu_\infty \cdot n_\infty$ for some $\nu_\infty \in (EG)^\times$. Then a representative of $h_{\mathfrak{A},E}([M, N; \xi])$ is given by the image of $[(\mu_v \cdot \nu_v^{-1})_v, \nu_\infty]$ in $J(K_1(FG)) \times K_1(EG)$, and a representative of $\partial^0(h_{\mathfrak{A},E}([M, N; \xi])) \in \mathrm{Cl}(\mathfrak{A})$ is given by the image of $(\mu_v \cdot \nu_v^{-1})_v \in J(K_1(FG))$.

Remark 5.6. As $\mathfrak{A}_v = F_v G$ when v is infinite (by convention), we see that

$$\frac{J(K_1(FG))}{\prod_v \mathrm{Det}(\mathfrak{A}_v)^\times} \simeq \frac{J_f(K_1(FG))}{\prod_{v \nmid \infty} \mathrm{Det}(\mathfrak{A}_v)^\times}.$$

Hence the infinite places of F in fact play no explicit role on the right-hand sides of the isomorphisms given by Theorem 5.4, and so these isomorphisms may be formulated using the finite idele group $J_f(K_1(FG))$ of $K_1(FG)$ instead of the full idele group $J(K_1(FG))$.

Lemma 5.7. *Suppose that v is a place of F and that E_v is any extension of F_v. Then there is an isomorphism*

$$K_0(\mathfrak{A}_v, E_v) \simeq \operatorname{Det}(E_vG)^\times / \operatorname{Det}(\mathfrak{A}_v)^\times.$$

Proof. This follows directly from the long exact sequence of relative K-theory (5-1) applied to $K_0(\mathfrak{A}_v, E_v)$, together with (5-3) and (5-4). □

For each place v of F, there is a localisation map on relative K-groups:

$$\lambda_v : K_0(\mathfrak{A}, E) \to K_0(\mathfrak{A}_v, E_v), \quad [M, N; \xi] \mapsto [M_v, N_v, \xi_v],$$

where ξ_v denotes the map obtained from ξ via extension of scalars from E to E_v. It is not hard to check that, in terms of the descriptions of $K_0(\mathfrak{A}, E)$ and $K_0(\mathfrak{A}_v, E_v)$ afforded by Theorem 5.4 and Lemma 5.7, the map λ_v is that induced by the homomorphism (which we denote by the same symbol λ_v)

$$\lambda_v : J(K_1(FG)) \times \operatorname{Det}(EG)^\times \to \operatorname{Det}(E_vG)^\times, \quad [(x_v)_v, x_\infty] \mapsto [x_v \cdot \operatorname{loc}_v(x_\infty)].$$

Definition 5.8. We define the idele group $J(K_0(\mathfrak{A}, E))$ of $K_0(\mathfrak{A}, E)$ to be the restricted direct product over all places v of F of the groups $K_0(\mathfrak{A}_v, E_v)$ with respect to the subgroups $K_0(\mathfrak{A}_v, O_{E_v})$.

We define the group of finite ideles $J_f(K_0(\mathfrak{A}, F^c))$ in a similar manner, but with the restricted direct product taken over all finite places of F.

Proposition 5.9. (a) *The homomorphism*

$$\lambda := \prod_v \lambda_v : K_0(\mathfrak{A}, E) \to \prod_v K_0(\mathfrak{A}_v, E_v)$$

is injective.

(b) *If F has no real places or if G admits no irreducible symplectic characters, then the homomorphism*

$$\lambda_f := \prod_{v \nmid \infty} \lambda_v : K_0(\mathfrak{A}, E) \to \prod_{v \nmid \infty} K_0(\mathfrak{A}_v, E_v)$$

is injective.

(c) *The image of λ lies in the idele group $J(K_0(\mathfrak{A}, E))$.*

Proof. (a) Suppose that $\alpha \in K_0(\mathfrak{A}, E)$ lies in the kernel of λ, and let

$$[(x_v)_v, x_\infty] \in J(K_1(FG)) \times \operatorname{Det}(EG)^\times$$

be a representative of α. Then for each v, we have

$$x_v \cdot \operatorname{loc}_v(x_\infty) \in \operatorname{Det}(\mathfrak{A}_v)^\times \subseteq \operatorname{Det}(F_vG)^\times. \tag{5-5}$$

Since $x_v \in \mathrm{Det}(F_vG)^\times \subseteq Z(F_vG)^\times$, we see that $\mathrm{loc}_v(x_\infty) \in Z(F_vG)^\times$ for each v. Hence $x_\infty \in Z(FG)^\times$, and so via the Hasse–Schilling norm theorem [Swan 1970, Theorem 7.6; Curtis and Reiner 1981, Theorem 7.8] we deduce that $x_\infty \in \mathrm{Det}(FG)^\times$. Hence α is also represented by the idele

$$[(\mathrm{loc}_v(x_\infty))_v, x_\infty^{-1}] \cdot [(x_v)_v, x_\infty] = [(x_v \cdot \mathrm{loc}_v(x_\infty))_v, 1],$$

and now (5-5) and Theorem 5.4(b) imply that $\alpha = 0$ in $K_0(\mathfrak{A}, E)$. Therefore λ is injective, as claimed.

(b) The proof of this assertion is virtually identical to that of part (a). Using the same notation as in the proof of part (a), we see that $\mathrm{loc}_v(x_\infty) \in \mathrm{Det}(F_vG)^\times \simeq Z(F_vG)^\times$ for each finite place v of F. This implies that $x_\infty \in Z(FG)^\times$. Under our hypotheses, we have that $\mathrm{Det}(FG)^\times \simeq Z(FG)^\times$, and so $x_\infty \in \mathrm{Det}(FG)^\times$. The remainder of the argument proceeds exactly as in the proof of part (a).

(c) If $\beta = [M, N; \xi] \in K_0(\mathfrak{A}, E)$, then for all but finitely many places v, the isomorphism $\xi_v : M \otimes_{O_F} E_v \xrightarrow{\sim} N \otimes_{O_F} E_v$ obtained from ξ via extension of scalars from E to E_v restricts to an isomorphism $M \otimes_{O_F} O_{E_v} \xrightarrow{\sim} N \otimes_{O_F} O_{E_v}$. Hence, for all but finitely many v, we have that $\lambda_v(\beta) \in K_0(\mathfrak{A}_v, O_{E_v})$, and so $\lambda(\beta) \in J(K_0(\mathfrak{A}, E))$, as asserted. □

6. Cohomological classes in relative K-groups

Recall that F is a number field and that G is a finite group upon which Ω_F acts trivially. In this section we shall explain how the set of realisable classes $\mathcal{R}(O_FG) \subseteq \mathrm{Cl}(O_FG)$ may be studied via imposing local cohomological conditions on elements of the relative K-group $K_0(O_FG, F^c)$.

Definition 6.1. We define maps Ψ and Ψ_v (for each place v of F) by

$$\Psi = \Psi_G : H_t^1(F, G) \to K_0(O_FG, F^c), \qquad [\pi] \mapsto [O_\pi, O_FG; \boldsymbol{r}_G],$$
$$\Psi_v = \Psi_{G,v} : H_t^1(F_v, G) \to K_0(O_{F_v}G, F_v^c), \quad [\pi_v] \mapsto [O_{\pi_v}, O_{F_v}G; \boldsymbol{r}_G].$$

We set

$$K\mathcal{R}(O_FG) := \mathrm{Im}(\Psi).$$

Definition 6.2. We define the pointed set of ideles $J(H_t^1(F, G))$ of $H_t^1(F, G)$ to be the restricted direct product over all places v of F of the pointed sets $H_t^1(F_v, G)$ with respect to the pointed subsets $H_{\mathrm{nr}}^1(F_v, G)$, and we write

$$\Psi^{\mathrm{id}} : J(H_t^1(F, G)) \to J(K_0(O_FG, F^c))$$

for the map afforded by the maps $\Psi_v : H_t^1(F_v, G) \to K_0(O_{F_v}G, F_v^c)$.

In general, $K\mathcal{R}(O_FG)$ is not a subgroup of $K_0(O_FG, F^c)$. However, although $H_{\mathrm{nr}}^1(F_v, G)$ is in general merely a pointed set and not a group, the following result holds.

Proposition 6.3. *Let v be any place of F, and write Ψ_v^{nr} for the restriction of Ψ_v to $H_{\mathrm{nr}}^1(F_v, G)$. Then $\mathrm{Im}(\Psi_v^{\mathrm{nr}})$ is a subgroup of $K_0(O_{F_v}G, F_v^c)$.*

Proof. If v is infinite, then $H_{\mathrm{nr}}^1(F_v, G) = 0$, and so $\mathrm{Im}(\Psi_v^{\mathrm{nr}}) = 0$. For finite v, the result follows from Proposition 3.7 and Lemma 5.7. □

Definition 6.4. We say that an element $x \in K_0(O_F G, F^c)$ is *cohomological* (respectively *cohomological at* v) if $x \in \mathrm{Im}(\Psi)$ (respectively $\lambda_v(x) \in \mathrm{Im}(\Psi_v)$). We say that x is *locally cohomological* if x is cohomological at v for all places v of F. We write

$$\mathrm{LC}(O_F G) := \lambda^{-1}(\mathrm{Im}(\Psi^{\mathrm{id}}))$$

for the subset of $K_0(O_F G, F^c)$ consisting of locally cohomological elements.

The long exact sequence of relative K-theory (5-1) applied to $K_0(O_F G, F^c)$ yields a long exact sequence

$$K_1(O_F G) \to K_1(F^c G) \xrightarrow{\partial^1} K_0(O_F G, F^c) \xrightarrow{\partial^0} \mathrm{Cl}(O_F G) \to 0, \tag{6-1}$$

where $\mathrm{Cl}(O_F G)$ denotes the locally free class group of $O_F G$. We set

$$\psi := \partial^0 \circ \Psi,$$

and we write

$$\mathcal{R}(O_F G) := \mathrm{Im}(\psi).$$

McCulloh has conjectured that $\mathcal{R}(O_F G)$ is always a subgroup of $\mathrm{Cl}(O_F G)$, and he has proved that this is true whenever G is abelian [McCulloh 1987, Corollary 6.20]. The following conjecture gives a precise characterisation of the image $K\mathcal{R}(O_F G)$ of Ψ.

Conjecture 6.5. An element of $K_0(O_F G, F^c)$ is cohomological if and only if it is locally cohomological. In other words, we have that

$$K\mathcal{R}(O_F G) = \mathrm{LC}(O_F G).$$

Let us now explain why Conjecture 6.5 implies that $\mathcal{R}(O_F G)$ is a subgroup of $\mathrm{Cl}(O_F G)$. In order to do this, we shall require the following result which is equivalent to a theorem of McCulloh when G is abelian, and whose proof relies on results contained in [McCulloh 1987; 2011]. Before stating the result, we remind the reader that $\prod_v \mathrm{Im}(\Psi_v^{\mathrm{nr}})$ is not merely a pointed set, but is in fact a subgroup of $J(K_0(O_F G, F^c))$ (see Proposition 6.3).

Theorem 6.6. *Let*

$$\overline{\Psi^{\mathrm{id}}} : J(H_t^1(F, G)) \to \frac{J(K_0(O_F G, F^c))}{\lambda[\partial^1(K_1(F^c G))] \cdot \prod_v \mathrm{Im}(\Psi_v^{\mathrm{nr}})}$$

denote the map of pointed sets given by the composition of the map Ψ^{id} *with the quotient homomorphism*

$$J(K_0(O_F G, F^c)) \to \frac{J(K_0(O_F G, F^c))}{\lambda[\partial^1(K_1(F^c G))] \cdot \prod_v \mathrm{Im}(\Psi_v^{\mathrm{nr}})}.$$

Then the image of $\overline{\Psi^{\mathrm{id}}}$ *is in fact a group. Hence it follows that*

$$\lambda[\partial^1(K_1(F^c G))] \cdot \mathrm{Im}(\Psi^{\mathrm{id}})$$

is a subgroup of $J(K_0(O_F G, F^c))$.

This theorem will be proved in Section 12. It implies the following result.

Theorem 6.7. *If Conjecture 6.5 holds, then $\mathcal{R}(O_F G)$ is a subgroup of $\mathrm{Cl}(O_F G)$.*

Proof. It follows from the exact sequence (6-1) that $\mathcal{R}(O_F G)$ is a subgroup of $\mathrm{Cl}(O_F G)$ if and only if $\partial^1(K_1(F^c G)) \cdot K\mathcal{R}(O_F G)$ is a subgroup of $K_0(O_F G, F^c)$. However, if Conjecture 6.5 is true, then Theorem 6.6 implies that

$$\partial^1(K_1(F^c G)) \cdot K\mathcal{R}(O_F G) = \partial^1(K_1(F^c G)) \cdot \mathrm{LC}(O_F G) \tag{6-2}$$

is the kernel of the homomorphism

$$K_0(O_F G, F^c) \xrightarrow{\lambda} J(K_0(O_F G, F^c)) \to \frac{J(K_0(O_F G, F^c))}{\lambda[\partial^1(K_1(F^c G))] \cdot \mathrm{Im}(\Psi^{\mathrm{id}})},$$

where the last arrow denotes the obvious quotient homomorphism. This implies the desired result. □

We conclude this section with the following result on unramified locally cohomological classes in $K_0(O_F G, F^c)$. This will be used in the proofs of Theorem 16.4 and Theorem E of the introduction (see Section 16 below).

Proposition 6.8. (*a*) *Let L be the maximal, abelian, everywhere unramified (including at all infinite places) extension of F of exponent $|G^{\mathrm{ab}}|$, and suppose that $y \in K_0(O_F G, F^c)$ lies in the kernel of the map*

$$\beta : K_0(O_F G, F^c) \xrightarrow{\lambda_F} J(K_0(O_F G, F^c)) \to \frac{J(K_0(O_F G, F^c))}{\prod_v \mathrm{Im}(\Psi_v^{\mathrm{nr}})}.$$

Then y lies in the kernel of the extension of scalars map

$$e_L : K_0(O_F G, F^c) \to K_0(O_L G, F^c).$$

Hence, if $(h_F^+, |G^{\mathrm{ab}}|) = 1$ (where h_F^+ denotes the narrow class number of F), then $L = F$, and so β is injective.

(*b*) *Suppose that G admits no nontrivial irreducible symplectic characters, or that F has no real places, and that $y \in K_0(O_F G, F^c)$ lies in the kernel of the map*

$$\beta_f : K_0(O_F G, F^c) \xrightarrow{\lambda_{f,F}} J_f(K_0(O_F G, F^c)) \to \frac{J_f(K_0(O_F G, F^c))}{\prod_{v \nmid \infty} \mathrm{Im}(\Psi_v^{\mathrm{nr}})}.$$

Then y lies in the kernel of the extension of scalars map

$$e_M : K_0(O_F G, F^c) \to K_0(O_M G, F^c),$$

where M is the maximal, abelian, unramified (at all finite places) extension of F of exponent $|G^{\mathrm{ab}}|$.

Hence if $(h_F, |G^{\mathrm{ab}}|) = 1$ then $L = F$, and so β_f is injective.

Proof. (a) Suppose that $y = [(y_v), y_\infty]$ lies in the kernel of β, and let E/F be the smallest Galois extension such that Ω_E fixes y_∞. For each place v of F, let $w(v)$ be the place of E afforded by our fixed choice of embedding $F^c \to F_v^c$.

As y lies in the kernel of β, we have that $y_v \cdot \mathrm{loc}_v(y_\infty) \in \mathrm{Im}(\Psi_v^{\mathrm{nr}})$ for each place v. Hence, for each v, $\mathrm{loc}_v(y_\infty) \in H(Z(F_v G))$ is an unramified G^{ab}-resolvend over F_v (see Proposition 3.6). It follows that, for each v, the extension $E_{w(v)}/F_v$ is unramified and that $[E_{w(v)} : F_v]$ divides $|G^{\mathrm{ab}}|$. This implies that E/F is unramified at all places v, and is of exponent dividing $|G^{\mathrm{ab}}|$. Hence $E \subseteq L$, and so $y_\infty \in \mathrm{Det}(LG)^\times$.

Now since $y_v \cdot \mathrm{loc}_v(y_\infty) \in \mathrm{Im}(\Psi_v^{\mathrm{nr}})$ for each place v, we see that in fact $y_v \cdot \mathrm{loc}_v(y_\infty) \in \mathrm{Det}(O_{L_v} G)^\times$. Hence $e_L(y)$ is in the kernel of the localisation map

$$\lambda_L : K_0(O_L G, F^c) \to J(K_0(O_L G, F^c)),$$

and since λ_L is injective (see Proposition 5.9(a)) it follows that $e_L(y) = 0$.

The final assertion now follows immediately.

(b) This proof is virtually identical to the proof of (a), except that here, because either G admits no irreducible symplectic characters or F has no real places, we may appeal to the injectivity of the localisation map $\lambda_{f,M}$ (see Proposition 5.9(b)) rather than that of λ_M. □

7. Local extensions I

The goal of this section is to describe how resolvends of normal integral bases of tamely ramified, nonarchimedean local extensions admit *Stickelberger factorisations* (see Definition 7.12). This reflects the fact that every tamely ramified G-extension of F_v is a compositum of an unramified extension of F_v and a twist of a totally ramified extension of F_v. All of the results in this section are based on unpublished notes of the second-named author.

For each finite place v of F, we fix a uniformiser ϖ_v of F_v, and we write q_v for the order of the residue field of F_v. We fix a compatible set of roots of unity $\{\zeta_m\}$, and a compatible set $\{\varpi_v^{1/m}\}$ of roots of ϖ_v. So, if m and n are any two positive integers, then we have $(\zeta_{mn})^m = \zeta_n$, and $(\varpi_v^{1/mn})^m = \varpi_v^{1/n}$.

Recall that F_v^{nr} (respectively F_v^t) denotes the maximal unramified (respectively tamely ramified) extension of F_v. Then

$$F_v^{\mathrm{nr}} = \bigcup_{\substack{m \geq 1 \\ (m, q_v) = 1}} F_v(\zeta_m) \quad \text{and} \quad F_v^t = \bigcup_{\substack{m \geq 1 \\ (m, q_v) = 1}} F_v(\zeta_m, \varpi_v^{1/m}).$$

The group $\Omega_v^{\mathrm{nr}} := \mathrm{Gal}(F_v^{\mathrm{nr}}/F_v)$ is topologically generated by a Frobenius element ϕ_v which may be chosen to satisfy

$$\phi_v(\zeta_m) = \zeta_m^{q_v} \quad \text{and} \quad \phi_v(\varpi_v^{1/m}) = \varpi_v^{1/m}$$

for each integer m coprime to q_v. Our choice of compatible roots of unity also uniquely specifies a topological generator σ_v of $\mathrm{Gal}(F_v^t/F_v^{\mathrm{nr}})$ by the conditions

$$\sigma_v(\varpi_v^{1/m}) = \zeta_m \cdot \varpi_v^{1/m} \quad \text{and} \quad \sigma_v(\zeta_m) = \zeta_m$$

for all integers m coprime to q_v. The group $\Omega_v^t := \mathrm{Gal}(F_v^t/F_v)$ is topologically generated by ϕ_v and σ_v, subject to the relation

$$\phi_v \cdot \sigma_v \cdot \phi_v^{-1} = \sigma_v^{q_v}. \tag{7-1}$$

While reading the remainder of this section (especially Proposition 7.7 below), it may be helpful for the reader to keep in mind the statement and proof of the following well-known result which provides some motivation for a number of subsequent constructions.

Proposition 7.1. *Set $L := F_v$. Let n be a positive integer with $(n, q_v) = 1$, and suppose that $\mu_n \subseteq L$. Set $E = L(\varpi_v^{1/n})$, $\Gamma = \mathrm{Gal}(E/L) = \mathbb{Z}/n\mathbb{Z}$, and $\beta = \sum_{i=0}^{n-1} \varpi_v^{i/n}$. Then $O_E = O_L\Gamma \cdot \beta$.*

Proof. We first observe that plainly $O_L\Gamma \cdot \beta \subseteq O_E$, as $\beta \in O_E$.

Let χ denote the Kummer character of Γ, defined by

$$\chi(\gamma) = \frac{\gamma(\varpi_v^{1/n})}{\varpi_v^{1/n}} \in \mu_n$$

for each $\gamma \in \Gamma$. Then $\hat{\Gamma} = \langle \chi \rangle$, and for each $0 \leq j \leq n-1$, we have

$$\Bigl(\sum_{\gamma} \chi^j(\gamma)\gamma^{-1}\Bigr) \cdot \beta = \Bigl(\sum_{\gamma} \chi^j(\gamma)\gamma^{-1}\Bigr) \cdot \Bigl(\sum_{i=0}^{n-1} \varpi_v^{i/n}\Bigr) = \sum_{i=0}^{n-1}\Bigl(\sum_{\gamma} \chi^j(\gamma) \cdot \chi^{-i}(\gamma) \cdot \varpi_v^{i/n}\Bigr) = n \cdot \varpi_v^{j/n}.$$

As $n \in O_L^\times$, we therefore see that $\{\varpi_v^{j/n}\}_{j=0}^{n-1} \subseteq O_L\Gamma \cdot \beta$, which implies that $O_E \subseteq O_L\Gamma \cdot \beta$. This implies the desired result. □

Definition 7.2. For each finite place v of F, we define

$$\Sigma_v(G) := \{s \in G \mid s^{q_v} \in c(s)\}$$

(recall that $c(s)$ denotes the conjugacy class of s in G). Plainly if $s \in \Sigma_v(G)$, then both $c(s)$ and $\langle s \rangle$ are subsets of $\Sigma_v(G)$. Let us also remark that if $s \in \Sigma_v(G)$, then the order $|s|$ of s is coprime to q_v.

Definition 7.3. If $s \in G$, we set

$$\beta_s := \frac{1}{|s|} \sum_{i=0}^{|s|-1} \varpi_v^{i/|s|};$$

note that β_s depends only upon $|s|$, and so in particular we have

$$\beta_s = \beta_{g^{-1}sg}$$

for every $g \in G$. We define $\varphi_{v,s} \in \mathrm{Map}(G, O_{F_v^c})$ by setting

$$\varphi_{v,s}(g) = \begin{cases} \sigma_v^i(\beta_s) & \text{if } g = s^i, \\ 0 & \text{if } g \notin \langle s \rangle. \end{cases}$$

Then

$$\boldsymbol{r}_G(\varphi_{v,s}) = \sum_{i=0}^{|s|-1} \varphi_{v,s}(s^i)s^{-i} = \sum_{i=0}^{|s|-1} \sigma_v^i(\beta_s)s^{-i}. \tag{7-2}$$

We note that for each $g \in G$, we have

$$\boldsymbol{r}_G(\varphi_{v,g^{-1}sg}) = g^{-1} \cdot \boldsymbol{r}_G(\varphi_{v,s}) \cdot g, \tag{7-3}$$

and so

$$\mathrm{Det}(\boldsymbol{r}_G(\varphi_{v,g^{-1}sg})) = \mathrm{Det}(\boldsymbol{r}_G(\varphi_{v,s})), \tag{7-4}$$

i.e., the element $\mathrm{Det}(\boldsymbol{r}_G(\varphi_{v,s}))$ depends only upon the conjugacy class $c(s)$ of s in G. We remark that it will be shown later as a consequence of properties of the Stickelberger pairing that $\mathrm{Det}(\boldsymbol{r}_G(\varphi_{v,s}))$ in fact determines the subgroup $\langle s\rangle$ of G up to conjugation (see Remark 4.2 and Proposition 10.5(b)).

We shall see that generators of inertia subgroups of tame Galois G-extensions of F_v lie in $\Sigma_v(G)$, and that the elements $\varphi_{v,s}$ for $s \in G$ with $(|s|, q_v) = 1$ may be used to construct normal integral basis generators of tame (and of course totally ramified) Galois G-extensions of F_v^{nr}.

In order to ease notation, we shall now set $L := F_v$ and $O := O_L$, and we shall drop the subscript v from our notation for the rest of this section.

Suppose now that L_π/L is a tamely ramified Galois G-extension of L, corresponding to $\pi \in \mathrm{Hom}(\Omega^t, G)$. We are going to describe McCulloh's [2011] decomposition of resolvends of normal integral basis generators of L_π/L (see also [Byott 1998, §6]). When G is abelian, this decomposition is an analogue of a version of Stickelberger's factorisation of Gauss sums.

Write $s := \pi(\sigma)$ and $t := \pi(\phi)$; then $t \cdot s \cdot t^{-1} = s^q$, and so $s \in \Sigma(G)$. We define $\pi_r, \pi_{\mathrm{nr}} \in \mathrm{Map}(\Omega^t, G)$ by setting

$$\pi_r(\sigma^m \phi^n) = \pi(\sigma^m) = s^m, \tag{7-5}$$

$$\pi_{\mathrm{nr}}(\sigma^m \phi^n) = \pi(\phi^n) = t^n. \tag{7-6}$$

If $\omega_i \in \Omega^t$ $(i = 1, 2)$ with $\omega_i = \sigma^{m_i} \cdot \phi^{n_i}$, then a straightforward calculation using (7-1) shows that

$$\omega_1 \cdot \omega_2 = \sigma^{m_1 + m_2 q^{n_1}} \cdot \phi^{n_1+n_2}.$$

This implies that $\pi_{\mathrm{nr}} \in \mathrm{Hom}(\Omega^{\mathrm{nr}}, G)$. Plainly we have

$$\pi(\omega) = \pi_r(\omega) \cdot \pi_{\mathrm{nr}}(\omega) \tag{7-7}$$

for every $\omega = \sigma^m \cdot \phi^n \in \Omega^t$. The map $\pi_{\mathrm{nr}} \in \mathrm{Hom}(\Omega^{\mathrm{nr}}, G)$ corresponds to an unramified Galois G-extension $L_{\pi_{\mathrm{nr}}}$ of L (see Remark 7.10 below for a more detailed discussion of this point). Since $L_{\pi_{\mathrm{nr}}}/L$ is unramified, $O_{\pi_{\mathrm{nr}}}$ is a free $O_L G$-module. Let a_{nr} be any normal integral basis generator of this extension. Note that $\boldsymbol{r}_G(a_{\mathrm{nr}}) \in H(OG)$, because $L_{\pi_{\mathrm{nr}}}/L$ is unramified (see Corollary 3.3(iv)).

Definition 7.4. Let $G(\pi_{\mathrm{nr}})$ denote the group G with Ω^t-action given by

$$\omega(g) = \pi_{\mathrm{nr}}(\omega) \cdot g \cdot \pi_{\mathrm{nr}}(\omega)^{-1}$$

for $\omega \in \Omega^t$ and $g \in G$.

Lemma 7.5. *The map π_r is a $G(\pi_{\mathrm{nr}})$-valued 1-cocycle of Ω^t.*

Proof. Suppose that $\omega_1, \omega_2 \in \Omega^t$. Then since $\pi_{\mathrm{nr}} \in \mathrm{Hom}(\Omega^{\mathrm{nr}}, G)$ and $\pi = \pi_r \cdot \pi_{\mathrm{nr}}$, a straightforward calculation shows that

$$\pi_r(\omega_1\omega_2) = \pi_r(\omega_1) \cdot \pi_{\mathrm{nr}}(\omega_1) \cdot \pi_r(\omega_2) \cdot \pi_{\mathrm{nr}}(\omega_1)^{-1},$$

and this establishes the desired result. □

Definition 7.6. We write ${}^{\pi_r}G(\pi_{\mathrm{nr}})$ for the set G endowed with the following action of Ω^t: for every $g \in G$ and $\omega \in \Omega^t$ we have

$$g^{\omega} = \pi_r(\omega) \cdot \pi_{\mathrm{nr}}(\omega) \cdot g \cdot \pi_{\mathrm{nr}}(\omega)^{-1}.$$

Lemma 7.5 implies that if $\omega_1, \omega_2 \in \Omega^t$, then

$$g^{(\omega_1\omega_2)} = (g^{\omega_2})^{\omega_1}.$$

We set

$$L_{\pi_r}(\pi_{\mathrm{nr}}) := \mathrm{Map}_{\Omega^t}({}^{\pi_r}G(\pi_{\mathrm{nr}}), L^t).$$

The algebra $(L^t G(\pi_{\mathrm{nr}}))^{\Omega^t}$ acts on $L_{\pi_r}(\pi_{\mathrm{nr}})$ via the rule

$$(\alpha \cdot a)(h) = \sum_{g \in G} \alpha_g \cdot a(h \cdot g)$$

for all $h \in G$ and $\alpha = \sum_{g \in G} \alpha_g \cdot g \in (L^t G(\pi_{\mathrm{nr}}))^{\Omega^t}$.

Proposition 7.7. (a) *Recall that $s \in \Sigma(G)$. We have that $\varphi_s \in L_{\pi_r}(\pi_{\mathrm{nr}})$.*

(b) *Set*

$$\mathfrak{A}(\pi_{\mathrm{nr}}) = (O_{L^c} G(\pi_{\mathrm{nr}}))^{\Omega^t},$$

and let $O_{\pi_r}(\pi_{\mathrm{nr}})$ be the integral closure of O_L in $L_{\pi_r}(\pi_{\mathrm{nr}})$. Then

$$\mathfrak{A}(\pi_{\mathrm{nr}}) \cdot \varphi_s = O_{\pi_r}(\pi_{\mathrm{nr}}).$$

(c) *For any $\alpha_r \in L_{\pi_r}(\pi_{\mathrm{nr}})$ and $\omega \in \Omega^t$, we have*

$$\boldsymbol{r}_G(\alpha_r)^{\omega} = \pi_{\mathrm{nr}}(\omega)^{-1} \cdot \boldsymbol{r}_G(\alpha_r) \cdot \pi(\omega).$$

Proof.

(a) Suppose that $\omega = \sigma^m \cdot \phi^n \in \Omega^t$. If $g \in G$ and $g \notin \langle s \rangle$, then we have that

$$\varphi_s(g^{\omega}) = 0 = \varphi_s(g)^{\omega}.$$

On the other hand, we also have

$$\varphi_s((s^i)^{\omega}) = \varphi_s((s^i)^{\sigma^m\phi^n}) = \varphi_s(s^m \cdot t^n \cdot s^i \cdot t^{-n}) = \varphi_s(s^{m+iq^n}) = \sigma^{m+iq^n}(\beta_s) = (\sigma^m \cdot \phi^n) \cdot \sigma^i(\beta_s) = \varphi_s(s^i)^{\omega}.$$

Hence $\varphi_s \in L_{\pi_r}(\pi_{\mathrm{nr}})$, as claimed.

(b) The proof of this assertion is very similar to that of [Byott 1998, Lemma 6.6], which is in turn an analogue of [McCulloh 1987, 5.4].

Set $H = \langle s \rangle$. Then Ω^t acts transitively on ${}^{\pi_r}H(\pi_{\mathrm{nr}}) \subseteq^{\pi_r} G(\pi_{\mathrm{nr}})$, and so the algebra

$$L_{\pi_r}(\pi_{\mathrm{nr}})^H := \mathrm{Map}_{\Omega^t}({}^{\pi_r}H(\pi_{\mathrm{nr}}), L^t)$$

may be identified with a subfield of L^t via identifying $b \in L_{\pi_r}(\pi_{\mathrm{nr}})^H$ with $x_b = b(\mathbf{1}) \in L^t$. We have that

$$x_b^{\sigma^m} = b(s^m) \quad \text{and} \quad x_b^{\phi} = x_b,$$

and so it follows that $L_{\pi_r}(\pi_{\mathrm{nr}})^H$ is the subfield of L^t consisting of those elements of L^t that are fixed by both ϕ and $\sigma^{|s|}$. This implies that $L_{\pi_r}(\pi_{\mathrm{nr}})^H = L[\varpi^{1/|s|}]$ (which in general will not be normal over L), and that the integral closure of O_L in $L_{\pi_r}(\pi_{\mathrm{nr}})^H$ is equal to $O_L[\varpi^{1/|s|}]$. Plainly $\beta_s \in O_L[\varpi^{1/|s|}]$ (as $|s|$ is invertible in O_L), and the element β_s corresponds to the element $\varphi_s|_H \in L_{\pi_r}(\pi_{\mathrm{nr}})^H$.

If we set $\mathfrak{A}(\pi_{\mathrm{nr}})_H := (O_{L^t}H(\pi_{\mathrm{nr}}))^{\Omega^t}$, then for each integer k with $0 \le k \le |s|-1$, it is not hard to check that

$$\left(\sum_{i=0}^{|s|-1} \zeta_{|s|}^{-ki} s^i\right)^{\phi} = \sum_{i=0}^{|s|-1} \zeta_{|s|}^{-ki} s^i,$$

and so we see that

$$\sum_{i=0}^{|s|-1} \zeta_{|s|}^{-ki} s^i \in \mathfrak{A}(\pi_{\mathrm{nr}})_H.$$

A straightforward computation (cf. [McCulloh 1987, 5.4]) also shows that

$$\left(\sum_{i=0}^{|s|-1} \zeta_{|s|}^{-ki} s^i\right) \cdot \beta_s = \varpi^{k/|s|}.$$

It therefore follows that $\mathfrak{A}(\pi_{\mathrm{nr}})_H \cdot \beta_s = O_L[\varpi^{1/|s|}]$, and this in turn implies that

$$\mathfrak{A}(\pi_{\mathrm{nr}}) \cdot \varphi_s = O_{\pi_r}(\pi_{\mathrm{nr}}),$$

as asserted.

(c) We have

$$\begin{aligned}
\boldsymbol{r}_G(\alpha_r)^{\omega} &= \sum_{g \in G} \alpha_r(g)^{\omega} \cdot g^{-1} \\
&= \sum_{g \in G} \alpha_r(g^{\omega}) \cdot g^{-1} \\
&= \sum_{g \in G} \alpha_r(\pi_r(\omega) \cdot \pi_{\mathrm{nr}}(\omega) \cdot g \cdot \pi_{\mathrm{nr}}^{-1}(\omega)) \cdot g^{-1} \\
&= \sum_{g \in G} \alpha_r(g) \cdot \pi_{\mathrm{nr}}(\omega)^{-1} \cdot g^{-1} \cdot \pi_r(\omega) \cdot \pi_{\mathrm{nr}}(\omega) \\
&= \pi_{\mathrm{nr}}(\omega)^{-1} \cdot \boldsymbol{r}_G(\alpha_r) \cdot \pi(\omega),
\end{aligned}$$

as claimed. □

Corollary 7.8. *For any $\alpha_r \in L_{\pi_r}(\pi_{\mathrm{nr}})$ and $\alpha_{\mathrm{nr}} \in L_{\pi_{\mathrm{nr}}}$, there is a unique $\alpha \in L_\pi$ such that*

$$\boldsymbol{r}_G(\alpha_{\mathrm{nr}}) \cdot \boldsymbol{r}_G(\alpha_r) = \boldsymbol{r}_G(\alpha).$$

Proof. Proposition 7.7(c) implies that, for any $\omega \in \Omega^t$, we have

$$[\boldsymbol{r}_G(\alpha_{\mathrm{nr}}) \cdot \boldsymbol{r}_G(\alpha_r)]^\omega = \boldsymbol{r}_G(\alpha_{\mathrm{nr}}) \cdot \boldsymbol{r}_G(\alpha_r) \cdot \pi(\omega),$$

and so $\boldsymbol{r}_G(\alpha_{\mathrm{nr}}) \cdot \boldsymbol{r}_G(\alpha_r) \in H(LG)$. As the map $\boldsymbol{r}_G$ is bijective, it follows that there is a unique $\alpha \in \mathrm{Map}(G, L^c)$ such that

$$\boldsymbol{r}_G(\alpha_{\mathrm{nr}}) \cdot \boldsymbol{r}_G(\alpha_r) = \boldsymbol{r}_G(\alpha),$$

and that $\alpha \in L_\pi$. □

Theorem 7.9. *If $a_{\mathrm{nr}} \in L_{\pi_{\mathrm{nr}}}$ is any normal integral basis generator of $L_{\pi_{\mathrm{nr}}}/L$, then the element $a \in L_\pi$ defined by*

$$\boldsymbol{r}_G(a_{\mathrm{nr}}) \cdot \boldsymbol{r}_G(\varphi_s) = \boldsymbol{r}_G(a) \tag{7-8}$$

is a normal integral basis generator of L_π/L.

Proof. The proof of this assertion is very similar to that of the analogous result in the abelian case described in [McCulloh 1987, (5.7), p. 283]. We first observe that plainly $O_L G \cdot a \subseteq O_\pi$ because $a_{\mathrm{nr}} \in O_{\pi_{\mathrm{nr}}}$ and $\varphi_s \in O_{\pi_r}(\pi_{\mathrm{nr}})$. Hence, to prove the desired result, it suffices to show that

$$\mathrm{disc}(O_L G \cdot a/O_L) = \mathrm{disc}(O_\pi/O_L).$$

This will in turn follow if we show that

$$\mathrm{disc}(O_{L^{\mathrm{nr}}} G \cdot a/O_{L^{\mathrm{nr}}}) = \mathrm{disc}(O_\pi/O_L) \cdot O_{L^{\mathrm{nr}}}.$$

Recall (see (2-3)) that we may write $L_\pi \simeq \bigoplus_{G/\pi(\Omega^t)} L^\pi$, where L^π is a field with $\mathrm{Gal}(L^\pi/L) \simeq \pi(\Omega^t)$. Under this last isomorphism, the inertia subgroup of $\mathrm{Gal}(L^\pi/L)$ is isomorphic to $\langle s \rangle$. The standard formula for tame field discriminants therefore yields

$$\mathrm{disc}(O^\pi/O_L) = \varpi^{(|s|-1)|\pi(\Omega^t)|/|s|} \cdot O_L$$

and so we have

$$\mathrm{disc}(O_\pi/O) = \varpi^{(|s|-1)|G|/|s|} \cdot O_L. \tag{7-9}$$

Now $\boldsymbol{r}_G(a_{\mathrm{nr}}) \in (O_{L^{\mathrm{nr}}} G)^\times$, and we see from the proof of Proposition 7.7(b) that

$$O_{L^{\mathrm{nr}}} G \cdot a = O_{L^{\mathrm{nr}}} G \cdot \varphi_s = O_{\pi_r}(\pi_{\mathrm{nr}}) \otimes_{O_L} O_{L^{\mathrm{nr}}} \simeq \bigoplus_{G/\langle s \rangle} O_{L^{\mathrm{nr}}}[\varpi^{1/|s|}].$$

Since

$$\mathrm{disc}(O_{L^{\mathrm{nr}}}[\varpi^{1/|s|}]/O_{L^{\mathrm{nr}}}) = \varpi^{|s|-1} \cdot O_{L^{\mathrm{nr}}},$$

it follows that

$$\mathrm{disc}(O_{L^{\mathrm{nr}}}G\cdot a/O_{L^{\mathrm{nr}}}) = \varpi^{(|s|-1)|G|/|s|}\cdot O_{L^{\mathrm{nr}}} = \mathrm{disc}(O_\pi/O)\cdot O_{L^{\mathrm{nr}}},$$

and this establishes the desired result. □

Remark 7.10. We caution the reader that $L_{\pi_{\mathrm{nr}}}$ is *not* in general equal to the maximal unramified subextension of L_π/L, even when L_π is a field. Suppose, for example, that L_π is a field, and write L_0 for the maximal unramified subextension of L_π/L. Set $f=[L_0:L]$. Then it is not hard to check that

$$L_{\pi_{\mathrm{nr}}} \simeq \prod_{i=1}^{|G|/f} L_0, \tag{7-10}$$

and so $L_{\pi_{\mathrm{nr}}}$ is a Galois algebra with "core field" L_0. If $\alpha\in O_{L_0}$ is such that $O_{L_0}=O_L[\mathrm{Gal}(L_0/L)]\cdot\alpha$, then we may take $a_{\mathrm{nr}}=(\alpha,0,\dots,0)$ under the identification given by (7-10).

Suppose further that L contains the $|s|$-th roots of unity, and that $L_\pi = L_0\cdot L(\varpi^{1/|s|})$. To ease notation, write $M:=L(\varpi^{1/|s|})$, and set $H=\langle s\rangle$. Then a calculation similar to (but simpler than) that given in the proof of Proposition 7.7(b) (see also Proposition 7.1) shows that $O_M = O_L[H]\cdot\beta_s$, and it may be shown by computing the coefficient of $\mathbf{1}_G$ on the left-hand side of (7-8) that $a=\alpha\cdot\beta_s$, as is of course well known.

Remark 7.11. Suppose that $s\in G$ with $(|s|,q)=1$. A straightforward computation (cf. the proofs of Propositions 7.1 and 7.7(b)) shows that for every $\omega\in\Omega_{L^{\mathrm{nr}}}$, we may write

$$\boldsymbol{r}_G(\varphi_s)^\omega = \boldsymbol{r}_G(\varphi_s)\cdot\tilde{\varphi}_s(\omega)$$

where $[\tilde{\varphi}_s]\in H^1_t(L^{\mathrm{nr}},G)$, and that φ_s is a normal integral basis generator of $L^{\mathrm{nr}}_{\tilde{\varphi}_s}/L^{\mathrm{nr}}$. We have that $[\tilde{\varphi}_{s_1}]=[\tilde{\varphi}_{s_2}]$ in $H^1_t(L^{\mathrm{nr}},G)$ if and only if $c(s_1)=c(s_2)$. It is easy to show that every element of $H^1_t(L^{\mathrm{nr}},G)$ is of the form $[\tilde{\varphi}_s]$ for some $s\in G$ with $(|s|,q)=1$ (cf. the proof of Proposition 7.1 again).

Definition 7.12. Let a be any normal integral basis generator of L_π/L. Theorem 7.9 implies that we may write

$$\boldsymbol{r}_G(a) = u\cdot\boldsymbol{r}_G(a_{\mathrm{nr}})\cdot\boldsymbol{r}_G(\varphi_s), \tag{7-11}$$

where $u\in(OG)^\times$ and a_{nr} is any normal integral basis generator of $L_{\pi_{\mathrm{nr}}}/L$. This may be viewed as being a nonabelian analogue of a version of Stickelberger's factorisation of abelian Gauss sums (see [Hilbert 1998, pages XXXV–XXXVI, and Theorems 135 and 136; McCulloh 1987, Introduction]), and so we call (7-11) a *Stickelberger factorisation* of $\boldsymbol{r}_G(a)$.

8. Local extensions II

Our goal in this section is to state certain results analogous to, (but very much simpler than), those in Section 7, for extensions of F_v where v is an infinite place of F. This section may therefore be viewed as being a "supplement at infinity" to Section 7 (cf. [Fröhlich 1984, Chapter I, §3]). We remind the reader that, if v is infinite, by convention, we set $O_{F_v}G = F_vG$ and $H^1_t(F_v, G) = H^1(F_v, G)$.

Suppose first that v is a complex place of F. Then

$$K_0(O_{F_v}G, F_v^c) = 0 \quad \text{and} \quad H^1(F_v, G) = 0,$$

and we set $\Sigma_v(G) = \{1\}$. As this case is totally degenerate, we therefore suppose henceforth in this section that v is real. We set $L = F_v \simeq \mathbb{R}$, and for the remainder of this section, we drop any further reference to v from our notation.

Set $\mathrm{Gal}(L^c/L) = \langle\sigma\rangle$, and fix a primitive fourth root of unity $\zeta_4 \in L^c$ (cf. the choice of compatible roots of unity made at the beginning of Section 7), so $L^c = L(\zeta_4)$.

Write

$$\Sigma(G) := \{s \in G \mid s^2 = e\}. \tag{8-1}$$

(Note that this set is in fact independent of v.) For each $s \in \Sigma(G)$, we set

$$\beta_s = \tfrac{1}{2}(1 + \zeta_4).$$

Define $\varphi_s \in \mathrm{Map}(G, L^c)$ by

$$\varphi_s(g) = \begin{cases} \sigma^i(\beta_s) & \text{if } g = s^i, \\ 0 & \text{if } g \notin \langle s\rangle. \end{cases}$$

Then it is easy to check that

$$\boldsymbol{r}_G(\varphi_s) = \beta_s \cdot e + \sigma(\beta_s) \cdot s = \tfrac{1}{2}[(1 + \zeta_4) \cdot e + (1 - \zeta_4) \cdot s].$$

Proposition 8.1. *Suppose that $\pi \in \mathrm{Hom}(\Omega_L, G)$ with $\pi(\sigma) = s$. Then $\varphi_s \in L_\pi$, and*

$$L_\pi = LG \cdot \varphi_s.$$

Proof. The first assertion follows directly from the definition of φ_s. The second is an immediate consequence of the fact that $\boldsymbol{r}_G(\varphi_s) \in (L^cG)^\times$, because

$$\tfrac{1}{2}((1 + \zeta_4) \cdot e + (1 - \zeta_4) \cdot s) \cdot \tfrac{1}{2}((1 - \zeta_4) \cdot e + (1 + \zeta_4) \cdot s) = 1. \qquad \square$$

Proposition 8.2. *Suppose that $\chi \in R_G$, and write*

$$\chi|_{\langle s\rangle} = a \cdot \mathbf{1} + b \cdot \varepsilon,$$

where ε denotes the unique nontrivial irreducible character of $\langle s\rangle$. Then

$$[\mathrm{Det}(\boldsymbol{r}_G(\varphi_s))](\chi) = (-1)^{b/2}.$$

Proof. This follows via a straightforward computation:

$$[\mathrm{Det}(\boldsymbol{r}_G(\varphi_s))](\chi) = \mathbf{1}(\boldsymbol{r}_G(\varphi_s))^a \cdot \varepsilon(\boldsymbol{r}_G(\varphi_s))^b = (\beta_s + \sigma(\beta_s))^a \cdot (\beta_s - \sigma(\beta_s))^b = 1^a \cdot \zeta_4^b = (-1)^{b/2}. \quad \square$$

Remark 8.3. In terms of the Stickelberger pairing $\langle -, -\rangle_G$ which will be introduced in the next section, Proposition 8.2 asserts that

$$[\mathrm{Det}(\boldsymbol{r}_G(\varphi_s))](\chi) = (-1)^{\langle \chi, s\rangle_G}.$$

9. The Stickelberger pairing

Definition 9.1. The *Stickelberger pairing* is a $\mathbb{Q}$-bilinear pairing

$$\langle -, -\rangle_G : \mathbb{Q}R_G \times \mathbb{Q}G \to \mathbb{Q} \tag{9-1}$$

that is defined as follows.

Let $\zeta_{|G|}$ be a fixed, primitive $|G|$-th root of unity (see the conventions established at the beginning of Section 7), and suppose first that G is abelian. Then if $\chi \in \mathrm{Irr}(G)$ and $g \in G$, we may write $\chi(g) = \zeta_{|G|}^r$ for some integer r. We define

$$\langle \chi, g\rangle_G = \left\{ \frac{r}{|G|} \right\},$$

where $\{x\}$ denotes the fractional part of $x \in \mathbb{Q}$, and we extend this to a pairing on $\mathbb{Q}R_G \times \mathbb{Q}G$ via linearity. For arbitrary finite G, the Stickelberger pairing is defined via reduction to the abelian case by setting

$$\langle \chi, g\rangle_G = \langle \chi|_{\langle g\rangle}, g\rangle_{\langle g\rangle}.$$

It is easy to check that both definitions agree when G is abelian.

We shall now explain a different way of expressing the Stickelberger pairing using the standard inner product on R_G. In order to do this, we must introduce some further notation.

For each $s \in G$, we set $m_s := |G|/|s|$. We define a character ξ_s of $\langle s\rangle$ by $\xi_s(s^i) = \zeta_{|G|}^{im_s}$; so ξ_s is a generator of the group of irreducible characters of $\langle s\rangle$. Then it follows from Definition 9.1 that

$$\langle \xi_s^\alpha, s^\beta\rangle_{\langle s\rangle} = \left\{ \frac{\alpha\beta}{|s|} \right\}.$$

Define

$$\Xi_s := \frac{1}{|s|} \sum_{j=1}^{|s|-1} j\xi_s^j.$$

Proposition 9.2. *Let $(-, -)_G$ denote the standard inner product on R_G, and suppose that $\chi \in R_G$ and $s \in G$. Then we have*

$$(\chi, \mathrm{Ind}_{\langle s\rangle}^G(\Xi_s))_G = \langle \chi, s\rangle_G.$$

Proof. Suppose that

$$\chi|_{\langle s\rangle} = \sum_{j=0}^{|s|-1} {}^{|s|-1} a_j \xi_s^j,$$

where $a_j \in \mathbb{Z}$ for each j. Then we have

$$\langle \chi, s\rangle_G = \sum_{j=0}^{|s|-1} a_j \langle \xi_s^j, s\rangle_{\langle s\rangle} = \sum_{j=0}^{|s|-1} a_j \left\{ \frac{j}{|s|} \right\} = \frac{1}{|s|} \sum_{j=0}^{|s|-1} a_j j.$$

On the other hand, via Frobenius reciprocity, we have

$$(\chi, \mathrm{Ind}_{\langle s\rangle}^{G}(\Xi_s))_G = (\chi|_{\langle s\rangle}, \Xi(s))_{\langle s\rangle} = \left(\sum_{j=0}^{|s|-1} a_j \xi_s^j, \frac{1}{|s|} \sum_{j=0}^{|s|-1} j \xi_s^j \right)_{\langle s\rangle} = \frac{1}{|s|} \sum_{j=0}^{|s|-1} a_j j = \langle \chi, s\rangle_G,$$

and this establishes the desired result. □

In order to apply Proposition 9.2, we shall require the following result concerning traces of sums of roots of unity.

Lemma 9.3. *Let $n > 1$ be an integer, and suppose that ζ is any primitive n-th root of unity. Write*

$$y := \sum_{i=1}^{n-1} i \cdot \zeta^i.$$

Then

$$\mathrm{Tr}_{\mathbb{Q}(\zeta)/\mathbb{Q}}(y) = -\tfrac{1}{2} n\phi(n),$$

where ϕ is the Euler ϕ-function. In particular, $\mathrm{Tr}_{\mathbb{Q}(\zeta)/\mathbb{Q}}(y) \neq 0$.

Proof. Each ζ^i is a primitive d-th root of unity for some divisor d of n, and so it follows that

$$y = \sum_{d\,|\,n} \sum_{\substack{1 \le r \le d-1 \\ (r,d)=1}} \frac{\mathrm{nr}}{d} \zeta^{nr/d}.$$

If $d \,|\, n$, then applying Möbius inversion to the identity $x^d - 1 = \prod_{m\,|\,d} \Phi_m(x)$ (where $\Phi_m(x)$ denotes the m-th cyclotomic polynomial) yields $\Phi_m(x) = \prod_{m\,|\,d} (x^{m\cdot} - 1)^{\mu(d/m)}$, whence it is not hard to show that $\mathrm{Tr}_{\mathbb{Q}(\varepsilon)/\mathbb{Q}}(\varepsilon) = \mu(d)$ for any primitive d-th root ε of unity. Hence $\mathrm{Tr}_{\mathbb{Q}(\zeta)/\mathbb{Q}}(\varepsilon) = \phi(n)\mu(d)/\phi(d)$, and so we have

$$\mathrm{Tr}_{\mathbb{Q}(\zeta)/\mathbb{Q}}(y) = \sum_{d\,|\,n} \sum_{\substack{1 \le r \le d-1 \\ (r,d)=1}} \frac{\mathrm{nr}}{d} \mathrm{Tr}_{\mathbb{Q}(\zeta)/\mathbb{Q}}(\zeta^{nr/d}) = n \sum_{d\,|\,n} \frac{\mu(d)}{d} \frac{\phi(n)}{\phi(d)} s(d),$$

where

$$s(d) = \begin{cases} 1 & \text{if } d = 1, \\ \sum_{\substack{1 \le i \le d-1 \\ (i,d)=1}} i & \text{if } d > 1. \end{cases}$$

It is well-known that

$$s(d) = \tfrac{1}{2} d\phi(d)$$

for any integer $d > 1$ (see, e.g., [Burton 2007, Theorem 7.7]). It therefore follows that

$$\mathrm{Tr}_{\mathbb{Q}(\zeta)/\mathbb{Q}}(y) = \tfrac{1}{2}n\phi(n) \sum_{\substack{d \mid n \\ d>1}} \mu(d) = -\tfrac{1}{2}n\phi(n),$$

as claimed. □

We can now state the following corollary to Proposition 9.2.

Corollary 9.4. *Suppose that s_1 and s_2 are elements of G.*

(i) *If $c(s_1) = c(s_2)$, then $\langle \chi, s_1\rangle_G = \langle \chi, s_2\rangle_G$ for all $\chi \in \mathbb{Q}R_G$.*

(ii) *If $\langle \chi, s_1\rangle_G = \langle \chi, s_2\rangle_G$ for all $\chi \in \mathbb{Q}R_G$, then $\langle s_1\rangle$ is conjugate to $\langle s_2\rangle$ in G.*

(iii) *We have that $\langle \chi, s_1\rangle_G = 0$ for all $\chi \in \mathbb{Q}R_G$ if and only if $s_1 = e$.*

Proof. (i) Let $\chi \in R_G$ and $s \in G$. It follows from the definition of the Stickelberger pairing that for fixed χ the value of $\langle \chi, s\rangle_G$ depends only upon the conjugacy class $c(s)$ of s in G. Hence, if $c(s_1) = c(s_2)$, then $\langle \chi, s_1\rangle_G = \langle \chi, s_2\rangle_G$ for all $\chi \in \mathbb{Q}R_G$.

(ii) To show this we use Proposition 9.2. We first note that a straightforward computation shows that the degree of the virtual character $\mathrm{Ind}_{\langle s\rangle}^G(\Xi_s)$ is equal to $|G|(|s|-1)/2|s|$, and so we see that $\mathrm{Ind}_{\langle s\rangle}^G(\Xi_s)$ determines $|s|$. Next, we remark that If $\{t_i\}$ is a set of representatives of $G/\langle s\rangle$, then for each $g \in G$, we have

$$[\mathrm{Ind}_{\langle s\rangle}^G(\Xi_s)](g) = \sum_{t_i^{-1}gt_i \in \langle s\rangle} \xi_s(t_i^{-1}gt_i), \tag{9-2}$$

and so the character $\mathrm{Ind}_{\langle s\rangle}^G(\Xi_s)$ vanishes on all elements of G that are not conjugate to an element of $\langle s\rangle$.

Proposition 9.2 implies that under our hypotheses, $\mathrm{Ind}_{\langle s_1\rangle}^G(\Xi_{s_1}) = \mathrm{Ind}_{\langle s_2\rangle}^G(\Xi_{s_2})$. Hence, to prove the desired result, it suffices to show that $[\mathrm{Ind}_{\langle s_1\rangle}^G(\Xi_{s_1})](s_1) \neq 0$, because then

$$[\mathrm{Ind}_{\langle s_2\rangle}^G(\Xi_{s_1})](s_1) = [\mathrm{Ind}_{\langle s_1\rangle}^G(\Xi_{s_1})](s_1) \neq 0,$$

which implies (since $|s_1| = |s_2|$) that s_1 is conjugate to a generator of $\langle s_2\rangle$.

Now if s_1^a is any generator of $\langle s_1\rangle$, then $\xi_{s_1}(s_1^a)$ is a primitive $|s_1|$-th root of unity, and we have

$$\xi_{s_1}(s_1^a) = \sum_{i=1}^{|s_1|-1} i\xi_{s_1}(s_1^a)^i.$$

Hence if ζ denotes any primitive $|s_1|$-th root of unity, Lemma 9.3 implies that

$$\mathrm{Tr}_{\mathbb{Q}(\zeta)\mathbb{Q}}(\xi_{s_1}(s_1^a)) = -\tfrac{1}{2}|s_1|\phi(|s_1|).$$

It follows from (9-2) that $\mathrm{Tr}_{\mathbb{Q}(\zeta)/\mathbb{Q}}[\mathrm{Ind}_{s_1}^G(\Xi_{s_1})](s_1)$ is equal to a nonzero multiple of $-|s_1|\phi(|s_1|)/2$, and so is nonzero. This in turn implies that $[\mathrm{Ind}_{s_1}^G(\Xi_{s_1})](s_1)$ is also nonzero, thereby establishing the desired result.

(iii) Proposition 9.2 implies that $\langle \chi, s_1 \rangle_G = 0$ for all $\chi \in \mathbb{Q}R_G$ if and only if $(\mathrm{Ind}_{\langle s_1 \rangle}^G(\Xi_{s_1}), \chi)_G = 0$ for all $\chi \in \mathbb{Q}R_G$. The latter condition holds if and only if $\mathrm{Ind}_{\langle s_1 \rangle}^G(\Xi_{s_1}) = 0$ and this happens if and only if $s_1 = e$. □

Remark 9.5. (a) The converse to Corollary 9.4(i) does not hold in general, e.g., it fails for the dihedral group D_{2p} of order $2p$, where $p > 3$ is a prime. (See [Siviero 2013, Chapter 3; 2016] for an explicit description of the Stickelberger pairing in this case.)

(b) Let $\chi_1, \dots, \chi_d$ (respectively $c_1, \dots, c_d$) be the set of irreducible characters (respectively conjugacy classes) of G. We refer the reader to [Bueno et al. 2016] for computations and conjectures concerning the rank of the $d \times d$-matrix $[\langle \chi_i, c_j \rangle_G]$ associated to the Stickelberger pairing $\langle -, - \rangle_G$ when G is cyclic.

10. The Stickelberger map and transpose homomorphisms

The Stickelberger map.

Definition 10.1. The *Stickelberger map*

$$\Theta = \Theta_G : \mathbb{Q}R_G \to \mathbb{Q}G \tag{10-1}$$

is defined by

$$\Theta(\chi) = \sum_{g \in G} \langle \chi, g \rangle_G \cdot g.$$

We write $G(-1)$ for the set G endowed with an action of Ω_F via the inverse cyclotomic character. Note that in general, for nonabelian G, this Ω_F-action is not an action on G via group automorphisms; it is only an action on the set G. However, it does induce an action on the additive group $\mathbb{Q}G(-1)$, which is all that we shall require.

The following proposition summarises some basic properties of the Stickelberger map.

Proposition 10.2. (a) *We have that $\Theta(\chi) \in Z(\mathbb{Q}G)$ for all $\chi \in R_G$, i.e., in fact*

$$\Theta : \mathbb{Q}R_G \to Z(\mathbb{Q}G).$$

(b) *Suppose that $\chi \in R_G$. Then $\Theta(\chi) \in \mathbb{Z}G$ if and only if $\chi \in A_G$. Hence Θ induces a homomorphism $A_G \to \mathbb{Z}G$.*

(c) *The map*

$$\Theta : \mathbb{Q}R_G \to \mathbb{Q}G(-1)$$

is Ω_F-equivariant.

Proof. The proofs of these assertions for arbitrary G are essentially the same as those in the case of abelian G. See [McCulloh 1987, Propositions 4.3 and 4.5].

(a) It follows from the definition of the Stickelberger pairing that if $\chi \in R_G$ and $g \in G$, then $\langle \chi, g \rangle_G$ is determined by the conjugacy class $c(g)$ of g in G. This implies that $\Theta(R_G) \subseteq Z(\mathbb{Q}G)$, as claimed.

(b) Suppose that $\chi \in R_G$ and $g \in G$. Write

$$\chi|_{\langle g\rangle} = \sum_{\eta} a_\eta \eta,$$

where the sum is over irreducible characters of $\langle g\rangle$, and set $\zeta_{|g|} := \zeta_{|G|}^{|G|/|g|}$. Then

$$(\det(\chi))(g) = \det(\chi|_{\langle g\rangle})(g) = \prod_{\eta} \eta(g)^{a_\eta} = \prod_{\eta} \zeta_{|g|}^{|g|\langle a_\eta \eta, g\rangle_{\langle g\rangle}} = \zeta_{|g|}^{|g|\sum_\eta \langle a_\eta \eta, g\rangle_{\langle g\rangle}} = \zeta_{|g|}^{|g|\langle \chi, g\rangle_G}.$$

It now follows that $\langle \chi, g\rangle_G \in \mathbb{Z}$ for all $g \in G$ if and only if $\chi \in \operatorname{Ker}(\det) = A_G$, as required.

(c) Let κ denote the cyclotomic character of Ω_F, and suppose that $\chi \in R_G$ is of degree one. Then, for each $g \in G$ and $\omega \in \Omega_F$, we have

$$\chi^{\omega}(g) = \chi(g^{\kappa(\omega)}),$$

and so

$$\langle \chi^{\omega}, g\rangle_G = \langle \chi, g^{\kappa(\omega)}\rangle_G. \tag{10-2}$$

It follows via bilinearity that (10-2) holds for all $\chi \in R_G$ and all $g \in G$. Hence, if we view $\Theta(\chi)$ as being an element of $\mathbb{Q}G(-1)$, then

$$\Theta(\chi^{\omega}) = \sum_{g\in G} \langle \chi^{\omega}, g\rangle_G \cdot g = \sum_{g\in G} \langle \chi, g^{\kappa(\omega)}\rangle_G \cdot g = \sum_{g\in G} \langle \chi, g\rangle_G \cdot g^{\kappa^{-1}(\omega)} = \Theta(\chi)^{\omega}. \qquad \square$$

Transpose Stickelberger homomorphisms. We see from Proposition 10.2 that dualising the homomorphism

$$\Theta : A_G \to Z(\mathbb{Z}G)$$

and twisting by the inverse cyclotomic character yields an Ω_F-equivariant *transpose Stickelberger homomorphism*

$$\Theta^t : \operatorname{Hom}(Z(\mathbb{Z}G(-1)), (F^c)^\times) \to \operatorname{Hom}(A_G, (F^c)^\times). \tag{10-3}$$

Composing (10-3) with the sequence of homomorphisms

$$\operatorname{Hom}(A_G, (F^c)^\times) \xrightarrow{\sim} Z(F^cG)^\times/G^{\text{ab}} \to \frac{\operatorname{Det}(F^cG)^\times}{\operatorname{Det}(O_FG)^\times} \to K_0(O_FG, F^c), \tag{10-4}$$

(where the first arrow is given by (4-6), the second via (the inverse of) (4-3), and the third is via the homomorphism ∂^1 of (6-1)) yields a homomorphism

$$K\Theta^t : \operatorname{Hom}(Z(\mathbb{Z}G(-1)), (F^c)^\times) \to K_0(O_FG, F^c). \tag{10-5}$$

Hence, if we write $\mathcal{C}(G(-1))$ for the set of conjugacy classes of G endowed with Ω_F-action via the inverse cyclotomic character, and set

$$\Lambda(O_FG) := \operatorname{Hom}_{\Omega_F}(Z(\mathbb{Z}G(-1)), O_{F^c}) = \operatorname{Map}_{\Omega_F}(\mathcal{C}(G(-1)), O_{F^c}) = Z(O_{F^c}[G(-1)])^{\Omega_F},$$
$$\Lambda(FG) := \operatorname{Hom}_{\Omega_F}(Z(\mathbb{Z}G(-1)), F^c) = \operatorname{Map}_{\Omega_F}(\mathcal{C}(G(-1)), F^c) = Z(F^c[G(-1)])^{\Omega_F},$$

then $K\Theta^t$ induces a homomorphism (which we denote by the same symbol):

$$K\Theta^t : \Lambda(FG)^\times \to K_0(O_F G, F^c).$$

For each place v of F, we may apply the discussion above with F replaced by F_v to obtain local versions

$$\Theta^t_v : \mathrm{Hom}(Z(\mathbb{Z}G(-1)), (F^c_v)^\times) \to \mathrm{Hom}(A_G, (F^c_v)^\times) \tag{10-6}$$

and

$$K\Theta^t_v : \Lambda(F_v G)^\times \to K_0(O_{F_v} G, F^c_v) \tag{10-7}$$

of the maps Θ^t and $K\Theta^t$ respectively. The homomorphism Θ^t commutes with local completion, and $K\Theta^t$ commutes with the localisation maps

$$\lambda_v : K_0(O_F G, F^c) \to K_0(O_{F_v} G, F^c_v).$$

Definition 10.3. We define the group of ideles $J(\Lambda(FG))$ of $\Lambda(FG)$ to be the restricted direct product over all places v of F of the groups $\Lambda(F_v G)^\times$ with respect to the subgroups $\Lambda(O_{F_v} G)^\times$.

For all finite places v of F not dividing the order of G, as $O_{F_v} G$ is an O_{F_v}-maximal order in $F_v G$, we have that (see Proposition 4.5(ii))

$$\Theta^t_v(\Lambda(O_{F_v} G)) \subseteq \mathrm{Hom}_{\Omega_{F_v}}(A_G, (O_{F^c_v})^\times) = \mathrm{Det}(\mathcal{H}(O_{F_v} G)),$$

and so

$$K\Theta^t_v(\Lambda(O_{F_v} G)) \subseteq K_0(O_{F_v} G, O_{F^c_v}).$$

It follows that the homomorphisms Θ^t_v combine to yield an idelic transpose Stickelberger homomorphism

$$K\Theta^t : J(\Lambda(FG)) \to J(K_0(O_F G, F^c)). \tag{10-8}$$

We shall see in the next subsection that the idelic homomorphism $K\Theta^t$ is closely related to the homomorphism

$$\Psi^{\mathrm{id}} : J(H^1_t(F, G)) \to J(K_0(O_F G, F^c))$$

of Definition 6.2.

Prime F-elements.

Definition 10.4. Let v be a place of F. For each element $s \neq e$ of $\Sigma_v(G)$ (see Definition 7.2 and (8-1)), define $f_{v,s} \in \Lambda(F_v G)^\times$ by

$$f_{v,s}(c) = \begin{cases} -1 & \text{if } v \text{ is real and } c = c(s), \\ \varpi_v & \text{if } v \text{ is finite and } c = c(s), \\ 1 & \text{otherwise.} \end{cases} \tag{10-9}$$

Observe that $f_{v,s}$ is Ω_{F_v}-equivariant because $s \in \Sigma_v(G)$ and so Ω_{F_v} fixes $c(s)$ when s is viewed as an element of $G(-1)$. The element $f_{v,s}$ depends only upon the conjugacy class $c(s)$ of s. For all places v of F, we define $f_{v,e} \in (\Lambda(F_vG))^\times$ to be the constant function $f_{v,e} = 1$.

Write

$$\boldsymbol{F}_v := \{f_{v,s} \mid s \in \Sigma_v(G)\},$$

and define the subset $\boldsymbol{F} \subset J(\Lambda(FG))$ of prime $\boldsymbol{F}$-elements by

$$f \in \boldsymbol{F} \iff f \in J(\Lambda(FG)) \text{ and } f_v \in \boldsymbol{F}_v \text{ for all places } v \text{ of } F.$$

Following [Byott 1998, Definition 7.1], we define the *support* $\mathrm{Supp}(f)$ of $f \in \boldsymbol{F}$ to be set of all places v of F for which $f_v \neq 1$. We say that f is *full* if, for each $s \in G$ there is a place v with $f_v = f_{v,s}$.

Our interest in the set $\boldsymbol{F}$, as well as the relationship between $K\Theta^t$ and Ψ^{id}, is explained by the following result.

Proposition 10.5. *Let v be a place of F.*

(a) *For each $s \in \Sigma_v(G)$, we have*

$$\mathrm{Det}(\boldsymbol{r}_G(\varphi_{v,s})) = K\Theta^t_v(f_{v,s})$$

in $K_0(O_{F_v}G, F_v^c)$.

(b) *Suppose that $s_1, s_2 \in \Sigma_v(G)$ with*

$$\mathrm{Det}(\boldsymbol{r}_G(\varphi_{v,s_1})) = \mathrm{Det}(\boldsymbol{r}_G(\varphi_{v,s_2})). \tag{10-10}$$

Then $\langle s_1\rangle$ is conjugate in G to $\langle s_2\rangle$.

(c) *Suppose that v is finite. Let $\pi_1, \pi_2 \in \mathrm{Hom}(\Omega_{F_v}, G)$ with $[\pi_i] \in H^1_t(F_v, G)$ for each i, and set $s_i = \pi_i(\sigma_v)$ (see (7-5)). Let a_i be a normal integral basis generator of $F_{v,\pi_i}/F_v$, and let*

$$\boldsymbol{r}_G(a_i) = u_i \cdot \boldsymbol{r}_G(a_{i,nr}) \cdot \boldsymbol{r}_G(\varphi_{s_i})$$

be a Stickelberger factorisation of $\boldsymbol{r}_G(a_i)$ (see Definition 7.12). Suppose that

$$\mathrm{Det}(\boldsymbol{r}_G(a_1)) \cdot \mathrm{Det}(\boldsymbol{r}_G(a_2))^{-1} \in \mathrm{Det}((O_{F_v^c}G)^\times). \tag{10-11}$$

Then

$$\mathrm{Det}(\boldsymbol{r}_G(\varphi_{s_1})) = \mathrm{Det}(\boldsymbol{r}_G(\varphi_{s_2}))$$

and for some integer m and some $h \in G$, the equality

$$\pi_1(\omega) = h \cdot \pi_2(\omega)^m \cdot h^{-1}$$

holds for all $\omega \in I_v$.

Proof.

(a) The proof of this assertion is very similar to that of [McCulloh 1987, Proposition 5.4].

It suffices to show that the equality

$$\mathrm{Det}(\boldsymbol{r}_G(\varphi_{v,s})) = \Theta_v^t(f_{v,s})$$

holds in $\mathrm{Hom}(A_G, (F_v^c)^\times)$.

Let $\chi \in R_G$, and write

$$\chi|_{<s>} = \sum_\eta a_\eta \eta,$$

where the sum is over irreducible characters η of $\langle s \rangle$.

Suppose first that v is finite. Using (7-2), we see that (cf. [McCulloh 1987, Proposition 5.4])

$$[\mathrm{Det}(\boldsymbol{r}_G(\varphi_{v,s}))](\chi) = \prod_\eta \biggl(\sum_{i=0}^{|s|-1} \sigma_v^i(\beta_s)\eta(s^{-i}) \biggr)^{a_\eta} = \varpi_v^{\langle \sum_\eta a_\eta \eta, s\rangle_{\langle s \rangle}} = \varpi_v^{\langle \chi, s \rangle_G}, \tag{10-12}$$

and so it follows that

$$[\mathrm{Det}(\boldsymbol{r}_G(\varphi_{v,s}))](\alpha) = \varpi_v^{\langle \alpha, s \rangle_G}$$

for all $\alpha \in A_G$.

If v is real, then the proof of Proposition 8.2 shows directly that

$$[\mathrm{Det}(\boldsymbol{r}_G(\varphi_{v,s}))](\chi) = (-1)^{\langle \chi, s \rangle_G},$$

and so we have

$$[\mathrm{Det}(\boldsymbol{r}_G(\varphi_{v,s}))](\alpha) = (-1)^{\langle \alpha, s \rangle_G}$$

for all $\alpha \in A_G$ in this case also.

Now suppose that v is either finite or real. If $\alpha \in A_G$, then we have

$$(\Theta_v^t(f_{v,s}))(\alpha) = f_{v,s}(\Theta(\alpha)) = f_{v,s}\biggl(\sum_{g \in G} \langle \alpha, g \rangle_G \cdot g \biggr) = \prod_{g \in G} f_{v,s}(g)^{\langle \alpha, g \rangle_G} = \begin{cases} \varpi_v^{\langle \alpha, s \rangle_G} & \text{if } v \text{ is finite,} \\ (-1)^{\langle \alpha, s \rangle_G} & \text{if } v \text{ is real.} \end{cases}$$

The desired result now follows.

(b) The proof of (a) above shows that if (10-10) holds, then

$$\langle \chi, s_1 \rangle_G = \langle \chi, s_2 \rangle_G$$

for every $\chi \in R_G$. It therefore follows from Corollary 9.4 that $\langle s_1 \rangle$ is conjugate in G to $\langle s_2 \rangle$.

(c) Observe that (10-11) holds if and only if

$$\mathrm{Det}(\boldsymbol{r}_G(\varphi_{s_1})) \cdot \mathrm{Det}(\boldsymbol{r}_G(\varphi_{s_2})^{-1}) \in \mathrm{Det}((O_{F_v^c} G)^\times), \tag{10-13}$$

and the proof of part (a) (see (10-12)) implies that (10-13) holds if and only if

$$\mathrm{Det}(\boldsymbol{r}_G(\varphi_{s_1})) = \mathrm{Det}(\boldsymbol{r}_G(\varphi_{s_2})).$$

Part (b) therefore implies that $\langle s_1\rangle$ and $\langle s_2\rangle$ are conjugate. Hence

$$s_1 = h \cdot s_2^m \cdot h^{-1}$$

for some $m \in \mathbb{Z}$ and $h \in G$, and so

$$\boldsymbol{r}_G(\varphi_{s_1}) = h \cdot \boldsymbol{r}_G(\varphi_{s_2^m}) \cdot h^{-1}$$

(see (7-3)).

For any $\omega \in \Omega_{F_v^{\mathrm{nr}}}$, we have

$$\pi_i(\omega) = \boldsymbol{r}_G(a_i)^{-1} \cdot \boldsymbol{r}_G(a_i)^{\omega} = \boldsymbol{r}_G(\varphi_{s_i})^{-1} \cdot \boldsymbol{r}_G(\varphi_{s_i})^{\omega}.$$

Applying the map $F_v^c G \to F_v^c G$ defined by $\sum_g a_g g \mapsto \sum_g a_g g^m$ to this equality (when $i = 2$) yields

$$\pi_2(\omega)^m = \boldsymbol{r}_G(\varphi_{s_2^m})^{-1} \cdot \boldsymbol{r}_G(\varphi_{s_2^m})^{\omega}.$$

The final assertion now follows. □

The Stickelberger pairing revisited. In this subsection we shall briefly describe an alternative definition of the Stickelberger pairing that involves a direct connection with resolvends of local normal integral basis generators. This will not be used in the sequel.

Let v be a finite place of F. There is a natural pairing

$$\{-,-\}_{G,v} : \mathrm{Irr}(G) \times H^1(F_v^{\mathrm{nr}}, G) \to \mathbb{Q}/\mathbb{Z}, \quad (\chi, [\pi]) \mapsto [v(\mathrm{Det}(\boldsymbol{r}_G(a(\pi)))(\chi))], \tag{10-14}$$

where $a(\pi)$ is any normal basis generator of $F_{v,\pi}^{\mathrm{nr}}/F_v^{\mathrm{nr}}$. Recall that every element of $H_t^1(F_v^{\mathrm{nr}}, G)$ is of the form $\tilde{\varphi}_{v,s}$ for some $s \in G$ with $v \nmid |s|$ (see Remark 7.11). The restriction of $\{-,-\}_{G,v}$ to $\mathrm{Irr}(G) \times H_t^1(F_v^{\mathrm{nr}}, G)$ yields a refined pairing

$$\{-,-\}_{G,v}^{(1)} : \mathrm{Irr}(G) \times H_t^1(F_v^{\mathrm{nr}}, G) \to \mathbb{Q}, \quad (\chi, \tilde{\varphi}_{v,s}) \mapsto v(\mathrm{Det}(\boldsymbol{r}_G(\varphi_{v,s}))(\chi)). \tag{10-15}$$

This leads to the following definition.

Definition 10.6. Suppose that v is finite and that $v \nmid |G|$. We define a pairing

$$[-,-]_{G,v} : \mathrm{Irr}(G) \times G \to \mathbb{Q}, \quad (\chi, g) \mapsto v(\mathrm{Det}(\boldsymbol{r}_G(\varphi_{v,g}))(\chi)), \tag{10-16}$$

and we extend this to a pairing on $\mathbb{Q}R_G \times \mathbb{Q}G$ via linearity.

Proposition 10.7. *Suppose that v is finite and that $v \nmid |G|$. Then for each $\chi \in \mathrm{Irr}(G)$ and $g \in G$, we have*

$$[\chi, g]_{G,v} = [\chi|_{\langle g\rangle}, g]_{\langle g\rangle,v}. \tag{10-17}$$

Proof. Set $H := \langle g \rangle$. The property (10-17) is a direct consequence of the fact that the restriction map $R_G \to R_H$ induces a homomorphism $\mathrm{Hom}(R_H, (F_v^c)^\times) \to \mathrm{Hom}(R_G, (F_v^c)^\times)$ such that the following diagram commutes:

$$\begin{tikzcd}
(F_v^c H)^\times \arrow[r, "\subseteq"] \arrow[d, "\mathrm{Det}"] & (F_v^c G)^\times \arrow[d, "\mathrm{Det}"] \\
\mathrm{Hom}(R_H, (F_v^c)^\times) \arrow[r] & \mathrm{Hom}(R_G, (F_v^c)^\times)
\end{tikzcd}$$

(see, e.g., [Fröhlich 1976, p. 436; 1984, p. 118]). □

Proposition 10.8. *Suppose that v is finite and that $v \nmid |G|$. Then for each $\chi \in \mathrm{Irr}(G)$ and $g \in G$, we have*

$$[\chi, g]_{G,v} = \langle \chi, g \rangle_G. \tag{10-18}$$

In particular, $[-,-]_{G,v}$ is independent of our choice of v.

Proof. Proposition 10.7 implies that we may assume that G is cyclic. The equality (10-18) may then be established via an argument identical to that used in the proof of Proposition 10.5(a) (see also [McCulloh 1987, Proposition 5.4]). □

11. Modified ray class groups

Definition 11.1. Let $\mathfrak{a}$ be an integral ideal of O_F. For each finite place v of F, recall that

$$U_{\mathfrak{a}}(O_{F_v^c}) := (1 + \mathfrak{a} O_{F_v^c}) \cap (O_{F_v^c})^\times.$$

We define

$$U'_{\mathfrak{a}}(\Lambda(O_{F_v} G)) \subseteq \Lambda(F_v G)^\times = \mathrm{Map}_{\Omega_{F_v}}(\mathcal{C}(G(-1)), (F_v^c)^\times)$$

by

$$U'_{\mathfrak{a}}(\Lambda(O_{F_v} G)) := \{ g_v \in \Lambda(F_v G)^\times | g_v(c) \in U_{\mathfrak{a}}(O_{F_v^c}) \quad \forall c \neq 1 \}$$

(with $g_v(1)$ allowed to be arbitrary).

Set

$$U'_{\mathfrak{a}}(\Lambda(O_F G)) := \Bigl(\prod_v U'_{\mathfrak{a}}(\Lambda(O_{F_v} G)) \Bigr) \cap J(\Lambda(FG)).$$

Definition 11.2. For each real place v of F, we define

$$\Lambda(F_v G)_+^\times := \{ g_v \in \Lambda(F_v G)^\times \mid g_v(c) \in \mathbb{R}_{>0}^\times \text{ for all } c \in \mathcal{C}(G(-1)) \}$$

(with $g_v(1)$ allowed to be arbitrary).

If v is complex, we set $\Lambda(F_v G)_+^\times := \Lambda(F_v G)^\times$. We define

$$U'_\infty(\Lambda(O_F G)) := \Bigl(\prod_{v \mid \infty} \Lambda(FG)^\times \Bigr) \cap J(\Lambda(FG)),$$

and

$$U'_\infty(\Lambda(O_F G))_+ := \Bigl(\prod_{v \mid \infty} \Lambda(FG)_+^\times\Bigr) \cap J(\Lambda(FG)).$$

Definition 11.3. The *modified ray class group modulo* $\mathfrak{a}$ of $\Lambda(O_F G)$ is defined by

$$\mathrm{Cl}'_{\mathfrak{a}}(\Lambda(O_F G)) := \frac{J(\Lambda(FG))}{\Lambda(FG)^\times \cdot U'_{\mathfrak{a}}(\Lambda(O_F G)) \cdot U'_\infty(\Lambda(O_F G))}.$$

The *modified narrow ray class group modulo* $\mathfrak{a}$ is defined by

$$\mathrm{Cl}'^{+}_{\mathfrak{a}}(\Lambda(O_F G)) := \frac{J(\Lambda(FG))}{\Lambda(FG)^\times \cdot U'_{\mathfrak{a}}(\Lambda(O_F G)) \cdot U'_\infty(\Lambda(O_F G))_+}.$$

We refer to the elements of $\mathrm{Cl}'_{\mathfrak{a}}(\Lambda(O_F G))$ (respectively $\mathrm{Cl}'^{+}_{\mathfrak{a}}(\Lambda(O_F G))$) as the *modified ray classes* (respectively *modified narrow ray classes*) of $\Lambda(O_F G)$ modulo $\mathfrak{a}$.

Remark 11.4. Fix a set of representatives T of $\Omega_F \backslash \mathcal{C}(G(-1))$, and for each $t \in T$, let $F(t)$ be the smallest extension of F such that $\Omega_{F(t)}$ fixes t. Then the Wedderburn decomposition of $\Lambda(FG)$ is given by

$$\Lambda(FG) = \mathrm{Map}_{\Omega_F}(\mathcal{C}(G(-1)), F^c) \simeq \prod_{t \in T} F(t), \tag{11-1}$$

where the isomorphism is induced by evaluation on the elements of T.

The group $\mathrm{Cl}'_{\mathfrak{a}}(\Lambda(O_F G))$ (respectively $\mathrm{Cl}'^{+}_{\mathfrak{a}}(\Lambda(O_F G))$) above is finite, and is isomorphic to the product of the ray class groups $\mathrm{Cl}_{\mathfrak{a}}(O_{F(t)})$ (respectively the narrow ray class groups $\mathrm{Cl}^{+}_{\mathfrak{a}}(O_{F(t)})$) modulo $\mathfrak{a}$ of the Wedderburn components $F(t)$ of $\Lambda(FG)$ with $t \neq 1$. There is a natural surjection

$$\mathrm{Cl}'^{+}_{\mathfrak{a}}(\Lambda(O_F G)) \to \mathrm{Cl}'_{\mathfrak{a}}(\Lambda(O_F G))$$

with kernel an elementary abelian 2-group.

If $|G|$ is odd, then (as no nontrivial element of G is conjugate to its inverse) $F(t)$ has no real places when $t \neq 1$, and so $\mathrm{Cl}_{\mathfrak{a}}(O_{F(t)}) = \mathrm{Cl}^{+}_{\mathfrak{a}}(O_{F(t)})$. Hence we have

$$\mathrm{Cl}'^{+}_{\mathfrak{a}}(\Lambda(O_F G)) = \mathrm{Cl}_{\mathfrak{a}}(\Lambda(O_F G))$$

whenever G is of odd order.

Proposition 11.5. *Let* $\mathfrak{a}$ *be any integral ideal of* O_F. *Then the inclusion* $\boldsymbol{F} \to J(\Lambda(FG))$ *induces a surjection* $\boldsymbol{F} \to \mathrm{Cl}'^{+}_{\mathfrak{a}}(\Lambda(O_F G))$. *In particular, each modified narrow ray class modulo* $\mathfrak{a}$ *of* $\Lambda(O_F G)$ *contains infinitely many elements of* $\boldsymbol{F}$.

Proof. Let $I(\Lambda(O_F G))$ denote the group of fractional ideals of $\Lambda(O_F G)$. Then via the Wedderburn decomposition (11-1) of $\Lambda(FG)$, we see that each fractional ideal $\mathfrak{B}$ in $\Lambda(O_F G)$ may be written in the form $\mathfrak{B} = (\mathfrak{B}_t)_{t \in T}$, where each $\mathfrak{B}_t$ is a fractional ideal of $O_{F(t)}$. For each conjugacy class $t \in T$, let $o(t)$ denote the Ω_F-orbit of t in $\mathcal{C}(G(-1))$, and write $|t|$ for the order of any element of t.

For each idele $\nu \in J(\Lambda(FG))$, let

$$\mathrm{co}(\nu) := [\mathrm{co}(\nu)_t]_{t\in T} \in I(\Lambda(O_FG)) \simeq \prod_{t\in T} I(O_{F(t)})$$

denote the ideal obtained by taking the idele content of ν. If v is a place of F, we view $\boldsymbol{F}_v$ as being a subset of $\boldsymbol{F}$ via the obvious embedding $\Lambda(F_vG)^\times \subseteq J(\Lambda(FG))$, and we set

$$\mathcal{F}_v := \{\mathrm{co}(f_v) \mid f_v \in \boldsymbol{F}_v\}.$$

Now suppose that v is finite, and consider the ideal

$$\mathrm{co}(f_{v,s}) = [\mathrm{co}(f_{v,s})_t]_{t\in T}$$

in $I(\Lambda(O_FG))$. If $c(s) \notin o(t)$, then it follows from the definition of $f_{v,s}$ that $\mathrm{co}(f_{v,s})_t = O_{F(t)}$. Suppose that $c(s) \in o(t)$. Since $s \in \Sigma_v(G)$, it follows that $v(|s|) = 0$ and that Ω_{F_v} fixes $c(s)$. Hence $F_v(t) = F_v$, and so we see that $\mathrm{co}(f_{v,s})_t$ is a prime ideal of $O_{F(t)}$ of degree one lying above v (cf. [McCulloh 1987, pp. 287–289]). Furthermore, if $t \in T$ and if v is a finite place of F that is totally split in $F(t)$, then $f_{v,s} \in \boldsymbol{F}_v$ for all $c(s) \in o(t)$.

We therefore deduce that if v is finite, the set $\mathcal{F}_v$ consists precisely of the invertible prime ideals $\mathfrak{p} = (\mathfrak{p}_t)_{t\in T}$ of $\Lambda(O_FG)$ with $\mathfrak{p}_{t_1}$ a prime of degree one above v in $F(t_1)$ for some $t_1 \in T$ with $v(|t_1|) = 0$ and $\mathfrak{p}_t = O_{F(t)}$ for all $t \neq t_1$. For every $t \in T$, the narrow ray class modulo $\mathfrak{a}$ of $F(t)$ contains infinitely many primes of degree one, and this implies that $\boldsymbol{F}$ surjects onto $\mathrm{Cl}_{\mathfrak{a}}'^{+}(\Lambda(O_FG))$ as claimed. □

Our next result describes a transpose Stickelberger homomorphism on modified narrow ray class groups $\mathrm{Cl}_{\mathfrak{a}}'^{+}(\Lambda(O_FG))$ for a suitable choice of $\mathfrak{a}$. Before stating it, we remind the reader that Proposition 6.3 implies that $\prod_v \mathrm{Im}(\Psi_v^{\mathrm{nr}})$ is a subgroup of $J(K_0(O_FG, F^c))$.

Proposition 11.6. *Let N be an integer, and set $\mathfrak{a} := N \cdot O_F$. Then if N is divisible by a sufficiently high power of $|G|$, the idelic transpose Stickelberger homomorphism*

$$K\Theta^t : J(\Lambda(FG)) \to J(K_0(O_FG, F^c))$$

induces a homomorphism

$$\Theta_{\mathfrak{a}}^t : \mathrm{Cl}_{\mathfrak{a}}'^{+}(\Lambda(O_FG)) \to \frac{J(K_0(O_FG, F^c))}{\lambda[\partial^1(K_1(F^cG))] \cdot \prod_v \mathrm{Im}(\Psi_v^{\mathrm{nr}})}.$$

Proof. To show this, we first observe that Proposition 4.6 implies that if N is divisible by a sufficiently high power of $|G|$ and v is any finite place of F, then we have

$$\Theta_v^t(U_{\mathfrak{a}}'(\Lambda(O_{F_v}G))) \subseteq \mathrm{Det}((O_{F_v}G)^\times/G) \subseteq \mathrm{Det}(\mathcal{H}(O_{F_v}G)) = \mathrm{Im}(\Psi_v^{\mathrm{nr}}),$$

and so it follows that

$$K\Theta^t(U_{\mathfrak{a}}'(\Lambda(O_FG))) \subseteq \prod_v \mathrm{Im}(\Psi_v^{\mathrm{nr}})$$

in $J(K_0(O_FG, F^c))$.

Suppose that v is a real place of F and that $h \in \Lambda(F_v G)_{+}^{\times}$. Then for each $\chi \in R_G$, we have (recalling that $\langle \chi, e \rangle_G = 0$)

$$\Theta_v^t(h)(\chi) = \prod_{g \in G} h(c(g))^{\langle \chi, g \rangle_G} > 0,$$

and so $\Theta_v^t(h) \in \mathrm{Hom}^{+}_{\Omega_{F_v}}(R_G, (F_v^c)^{\times})$. This implies that $K\Theta^t(h) = 1$ in $K_0(O_{F_v}G, F_v^c)$, and therefore $K\Theta^t(U'_{\infty}(\Lambda(O_F G))) = 1$ in $J(K_0(O_F G, F^c))$.

It now follows that $K\Theta^t$ induces a homomorphism

$$\Theta^t_{\mathfrak{a}} : \mathrm{Cl}'^{+}_{\mathfrak{a}}(\Lambda(O_F G)) \to \frac{J(K_0(O_F G, F^c))}{\lambda[\partial^1(K_1(F^c G))] \cdot \prod_v \mathrm{Im}(\Psi_v^{\mathrm{nr}})},$$

as claimed. □

12. Proof of Theorem 6.6

In this section we shall prove Theorem 6.6. Recall that we wish to show that if

$$\overline{\Psi^{\mathrm{id}}} : J(H_t^1(F, G)) \to \frac{J(K_0(O_F G, F^c))}{\lambda[\partial^1(K_1(F^c G))] \cdot \prod_v \mathrm{Im}(\Psi_v^{\mathrm{nr}})}$$

denotes the map of pointed sets given by the composition of the map Ψ^{id} with the quotient homomorphism

$$q_1 : J(K_0(O_F G, F^c)) \to \frac{J(K_0(O_F G, F^c))}{\lambda[\partial^1(K_1(F^c G))] \cdot \prod_v \mathrm{Im}(\Psi_v^{\mathrm{nr}})},$$

then the image of $\overline{\Psi^{\mathrm{id}}}$ is in fact a group.

To show this, we choose an ideal $\mathfrak{a} = N \cdot O_F$ as in Proposition 11.6, and we consider the diagram

$$\begin{tikzcd}
 & & J(H_t^1(F, G)) \arrow[d, "\Psi^{\mathrm{id}}"'] \\
\boldsymbol{F} \arrow[r, "\subset"] \arrow[d, "q_2"'] & J(\Lambda(FG)) \arrow[r, "K\Theta^t"] \arrow[d, "q_2"'] & J(K_0(O_F G, F^c)) \arrow[d, "q_1"'] \\
\mathrm{Cl}'^{+}_{\mathfrak{a}}(\Lambda(O_F G)) \arrow[r, equal] & \mathrm{Cl}'^{+}_{\mathfrak{a}}(\Lambda(O_F G)) \arrow[r, "\Theta^t_{\mathfrak{a}}"] & \dfrac{J(K_0(O_F G, F^c))}{\lambda[\partial^1(K_1(F^c G))] \cdot \prod_v \mathrm{Im}(\Psi_v^{\mathrm{nr}})}
\end{tikzcd} \tag{12-1}$$

Here q_2 denotes the obvious quotient map. Proposition 11.6 shows that the right-hand square commutes, and Proposition 11.5 shows that the left-most vertical arrow is surjective.

It follows from Proposition 10.5(a) that

$$q_1[K\Theta^t(\boldsymbol{F})] = q_1[\Psi^{\mathrm{id}}(J(H_t^1(F, G)))] = \mathrm{Im}\,\overline{\Psi^{\mathrm{id}}}.$$

On the other hand, we also have that

$$q_1[K\Theta^t(\boldsymbol{F})] = \Theta^t_{\mathfrak{a}}(\mathrm{Cl}'^{+}_{\mathfrak{a}}(\Lambda(O_F G))),$$

which is a group. It therefore follows that $\mathrm{Im}(\overline{\Psi^{\mathrm{id}}})$ is indeed a group, as claimed.

This completes the proof of Theorem 6.6. □

13. Realisable classes from field extensions

In this section, after first proving that the kernel of Ψ is finite, we explain how a slightly weaker form of Conjecture B implies that every element of $\mathcal{R}(O_F G)$ may be realised by the ring of integers of a tame field (as opposed to merely a Galois algebra) G-extension of F.

Recall that G' denotes the derived subgroup of G, and note that we may view $H^1(F, G')$ and $H^1(F_v, G')$ as being pointed subsets of $H^1(F, G)$ and $H^1(F_v, G)$ respectively via taking Galois cohomology of the exact sequence of groups

$$0 \to G' \to G \to G^{\mathrm{ab}} \to 0.$$

Recall also that we write $H^1_{\mathrm{fnr}}(F, G')$ for the set of isomorphism classes of G'-Galois F-algebras that are unramified at all finite places of F.

Proposition 13.1. (a) *Let v be a finite place of F. Then* $\mathrm{Ker}(\Psi_v) \subseteq H^1_{\mathrm{nr}}(F_v, G')$.

(b) *Suppose that* $[\pi] \in \mathrm{Ker}(\Psi)$. *Then* $[\pi] \in H^1_{\mathrm{fnr}}(F, G') \subseteq H^1(F, G)$. *We have that* $\mathrm{Ker}(\Psi)$ *is finite.*

(c) *Suppose that $F/\mathbb{Q}$ is at most tamely ramified at all primes dividing $|G|$. Then* $H^1_{\mathrm{nr}}(F, G') \subseteq \mathrm{Ker}(\Psi)$.

(d) *Suppose that G has no irreducible symplectic characters or that F has no real places. Suppose also that $F/\mathbb{Q}$ is at most tamely ramified at all primes dividing $|G|$. Then* $\mathrm{Ker}(\Psi) = H^1_{\mathrm{fnr}}(F, G')$.

Proof.

(a) Let v be a finite place of F. Suppose that $[\pi_v] \in H^1_t(F_v, G)$, and that $O_{\pi_v} = O_{F_v} G \cdot a_v$. Recall (see Sections 5 and 6) that we have

$$\Psi_v : H^1_t(F_v, G) \to K_0(O_{F_v} G, F_v^c) \simeq \frac{\mathrm{Det}(F_v^c G)^\times}{\mathrm{Det}(O_{F_v} G)^\times},$$

and that $\Psi_v([\pi_v]) = [\mathrm{Det}(\boldsymbol{r}_G(a_v))]$ (see also Definition 4.1 and Remark 4.2). It follows that $\Psi_v([\pi_v]) = 0$ if and only if $\mathrm{Det}(\boldsymbol{r}_G(a_v)) \in \mathrm{Det}(O_{F_v} G)^\times$.

Hence, if $\Psi_v([\pi_v]) = 0$, then for each $\omega \in \Omega_{F_v}$, we have

$$\mathrm{Det}(\boldsymbol{r}_G(a_v)^{-1}) \cdot \mathrm{Det}(\boldsymbol{r}_G(a_v))^\omega = 1,$$

and so we deduce from (3-8) that $[\pi_v]$ lies in the kernel of the natural map $H^1(F_v, G) \to H^1(F_v, G^{\mathrm{ab}})$ of pointed sets. This implies that $[\pi_v] \in H^1(F_v, G')$. Finally, we see from (7-11) and Proposition 10.5(c) that $\mathrm{Det}(\boldsymbol{r}_G(a_v)) \in \mathrm{Det}((O_{F_v} G)^\times)$ only if $[\pi_v] \in H^1_{\mathrm{nr}}(F_v, G)$. We now conclude that if $[\pi_v] \in \mathrm{Ker}(\Psi_v)$, then $[\pi_v] \in H^1_{\mathrm{nr}}(F_v, G')$. This establishes part (a).

(b) Suppose that $[\pi] \in H^1(F, G)$ satisfies $\Psi([\pi]) = 0$. Then $\Psi_v(\mathrm{loc}_v([\pi])) = 0$ for each place v, and so it follows from part (a) that $\mathrm{loc}_v([\pi]) \in H^1_{\mathrm{nr}}(F_v, G')$ for all finite places v of F. Therefore $[\pi] \in H^1(F, G')$, and π is unramified at each finite place of F, i.e., $[\pi] \in H^1_{\mathrm{fnr}}(F, G')$. As there are only finitely many unramified extensions of F of bounded degree, it follows that $H^1_{\mathrm{fnr}}(F, G')$ is finite, and so $\mathrm{Ker}(\Psi)$ is finite, as claimed.

(c) Suppose that $[\pi] \in H^1_{\mathrm{nr}}(F, G') \subseteq H^1_t(F, G)$, and write $O_{\pi_v} = O_{F_v}G \cdot a_v$ for each finite place v of F. As π is unramified at v, it follows that $\mathrm{Det}(\boldsymbol{r}_G(a_v)) \in \mathrm{Det}(O_{F_v^{\mathrm{nr}}}G)^\times$. Since $\mathrm{loc}_v([\pi])$ lies in the kernel of the natural map $H^1(F_v, G) \to H^1(F_v, G^{\mathrm{ab}})$, we see from the diagram (3-8) that the image of $\mathrm{Det}(\boldsymbol{r}_G(a_v))$ in $Z(F_vG)^\times \backslash \mathcal{H}(Z(F_vG))$ is trivial, and so in fact $\mathrm{Det}(\boldsymbol{r}_G(a_v)) \in [\mathrm{Det}(O_{F_v^{\mathrm{nr}}}G)^\times]^{\Omega_{F_v}}$. Note that $\mathrm{Det}(\boldsymbol{r}_G(a_v))$ is defined over the finite, unramified extension $F_v^{\pi_v}$ of F_v (see (2-2)). Let L denote an arbitrary finite, unramified extension of F_v.

If $v \nmid |G|$, then O_LG is an O_L-maximal order in LG, and we have (see (4-12))

$$[\mathrm{Det}(O_LG)^\times]^{\Omega_{F_v}} \simeq [\mathrm{Hom}_{\Omega_L}(R_G, (O_{F_v^c})^\times)]^{\Omega_{F_v}} \simeq \mathrm{Hom}_{\Omega_{F_v}}(R_G, (O_{F_v^c})^\times) \simeq \mathrm{Det}(O_{F_v}G)^\times.$$

If $v \mid |G|$, then because $F/\mathbb{Q}$ is at most tamely ramified at all primes dividing $|G|$, it follows from M. J. Taylor's fixed point theorem for group determinants [1984, Chapter VIII] that

$$[\mathrm{Det}(O_LG)^\times]^{\Omega_{F_v}} = \mathrm{Det}(O_{F_v}G)^\times.$$

Hence, for each finite place v of F, we see that $\mathrm{Det}(\boldsymbol{r}_G(a_v)) \in \mathrm{Det}(O_{F_v}G)^\times$, and so $\Psi_v([\pi_v]) = 0$ (cf. part (a) above).

Since $H^1_{\mathrm{nr}}(F_v, G) = 0$ for all infinite places of F, it follows that $\Psi_v([\pi_v]) = 0$ for all places v of F. This in turn implies that $\lambda(\Psi([\pi])) = 0$. As the localisation map λ is injective (see Proposition 5.9(a)), it follows that $\Psi([\pi]) = 0$. Hence $H^1_{\mathrm{nr}}(F, G') \subseteq \mathrm{Ker}(\Psi)$, as claimed.

(d) The proof of this assertion is very similar to that of part (c) above, and so here we shall be brief. Suppose that $[\pi] \in H^1_{\mathrm{fnr}}(F, G')$. Arguing exactly as in part (c), we see that $\Psi_v([\pi]_v) = 0$ for all finite places v of F, which in turn implies that $\lambda_f(\Psi([\pi])) = 0$. Under our hypotheses, Proposition 5.9(b) implies that the localisation map λ_f is injective, and so $\Psi([\pi]) = 0$. Hence we see that $H^1_{\mathrm{fnr}}(F, G') \subseteq \mathrm{Ker}(\Psi)$, and so it follows from part (b) above that in fact $H^1_{\mathrm{fnr}}(F, G') = \mathrm{Ker}(\Psi)$, as asserted. □

Definition 13.2. Suppose that $x \in \mathrm{LC}(O_FG)$ (see Definition 6.4). We say that x is *unramified* (respectively *ramified*) at a place v of F if $\lambda_v(x) \in \mathrm{Im}(H^1_{\mathrm{nr}}(F_v, G))$ (respectively if $\lambda_v(x) \notin \mathrm{Im}(H^1_{\mathrm{nr}}(F_v, G))$).

If S is any finite set of places of F, we denote the set of $x \in \mathrm{LC}(O_FG)$ that are unramified at all places in S by $\mathrm{LC}(O_FG)_S$.

Before stating our next result, it will be helpful to introduce the following notation. Suppose that $x \in \mathrm{LC}(O_FG)$ and let $[(x_v)_v, x_\infty] \in J(K_1(FG)) \times \mathrm{Det}(F^cG)^\times$ be a representative of x. Then $\lambda(x) \in J(K_0(O_FG, F^c))$ is represented by the element $(x_v \cdot \mathrm{loc}_v(x_\infty)) \in \prod_v \mathrm{Det}(F_v^cG)^\times$. Hence it follows from Theorem 7.9 and Proposition 10.5(a) that we have an equality

$$[(x_v \cdot \mathrm{loc}_v(x_\infty))] = [a(x)] \cdot K\Theta^t(f(x)) \tag{13-1}$$

in $J(K_0(O_FG, F^c))$, where $a(x) = (a(x)_v) \in \prod_v \mathrm{Det}(\mathcal{H}(O_{F_v}G))$ and $f(x) \in \boldsymbol{F}$.

Definition 13.3. We say that $x \in \mathrm{LC}(O_F G)$ is *fully ramified* if $f(x)$ is full (see Definition 10.4 — note in particular that this does *not* mean that x is ramified at all places of F, which would of course be absurd!).

Let us also recall that $\partial^0(x) \in \mathrm{Cl}(O_F G)$ is represented by the idele $(x_v)_v \in J(K_1(FG))$ (see Remark 5.5).

Proposition 13.4. *Suppose that S is any finite set of places of F, and that $x \in \mathrm{LC}(O_F G)$. Then there exist infinitely many $y \in \mathrm{LC}(O_F G)_S$ with $\partial^0(y) = \partial^0(x)$ in $\mathrm{Cl}(O_F G)$. Hence we have*

$$\partial^0(LC(O_F G)) = \partial^0(LC(O_F G)_S). \tag{13-2}$$

Proof. Let $\mathfrak{a}$ be an ideal of F chosen as in Proposition 11.6 (so $\mathfrak{a}$ is divisible by a sufficiently high power of $|G|$ for the homomorphism $\Theta^t_{\mathfrak{a}}$ to be defined). Proposition 11.5 implies that there are infinitely many choices of $g \in \boldsymbol{F}$ such that $\mathrm{Supp}(g)$ is disjoint from S and g lies in the same modified narrow ray class modulo $\mathfrak{a}$ as $f(x)$, i.e.,

$$f(x) \equiv g \pmod{\Lambda(FG)^\times \cdot U'_{\mathfrak{a}}(\Lambda(O_F G)) \cdot U'_\infty(\Lambda(O_F G))_+}.$$

Hence for any such g, we have

$$K\Theta^t(f(x)) = K\Theta^t(\beta \cdot b \cdot g)$$

where $\beta \in \Lambda(FG)^\times$ and $b = (b_v) \in U'_{\mathfrak{a}}(\Lambda(O_F G)) \cdot U'_\infty(\Lambda(O_F G))_+$. Now $K\Theta^t(\beta) \in \partial^1(K_1(F^c G))$ (see (10-3)–(10-5)), while $K\Theta^t(b)$ lies in the image of $\prod_v \mathrm{Det}(\mathcal{H}(O_{F_v} G))$ in $J(K_0(O_F G, F^c))$, by virtue of our choice of $\mathfrak{a}$. We therefore see from (13-1) that we have the equality

$$[(x_v \cdot \mathrm{loc}_v(x_\infty))] \cdot K\Theta^t(\beta)^{-1} = [a(x)] \cdot K\Theta^t(b) \cdot K\Theta^t(g)$$

in $J(K_0(O_F G, F^c))$. Then the class

$$y = [(x_v \cdot \mathrm{loc}_v(x_\infty))] \cdot K\Theta^t(\beta)^{-1}$$

in $J(K_0(O_F G, F^c))$ satisfies the desired conditions.

The final assertion follows immediately from the exact sequence (6-1). □

Proposition 13.5. *Suppose that S is any finite set of places of F, and that $x \in \mathrm{LC}(O_F G)$. Then there exist infinitely many $y \in \mathrm{LC}(O_F G)_S$ such that y is fully ramified and $\partial^0(y) = \partial^0(x)$ in $\mathrm{Cl}(O_F G)$.*

Proof. This is a generalisation of [McCulloh 1983, Proposition 6.14], and it may be proved in the same way as [Byott 1998, Proposition 7.4].

We begin by constructing a full element h of $\boldsymbol{F}$ as follows. Let M/F be a finite Galois extension such that Ω_M acts trivially on $\mathcal{C}(G(-1))$. For each $s \in G$, choose a place $v(s)$ of F that splits completely in M/F; the Chebotarev density theorem implies that this may be done so that the places $v(s)$ are distinct and disjoint from S. Then the element $h = \prod_{s \in G} f_{v(s),s}$ is full.

Next, we choose an ideal $\mathfrak{a}$ of F as in Proposition 11.6 and observe that Proposition 11.5 implies that there are infinitely many choices of $g \in \boldsymbol{F}$ with $\mathrm{Supp}(g)$ disjoint from $S \cup \mathrm{Supp}(h)$ such that g lies in the same modified narrow ray class of $\Lambda(O_F G)$ modulo $\mathfrak{a}$ as $f(x) \cdot h^{-1}$. Then, for any such g, we have that

$$f(x) \equiv g \cdot h \pmod{\Lambda(FG)^{\times} \cdot U_{\mathfrak{a}}(\Lambda(O_F G)) \cdot U'_{\infty}(\Lambda(O_F G))_{+}},$$

and $g \cdot h \in \boldsymbol{F}$ is full. Now exactly as in the proof of Proposition 13.4 we may replace $f(x)$ by $g \cdot h$ in (13-1), changing the other terms in the equality as needed, to obtain $y \in K_0(O_F G, F^c)$ satisfying the stated conditions. □

Theorem 13.6. *Let S be any finite set of places of F, and suppose that Conjecture B holds for* $\mathrm{LC}(O_F G)_S$, *i.e., that*

$$\mathrm{LC}(O_F G)_S \subseteq K\mathcal{R}(O_F G) = \mathrm{Im}(\Psi). \tag{13-3}$$

Then $\mathcal{R}(O_F G)$ is a subgroup of $\mathrm{Cl}(O_F G)$. *If $c \in \mathcal{R}(O_F G)$, then there exist infinitely many $[\pi] \in H^1_t(F, G)$ such that F_π is a field and $(O_\pi) = c$. The extensions F_π/F may be chosen to have ramification disjoint from S.*

Proof. To prove the first assertion, it suffices to show that, under the given hypotheses, we have

$$\partial^0(\mathrm{LC}(O_F G)) = \mathcal{R}(O_F G) \tag{13-4}$$

(see the proof of Theorem 6.7, especially (6-2)).

We plainly have $\mathcal{R}(O_F G) \subseteq \partial^0(\mathrm{LC}(O_F G))$. Suppose that $x \in \mathrm{LC}(O_F G)$, and set $c_x = \partial^0(x)$. Then Proposition 13.5 implies that there exists $y \in \mathrm{LC}(O_F G)_S$ with $\partial^0(y) = c_x$. By hypothesis, we have $y \in \mathrm{Im}(\Psi)$, and so $\partial^0(y) = c_x \in \mathcal{R}(O_F G)$. This implies that $\partial^0(\mathrm{LC}(O_F G)) \subseteq \mathcal{R}(O_F G)$. Hence (13-4) holds, and so $\mathcal{R}(O_F G)$ is a subgroup of $\mathrm{Cl}(O_F G)$, as claimed.

Next, we observe that if $c \in \mathcal{R}(O_F G)$, then (13-4) and Proposition 13.5 imply that there are infinitely many $x \in \mathrm{LC}(O_F G)_S$ such that x is fully ramified and $\partial^0(x) = c$. For each such x, our hypotheses imply that there exists $\pi_x \in \mathrm{Hom}(\Omega_F, G)$ with $[\pi_x] \in H^1_t(F, G)$ and $\Psi([\pi_x]) = x$. The set of primes that ramify in F_{π_x}/F is equal to $\mathrm{Supp}(f(x))$, and so F_{π_x}/F has ramification disjoint from S. As $f(x)$ is full, we see that for each nonidentity element $s \in G$, there is a place $v(s) \in \mathrm{Supp}(f(x))$ such that $\pi_x(\sigma_{v(s)}) \in c(s)$ (see (7-5) and Proposition 10.5(a) and (b)). Hence $\mathrm{Im}(\pi_x)$ has nontrivial intersection with every conjugacy class of G and so is equal to the whole of G, by a lemma of Jordan (see [Serre 2003, p. 435, Theorem 4']). Therefore π_x is surjective, and so F_{π_x} is a field. This establishes the result. □

14. Abelian groups

In this section we shall prove that Conjecture 6.5 holds for abelian groups. We shall also show that the map Ψ is injective in this case.

Let G be abelian, and suppose that L is any finite extension of F or of F_v for some place v of F. As G is abelian, the reduced norm map induces isomorphisms

$$(LG)^{\times} \simeq \mathrm{Det}(LG)^{\times}, \quad (O_L G)^{\times} \simeq \mathrm{Det}(O_L G)^{\times}, \quad (L^c G)^{\times} \simeq \mathrm{Det}(L^c G)^{\times}. \tag{14-1}$$

For each finite place v of F, Lemma 5.7 and (14-1) imply that there are isomorphisms

$$K_0(O_{F_v}G,\, F_v^c) \simeq \frac{\mathrm{Det}(F_v^c G)^\times}{\mathrm{Det}(O_{F_v}G)^\times} \simeq \frac{(F_v^c G)^\times}{(O_{F_v}G)^\times}.$$

Proposition 14.1. *Let G be abelian and suppose that v is a finite place of F. Then the map Ψ_v is injective.*

Proof. Suppose that $[\pi_{v,i}] \in H_t^1(F_v, G)$ $(i = 1, 2)$, with $O_{\pi_{v,i}} = O_{F_v}G \cdot a_{v,i}$. Then $\Psi_v([\pi_{v,i}]) = [\boldsymbol{r}_G(a_{v,i})]$ in $(F_v^c G)^\times/(O_{F_v}G)^\times$. Hence if $\Psi([\pi_{v,1}]) = \Psi([\pi_{v,2}])$, then we have $\boldsymbol{r}_G(a_{v,1}) \cdot \boldsymbol{r}_G(a_{v,2})^{-1} \in (O_{F_v}G)^\times$. This implies that $[\pi_{1,v}] = [\pi_{2,v}]$ in $H_t^1(F_v, G)$, and so it follows that Ψ_v is injective, as claimed. □

Again because G is abelian, the pointed set of resolvends $H_t(LG)$ is an abelian group, and the exact sequences (3-3) and (3-4) show that there is an isomorphism

$$\tau : H_t^1(L, G) \xrightarrow{\sim} \frac{H_t(LG)}{(LG)^\times} \tag{14-2}$$

defined as follows: if $[\pi] \in H_t^1(L, G)$ with $L_\pi = LG \cdot b_\pi$, then $\tau([\pi]) = [\boldsymbol{r}_G(b_\pi)]$.

Note also that Theorem 5.4(b) and (14-1) imply that $K_0(O_F G,\, F^c)$ is isomorphic to the cokernel of the homomorphism

$$\Delta_{O_F G, F^c} : (FG)^\times \to \frac{J(FG)}{\prod_v (O_{F_v}G)^\times} \times (F^c G)^\times$$

induced by

$$(FG)^\times \to J(FG) \times (F^c G)^\times, \quad x \mapsto ((\mathrm{loc}_v(x))_v, x^{-1}).$$

Theorem 14.2. *Conjecture 6.5 is true when G is abelian.*

Proof. Suppose that $x \in \mathrm{LC}(O_F G)$, and let $[(x_v)_v, x_\infty] \in J(FG) \times (F^c G)^\times$ be a representative of x. We shall explain how to construct an element $[\pi] \in H_t^1(F, G)$ such that $\lambda_v(x) = \lambda_v(\Psi([\pi]))$ for all finite places v of F. Since G is abelian, and therefore admits no nontrivial irreducible symplectic characters, this will imply that $x = \Psi([\pi])$ (see Proposition 5.9(b)).

For each v, we have that $x_v \cdot \mathrm{loc}_v(x_\infty) \in H_t(F_v G)$. As $x_v \in (F_v G)^\times$, this implies that $\mathrm{loc}_v(x_\infty) \in H_t(F_v G)$ for each v. It follows from Proposition 2.3 that $x_\infty \in H(FG)$, and we see in addition that in fact $x_\infty \in H_t(FG)$. Hence x_∞ is the resolvend of a normal basis generator of a tame extension F_π/F. Set $\pi_v := \mathrm{loc}_v(\pi)$. Then for each finite v, we have

$$\tau(\Psi_v^{-1}(\lambda_v(x))) = [\mathrm{loc}_v(x_\infty)] = \tau([\pi_v])$$

in $H_t(F_v G)/(F_v G)^\times$, which in turn implies that

$$\lambda_v(x) = \Psi_v([\pi_v]) = \lambda_v(\Psi([\pi])).$$

Hence $x = \Psi([\pi])$, as required. □

Proposition 14.3. *If G is abelian, then the map Ψ is injective.*

Proof. Let $[\pi] \in H_t^1(F_v, G)$, and suppose that $[(x_v)_v, x_\infty] \in J(K_1(FG)) \times (F^c G)^\times$ is a representative of $\Psi([\pi])$. Then it follows from the proof of Theorem 14.2 that $\tau([\pi]) = x_\infty$ in $H_t(FG)/(FG)^\times$. Since τ is an isomorphism, we deduce that Ψ is injective. □

15. Neukirch's lifting theorem

Our main purpose in this section is to describe certain results, mainly from [Neukirch 1979], that will be used in the proof of Theorem E. We refer the reader to [Neukirch 1979; 2008, IX.5] for full details regarding these topics.

Let D be an arbitrary finite group. Consider the category $\mathcal{D}$ of homomorphisms $\eta : \mathcal{G} \to D$ of arbitrary profinite groups $\mathcal{G}$ into D in which a morphism between two objects $\eta_1 : \mathcal{G}_1 \to D$ and $\eta_2 : \mathcal{G}_2 \to D$ is defined to be a homomorphism $\nu : \mathcal{G}_1 \to \mathcal{G}_2$ such that $\eta_1 = \eta_2 \circ \nu$. We say that two such morphisms $\nu_i : \mathcal{G}_1 \to \mathcal{G}_2$ $(i = 1, 2)$ are *equivalent* if there is an element $k \in \mathrm{Ker}(\eta_2)$ such that $\nu_1(\omega) = k \cdot \nu_2(\omega) \cdot k^{-1}$ for all $\omega \in \mathcal{G}_1$. Write $\mathcal{H}om_D(\mathcal{G}_1, \mathcal{G}_2)$ for the set of equivalence classes of homomorphisms $\mathcal{G}_1 \to \mathcal{G}_2$, and $\mathcal{H}om_D(\mathcal{G}_1, \mathcal{G}_2)_{\mathrm{epi}}$ for the subset of $\mathcal{H}om_D(\mathcal{G}_1, \mathcal{G}_2)$ consisting of equivalence classes of surjective homomorphisms.

Suppose now that we have an exact sequence

$$0 \to B \to G \xrightarrow{q} D \to 0$$

with B abelian, and that L is a number field or a local field. Let $h : \Omega_L \to D$ be a fixed homomorphism. We view $\Omega_L \xrightarrow{h} D$ and $G \xrightarrow{q} D$ as being elements of $\mathcal{D}$. The group D acts on B via inner automorphisms, and this in turn induces an action of Ω_L on B via h. We write $L(B)$ for the smallest extension of L such that $\Omega_{L(B)}$ fixes B (i.e., $L(B)$ is the field of definition of B).

It may be shown that the group $H^1(L, B)$ acts on $\mathcal{H}om_D(\Omega_L, G)$ in the following way. Let $z \in Z^1(L, B)$ be any 1-cocycle representing $[z] \in H^1(L, B)$, and let $\nu \in \mathrm{Hom}(\Omega_L, G)$ be any homomorphism, representing an element $[\nu] \in \mathcal{H}om_D(\Omega_L, G)$. Define $z \cdot \nu : \Omega_L \to G$ by

$$(z \cdot \nu)(\omega) = z(\omega) \cdot \nu(\omega)$$

for all $\omega \in \Omega_L$. It is not hard to check that

$$h = q \circ (z \cdot \nu),$$

and that the element $[z \cdot \nu] \in \mathcal{H}om_D(\Omega_L, G)$ is independent of the choices of z and ν. It may also be shown that $\mathcal{H}om_D(\Omega_L, G)$ is a principal homogeneous space over $H^1(L, B)$.

For a number field F, and a finite place v of F, we let $\mathcal{H}om_D(\Omega_{F_v}, G)_{\mathrm{nr}}$ denote the set of classes of homomorphisms $\Omega_{F_v} \to G$ that are trivial on I_v. We write $J_f(\mathcal{H}om_D(\Omega_F, G))$ for the restricted direct product over all finite places of F of the sets $\mathcal{H}om_D(\Omega_{F_v}, G)$ with respect to the subsets $\mathcal{H}om_D(\Omega_{F_v}, G)_{\mathrm{nr}}$.

Now we can state Neukirch's lifting theorem.

Theorem 15.1. *Let F be a number field and let $h : \Omega_F \to D$ be a fixed, surjective homomorphism. Suppose that*

$$0 \to B \to G \xrightarrow{q} D \to 0$$

is an exact sequence for which B is a simple Ω_F-module. (*This implies that $l \cdot B = 0$ for a unique prime l.*) *Assume that the field of definition $F(B)$ of B contains no nontrivial l-th roots of unity, and that $J_f(\mathcal{H}om_D(\Omega_F, G)) \neq \varnothing$. Let S be any finite set of finite places of F. Then the natural map*

$$\mathcal{H}om_D(\Omega_F, G)_{\text{epi}} \to \prod_{v \in S} \mathcal{H}om_D(\Omega_{F_v}, G)$$

is surjective.

Proof. This is [Neukirch 1979, Main Theorem, p. 148]. □

The following result implies that $\mathcal{H}om_D(\Omega_{F_v}, G) \neq \varnothing$ for all but finitely many v.

Proposition 15.2 [Neukirch 1979, Lemma 5]. *Let F be a number field, and let v be a finite place of F. Suppose that $\mathcal{G}_1 \to \mathcal{G}_2$ is a surjective homomorphism of arbitrary profinite groups, and that there exists an unramified homomorphism $h_v : \Omega_{F_v} \to \mathcal{G}_2$. Then $\mathcal{H}om_{\mathcal{G}_2}(\Omega_{F_v}, \mathcal{G}_1)_{\text{nr}} \neq \varnothing$, and so $\mathcal{H}om_{\mathcal{G}_2}(\Omega_{F_v}, \mathcal{G}_1) \neq \varnothing$ also.*

Proof. If h_v is unramified, then h_v factors through $\Omega_{F_v}/I_v \simeq \hat{\mathbb{Z}}$, and a map $\hat{\mathbb{Z}} \to \mathcal{G}_2$ may always be lifted to a map $\hat{\mathbb{Z}} \to \mathcal{G}_1$ by lifting the image of a topological generator of $\hat{\mathbb{Z}}$. □

We now turn to two results of a local-global nature that will play a role in the proof of Theorem 16.4. In order to describe them, we let Γ be a finite abelian group equipped with an action of Ω_F such that Γ is a simple Ω_F-module. Then $l \cdot \Gamma = 0$ for a unique prime l. Write $F(\Gamma)$ for the field of definition of Γ.

Theorem 15.3. *Let M/F be a Galois extension with $F(\Gamma) \subseteq M$ and $\mu_l \nsubseteq M$, and let $\mathcal{N}/M$ be a finite abelian extension. Let S be a finite set of finite places of F, and suppose given an element $y_v \in H^1(F_v, \Gamma)$ for each $v \in S$. Then there exists an element $z \in H^1(F, \Gamma)$ satisfying the following local conditions*:

(i) *$z_v = y_v$ for each $v \in S$.*

(ii) *If $v \notin S$, then z_v is cyclic* (*i.e., is trivialised by a cyclic extension of F_v*), *and if z_v is ramified, then v splits completely in $\mathcal{N}/F$.*

Proof. This is [Neukirch 1979, Theorem 1]. □

In order to state our next result, we introduce the following notation.

Definition 15.4. Let $T := \{v_1, \dots, v_r\}$ be any finite set of finite places of F containing all places that ramify in $F(\Gamma)/F$ and all places above l. Let $\mathfrak{p}_i$ denote the prime ideal of F corresponding to v_i. Proposition 4.8 implies that we may choose an integer $N = N(T)$ such that for each $1 \leq i \leq r$ and for every place w of $F(\Gamma)$ lying above v_i, we have

$$\operatorname{Hom}_{\Omega_{F(\Gamma)_w}}(A_\Gamma, U_{\mathfrak{p}_i^N}(O_{F(\Gamma)_w^c})) \subseteq \operatorname{rag}[\operatorname{Hom}_{\Omega_{F(\Gamma)_w}}(R_\Gamma, O^\times_{F(\Gamma)_w^c})].$$

Set

$$\mathfrak{a} = \mathfrak{a}(T) = \prod_{i=1}^{r} \mathfrak{p}_i.$$

Let $F(\mathfrak{a}^N)$ denote the ray class field of F modulo $\mathfrak{a}^N$.

Theorem 15.5. *Let $v \notin T$ be any finite place of F that splits completely in $F(\mathfrak{a}^N)$, and suppose that s is any nontrivial element of Γ. Then there is an element $b = b(v; s) \in H^1(F, \Gamma)$ satisfying the following local conditions*:

(i) $\mathrm{loc}_{v_i}(b) = 0$ *for* $1 \leq i \leq r$.

(ii) $b|_{I_v} = \tilde{\varphi}_{v,s}$ *(see Remark 7.11).*

(iii) *b is unramified away from v.*

Proof. Let $\mathfrak{p}$ be the prime ideal of F corresponding to v. Our hypotheses on v imply that $\mathfrak{p}$ is principal, with $\mathfrak{p} \equiv 1 \pmod{\mathfrak{a}^N}$. Set $M := F(\Gamma)$. As Γ is abelian, we have that $\mathcal{H}(M\Gamma) \simeq \mathrm{Hom}_{\Omega_M}(A_\Gamma, (M^c)^\times)$ (see (4-6)). Let ϖ be a generator of $\mathfrak{p}$, and define $\rho \in \mathrm{Hom}_{\Omega_M}(A_\Gamma, (M^c)^\times)$ by

$$\rho(\alpha) = \varpi^{\langle \alpha, s \rangle_\Gamma}.$$

(This homomorphism is Ω_M-equivariant because Ω_M fixes Γ.) Then ρ is the reduced resolvend of a normal basis generator of an extension $M_{\pi(\rho)}/M$ corresponding to $[\pi(\rho)] \in H^1(M, \Gamma)$. Since $\mathfrak{p} \equiv 1 \pmod{\mathfrak{a}^N}$, for each place w of M lying above a place v_i in T, we have

$$\mathrm{loc}_w(\rho) \in \mathrm{Hom}_{\Omega_{M_w}}(A_\Gamma, U_{\mathfrak{p}_i^N}(O_{M_w^c})) \subseteq \mathrm{rag}[\mathrm{Hom}_{\Omega_{M_w}}(R_\Gamma, O_{M_w^c}^\times)],$$

and so it follows that $\mathrm{loc}_w(\pi(\rho)) = 0$ (see (4-7)). In particular, $\pi(\rho)$ is unramified at all places above T.

For all places w' of M not lying above T or v we have that

$$\mathrm{loc}_{w'}(\rho) \in \mathrm{Hom}_{\Omega_{M_{w'}}}(A_\Gamma, O_{M_{w'}^c}^\times),$$

and so $\pi(\rho)$ is unramified at w'. This implies that $\pi(\rho)$ is unramified away from v, since we have already seen that $\pi(\rho)$ does not ramify at any place above T. It is also easy to see that

$$b\,|_{I_{w(v)}} = \tilde{\varphi}_{w(v),s}$$

for any place $w(v)$ of M lying above v (cf. the proof of Proposition 10.5(a)).

As $\varpi \in F$, we have that $\pi(\rho) \in H^1(M, \Gamma)^{\mathrm{Gal}(M/F)}$. Since $\Gamma^{\Omega_F} = 0$ (because Γ is a simple Ω_F-module), the restriction map $H^1(F, \Gamma) \to H^1(M, \Gamma)$ is injective and induces an isomorphism $H^1(F, \Gamma) \simeq H^1(M, \Gamma)^{\mathrm{Gal}(M/F)}$. Hence $\pi(\rho)$ is the image of an element $b \in H^1(F, \Gamma)$ satisfying the conditions (i), (ii) and (iii) of the theorem. □

16. Soluble groups

In this section we shall use Neukirch's lifting theorem to prove a result (see Theorem 16.4 below) that implies Theorem E of the introduction. In order to describe this result, it will be helpful to formulate the following definition.

Definition 16.1 (Property R). Let S be any finite (possibly empty) set of places of F. We shall say that $\mathrm{LC}(O_F G)_S$ *satisfies Property R* if the following holds: Suppose given any fully ramified $x \in \mathrm{LC}(O_F G)_S$. For each finite place v of F, suppose also given a homomorphism $\pi_{v,x} \in \mathrm{Hom}(\Omega_{F_v}, G)$ such that $[\pi_{v,x}] \in H^1_t(F_v, G)$ and $\lambda_v(x) = \Psi_v([\pi_{v,x}])$. (Note that in general, such a choice of $\pi_{v,x}$ is not unique.) Then there exists $\Pi \in \mathrm{Hom}(\Omega_F, G)$ with $[\Pi] \in H^1_t(F, G)$ such that

(a) $x = \Psi([\Pi])$,

(b) $\Pi|_{I_v} = \pi_{v,x}|_{I_v}$ for each finite place v of F.

(So in particular, x is cohomological.)

Proposition 16.2. *If G is abelian, then* $\mathrm{LC}(O_F G)$ *satisfies Property R.*

Proof. We shall in fact prove a slightly stronger result. Suppose that G is abelian, and let $x \in \mathrm{LC}(O_F G)$. (Note that we do not assume that x is fully ramified.) Then Theorem 14.2 implies that x is cohomological. As G is abelian, the maps Ψ and Ψ_v are injective (see Propositions 14.1 and 14.3). Hence it follows that there is a unique $[\Pi] \in H^1_t(F, G)$ such that $x = \Psi([\Pi])$, and a unique $[\pi_{v,x}] \in H^1_t(F_v, G)$ such that $\lambda_v(x) = \Psi_v([\pi_{v,x}])$. We therefore see that

$$\lambda_v(x) = \Psi_v([\Pi_v]) = \Psi([\pi_{v,x}]),$$

and so $\Pi_v = \pi_{v,x}$. This implies that $\mathrm{LC}(O_F G)$ satisfies Property R. □

Theorem 16.3. *Suppose that* $\mathrm{LC}(O_F G)_S$ *satisfies Property R. Then $\mathcal{R}(O_F G)$ is a subgroup of* $\mathrm{Cl}(O_F G)$. *If $c \in \mathcal{R}(O_F G)$, then there exist infinitely many $[\pi] \in H^1_t(F, G)$ such that F_π is a field and $(O_\pi) = c$. The extensions F_π/F may be chosen to have ramification disjoint from S.*

Proof. This is an immediate consequence of Theorem 13.6. □

Our proof of Theorem E rests on the following result.

Theorem 16.4. *Suppose that there is an exact sequence*

$$0 \to B \to G \to D \to 0,$$

where B is an abelian minimal normal subgroup of G with $l \cdot B = 0$ for an odd prime l. Let S be any finite set of finite places of F containing all places dividing $|G|$. Assume that the following conditions hold:

(i) *The set* $\mathrm{LC}(O_F D)_S$ *satisfies Property R.*

(ii) *We have* $(|G|, h_F) = 1$, *where h_F denotes the class number of F.*

(iii) *Either G admits no irreducible symplectic characters*, *or F has no real places.*

(iv) *The field F contains no nontrivial l-th roots of unity.*

Then $\mathrm{LC}(O_F G)_S$ *satisfies Property R.*

Proof. We shall establish this result in several steps, one of which crucially involves Neukirch's lifting theorem (see Theorem 15.1).

Suppose that $x \in \mathrm{LC}(O_F G)_S$ is fully ramified. For each finite place v of F, choose $\pi_{v,x} \in \mathrm{Hom}(\Omega_{F_v}, G)$ such that $[\pi_{v,x}] \in H^1_t(F_v, G)$ with

$$\lambda_v(x) = \Psi_v([\pi_{v,x}]).$$

The choice of $\pi_{v,x}$ is not unique. However, if $a(\pi_{v,x})$ is any normal integral basis generator of $F_{\pi_{v,x}}/F_v$, with Stickelberger factorisation (see Definition 7.12)

$$\boldsymbol{r}_G(a(\pi_{v,x})) = u(a(\pi_{v,x})) \cdot \boldsymbol{r}_G(a_{\mathrm{nr}}(\pi_{v,x})) \cdot \boldsymbol{r}_G(\varphi(\pi_{v,x})), \tag{16-1}$$

then Proposition 10.5(c) implies that $\mathrm{Det}(\boldsymbol{r}_G(\varphi(\pi_{v,x})))$ is independent of the choice of $\pi_{v,x}$. Hence, if $\varphi(\pi_{v,x}) = \varphi_{v,s}$, say, then it follows from Proposition 10.5(b) that the subgroup $\langle s \rangle$ of G (up to conjugation) and the determinant $\mathrm{Det}(\boldsymbol{r}_G(\varphi_{v,s}))$ of the resolvend $\boldsymbol{r}_G(\varphi_{v,s})$ do not depend upon the choice of $\pi_{v,x}$.

We write $q : G \to D$ for the obvious quotient map, and we use the same symbol q for the induced maps

$$K_0(O_F G, F^c) \to K_0(O_F D, F^c), \quad H^1(F, G) \to H^1(F, D), \quad H^1(F_v, G) \to H^1(F_v, D).$$

Set

$$\bar{x} := q(x), \quad \pi_{v,\bar{x}} := q(\pi_{v,x}).$$

Then $\bar{x} \in \mathrm{LC}(O_F D)_S$ with

$$\lambda_v(\bar{x}) = \Psi_{D,v}(\pi_{v,\bar{x}})$$

for each finite place v of F, and $\bar{x}$ is fully ramified.

By hypothesis, $\mathrm{LC}(O_F D)_S$ satisfies Property R, and so there exists $\rho \in \mathrm{Hom}(\Omega_F, D)$ with $[\rho] \in H^1_t(F, D)$ such that

$$\bar{x} = \Psi_D([\rho]) \tag{16-2}$$

and

$$\rho|_{I_v} = \pi_{v,\bar{x}}|_{I_v} \tag{16-3}$$

for each finite place v of F. Hence, for each such v, we have that

$$\mathrm{Det}(\boldsymbol{r}_D(\varphi(\rho_v))) = \mathrm{Det}(\boldsymbol{r}_D(\varphi(\pi_{v,\bar{x}}))),$$

using the notation established in (16-1) above concerning Stickelberger factorisations. As $\bar{x}$ is fully ramified, we see from the proof of Theorem 13.6 that ρ is surjective, and so F_ρ is a field. We also see that, as $\bar{x} \in \mathrm{LC}(O_F D)_S$, the extension F_ρ/F is unramified at all places dividing $|D|$. Furthermore, if $v \mid l$ (so $v \in S$), then since $\pi_{v,x}$ is unramified, the same is true of $\pi_{v,\bar{x}}$, and so F_ρ/F is also unramified at v. Hence, as $F \cap \mu_l = \{1\}$ by hypothesis, it follows that $F_\rho \cap \mu_l = \{1\}$ also.

For each finite place v of F, we are now going to use the fact that $x \in \mathrm{LC}(O_F G)$ to construct a lift $\tilde{\rho}_v \in \mathrm{Hom}(\Omega_{F_v}, G)$ of ρ_v such that $[\tilde{\rho}_v] \in H^1_t(F_v, G)$ with

$$\tilde{\rho}_v|_{I_v} = \pi_{v,x}|_{I_v}. \tag{16-4}$$

To do this, we first observe that if $\varphi(\pi_{v,x}) = \varphi_{v,s}$, then $\varphi(\pi_{v,\bar{x}}) = \varphi_{v,\bar{s}}$, where $\bar{s} = q(s)$, and so we have

$$\varphi(\rho_v) = \varphi(\pi_{v,\bar{x}}) = \varphi_{v,\bar{s}}$$

(see (16-3)).

Next, we write

$$\rho_v = \rho_{v,r} \cdot \rho_{v,nr},$$

with $[\rho_{v,nr}] \in H^1_{\mathrm{nr}}(F_v, D)$ (see (7-7)). Since $\rho_{v,nr}$ is unramified, Proposition 15.2 implies that $[\rho_{v,nr}]$ may be lifted to $[\tilde{\rho}_{v,nr}] \in H^1_{\mathrm{nr}}(F_v, G)$. Let $a(\tilde{\rho}_{v,nr})$ be a normal integral basis generator of $F_{\tilde{\rho}_{v,nr}}/F_v$. Then $\boldsymbol{r}_G(a(\tilde{\rho}_{v,nr})) \cdot \boldsymbol{r}_G(\varphi_{v,s})$ is the resolvend of a normal integral basis generator of a tame Galois G-extension $F_{\tilde{\rho}_v}/F_v$ such that $q([\tilde{\rho}_v]) = \rho_v$ (see Corollary 7.8 and Theorem 7.9). As $\varphi(\pi_{v,x}) = \varphi_{v,s}$, we see from the construction of $\tilde{\rho}$ that

$$\tilde{\rho}_v|_{I_v} = \pi_{v,x}|_{I_v} = \tilde{\varphi}_{v,s},$$

where $[\tilde{\varphi}_{v,s}] \in H^1_t(I_v, G)$ is defined in Remark 7.11. The map $\tilde{\rho}_v$ is our desired lift of ρ_v.

We are now ready to apply the results contained in Section 15. Consider the following diagram:

$$\begin{tikzcd}
0 \arrow[r] & B \arrow[r] & G \arrow[r, "q"] & D \arrow[r] & 0 \\
& & & \Omega_F \arrow[u, "\rho"'] &
\end{tikzcd}$$

The group D acts on B via inner automorphisms, and we view B as being an Ω_F-module via ρ. Then B is a simple Ω_F-module because B is a minimal normal subgroup of G and ρ is surjective. The field of definition $F(B)$ of B is contained in the field F_ρ, and so in particular $F(B)$ contains no nontrivial l-th roots of unity. We are going to construct an element $\Pi \in \mathcal{H}om_D(\Omega_F, G)$ such that

$$\Pi|_{I_v} = \pi_{v,x}|_{I_v}$$

for each finite place v of F. This will be accomplished in the following three steps:

I. We begin by observing that our construction above of a lift $\tilde{\rho}_v$ of ρ_v for each finite v shows that $J_f(\mathcal{H}om_D(\Omega_F, G))$ is nonempty. Let $\mathcal{S}$ be the set of finite places v of F at which x is ramified or $v \mid |G|$. Theorem 15.1 implies that there exists $\Pi_1 \in \mathcal{H}om_D(\Omega_F, G)$ such that $\Pi_{1,v} = \tilde{\rho}_v$ for all $v \in \mathcal{S}$. Observe that Π_1 is unramified at all $v \mid |G|$ because $\tilde{\rho}_v$ is unramified at these places (see (16-4)). Note also that Π_1 may well be ramified outside $\mathcal{S}$.

II. Recall that $\mathcal{H}om_D(\Omega_F, G)$ (respectively $\mathcal{H}om_D(\Omega_{F_v}, G)$ for each finite v) is a principal homogeneous space over $H^1(F, B)$ (respectively $H^1(F_v, B)$). Let $\mathcal{S}_1$ denote the set of finite places $v \notin \mathcal{S}$ of F at which Π_1 is ramified. For each $v \in \mathcal{S}_1$, choose $y_v \in H^1(F_v, B)$ so that $y_v \cdot \Pi_{1,v} \in \mathcal{H}om_D(\Omega_{F_v}, G)$ is unramified.

Now apply Definition 15.4 (with $\Gamma = B$ and $T = \mathcal{S}$) to obtain an ideal $\mathfrak{a} = \mathfrak{a}(\mathcal{S})$ and an integer $N = N(\mathcal{S})$ as described there. Theorem 15.3 implies that there exists an element $z \in H^1(F, B)$ such that:

(z1) $z_v = y_v$ for all $v \in \mathcal{S}_1$.

(z2) $z_v = 1$ for all $v \in \mathcal{S}$.

(z3) If $v \notin \mathcal{S} \cup \mathcal{S}_1$, then z_v is cyclic, and if z_v is ramified, then v splits completely in $(F(B) \cdot F(\mathfrak{a}^N))/F$, where $F(\mathfrak{a}^N)$ denotes the ray class field of F modulo $\mathfrak{a}^N$.

Set $\Pi_2 := z \cdot \Pi_1 \in \mathcal{H}om_D(\Omega_F, G)$. Note that, as z might possibly be ramified, the homomorphism Π_2 might be ramified outside $\mathcal{S}$. We shall eliminate any such potential ramification in the third and final step.

III. Let $\mathcal{S}_z$ be the set of places of F at which z is ramified (so $\mathcal{S} \cap \mathcal{S}_z = \varnothing$). We see from (z3) that each $v \in \mathcal{S}_z$ is totally split in $F(\mathfrak{a}^N)/F$. Hence Theorem 15.5 implies that for each $v \in \mathcal{S}_z$, we may choose $b(v) \in H^1(F, B)$ such that:

(b1) $b(v)_w = 1$ for all $w \in \mathcal{S}$.

(b2) $b(v)|_{I_v} = z_v^{-1}|_{I_v}$.

(b3) $b(v)$ is unramified away from v.

Set

$$\Pi := \left[\left(\prod_{v \in S_z} b(v)\right) \cdot z\right] \cdot \Pi_2.$$

Then it follows directly from the construction of Π that we have

$$\Pi|_{I_v} = \pi_{v,x}|_{I_v} \tag{16-5}$$

for all finite places v of F.

We claim that

$$x = \Psi(\Pi).$$

To show this, let $\tau = \Psi(\Pi)^{-1} \cdot x$. We see from (16-5) that

$$\lambda_v(\tau) \in \mathrm{Im}(\Psi_v^{\mathrm{nr}})$$

for every finite place v of F. As either G admits no irreducible symplectic characters or F has no real places, and as $(h_F, |G|) = 1$ by hypothesis, Proposition 6.8(b) implies that $\tau = 0$. Hence $x = \Psi(\Pi)$, as claimed.

This completes the proof that $\mathrm{LC}(O_F G)_S$ satisfies Property R. □

Theorem 16.4 (in conjunction with Proposition 16.2) yields an abundant supply of groups G for which $\mathrm{LC}(O_F G)_S$ satisfies Property R (for a suitable choice of S), and therefore also for which Theorem 16.3 holds. Here is an example of this.

Theorem 16.5. *Let G be of odd order. Suppose that $(|G|, h_F) = 1$ and that F contains no nontrivial $|G|$-th roots of unity. Let S be any finite set of finite places of F containing all places dividing $|G|$. Then $\mathrm{LC}(O_F G)_S$ satisfies Property R.*

Proof. We shall establish this result by induction on the order of G. We first note that Proposition 16.2 implies that the theorem holds if G is abelian.

Suppose now that G is an arbitrary finite group of odd order. As $|G|$ is odd, a well-known theorem of Feit and Thompson [1963] implies that G is soluble. Hence G has an abelian minimal normal subgroup B such that $l \cdot B = 0$ for some odd prime l (see, e.g., [Rotman 1995, Theorem 5.24]), and there is an exact sequence

$$0 \to B \to G \to D \to 0$$

with D soluble. As $|G|$ is odd, G admits no nontrivial irreducible symplectic characters. We may therefore suppose by induction on the order of G that $\mathrm{LC}(O_F D)_S$ satisfies Property R. The desired result now follows from Theorem 16.4. □

Remark 16.6. It follows from Theorem 14.2 that in Theorem 16.4, we may take D to be a finite abelian group of arbitrary order (subject of course to the obvious constraint that the number field F is such that all other conditions of Theorem 16.4 are satisfied). This enables one to show that Property R holds for many nonabelian groups of even order (e.g., S_3). However, if for example G is a nonabelian 2-group (e.g., H_8), then because $\mu_2 \subseteq F$ for any number field F, we can no longer appeal to Neukirch's lifting theorem, and our proof of Theorem 16.4 fails. It appears very likely that new ideas are needed to establish Property R in such cases (see also the remarks contained in the final paragraph of [Neukirch 1979, Introduction], where a similar difficulty is briefly discussed in the context of the inverse Galois problem for finite groups).

We can now prove Theorem E of the introduction.

Theorem 16.7. *Let G be of odd order and suppose that $(|G|, h_F) = 1$, where h_F denotes the class number of F. Suppose also that F contains no nontrivial $|G|$-th roots of unity. Then $\mathcal{R}(O_F G)$ is a subgroup of $\mathrm{Cl}(O_F G)$. If $c \in \mathcal{R}(O_F G)$, then there exist infinitely many $[\pi] \in H_t^1(F, G)$ such that F_π is a field and $(O_\pi) = c$. The extensions F_π / F may be chosen to have ramification disjoint from any finite set S of places of F.*

Proof. This is an immediate consequence of Theorems 16.5 and 16.3. □

Acknowledgements

We are very grateful indeed to Andrea Siviero for his extremely detailed comments on an earlier draft of this paper, and to Ruth Sergel for her most perceptive remarks at a critical stage of this project. We heartily thank Nigel Byott and Cindy Tsang for the many very helpful comments, questions and corrections that we received from them, and Bob Guralnick for a very helpful remark concerning the proof of Corollary 9.4. We are also extremely grateful to the anonymous referees whose very careful reading of the manuscript led us to correct and considerably strengthen our original results, and to significantly improve our exposition.

References

[Agboola 2012] A. Agboola, "On counting rings of integers as Galois modules", *J. Reine Angew. Math.* **663** (2012), 1–31. MR Zbl

[Agboola and Burns 1998] A. Agboola and D. Burns, "On the Galois structure of equivariant line bundles on curves", *Amer. J. Math.* **120**:6 (1998), 1121–1163. MR Zbl

[Agboola and Burns 2001] A. Agboola and D. Burns, "Grothendieck groups of bundles on varieties over finite fields", *K-Theory* **23**:3 (2001), 251–303. MR Zbl

[Agboola and Burns 2006] A. Agboola and D. Burns, "On twisted forms and relative algebraic K-theory", *Proc. London Math. Soc.* (3) **92**:1 (2006), 1–28. MR Zbl

[Bueno et al. 2016] M. I. Bueno, S. Furtado, J. Karkoska, K. Mayfield, R. Samalis, and A. Telatovich, "The kernel of the matrix $[ij \pmod n]$ when n is prime", *Involve* **9**:2 (2016), 265–280. MR Zbl

[Burton 2007] D. M. Burton, *Elementary number theory*, McGraw-Hill, Boston, 2007.

[Byott 1998] N. P. Byott, "Tame realisable classes over Hopf orders", *J. Algebra* **201**:1 (1998), 284–316. MR Zbl

[Byott and Sodaïgui 2005] N. P. Byott and B. Sodaïgui, "Realizable Galois module classes for tetrahedral extensions", *Compos. Math.* **141**:3 (2005), 573–582. MR Zbl

[Byott et al. 2006] N. P. Byott, C. Greither, and B. Sodaïgui, "Classes réalisables d'extensions non abéliennes", *J. Reine Angew. Math.* **601** (2006), 1–27. MR Zbl

[Chinburg 1994] T. Chinburg, "Galois structure of de Rham cohomology of tame covers of schemes", *Ann. of Math.* (2) **139**:2 (1994), 443–490. MR Zbl

[Curtis and Reiner 1981] C. W. Curtis and I. Reiner, *Methods of representation theory, I: With applications to finite groups and orders*, Wiley, New York, 1981. MR Zbl

[Curtis and Reiner 1987] C. W. Curtis and I. Reiner, *Methods of representation theory, II: With applications to finite groups and orders*, Wiley, New York, 1987. MR Zbl

[Farhat and Sodaïgui 2015] M. Farhat and B. Sodaïgui, "Classes réalisables d'extensions non abéliennes de degré p^3", *J. Number Theory* **152** (2015), 55–89. MR Zbl

[Feit and Thompson 1963] W. Feit and J. G. Thompson, "Solvability of groups of odd order", *Pacific J. Math.* **13** (1963), 775–1029. MR Zbl

[Fröhlich 1976] A. Fröhlich, "Arithmetic and Galois module structure for tame extensions", *J. Reine Angew. Math.* **286/287** (1976), 380–440. MR Zbl

[Fröhlich 1983] A. Fröhlich, *Galois module structure of algebraic integers*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)] **1**, Springer, 1983. MR Zbl

[Fröhlich 1984] A. Fröhlich, *Classgroups and Hermitian modules*, Progress in Mathematics **48**, Birkhäuser, Boston, 1984. MR Zbl

[Hilbert 1998] D. Hilbert, *The theory of algebraic number fields*, Springer, 1998. MR Zbl

[Malle 2002] G. Malle, "On the distribution of Galois groups", *J. Number Theory* **92**:2 (2002), 315–329. MR Zbl

[McCulloh 1983] L. R. McCulloh, "Galois module structure of elementary abelian extensions", *J. Algebra* **82**:1 (1983), 102–134. MR Zbl

[McCulloh 1987] L. R. McCulloh, "Galois module structure of abelian extensions", *J. Reine Angew. Math.* **375/376** (1987), 259–306. MR Zbl

[McCulloh 2011] L. R. McCulloh, "On realisable classes for non-abelian extensions", lecture in Luminy, March 22 2011.

[McCulloh 2012] L. R. McCulloh, "From Galois module classes to Steinitz classes", preprint, 2012. arXiv

[Neukirch 1979] J. Neukirch, "On solvable number fields", *Invent. Math.* **53**:2 (1979), 135–164. MR Zbl

[Neukirch et al. 2008] J. Neukirch, A. Schmidt, and K. Wingberg, *Cohomology of number fields*, 2nd ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] **323**, Springer, 2008. MR Zbl

[Rotman 1995] J. J. Rotman, *An introduction to the theory of groups*, 4th ed., Graduate Texts in Mathematics **148**, Springer, 1995. MR Zbl

[Serre 1997] J.-P. Serre, *Galois cohomology*, Springer, 1997. MR Zbl

[Serre 2003] J.-P. Serre, "On a theorem of Jordan", *Bull. Amer. Math. Soc.* (*N.S.*) **40**:4 (2003), 429–440. MR Zbl

[Shafarevich 1954] I. R. Shafarevich, "Construction of fields of algebraic numbers with given solvable Galois group", *Izv. Akad. Nauk SSSR. Ser. Mat.* **18** (1954), 525–578. In Russian; translated in *Izv. Akad. Nauk SSSR. Ser. Mat.* **18** (1954), 525–578. MR Zbl

[Siviero 2013] A. Siviero, *Class invariants for tame Galois algebras*, Ph.D. thesis, Université de Bordeaux and Universiteit Leiden, 2013.

[Siviero 2016] A. Siviero, "Realisable classes, Stickelberger subgroup and its behaviour under change of the base field", pp. 69–92 in *Publications mathématiques de Besançon: algèbre et théorie des nombres, 2015*, Publ. Math. Besançon Algèbre Théorie Nr. **2015**, Presses Univ., Franche-Comté, Besançon, 2016. MR Zbl

[Swan 1968] R. G. Swan, *Algebraic K-theory*, Lecture Notes in Mathematics **76**, Springer, 1968. MR Zbl

[Swan 1970] R. G. Swan, *K-theory of finite groups and orders*, Lecture Notes in Mathematics **149**, Springer, 1970. MR Zbl

[Taylor 1984] M. Taylor, *Classgroups of group rings*, London Mathematical Society Lecture Note Series **91**, Cambridge University Press, 1984. MR Zbl

[Tsang 2016] C. Tsang, "On the Galois module structure of the square root of the inverse different in abelian extensions", *J. Number Theory* **160** (2016), 759–804. MR Zbl

[Tsang 2017] C. Tsang, "On the realizable classes of the square root of the inverse different in the unitary class group", *Int. J. Number Theory* **13**:4 (2017), 913–932. MR Zbl

[Wright 1989] D. J. Wright, "Distribution of discriminants of abelian extensions", *Proc. London Math. Soc.* (3) **58**:1 (1989), 17–50. MR Zbl

Communicated by Bjorn Poonen
Received 2015-08-28 Revised 2018-03-23 Accepted 2018-07-02

agboola@math.ucsb.edu *Department of Mathematics, University of California, Santa Barbara, CA, United States*

mcculloh@math.uiuc.edu *Department of Mathematics, University of Illinois, Urbana, IL, United States*

ALGEBRA AND NUMBER THEORY 12:8 (2018)
dx.doi.org/10.2140/ant.2018.12.1887

Categorical representations and KLR algebras

Ruslan Maksimau

We prove that the KLR algebra associated with the cyclic quiver of length e is a subquotient of the KLR algebra associated with the cyclic quiver of length $e+1$. We also give a geometric interpretation of this fact. This result has an important application in the theory of categorical representations. We prove that a category with an action of $\widetilde{\mathfrak{sl}}_{e+1}$ contains a subcategory with an action of $\widetilde{\mathfrak{sl}}_e$. We also give generalizations of these results to more general quivers and Lie types.

1. Introduction

Consider the complex affine Lie algebra $\widetilde{\mathfrak{sl}}_e = \mathfrak{sl}_e[t, t^{-1}] \oplus \mathbb{C}\mathbf{1}$. In this paper, we study categorical representations of $\widetilde{\mathfrak{sl}}_e$. Our goal is to relate the notion of a categorical representation of $\widetilde{\mathfrak{sl}}_e$ with the notion of a categorical representation of $\widetilde{\mathfrak{sl}}_{e+1}$.

The Lie algebra $\widetilde{\mathfrak{sl}}_e$ has generators e_i, f_i for $i \in [0, e-1]$. Let $\alpha_0, \dots, \alpha_{e-1}$ be the simple roots of $\widetilde{\mathfrak{sl}}_e$. Fix $k \in [0, e-1]$. Consider the following inclusion of Lie algebras $\widetilde{\mathfrak{sl}}_e \subset \widetilde{\mathfrak{sl}}_{e+1}$:

$$e_r \mapsto \begin{cases} e_r & \text{if } r \in [0, k-1], \\ [e_k, e_{k+1}] & \text{if } r = k, \\ e_{r+1} & \text{if } r \in [k+1, e-1], \end{cases} \qquad f_r \mapsto \begin{cases} f_r & \text{if } r \in [0, k-1], \\ [f_{k+1}, f_k] & \text{if } r = k, \\ f_{r+1} & \text{if } r \in [k+1, e-1]. \end{cases} \tag{1}$$

It is clear that each $\widetilde{\mathfrak{sl}}_{e+1}$-module can be restricted to the subalgebra $\widetilde{\mathfrak{sl}}_e$ of $\widetilde{\mathfrak{sl}}_{e+1}$. So it is natural to ask if we can do the same with categorical representations.

First, we recall the notion of a categorical representation. Let $\boldsymbol{k}$ be a field. Let $\mathcal{C}$ be an abelian Hom-finite $\boldsymbol{k}$-linear category that admits a direct sum decomposition $\mathcal{C} = \bigoplus_{\mu \in \mathbb{Z}^e} \mathcal{C}_\mu$. A categorical representation of $\widetilde{\mathfrak{sl}}_e$ in $\mathcal{C}$ is a pair of biadjoint functors $E_i, F_i \colon \mathcal{C} \to \mathcal{C}$ for $i \in [0, e-1]$ satisfying a list of axioms. The main axiom is that for each positive integer d there is an algebra homomorphism

MSC2010: primary 16G99; secondary 17B67, 18E10.
Keywords: KLR algebra, categorical representation, Hecke algebra, affine Lie algebra, quiver variety, flag variety.

$R_d(A^{(1)}_{e-1}) \to \mathrm{End}(F^d)^{\mathrm{op}}$, where $F = \bigoplus_{i=0}^{e-1} F_i$ and $R_d(A^{(1)}_{e-1})$ is the KLR algebra of rank d associated with the quiver $A^{(1)}_{e-1}$ (i.e., with the cyclic quiver of length e).

Let $\bar{\mathcal{C}}$ be an abelian Hom-finite $\boldsymbol{k}$-linear category. Assume that $\bar{\mathcal{C}} = \bigoplus_{\mu\in\mathbb{Z}^{e+1}} \bar{\mathcal{C}}_\mu$ has a structure of a categorical representation of $\widetilde{\mathfrak{sl}}_{e+1}$ with respect to functors $\bar{E}_i$, $\bar{F}_i$ for $i \in [0, e]$. We want to restrict the action of $\widetilde{\mathfrak{sl}}_{e+1}$ on $\bar{\mathcal{C}}$ to $\widetilde{\mathfrak{sl}}_e$. The most obvious way to do this is to define new functors $E_i, F_i : \bar{\mathcal{C}} \to \bar{\mathcal{C}}$, for $i \in [0, e-1]$, from the functors $\bar{E}_i, \bar{F}_i : \bar{\mathcal{C}} \to \bar{\mathcal{C}}$, for $i \in [0, e]$, by the same formulas as in (1). Of course, this makes no sense because the notion of a commutator of two functors does not exist. However, we are able to get a structure of a categorical representation on a subcategory $\mathcal{C} \subset \bar{\mathcal{C}}$ (and not on the category $\bar{\mathcal{C}}$ itself). We do this in the following way.

Assume additionally that the category $\bar{\mathcal{C}}_\mu$ is zero whenever μ has a negative entry. For each e-tuple $\mu = (\mu_1, \dots, \mu_e) \in \mathbb{Z}^e$ we consider the $(e+1)$-tuple $\bar{\mu} = (\mu_1, \dots, \mu_k, 0, \mu_{k+1}, \dots, \mu_e)$ and we set $\mathcal{C}_\mu = \bar{\mathcal{C}}_{\bar{\mu}}$,

$$\mathcal{C} = \bigoplus_{\mu\in\mathbb{Z}^e} \mathcal{C}_\mu.$$

Next, consider the endofunctors of $\mathcal{C}$ given by

$$E_i = \begin{cases} \bar{E}_i|_{\mathcal{C}} & \text{if } 0 \leqslant i < k, \\ \bar{E}_k\bar{E}_{k+1}|_{\mathcal{C}} & \text{if } i = k, \\ \bar{E}_{i+1}|_{\mathcal{C}} & \text{if } k < i < e, \end{cases} \qquad F_i = \begin{cases} \bar{F}_i|_{\mathcal{C}} & \text{if } 0 \leqslant i < k, \\ \bar{F}_{k+1}\bar{F}_k|_{\mathcal{C}} & \text{if } i = k, \\ \bar{F}_{i+1}|_{\mathcal{C}} & \text{if } k < i < e. \end{cases}$$

The following theorem holds.

Theorem 1.1. *The category $\mathcal{C}$ has the structure of a categorical representation of $\widetilde{\mathfrak{sl}}_e$ with respect to the functors $E_0, \dots, E_{e-1}, F_0, \dots, F_{e-1}$.* □

Let us explain our motivation for proving Theorem 1.1 (see [Maksimau 2015b] for more details). Let O^ν_{-e} be the parabolic category $\mathcal{O}$ for $\widehat{\mathfrak{gl}}_N = \mathfrak{gl}_N[t, t^{-1}] \oplus \mathbb{C}\mathbf{1} \oplus \mathbb{C}\partial$ with parabolic type ν at level $-e-N$. By [Rouquier et al. 2016], there is a categorical representation of $\widetilde{\mathfrak{sl}}_e$ in O^ν_{-e}. Now we apply Theorem 1.1 to $\bar{\mathcal{C}} = O^\nu_{-(e+1)}$. It happens that in this case the subcategory $\mathcal{C} \subset \bar{\mathcal{C}}$ defined as above is equivalent to O^ν_{-e}. This allows us to compare the categorical representations in the category $\mathcal{O}$ for $\widehat{\mathfrak{gl}}_N$ for two different (negative) levels.

A result similar to Theorem 1.1 has recently appeared in [Riche and Williamson 2018], where it is applied in the following way. It is known from [Chuang and Rouquier 2008] that there is a categorical representation of $\widetilde{\mathfrak{sl}}_p$ in the category $\mathrm{Rep}(\mathrm{GL}_n(\bar{\mathbb{F}}_p))$ of finite dimensional algebraic representations of $\mathrm{GL}_n(\bar{\mathbb{F}}_p)$. Riche and Williamson used this fact to construct a categorical representation of the Hecke category on the principal block $\mathrm{Rep}_0(\mathrm{GL}_n(\bar{\mathbb{F}}_p))$ of $\mathrm{Rep}(\mathrm{GL}_n(\bar{\mathbb{F}}_p))$ for $p > n$. Their proof is in two steps. First they show that the action of $\widetilde{\mathfrak{sl}}_p$ on $\mathrm{Rep}(\mathrm{GL}_n(\bar{\mathbb{F}}_p))$ induces an action of $\widetilde{\mathfrak{sl}}_n$ on some full subcategory of $\mathrm{Rep}(\mathrm{GL}_n(\bar{\mathbb{F}}_p))$. The second step is to show that the action of $\widetilde{\mathfrak{sl}}_n$ constructed on the first step induces an action of the Hecke category on $\mathrm{Rep}_0(\mathrm{GL}_n(\bar{\mathbb{F}}_p))$. The first step of their proof is essentially $p-n$ consecutive applications of Theorem 1.1.

The main difficulty in proving Theorem 1.1 is showing that the action of the KLR algebra $R_d(A^{(1)}_e)$ on

$\bar{F}^d$, where $\bar{F}=\bigoplus_{i=0}^{e}\bar{F}_i$, yields an action of the KLR algebra $R_d(A^{(1)}_{e-1})$ on F^d. So, to prove the theorem, we need to compare the KLR algebra $R_d(A^{(1)}_{e})$ with the KLR algebra $R_d(A^{(1)}_{e-1})$. This is done in Section 2.

We introduce the abbreviations $\Gamma=A^{(1)}_{e-1}$ and $\overline{\Gamma}=A^{(1)}_{e}$. Let $\alpha=\sum_{i=0}^{e-1}d_i\alpha_i$ be a dimension vector of the quiver Γ. We consider the dimension vector $\bar{\alpha}$ of $\overline{\Gamma}$ defined by

$$\bar{\alpha}=\sum_{i=0}^{k}d_i\alpha_i+\sum_{i=k+1}^{e}d_{i-1}\alpha_i.$$

Let $R_\alpha(\Gamma)$ and $R_{\bar{\alpha}}(\overline{\Gamma})$ be the KLR algebras associated with the quivers Γ and $\overline{\Gamma}$ and the dimension vectors α and $\bar{\alpha}$. The algebra $R_{\bar{\alpha}}(\overline{\Gamma})$ contains idempotents $e(\boldsymbol{i})$ parametrized by certain sequences $\boldsymbol{i}$ of vertices of $\overline{\Gamma}$. In Section 2D we consider some sets of such sequences $\bar{I}^{\bar{\alpha}}_{\mathrm{ord}}$ and $\bar{I}^{\bar{\alpha}}_{\mathrm{un}}$. Set $\boldsymbol{e}=\sum_{\boldsymbol{i}\in\bar{I}^{\bar{\alpha}}_{\mathrm{ord}}}e(\boldsymbol{i})\in R_{\bar{\alpha}}(\overline{\Gamma})$ and

$$S_{\bar{\alpha}}(\overline{\Gamma})=\boldsymbol{e}R_{\bar{\alpha}}(\overline{\Gamma})\boldsymbol{e}\Big/\sum_{\boldsymbol{i}\in\bar{I}^{\bar{\alpha}}_{\mathrm{un}}}\boldsymbol{e}R_{\bar{\alpha}}(\overline{\Gamma})e(\boldsymbol{i})R_{\bar{\alpha}}(\overline{\Gamma})\boldsymbol{e}.$$

The main result of Section 2 is the following theorem.

Theorem 1.2. *There is an algebra isomorphism* $R_\alpha(\Gamma)\simeq S_{\bar{\alpha}}(\overline{\Gamma})$. □

The paper has the following structure. In Section 2 we study KLR algebras. In particular, we prove Theorem 1.2. In Section 3 we study categorical representations. We prove our main result about categorical representations (Theorem 1.1). We also generalize this theorem to arbitrary symmetric Kac–Moody Lie algebras. In Appendix A we give a geometric construction of the isomorphism in Theorem 1.2. In Appendix B, we give some versions of Theorems 1.1 and 1.2 in type A over a local ring.

It is important to emphasize the relation between the present paper and [Maksimau 2015b]. That preprint contains (an earlier version of) the results of the present paper and an application of these results to the category $\mathcal{O}$ for $\widehat{\mathfrak{gl}}_N$. The preprint is expected to be published as two different papers. The present paper is the first of them. It contains the results of the preprint about KLR algebras and categorical representations. The second paper will give an application of the results of the first paper to the affine category $\mathcal{O}$.

2. KLR algebras

For a noetherian ring A we denote by $\mathrm{mod}\,(A)$ the abelian category of left finitely generated A-modules. We denote by $\mathbb{N}$ the set of nonnegative integers.

2A. ***Kac–Moody algebras associated with a quiver.*** Let $\Gamma=(I,H)$ be a quiver without 1-loops with the set of vertices I and the set of arrows H. For $i,j\in I$ let $h_{i,j}$ be the number of arrows from i to j and set also $a_{i,j}=2\delta_{i,j}-h_{i,j}-h_{j,i}$. Let $\mathfrak{g}_I$ be the Kac–Moody algebra over $\mathbb{C}$ associated with the matrix $(a_{i,j})$. Denote by e_i, f_i for $i\in I$ the Serre generators of $\mathfrak{g}_I$.

Remark 2.1. By the Kac–Moody Lie algebra associated with the Cartan matrix $(a_{i,j})$ we understand the Lie algebra with the set of generators e_i, f_i, h_i, $i\in I$, modulo the defining relations

$$\begin{aligned}
[h_i, h_j] &= 0,\\
[h_i, e_j] &= a_{i,j} e_j,\\
[h_i, f_j] &= -a_{i,j} e_j,\\
[e_i, f_j] &= \delta_{i,j} h_i,\\
(\mathrm{ad}(e_i))^{1-a_{i,j}}(e_j) &= 0 \quad i \neq j,\\
(\mathrm{ad}(f_i))^{1-a_{i,j}}(f_j) &= 0 \quad i \neq j.
\end{aligned}$$

In particular, if $(a_{i,j})$ is the affine Cartan matrix of type $A^{(1)}_{e-1}$, then we get the Lie algebra $\widetilde{\mathfrak{sl}}_e(\mathbb{C}) = \mathfrak{sl}_e(\mathbb{C}) \otimes \mathbb{C}[t, t^{-1}] \oplus \mathbb{C}\mathbf{1}$ (not $\mathfrak{sl}_e(\mathbb{C}) \otimes \mathbb{C}[t, t^{-1}] \oplus \mathbb{C}\mathbf{1} \oplus \mathbb{C}\partial$).

For each $i \in I$, let α_i be the simple root corresponding to e_i. Set

$$Q_I = \bigoplus_{i \in I} \mathbb{Z}\alpha_i \quad \text{and} \quad Q_I^+ = \bigoplus_{i \in I} \mathbb{N}\alpha_i.$$

For $\alpha = \sum_{i \in I} d_i \alpha_i \in Q_I^+$ denote by $|\alpha|$ its height, i.e., we have $|\alpha| = \sum_{i \in I} d_i$. Set $I^\alpha = \{\boldsymbol{i} = (i_1, \dots, i_{|\alpha|}) \in I^{|\alpha|} : \sum_{r=1}^{|\alpha|} \alpha_{i_r} = \alpha\}$.

2B. ***Doubled quiver.*** Let $\Gamma = (I, H)$ be a quiver without 1-loops. Fix a decomposition $I = I_0 \sqcup I_1$ such that there are no arrows between the vertices in I_1. In this section we define a *doubled quiver* $\overline{\Gamma} = (\bar{I}, \overline{H})$ associated with (Γ, I_0, I_1). The idea is to "double" each vertex in the set I_1 (we do not touch the vertices from I_0). We replace each vertex $i \in I_1$ by a couple of vertices i^1 and i^2 with an arrow $i^1 \to i^2$. Each arrow entering i should be replaced by an arrow entering to i^1, each arrow coming from i should be replaced by an arrow coming from i^2.

Now we describe the construction of $\overline{\Gamma} = (\bar{I}, \overline{H})$ formally. Let $\bar{I}_0$ be a set that is in bijection with I_0. Let i^0 be the element of $\bar{I}_0$ associated with an element $i \in I_0$. Similarly, let $\bar{I}_1$ and $\bar{I}_2$ be sets that are in bijection with I_1. Denote by i^1 and i^2 the elements of $\bar{I}_1$ and $\bar{I}_2$ respectively that correspond to an element $i \in I_1$. Put $\bar{I} = \bar{I}_0 \sqcup \bar{I}_1 \sqcup \bar{I}_2$. We define $\overline{H}$ in the following way. The set $\overline{H}$ contains 4 types of arrows:

- An arrow $i^0 \to j^0$ for each arrow $i \to j$ in H with $i, j \in I_0$.
- An arrow $i^0 \to j^1$ for each arrow $i \to j$ in H with $i \in I_0$, $j \in I_1$.
- An arrow $i^2 \to j^0$ for each arrow $i \to j$ in H with $i \in I_1$, $j \in I_0$.
- An arrow $i^1 \to i^2$ for each vertex $i \in I_1$.

Set $I^\infty = \coprod_{d \in \mathbb{N}} I^d$ and $\bar{I}^\infty = \coprod_{d \in \mathbb{N}} \bar{I}^d$, where I^d and $\bar{I}^d$ are the cartesian products. The concatenation yields a monoid structure on I^∞ and $\bar{I}^\infty$. Let $\phi \colon I^\infty \to \bar{I}^\infty$ be the unique morphism of monoids such that for $i \in I \subset I^\infty$ we have

$$\phi(i) = \begin{cases} i^0 & \text{if } i \in I_0,\\ (i^1, i^2) & \text{if } i \in I_1. \end{cases}$$

There is a unique $\mathbb{Z}$-linear map $\phi\colon\ Q_I \to Q_{\bar I}$ such that $\phi(I^\alpha) \subset \bar{I}^{\phi(\alpha)}$ for each $\alpha \in Q_I^+$. It is given by

$$\phi(\alpha_i) = \begin{cases} \alpha_{i^0} & \text{if } i \in I_0, \\ \alpha_{i^1} + \alpha_{i^2} & \text{if } i \in I_1. \end{cases}$$

2C. *KLR algebras.* Let $\boldsymbol{k}$ be a field. Let $\Gamma = (I, H)$ be a quiver without 1-loops. For $r \in [1, d-1]$ let s_r be the transposition $(r, r+1) \in \mathfrak{S}_d$. For $\boldsymbol{i} = (i_1, \dots, i_d) \in I^d$ set $s_r(\boldsymbol{i}) = (i_1, \dots, i_{r-1}, i_{r+1}, i_r, i_{r+2}, \dots, i_d)$. For $i, j \in I$ we set

$$Q_{i,j}(u, v) = \begin{cases} 0 & \text{if } i = j, \\ (v-u)^{h_{i,j}}(u-v)^{h_{j,i}} & \text{else.} \end{cases}$$

Definition 2.2. Assume that the quiver Γ is finite. The *KLR-algebra* $R_{d,\boldsymbol{k}}(\Gamma)$ is the $\boldsymbol{k}$-algebra with the set of generators $\tau_1, \dots, \tau_{d-1}, x_1, \dots, x_d, e(\boldsymbol{i})$ where $\boldsymbol{i} \in I^d$, modulo the following defining relations:

$$\begin{aligned}
e(\boldsymbol{i})e(\boldsymbol{j}) &= \delta_{\boldsymbol{i},\boldsymbol{j}}e(\boldsymbol{i}),\\
\sum_{\boldsymbol{i} \in I^d} e(\boldsymbol{i}) &= 1,\\
x_r e(\boldsymbol{i}) &= e(\boldsymbol{i})x_r,\\
\tau_r e(\boldsymbol{i}) &= e(s_r(\boldsymbol{i}))\tau_r,\\
x_r x_s &= x_s x_r,\\
\tau_r x_{r+1} e(\boldsymbol{i}) &= (x_r\tau_r + \delta_{i_r,i_{r+1}})e(\boldsymbol{i}),\\
x_{r+1}\tau_r e(\boldsymbol{i}) &= (\tau_r x_r + \delta_{i_r,i_{r+1}})e(\boldsymbol{i}),\\
\tau_r x_s &= x_s\tau_r \quad \text{if } s \neq r, r+1,\\
\tau_r\tau_s &= \tau_s\tau_r \quad \text{if } |r-s| > 1,\\
\tau_r^2 e(\boldsymbol{i}) &= \begin{cases} 0 & \text{if } i_r = i_{r+1},\\ Q_{i_r,i_{r+1}}(x_r, x_{r+1})e(\boldsymbol{i}) & \text{else,} \end{cases}\\
(\tau_r\tau_{r+1}\tau_r - \tau_{r+1}\tau_r\tau_{r+1})e(\boldsymbol{i}) &= \begin{cases} (x_{r+2}-x_r)^{-1}(Q_{i_r,i_{r+1}}(x_{r+2}, x_{r+1}) - Q_{i_r,i_{r+1}}(x_r, x_{r+1}))e(\boldsymbol{i}) & \text{if } i_r = i_{r+2},\\ 0 & \text{else.} \end{cases}
\end{aligned}$$

for each $\boldsymbol{i}$, $\boldsymbol{j}$, r and s. We may write $R_{d,\boldsymbol{k}} = R_{d,\boldsymbol{k}}(\Gamma)$. The algebra $R_{d,\boldsymbol{k}}$ admits a $\mathbb{Z}$-grading such that $\deg e(\boldsymbol{i}) = 0$, $\deg x_r = 2$ and $\deg \tau_s e(\boldsymbol{i}) = -a_{i_s,i_{s+1}}$, for each $1 \leqslant r \leqslant d$, $1 \leqslant s < d$ and $\boldsymbol{i} \in I^d$.

For each $\alpha \in Q_I^+$ such that $|\alpha| = d$ set $e(\alpha) = \sum_{\boldsymbol{i} \in I^\alpha} e(\boldsymbol{i}) \in R_{d,\boldsymbol{k}}$. It is a homogeneous central idempotent of degree zero. We have the following decomposition into a sum of unitary $\boldsymbol{k}$-algebras $R_{d,\boldsymbol{k}} = \bigoplus_{|\alpha|=d} R_{\alpha,\boldsymbol{k}}$, where $R_{\alpha,\boldsymbol{k}} = e(\alpha)R_{d,\boldsymbol{k}}$.

Let $\boldsymbol{k}_d^{(I)}$ be the direct sum of copies of the ring $\boldsymbol{k}_d[x] := \boldsymbol{k}[x_1, \dots, x_d]$ labeled by I^d. We write

$$\boldsymbol{k}_d^{(I)} = \bigoplus_{\boldsymbol{i} \in I^d} \boldsymbol{k}_d[x]e(\boldsymbol{i}), \tag{2}$$

where $e(\boldsymbol{i})$ is the idempotent of the ring $\boldsymbol{k}_d^{(I)}$ projecting to the component $\boldsymbol{i}$. A polynomial in $\boldsymbol{k}_d[x]$ can be considered as an element of $\boldsymbol{k}_d^{(I)}$ via the diagonal inclusion. For each $i, j \in I$ fix a polynomial $P_{i,j}(u,v)$ such that we have $Q_{i,j}(u,v) = P_{i,j}(u,v)P_{j,i}(v,u)$.

Denote by ∂_r the Demazure operator on $\boldsymbol{k}_d[x]$, i.e., we have

$$\partial_r(f) = (x_r - x_{r+1})^{-1}(s_r(f) - f).$$

The following is proved in [Rouquier 2008, §3.2].

Proposition 2.3. *The algebra $R_{d,\boldsymbol{k}}$ has a faithful representation in the vector space $\boldsymbol{k}_d^{(I)}$ such that the element $e(\boldsymbol{i})$ acts by projection to $\boldsymbol{k}_d^{(I)}e(\boldsymbol{i})$, the element x_r acts by multiplication by x_r and such that for $f \in \boldsymbol{k}_d[x]$ we have*

$$\tau_r \cdot fe(\boldsymbol{i}) = \begin{cases} \partial_r(f)e(\boldsymbol{i}) & \text{if } i_r = i_{r+1}, \\ P_{i_r,i_{r+1}}(x_{r+1}, x_r)s_r(f)e(s_r(\boldsymbol{i})) & \text{otherwise.} \end{cases} \tag{3}$$

We will always choose $P_{i,j}$ in the following way:

$$P_{i,j}(u,v) = (u-v)^{h_{j,i}}.$$

Remark 2.4. There is an explicit construction of a basis of a KLR algebra (see [Khovanov and Lauda 2009, Theorem 2.5]). Assume $\boldsymbol{i}, \boldsymbol{j} \in I^\alpha$. Set $\mathfrak{S}_{\boldsymbol{i},\boldsymbol{j}} = \{w \in \mathfrak{S}_d : w(\boldsymbol{i}) = \boldsymbol{j}\}$. For each permutation $w \in \mathfrak{S}_{\boldsymbol{i},\boldsymbol{j}}$ fix a reduced expression $w = s_{p_1} \cdots s_{p_r}$ and set $\tau_w = \tau_{p_1} \cdots \tau_{p_r}$. Then the vector space $e(\boldsymbol{j})R_{\alpha,\boldsymbol{k}}e(\boldsymbol{i})$ has a basis $\{\tau_w x_1^{a_1} \cdots x_d^{a_d} e(\boldsymbol{i}) : w \in \mathfrak{S}_{\boldsymbol{i},\boldsymbol{j}}, a_1, \dots, a_d \in \mathbb{N}\}$. Note that the element τ_w depends on the reduced expression of w. Moreover, if we change the reduced expression of w, then the element $\tau_w e(\boldsymbol{i})$ is changed only by a linear combination of monomials of the form $\tau_{q_1} \cdots \tau_{q_t} x_1^{b_1} \cdots x_d^{b_d} e(\boldsymbol{i})$ with $t < \ell(w)$. Note also that if $s_{p_1} \cdots s_{p_r}$ is not a reduced expression, then the element $\tau_{p_1} \cdots \tau_{p_r} e(\boldsymbol{i})$ may be written as a linear combination of monomials of the form $\tau_{q_1} \cdots \tau_{q_t} x_1^{b_1} \cdots x_d^{b_d} e(\boldsymbol{i})$ with $t < r$. Moreover, in both situations above, the linear combination can be chosen in such a way that for each monomial $\tau_{q_1} \cdots \tau_{q_t} x_1^{b_1} \cdots x_d^{b_d} e(\boldsymbol{i})$ in the linear combination, the expression $s_{q_1} \cdots s_{q_t}$ is reduced.

Remark 2.5. The algebra $R_{d,\boldsymbol{k}}$ in Definition 2.2 is well defined only for a finite quiver because of the second relation. However, the algebra $R_{\alpha,\boldsymbol{k}}$ is well defined even if the quiver is infinite because each α uses a finite set of vertices. Thus, for an infinite quiver we can define $R_{d,\boldsymbol{k}}$ as $R_{d,\boldsymbol{k}} = \bigoplus_{|\alpha|=d} R_{\alpha,\boldsymbol{k}}$. However, in this case the algebra $R_{d,\boldsymbol{k}}$ is not unitary.

2D. ***Balanced KLR algebras.*** From now on the quiver Γ is assumed to be finite. Fix a decomposition $I = I_0 \sqcup I_1$ as in Section 2B and consider the quiver $\bar{\Gamma} = (\bar{I}, \bar{H})$ as in Section 2B. Recall the decomposition $\bar{I} = \bar{I}_0 \sqcup \bar{I}_1 \sqcup \bar{I}_2$. In this section we work with the KLR algebra associated with the quiver $\bar{\Gamma}$.

We say that a sequence $\boldsymbol{i} = (i_1, i_2, \dots, i_d) \in \bar{I}^d$ is *unordered* if there is an index $r \in [1, d]$ such that the number of elements from $\bar{I}_2$ in the sequence $(i_1, i_2, \dots, i_r)$ is strictly greater than the number of elements from $\bar{I}_1$. We say that it is *well-ordered* if for each index a such that $i_a = i^1$ for some $i \in I_1$, we have $a < d$ and $i_{a+1} = i^2$. We denote by $\bar{I}^\alpha_{\mathrm{ord}}$ and $\bar{I}^\alpha_{\mathrm{un}}$ the subsets of well-ordered and unordered sequences in $\bar{I}^\alpha$.

The map ϕ from Section 2B yields a bijection

$$\phi\colon Q_I^+ \to \bigl\{\alpha = \sum_{i\in\bar I} d_i\alpha_i \in Q_{\bar I}^+ : d_{i^1} = d_{i^2}, \forall i \in I_1\bigr\}, \quad \alpha \mapsto \bar\alpha.$$

Fix $\alpha \in Q_I^+$. Set $\boldsymbol{e} = \sum_{\boldsymbol{i}\in \bar I^{\bar\alpha}_{\mathrm{ord}}} e(\boldsymbol{i}) \in R_{\bar\alpha,\boldsymbol{k}}(\bar\Gamma)$.

Definition 2.6. For $\alpha \in Q_I^+$, the *balanced KLR algebra* is the algebra

$$S_{\bar\alpha,\boldsymbol{k}}(\bar\Gamma) = \boldsymbol{e}R_{\bar\alpha,\boldsymbol{k}}(\bar\Gamma)\boldsymbol{e} \Big/ \sum_{\boldsymbol{i}\in \bar I^{\bar\alpha}_{\mathrm{un}}} \boldsymbol{e}R_{\bar\alpha,\boldsymbol{k}}(\bar\Gamma)e(\boldsymbol{i})R_{\bar\alpha,\boldsymbol{k}}(\bar\Gamma)\boldsymbol{e}.$$

We may write $S_{\bar\alpha,\boldsymbol{k}}(\bar\Gamma) = S_{\bar\alpha,\boldsymbol{k}}$.

Remark 2.7. Assume that $\boldsymbol{i} = (i_1, \dots, i_d) \in \bar I^{\bar\alpha}_{\mathrm{ord}}$. Let a be an index such that $i_a \in \bar I_1$. We have the relation $\tau_a^2 e(\boldsymbol{i}) = (x_{a+1} - x_a)e(\boldsymbol{i})$ in $R_{\bar\alpha,\boldsymbol{k}}$. Moreover, we have $\tau_a^2 e(\boldsymbol{i}) = \tau_a e(s_a(\boldsymbol{i}))\tau_a e(\boldsymbol{i})$ and $s_a(\boldsymbol{i})$ is unordered. Thus we have $x_a e(\boldsymbol{i}) = x_{a+1}e(\boldsymbol{i})$ in $S_{\bar\alpha,\boldsymbol{k}}$.

2E. *The polynomial representation of $S_{\bar\alpha,\boldsymbol{k}}$.*

We assume $\alpha = \sum_{i\in I} d_i\alpha_i \in Q_I^+$. Let $\boldsymbol{i} = (i_1, \dots, i_d) \in \bar I^{\bar\alpha}_{\mathrm{ord}}$. Denote by $J(\boldsymbol{i})$ the ideal of the polynomial ring $\boldsymbol{k}_d[x]e(\boldsymbol{i}) \subset \boldsymbol{k}_d^{(\bar I)}$ generated by the set

$$\{(x_r - x_{r+1})e(\boldsymbol{i}) : i_r \in \bar I_1\}.$$

Lemma 2.8. *Assume that $\boldsymbol{i} \in \bar I^{\bar\alpha}_{\mathrm{ord}}$ and $\boldsymbol{j} \in \bar I^{\bar\alpha}_{\mathrm{un}}$. Then each element of $e(\boldsymbol{i})R_{\bar\alpha,\boldsymbol{k}}e(\boldsymbol{j})$ maps $\boldsymbol{k}_d[x]e(\boldsymbol{j})$ to $J(\boldsymbol{i})$.*

Proof. We will prove by induction on k that for all $\boldsymbol{i} \in \bar I^{\bar\alpha}_{\mathrm{ord}}$ and $\boldsymbol{j} \in \bar I^{\bar\alpha}_{\mathrm{un}}$ and all $p_1, \dots, p_k$ such that the permutation $w = s_{p_1} \cdots s_{p_k} \in \mathfrak{S}_d$ satisfies $w(\boldsymbol{j}) = \boldsymbol{i}$, the monomial $\tau_{p_1} \cdots \tau_{p_k}$ maps $\boldsymbol{k}_d[x]e(\boldsymbol{j})$ to $J(\boldsymbol{i})$.

Assume $k = 1$. Write $p = p_1$. Let us write $\boldsymbol{i} = (i_1, \dots, i_d)$ and $\boldsymbol{j} = (j_1, \dots, j_d)$. Then we have $\boldsymbol{i} = s_p(\boldsymbol{j})$. By assumptions on $\boldsymbol{i}$ and $\boldsymbol{j}$ we know that there exists $i \in I_1$ such that $i_p = j_{p+1} = i^1$ and $i_{p+1} = j_p = i^2$. In this case the statement is obvious because τ_p maps $fe(\boldsymbol{j}) \in \boldsymbol{k}_d[x]e(\boldsymbol{j})$ to $(x_{p+1} - x_p)s_p(f)e(\boldsymbol{i})$ by (3).

Now consider a monomial $\tau_{p_1} \cdots \tau_{p_k}$ such that the permutation $w = s_{p_1} \cdots s_{p_k}$ satisfies $w(\boldsymbol{j}) = \boldsymbol{i}$ and assume that the statement is true for all such monomials of smaller length. By assumptions on $\boldsymbol{i}$ and $\boldsymbol{j}$ there is an index $r \in [1, d]$ such that $i_r = i^1$ for some $i \in I_1$ and $w^{-1}(r+1) < w^{-1}(r)$. Thus w has a reduced expression of the form $w = s_r s_{r_1} \cdots s_{r_h}$. This implies that $\tau_{p_1} \cdots \tau_{p_k} e(\boldsymbol{j})$ is equal to a monomial of the form $\tau_r\tau_{r_1} \cdots \tau_{r_h} e(\boldsymbol{j})$ modulo monomials of the form $\tau_{q_1} \cdots \tau_{q_t} x_1^{b_1} \cdots x_d^{b_d} e(\boldsymbol{j})$ with $t < k$, see Remark 2.4. As the sequence $s_r(\boldsymbol{i})$ is unordered, the case $k = 1$ and the induction hypothesis imply the statement. □

Lemma 2.9. *Assume that $\boldsymbol{i}, \boldsymbol{j} \in \bar I^{\bar\alpha}_{\mathrm{ord}}$. Then each element of $e(\boldsymbol{i})R_{\bar\alpha,\boldsymbol{k}}e(\boldsymbol{j})$ maps $J(\boldsymbol{j})$ into $J(\boldsymbol{i})$.*

Proof. Take $y \in e(\boldsymbol{i})R_{\bar\alpha,\boldsymbol{k}}e(\boldsymbol{j})$. We must prove that for each $r \in [1, d]$ such that $j_r = i^1$ for some $i \in I_1$ and each $f \in \boldsymbol{k}_d[x]$ we have $y((x_r - x_{r+1})fe(\boldsymbol{j})) \in J(\boldsymbol{i})$. We have $(x_r - x_{r+1})fe(\boldsymbol{j}) = -\tau_r^2(fe(\boldsymbol{i}))$ (see Remark 2.7). This implies

$$y((x_r - x_{r+1})fe(\boldsymbol{j})) = -y\tau_r^2(fe(\boldsymbol{j})) = -y\tau_r e(s_r(\boldsymbol{j}))(\tau_r(fe(\boldsymbol{j}))).$$

Thus Lemma 2.8 implies the statement because the sequence $s_r(\boldsymbol{j})$ is unordered. □

The representation of $R_{\bar{\alpha},\boldsymbol{k}}$ on

$$\boldsymbol{k}_{\bar{\alpha}}^{(\bar{I})} := \bigoplus_{\boldsymbol{i}\in \bar{I}^{\bar{\alpha}}} \boldsymbol{k}_{|\bar{\alpha}|}[x]e(\boldsymbol{i})$$

yields a representation of $\boldsymbol{e}R_{\bar{\alpha},\boldsymbol{k}}\boldsymbol{e}$ on

$$\boldsymbol{k}_{\bar{\alpha},\mathrm{ord}}^{(\bar{I})} := \bigoplus_{\boldsymbol{i}\in \bar{I}^{\bar{\alpha}}_{\mathrm{ord}}} \boldsymbol{k}_{|\bar{\alpha}|}[x]e(\boldsymbol{i}).$$

Set $J_{\bar{\alpha},\mathrm{ord}} = \bigoplus_{\boldsymbol{i}\in \bar{I}^{\bar{\alpha}}_{\mathrm{ord}}} J(\boldsymbol{i})$. From Lemmas 2.8 and 2.9 we deduce the following.

Lemma 2.10. *The representation of $R_{\bar{\alpha},\boldsymbol{k}}$ on $\boldsymbol{k}_{\bar{\alpha}}^{(\bar{I})}$ factors through a representation of $S_{\bar{\alpha},\boldsymbol{k}}$ on $\boldsymbol{k}_{\bar{\alpha},\mathrm{ord}}^{(\bar{I})}/J_{\bar{\alpha},\mathrm{ord}}$. This representation is faithful.*

Proof. The faithfulness is proved in the proof of Theorem 2.12. □

2F. *The comparison of the polynomial representations.* Fix $\alpha \in Q_I^+$. Set $d = |\alpha|$ and $\bar{d} = |\bar{\alpha}|$. For each sequence $\boldsymbol{i} = (i_1, \dots, i_d) \in I^\alpha$ and $r \in [1, d]$ we denote by r' or $r'_{\boldsymbol{i}}$ the positive integer such that $r' - 1$ is the length of the sequence $\phi(i_1, \dots, i_{r-1}) \in \bar{I}^\infty$.

For $r \in [1, d]$ and $r \in [1, d-1]$ consider the element $x_r^* \in S_{\bar{\alpha},\boldsymbol{k}}$ and $\tau_r^* \in S_{\bar{\alpha},\boldsymbol{k}}$, respectively, such that for each $\boldsymbol{i} \in I^\alpha$ we have

$$x_r^* e(\phi(\boldsymbol{i})) = x_{r'}e(\phi(\boldsymbol{i})), \quad \tau_r^* e(\phi(\boldsymbol{i})) = \begin{cases} \tau_{r'}e(\phi(\boldsymbol{i})), & \text{if } i_r, i_{r+1} \in I_0, \\ \tau_{r'}\tau_{r'+1}e(\phi(\boldsymbol{i})) & \text{if } i_r \in I_1, i_{r+1} \in I_0, \\ \tau_{r'+1}\tau_{r'}e(\phi(\boldsymbol{i})) & \text{if } i_r \in I_0, i_{r+1} \in I_1, \\ \tau_{r'+1}\tau_{r'+2}\tau_{r'}\tau_{r'+1}e(\phi(\boldsymbol{i})) & \text{if } i_r, i_{r+1} \in I_1, i_r \neq i_{r+1}, \\ -\tau_{r'+1}\tau_{r'+2}\tau_{r'}\tau_{r'+1}e(\phi(\boldsymbol{i})) & \text{if } i_r = i_{r+1} \in I_1. \end{cases}$$

For each $\boldsymbol{i} \in I^\alpha$ we have the algebra isomorphism

$$\boldsymbol{k}_d[x]e(\boldsymbol{i}) \simeq \boldsymbol{k}_{\bar{d}}[x]e(\phi(\boldsymbol{i}))/J(\phi(\boldsymbol{i})), \quad x_r e(\boldsymbol{i}) \mapsto x_{r'}e(\phi(\boldsymbol{i})).$$

We will always identify $\boldsymbol{k}_\alpha^{(I)}$ with $\boldsymbol{k}_{\bar{\alpha},\mathrm{ord}}^{(\bar{I})}/J_{\bar{\alpha},\mathrm{ord}}$ via this isomorphism.

Lemma 2.11. *The action of the elements $e(\boldsymbol{i})$, $x_r e(\boldsymbol{i})$ and $\tau_r e(\boldsymbol{i})$ of $R_{\alpha,\boldsymbol{k}}$ on $\boldsymbol{k}_\alpha^{(I)}$ is the same as the action of the elements $e(\phi(\boldsymbol{i}))$, $x_r^* e(\phi(\boldsymbol{i}))$, $\tau_r^* e(\phi(\boldsymbol{i}))$ of $S_{\bar{\alpha},\boldsymbol{k}}$ on $\boldsymbol{k}_{\bar{\alpha},\mathrm{ord}}^{(\bar{I})}/J_{\bar{\alpha},\mathrm{ord}}$.*

Proof. The proof is based on the observation that by construction for each $i \in I_1$ and $j \in I_0$ we have

$$P_{i^1,j^0}(u,v)P_{i^2,j^0}(u,v) = P_{i,j}(u,v), \quad P_{j^0,i^1}(u,v)P_{j^0,i^2}(u,v) = P_{j,i}(u,v). \tag{4}$$

For each $\boldsymbol{i} \in I^\alpha$, we write $\phi(\boldsymbol{i}) = (i'_1, i'_2, \dots, i'_{\bar{d}})$. The only difficult part concerns the operator $\tau_r e(\boldsymbol{i})$ when at least one of the elements i_r or i_{r+1} is in I_1. Assume that $i_r \in I_1$ and $i_{r+1} \in I_0$. In this case we have

$$i'_{r'} = (i_r)^1 \in \bar{I}_1, \quad i'_{r'+1} = (i_r)^2 \in \bar{I}_2, \quad i'_{r'+2} = (i_{r+1})^0 \in \bar{I}_0.$$

In particular, the element $i'_{r'+2}$ is different from $i'_{r'}$ and $i'_{r'+1}$. Then, by (3), for each $f \in \boldsymbol{k}_{\bar{d}}[x]$ the element $\tau_r^* e(\phi(\boldsymbol{i})) = \tau_{r'}\tau_{r'+1}e(\phi(\boldsymbol{i}))$ maps $fe(\phi(\boldsymbol{i})) \in \boldsymbol{k}_{\bar{\alpha},\mathrm{ord}}^{(\bar{I})}/J_{\bar{\alpha},\mathrm{ord}}$ to

$$\begin{aligned}P_{i'_{r'},i'_{r'+2}}(x_{r'+1},x_{r'})s_{r'}(P_{i'_{r'+1},i'_{r'+2}}(x_{r'+2},x_{r'+1})s_{r'+1}(f))e(s_{r'}s_{r'+1}(\phi(\boldsymbol{i})))&\\ &= P_{i'_{r'},i'_{r'+2}}(x_{r'+1},x_{r'})P_{i'_{r'+1},i'_{r'+2}}(x_{r'+2},x_{r'})s_{r'}s_{r'+1}(f)e(\phi(s_r(\boldsymbol{i})))\\ &= P_{i_r,i_{r+1}}(x_{r'+1},x_{r'})s_{r'}s_{r'+1}(f)e(\phi(s_r(\boldsymbol{i}))),\end{aligned}$$

where the last equality holds by (4). Thus we see that the action of $\tau_r^*e(\phi(\boldsymbol{i}))$ on the polynomial representation is the same as the action of $\tau_r e(\boldsymbol{i})$. The case when $i_r \in I_0$ and $i_{r+1} \in I_1$ can be done similarly.

Assume now that $i_r \neq i_{r+1}$ are both in I_1. By the assumption on the quiver Γ (see Section 2B), there are no arrows in Γ between i_r and i_{r+1}. Thus there are no arrows in $\overline{\Gamma}$ between any of the vertices $(i_r)^1 = i'_{r'}$ or $(i_r)^2 = i'_{r'+1}$ and any of the vertices $(i_{r+1})^1 = i'_{r'+2}$ or $(i_{r+1})^2 = i'_{r'+3}$. Then, by (3), for each $f \in \boldsymbol{k}_{\bar{d}}[x]$ the element $\tau_r^*e(\boldsymbol{i}) = \tau_{r'+1}\tau_{r'+2}\tau_{r'}\tau_{r'+1}e(\phi(\boldsymbol{i}))$ maps $fe(\phi(\boldsymbol{i}))$ to

$$s_{r'+1}s_{r'+2}s_{r'}s_{r'+1}(f)e(\phi(s_r(\boldsymbol{i}))).$$

Thus we see that the action of $\tau_r^*e(\phi(\boldsymbol{i}))$ on the polynomial representation is the same as that of $\tau_r e(\boldsymbol{i})$.

Finally, assume that $i_r = i_{r+1} \in I_1$. In this case we have

$$(i_r)^1 = i'_{r'} = (i_{r+1})^1 = i'_{r'+2} \quad \text{and} \quad (i_r)^2 = i'_{r'+1} = (i_{r+1})^2 = i'_{r'+3}.$$

Then, by (3), for each $f \in \boldsymbol{k}_{\bar{d}}[x]$ the element $\tau_r^*e(\phi(\boldsymbol{i})) = -\tau_{r'+1}\tau_{r'+2}\tau_{r'}\tau_{r'+1}e(\phi(\boldsymbol{i}))$ maps $fe(\phi(\boldsymbol{i}))$ to

$$s_{r'+1}\partial_{r'+2}\partial_{r'}(x_{r'+1}-x_{r'+2})s_{r'+1}(f)e(\phi(s_r(\boldsymbol{i}))),$$

where ∂_r is the Demazure operator (see the definition before Proposition 2.3). To prove that this gives the same result as for $\tau_r e(\boldsymbol{i})$, it is enough to check this on monomials $x_r^n x_{r+1}^m e(\boldsymbol{i})$. Assume for simplicity that $n \geqslant m$. The situation $n \leqslant m$ can be treated similarly. The element $\tau_r e(\boldsymbol{i})$ maps this monomial to

$$\partial_r(x_r^n x_{r+1}^m)e(\boldsymbol{i}) = -\sum_{a=m}^{n-1} x_r^a x_{r+1}^{n+m-1-a}e(\boldsymbol{i}).$$

Here the symbol $\sum_{a=x}^{y}$ means 0 when $y = x-1$. The element $\tau_r^*e(\phi(\boldsymbol{i}))$ maps $x_{r'+1}^n x_{r'+2}^m e(\phi(\boldsymbol{i}))$ to $s_{r'+1}\partial_{r'+2}\partial_{r'}[x_{r'+1}^{m+1}x_{r'+2}^n - x_{r'+1}^m x_{r'+2}^{n+1}]e(\phi(\boldsymbol{i}))$, which equals

$$\begin{aligned}&s_{r'+1}\Bigg[-\bigg(\sum_{a=0}^{m} x_{r'}^a x_{r'+1}^{m-a}\bigg)\bigg(\sum_{b=0}^{n-1} x_{r'+2}^b x_{r'+3}^{n-1-b}\bigg) + \bigg(\sum_{a=0}^{m-1} x_{r'}^a x_{r'+1}^{m-1-a}\bigg)\bigg(\sum_{b=0}^{n} x_{r'+2}^b x_{r'+3}^{n-b}\bigg)\Bigg]e(\phi(\boldsymbol{i}))\\ &= \Bigg[-\bigg(\sum_{a=0}^{m} x_{r'}^a x_{r'+2}^{m-a}\bigg)\bigg(\sum_{b=0}^{n-1} x_{r'+1}^b x_{r'+3}^{n-1-b}\bigg) + \bigg(\sum_{a=0}^{m-1} x_{r'}^a x_{r'+2}^{m-1-a}\bigg)\bigg(\sum_{b=0}^{n} x_{r'+1}^b x_{r'+3}^{n-b}\bigg)\Bigg]e(\phi(\boldsymbol{i}))\\ &= \Bigg[-x_{r'}^m\bigg(\sum_{b=0}^{n-1} x_{r'+1}^b x_{r'+3}^{n-1-b}\bigg) + x_{r'+1}^n\bigg(\sum_{a=0}^{m-1} x_{r'}^a x_{r'+2}^{m-1-a}\bigg)\Bigg]e(\phi(\boldsymbol{i}))\\ &= \Bigg[-x_{r'+1}^m\bigg(\sum_{b=0}^{n-1} x_{r'+1}^b x_{r'+2}^{n-1-b}\bigg) + x_{r'+1}^n\bigg(\sum_{a=0}^{m-1} x_{r'+1}^a x_{r'+2}^{m-1-a}\bigg)\Bigg]e(\phi(\boldsymbol{i})) = -\bigg(\sum_{a=m}^{n-1} x_{r'+1}^a x_{r'+2}^{m+n-1-a}\bigg)e(\phi(\boldsymbol{i})).\end{aligned}$$

Here the first equality follows from the following property of the Demazure operator

$$\partial_r(x_{r+1}^n) = -\partial_r(x_r^n) = \sum_{a=0}^{n-1} x_r^a x_{r+1}^{n-1-a},$$

the fourth equality follows from Remark 2.7. Other equalities are obtained by elementary manipulations with sums. □

2G. *The isomorphism Φ.*

Theorem 2.12. *For each $\alpha \in Q_I^+$, there is an algebra isomorphism $\Phi_{\alpha,\boldsymbol{k}}\colon R_{\alpha,\boldsymbol{k}} \to S_{\bar\alpha,\boldsymbol{k}}$ such that*

$$\begin{aligned} e(\boldsymbol{i}) &\mapsto e(\phi(\boldsymbol{i})),\\ x_r e(\boldsymbol{i}) &\mapsto x_r^* e(\phi(\boldsymbol{i})),\\ \tau_r e(\boldsymbol{i}) &\mapsto \tau_r^* e(\phi(\boldsymbol{i})).\end{aligned}$$

Proof. By Proposition 2.3, the representation $\boldsymbol{k}_\alpha^{(I)}$ of $R_{\alpha,\boldsymbol{k}}$ is faithful. Now, in view of Lemma 2.11, it is enough to prove the following two facts:

- The elements $e(\phi(\boldsymbol{i}))$, x_r^*, τ_r^* generate $S_{\bar\alpha,\boldsymbol{k}}$.
- The representation $\boldsymbol{k}_{\bar\alpha,\mathrm{ord}}^{(\bar I)}/J_{\bar\alpha,\mathrm{ord}}$ of $S_{\bar\alpha,\boldsymbol{k}}$ is faithful.

Fix $\boldsymbol{i}, \boldsymbol{j} \in I^\alpha$. Set $\boldsymbol{i}' = (i'_1, \dots, i'_{\bar d}) = \phi(\boldsymbol{i})$ and $\boldsymbol{j}' = \phi(\boldsymbol{j})$. Let $\boldsymbol{B}$ and $\boldsymbol{B}'$ be the bases of $e(\boldsymbol{j})R_{\alpha,\boldsymbol{k}}e(\boldsymbol{i})$ and $e(\boldsymbol{j}')R_{\bar\alpha,\boldsymbol{k}}e(\boldsymbol{i}')$, respectively, as in Remark 2.4. These bases depend on some choices of reduced expressions. We will make some special choices later. For each element $b = \tau_w x_1^{a_1} \cdots x_d^{a_d} e(\boldsymbol{i}) \in \boldsymbol{B}$ we construct an element $b^* \in e(\boldsymbol{j}')S_{\bar\alpha,\boldsymbol{k}}e(\boldsymbol{i}')$ that acts by the same operator on the polynomial representation. We set

$$b^* = \tau_{p_1}^* \cdots \tau_{p_k}^* (x_1^*)^{a_1} \cdots (x_d^*)^{a_d} e(\boldsymbol{i}') \in e(\boldsymbol{j}')S_{\bar\alpha,\boldsymbol{k}}e(\boldsymbol{i}'),$$

where $w = s_{p_1} \cdots s_{p_k}$ is a reduced expression (as we said above, some special choice of reduced expressions will be fixed later).

Let us call the permutation $w \in \mathfrak{S}_{\boldsymbol{i}',\boldsymbol{j}'}$ *balanced* if we have $w(a+1) = w(a)+1$ for each a such that $i'_a = i^1$ for some $i \in I$ (and thus $i'_{a+1} = i^2$). Otherwise we say that w is *unbalanced*. There exists a unique map $u\colon \mathfrak{S}_{\boldsymbol{i},\boldsymbol{j}} \to \mathfrak{S}_{\boldsymbol{i}',\boldsymbol{j}'}$ such that for each $w \in \mathfrak{S}_{\boldsymbol{i},\boldsymbol{j}}$ the permutation $u(w)$ is balanced and $w(r) < w(t)$ if and only if $u(w)(r') < u(w)(t')$ for each $r, t \in [1, d]$, where $r' = r'_{\boldsymbol{i}}$ and $t' = t'_{\boldsymbol{i}}$ are as in Section 2F. The image of u is exactly the set of all balanced permutations in $\mathfrak{S}_{\boldsymbol{i}',\boldsymbol{j}'}$.

Assume that $w \in \mathfrak{S}_{\boldsymbol{i}',\boldsymbol{j}'}$ is unbalanced. We claim that there exists an index a such that $i'_a \in \bar I_1$ and $w(a) > w(a+1)$. Indeed, let J be the set of indices $a \in [1, \bar d]$ such that $i'_a \in \bar I_1$. As $\boldsymbol{j}'$ is well-ordered, we have $\sum_{a \in J}(w(a+1) - w(a)) = \#J$. As w is unbalanced, not all summands in this sum are equal to 1. Then one of the summands must be negative. Let $a \in J$ be an index such that $w(a) > w(a+1)$. We can assume that the reduced expression of w is of the form $w = s_{p_1} \cdots s_{p_k} s_a$. In this case the element $\tau_w e(\boldsymbol{i}')$ is zero in $S_{\bar\alpha,\boldsymbol{k}}$ because the sequence $s_a(\boldsymbol{i}')$ is unordered.

Assume that $w \in \mathfrak{S}_{\boldsymbol{i}',\boldsymbol{j}'}$ is balanced. Thus, there exists some $\tilde{w} \in \mathfrak{S}_{\boldsymbol{i},\boldsymbol{j}}$ such that $u(\tilde{w}) = w$. We choose an arbitrary reduced expression $\tilde{w} = s_{p_1} \cdots s_{p_k}$ and we choose the reduced expression $w = s_{q_1} \cdots s_{q_r}$ of w obtained from the reduced expression of $\tilde{w}$ in the following way. For $t \in \{1, \dots, k\}$ set $\boldsymbol{i}^t = s_{p_{t+1}} \cdots s_{p_k}(\boldsymbol{i})$ (in particular, we have $\boldsymbol{i}^k = \boldsymbol{i}$). We write $\boldsymbol{i}^t = (i^t_1, \dots, i^t_d)$. We construct the reduced expression of w as $w = \hat{s}_{p_1} \cdots \hat{s}_{p_k}$, where for $a = p_t$ we have

$$\hat{s}_a = \begin{cases} s_{a'} & \text{if } i^t_a, i^t_{a+1} \in I_0, \\ s_{a'+1}s_{a'} & \text{if } i^t_a \in I_0 \text{ and } i^t_{a+1} \in I_1, \\ s_{a'}s_{a'+1} & \text{if } i^t_a \in I_1 \text{ and } i^t_{a+1} \in I_0, \\ s_{a'+1}s_{a'}s_{a'+2}s_{a'+1} & \text{if } i^t_a, i^t_{a+1} \in I_1, \end{cases}$$

where $a' = a'_{\boldsymbol{i}^r}$ is as in Section 2F. Let us explain why the obtained expression of w is reduced. The fact that the expression $\tilde{w} = s_{p_1} \cdots s_{p_k}$ is reduced means the following. When we apply the transpositions $s_{p_k}, s_{p_{k-1}}, \dots, s_{p_1}$ consecutively to the d-tuple $(1, 2, \dots, d)$, if two elements of the set $\{1, 2, \dots, d\}$ are exchanged once by some s, then these two elements are never exchanged again by another s later. It is clear that the expression $w = s_{q_1} \cdots s_{q_r} = \hat{s}_{p_1} \cdots \hat{s}_{p_k}$ inherits the same property from $\tilde{w} = s_{p_1} \cdots s_{p_k}$ because for each $a, b \in \{1, 2, \dots, d\}$, $a \neq b$ we have the following (we set $a' = a'_{\boldsymbol{i}}$ and $b' = b'_{\boldsymbol{i}}$):

- If $i_a, i_b \in I_0$, then if the reduced expression of $\tilde{w}$ exchanges a and b exactly once or never exchanges them then the expression of w exchanges a' and b' exactly once or never exchanges them, respectively.
- If $i_a \in I_0$ and $i_b \in I_1$, then if the reduced expression of $\tilde{w}$ exchanges a and b exactly once or never exchanges them then the expression of w exchanges a' and b' exactly once or never exchanges them, respectively, and it also exchanges a' with $b' + 1$ exactly once or, respectively, never exchanges them.
- If $i_a \in I_1$ and $i_b \in I_0$, then if the reduced expression of $\tilde{w}$ exchanges a and b exactly once or never exchanges them then the expression of w exchanges a' and b' exactly once or never exchanges them, respectively, and it also exchanges $a' + 1$ with b' exactly once or, respectively, never exchanges them.
- If $i_a, i_b \in I_1$, then if the reduced expression of $\tilde{w}$ exchanges a and b exactly once or never exchanges them then the expression of w exchanges a' and b' exactly once or never exchanges them, respectively, and the same thing for a' and $b' + 1$, for $a' + 1$ and b', and for $a' + 1$ and $b' + 1$.

If the reduced expressions are chosen as above, then the element $\tau_w e(\boldsymbol{i}') = \tau_{q_1} \cdots \tau_{q_r} e(\boldsymbol{i}') \in S_{\alpha,\boldsymbol{k}}$ is equal to $\pm(\tau_{p_1} \cdots \tau_{p_k} e(\boldsymbol{i}))^*$.

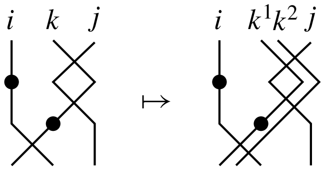
The discussion above shows that the image of an element $b' \in \boldsymbol{B}'$ in $e(\boldsymbol{j}')S_{\bar{\alpha},\boldsymbol{k}}e(\boldsymbol{i}')$ is either zero or of the form $\pm b^*$ for some $b \in \boldsymbol{B}$. Moreover, each b^* for $b \in \boldsymbol{B}$ can be obtained in such a way. Now we get the following:

- The elements $e(\phi(\boldsymbol{i}))$, x^*_r, and τ^*_r generate $S_{\bar{\alpha},\boldsymbol{k}}$ because the image of each element of $\boldsymbol{B}'$ in $e(\boldsymbol{j}')S_{\bar{\alpha},\boldsymbol{k}}e(\boldsymbol{i}')$ is either zero or a monomial in $e(\phi(\boldsymbol{i}))$, x^*_r, and τ^*_r.

- The representation $\boldsymbol{k}^{(\bar{I})}_{\bar{\alpha},\mathrm{ord}}/J_{\bar{\alpha},\mathrm{ord}}$ of $S_{\bar{\alpha},\boldsymbol{k}}$ is faithful because the spanning set $\{b^*: b \in \boldsymbol{B}\}$ of $e(\boldsymbol{j}')S_{\bar{\alpha},\boldsymbol{k}}e(\boldsymbol{i}')$ acts on the polynomial representation by linearly independent operators (because the polynomial representation of $R_{\alpha,\boldsymbol{k}}$ in Proposition 2.3 is faithful). □

Remark 2.13. (a) Note that Theorem 2.12 also remains true for an infinite quiver Γ because α is supported on a finite number of vertices (see also Remark 2.5).

(b) The formulas that define the isomorphism $\Phi_{\alpha,\boldsymbol{k}}$ become more natural if we look at them from the point of view of Khovanov–Lauda diagrams (see [Khovanov and Lauda 2009]). Diagrammatically, the isomorphism $\Phi_{\alpha,\boldsymbol{k}}$ looks in the following way. It sends a diagram representing an element of $R_{\alpha,\boldsymbol{k}}$ to the diagram (sometimes with a sign) obtained by replacing each strand with label $k \in I_1$ by two parallel strands with labels k^1 and k^2 (if there is a dot on the strand with label k, it should be moved to the strand with label k^1). For example, if $i, j \in I_0$ and $k \in I_1$, we have:



3. Categorical representations

3A. ***The standard representation of*** $\widetilde{\mathfrak{sl}}_e$**.** Consider the affine Lie algebra $\widetilde{\mathfrak{sl}}_e = \mathfrak{sl}_e \otimes \mathbb{C}[t, t^{-1}] \oplus \mathbb{C}\mathbf{1}$, defined over $\mathbb{C}$. Let e_i, f_i and h_i for $i = 0, 1, \dots, e-1$, be the standard generators of $\widetilde{\mathfrak{sl}}_e$ (see Remark 2.1). Let V_e be a $\mathbb{C}$-vector space with canonical basis $\{v_1, \dots, v_e\}$ and set $U_e = V_e \otimes \mathbb{C}[z, z^{-1}]$. The vector space U_e has a basis $\{u_r : r \in \mathbb{Z}\}$ where $u_{a+eb} = v_a \otimes z^{-b}$ for $a \in [1, e]$, $b \in \mathbb{Z}$. It has a structure of an $\widetilde{\mathfrak{sl}}_e$-module such that

$$f_i(u_r) = \delta_{i\equiv r} u_{r+1} \quad \text{and} \quad e_i(u_r) = \delta_{i\equiv r-1} u_{r-1}.$$

Let $\{v'_1, \dots, v'_{e+1}\}$ and $\{u'_r : r \in \mathbb{Z}\}$ denote the bases of V_{e+1} and U_{e+1}.

Fix an integer $0 \leqslant k < e$. Consider the following inclusion of vector spaces

$$V_e \subset V_{e+1},\, v_r \mapsto \begin{cases} v'_r & \text{if } r \leqslant k, \\ v'_{r+1} & \text{if } r > k. \end{cases}$$

It yields an inclusion $\mathfrak{sl}_e \subset \mathfrak{sl}_{e+1}$ such that

$$e_r \mapsto \begin{cases} e_r & \text{if } r \in [1, k-1], \\ [e_k, e_{k+1}] & \text{if } r = k, \\ e_{r+1} & \text{if } r \in [k+1, e-1], \end{cases}$$

$$f_r \mapsto \begin{cases} f_r & \text{if } r \in [1, k-1], \\ [f_{k+1}, f_k] & \text{if } r = k, \\ f_{r+1} & \text{if } r \in [k+1, e-1], \end{cases}$$

$$h_r \mapsto \begin{cases} h_r & \text{if } r \in [1, k-1], \\ h_k + h_{k+1} & \text{if } r = k, \\ h_{r+1} & \text{if } r \in [k+1, e-1]. \end{cases}$$

This inclusion lifts uniquely to an inclusion $\widetilde{\mathfrak{sl}}_e \subset \widetilde{\mathfrak{sl}}_{e+1}$ such that

$$e_0 \mapsto \begin{cases} e_0 & \text{if } k \neq 0, \\ [e_0, e_1] & \text{else}, \end{cases}$$

$$f_0 \mapsto \begin{cases} f_0 & \text{if } k \neq 0, \\ [f_1, f_0] & \text{else}, \end{cases}$$

$$h_0 \mapsto \begin{cases} h_0 & \text{if } k \neq 0, \\ h_0 + h_1 & \text{else}. \end{cases}$$

Consider the inclusion $U_e \subset U_{e+1}$ such that $u_r \mapsto u'_{\Upsilon(r)}$, where Υ is defined in (8).

Lemma 3.1. *The embeddings $V_e \subset V_{e+1}$ and $U_e \subset U_{e+1}$ are compatible with the actions of $\mathfrak{sl}_e \subset \mathfrak{sl}_{e+1}$ and $\widetilde{\mathfrak{sl}}_e \subset \widetilde{\mathfrak{sl}}_{e+1}$, respectively.* □

3B. *Type A quivers.* Let $\Gamma_\infty = (I_\infty, H_\infty)$ be the quiver with the set of vertices $I_\infty = \mathbb{Z}$ and the set of arrows $H_\infty = \{i \to i+1 : i \in I_\infty\}$. Assume that $e > 1$ is an integer. Let $\Gamma_e = (I_e, H_e)$ be the quiver with the set of vertices $I_e = \mathbb{Z}/e\mathbb{Z}$ and the set of arrows $H_e = \{i \to i+1 : i \in I_e\}$. Then $\mathfrak{g}_{I_e}$ is the Lie algebra $\widetilde{\mathfrak{sl}}_e = \mathfrak{sl}_e \otimes \mathbb{C}[t, t^{-1}] \oplus \mathbb{C}\mathbf{1}$ (see Remark 2.1).

Assume that $\Gamma = (I, H)$ is a quiver whose connected components are of the form Γ_e, with $e \in \mathbb{N}$, $e > 1$ or $e = \infty$. For $i \in I$ denote by $i+1$ and $i-1$ the (unique) vertices in I such that there are arrows $i \to i+1$ and $i-1 \to i$.

Let X_I be the free abelian group with basis $\{\varepsilon_i : i \in I\}$. Set also

$$X_I^+ = \bigoplus_{i \in I} \mathbb{N}\varepsilon_i. \tag{5}$$

Let us also consider the following additive map

$$\iota\colon Q_I \to X_I, \quad \alpha_i \mapsto \varepsilon_i - \varepsilon_{i+1}.$$

We may omit the symbol ι and write α instead of $\iota(\alpha)$. Let ϕ denote also the unique additive embedding

$$\phi\colon X_I \to X_{\bar{I}}, \quad \varepsilon_i \mapsto \varepsilon_{i'}, \tag{6}$$

where

$$i' = \begin{cases} i^0 & \text{if } i \in I_0, \\ i^1 & \text{if } i \in I_1. \end{cases}$$

3C. *Categorical representations.* Let $\Gamma = (I, H)$ be a quiver as in Section 3B. Let $\boldsymbol{k}$ be a field. Assume that $\mathcal{C}$ is a Hom-finite $\boldsymbol{k}$-linear abelian category.

Definition 3.2. A $\mathfrak{g}_I$-categorical representation (E, F, x, τ) in $\mathcal{C}$ is the following data:

(1) a decomposition $\mathcal{C} = \bigoplus_{\mu \in X_I} \mathcal{C}_\mu$,

(2) a pair of biadjoint exact endofunctors (E, F) of $\mathcal{C}$,

(3) morphisms of functors $x\colon F \to F$ and $\tau\colon F^2 \to F^2$,

(4) decompositions $E = \bigoplus_{i \in I} E_i$ and $F = \bigoplus_{i \in I} F_i$,

satisfying the following conditions:

(a) We have $E_i(\mathcal{C}_\mu) \subset \mathcal{C}_{\mu+\alpha_i}$, $F_i(\mathcal{C}_\mu) \subset \mathcal{C}_{\mu-\alpha_i}$.

(b) For each $d \in \mathbb{N}$ there is an algebra homomorphism $\psi_d\colon R_{d,\boldsymbol{k}} \to \mathrm{End}(F^d)^{\mathrm{op}}$ such that $\psi_d(e(\boldsymbol{i}))$ is the projector to $F_{i_d} \cdots F_{i_1}$, where $\boldsymbol{i} = (i_1, \dots, i_d)$ and

$$\psi_d(x_r) = F^{d-r} x F^{r-1} \quad \text{and} \quad \psi_d(\tau_r) = F^{d-r-1} \tau F^{r-1}.$$

(c) For each $M \in \mathcal{C}$ the endomorphism of $F(M)$ induced by x is nilpotent.

Remark 3.3. (a) For a pair of adjoint functors (E, F) we have an isomorphism $\mathrm{End}(E^d) \simeq \mathrm{End}(F^d)^{\mathrm{op}}$. In particular, the algebra homomorphism $R_{d,\boldsymbol{k}} \to \mathrm{End}(F^d)^{\mathrm{op}}$ in Definition 3.2 yields an algebra homomorphism $R_{d,\boldsymbol{k}} \to \mathrm{End}(E^d)$.

(b) If the quiver Γ is infinite, the direct sums in (4) should be understood in the following way. For each object $M \in \mathcal{C}$, there is only a finite number of $i \in I$ such that $E_i(M)$ and $F_i(M)$ are nonzero.

3D. *From $\widetilde{\mathfrak{sl}}_{e+1}$-categorical representations to $\widetilde{\mathfrak{sl}}_e$-categorical representations.* As in Section 3A, we fix $0 \leqslant k < e$. Only in Section 3D, we assume that $\Gamma = (I, H)$ and $\overline{\Gamma} = (\bar{I}, \overline{H})$ are fixed as in as in Section B2 (i.e., we have $\Gamma = \Gamma_e$, $I_1 = \{k\}$ and we identity $\overline{\Gamma}$ with Γ_{e+1}).

Let $\overline{\mathcal{C}}$ be a Hom-finite abelian $\boldsymbol{k}$-linear category. Let

$$\overline{E} = \overline{E}_0 \oplus \overline{E}_1 \oplus \cdots \oplus \overline{E}_e \quad \text{and} \quad \overline{F} = \overline{F}_0 \oplus \overline{F}_1 \oplus \cdots \oplus \overline{F}_e$$

be endofunctors defining a $\widetilde{\mathfrak{sl}}_{e+1}$-categorical representation in $\overline{\mathcal{C}}$. Let $\overline{\psi}_d\colon R_{d,\boldsymbol{k}} \to \mathrm{End}(\overline{F}^d)^{\mathrm{op}}$ be the corresponding algebra homomorphism. We set $\overline{F}_{\boldsymbol{i}} = \overline{F}_{i_d} \cdots \overline{F}_{i_1}$ for any tuple $\boldsymbol{i} = (i_1, \dots, i_d) \in \bar{I}^d$ and $\overline{F}_{\bar{\alpha}} = \bigoplus_{\boldsymbol{i} \in \bar{I}^{\bar{\alpha}}} \overline{F}_{\boldsymbol{i}}$ for any element $\bar{\alpha} \in Q^+_{\bar{I}}$. If $|\bar{\alpha}| = d$ let $\overline{\psi}_{\bar{\alpha}}\colon R_{\bar{\alpha},\boldsymbol{k}} \to \mathrm{End}(\overline{F}_{\bar{\alpha}})^{\mathrm{op}}$ be the $\bar{\alpha}$-component of $\overline{\psi}_d$.

Now, recall the notation $X^+_{\bar{I}}$ from (5). Assume that we have

$$\overline{\mathcal{C}}_\mu = 0, \quad \forall \mu \in X_{\bar{I}} \backslash X^+_{\bar{I}}. \tag{7}$$

For $\mu \in X^+_I$ set $\mathcal{C}_\mu = \overline{\mathcal{C}}_{\phi(\mu)}$, where the map ϕ is as in (6). Let $\mathcal{C} = \bigoplus_{\mu \in X^+_I} \mathcal{C}_\mu$.

Remark 3.4. (a) $\mathcal{C}$ is stable by $\overline{F}_i$, $\overline{E}_i$ for each $i \neq k, k+1$,

(b) $\mathcal{C}$ is stable by $\overline{F}_{k+1}\overline{F}_k$, $\overline{E}_k\overline{E}_{k+1}$,

(c) $\overline{F}_{i_d}\overline{F}_{i_{d-1}}\cdots\overline{F}_{i_1}(M) = 0$ for each $M \in \mathcal{C}$ whenever the sequence $(i_1, \dots, i_d)$ is unordered (see Section 2D).

Consider the following endofunctors of $\mathcal{C}$:

$$E_i = \begin{cases} \overline{E}_i|_{\mathcal{C}} & \text{if } 0 \leqslant i < k, \\ \overline{E}_k\overline{E}_{k+1}|_{\mathcal{C}} & \text{if } i = k, \\ \overline{E}_{i+1}|_{\mathcal{C}} & \text{if } k < i < e, \end{cases} \quad \text{and} \quad F_i = \begin{cases} \overline{F}_i|_{\mathcal{C}} & \text{if } 0 \leqslant i < k, \\ \overline{F}_{k+1}\overline{F}_k|_{\mathcal{C}} & \text{if } i = k, \\ \overline{F}_{i+1}|_{\mathcal{C}} & \text{if } k < i < e. \end{cases}$$

Similarly to the notations above we set $F_{\boldsymbol{i}} = F_{i_d}\cdots F_{i_1}$ for any tuple $\boldsymbol{i} = (i_1, \dots, i_d) \in I^d$ and $F_\alpha = \bigoplus_{\boldsymbol{i}\in I^\alpha} F_{\boldsymbol{i}}$ for any element $\alpha \in Q_I^+$. Note that we have $F_{\boldsymbol{i}} = \overline{F}_{\phi(\boldsymbol{i})}|_{\mathcal{C}}$ for each $\boldsymbol{i} \in I^\alpha$.

Let $\alpha \in Q_I^+$ and $\bar{\alpha} = \phi(\alpha)$. Note that we have

$$F_\alpha = \bigoplus_{\boldsymbol{i}\in \bar{I}^{\bar{\alpha}}_{\mathrm{ord}}} \overline{F}_{\boldsymbol{i}}|_{\mathcal{C}}.$$

The homomorphism $\overline{\psi}_{\bar{\alpha}}$ yields a homomorphism $\boldsymbol{e}R_{\bar{\alpha},\boldsymbol{k}}\boldsymbol{e} \to \mathrm{End}(F_\alpha)^{\mathrm{op}}$, where $\boldsymbol{e} = \sum_{\boldsymbol{i}\in\bar{I}^{\bar{\alpha}}_{\mathrm{ord}}} e(\boldsymbol{i})$. By (c), the homomorphism $\boldsymbol{e}R_{\bar{\alpha},\boldsymbol{k}}\boldsymbol{e} \to \mathrm{End}(F_\alpha)^{\mathrm{op}}$ factors through a homomorphism $S_{\bar{\alpha},\boldsymbol{k}} \to \mathrm{End}(F_\alpha)^{\mathrm{op}}$. Let us call it $\overline{\psi}'_{\bar{\alpha}}$. Then we can define an algebra homomorphism $\psi_\alpha\colon R_{\alpha,\boldsymbol{k}} \to \mathrm{End}(F_\alpha)^{\mathrm{op}}$ by setting $\psi_\alpha = \overline{\psi}'_{\bar{\alpha}} \circ \Phi_{\alpha,\boldsymbol{k}}$.

Now, Theorem 2.12 implies the following result.

Theorem 3.5. *For each category $\overline{\mathcal{C}}$, defined as above, that satisfies* (7), *we have a categorical representation of $\widetilde{\mathfrak{sl}}_e$ in the subcategory $\mathcal{C}$ of $\overline{\mathcal{C}}$ given by functors F_i and E_i and the algebra homomorphisms $\psi_\alpha\colon R_{\alpha,\boldsymbol{k}} \to \mathrm{End}(F_\alpha)^{\mathrm{op}}$.* □

Now, we describe the example that motivated us to prove Theorem 3.5. See [Maksimau 2015b] for details.

Example 3.6. Let U_e and V_e be as in Section 3A. Fix $\nu = (\nu_1, \dots, \nu_l) \in \mathbb{N}^l$ and put $N = \sum_{r=1}^l \nu_r$. Set $\wedge^\nu U_e = \wedge^{\nu_1} U_e \otimes \cdots \otimes \wedge^{\nu_l} U_e$.

Let O^ν_{-e} be the parabolic category $\mathcal{O}$ for $\widehat{\mathfrak{gl}}_N$ with parabolic type ν at level $-e-N$. The categorical representation of $\widetilde{\mathfrak{sl}}_e$ in O^ν_{-e} (constructed in [Rouquier et al. 2016]) yields an $\widetilde{\mathfrak{sl}}_e$-module structure on the (complexified) Grothendieck group $[O^\nu_{-e}]$ of O^ν_{-e}. This module is isomorphic to $\wedge^\nu U_e$.

Let us apply Theorem 1.1 to $\overline{\mathcal{C}} = O^\nu_{-(e+1)}$. It happens that in this case the subcategory $\mathcal{C} \subset \overline{\mathcal{C}}$ defined as above is equivalent to O^ν_{-e}. The embedding of categories $O^\nu_{-e} \subset O^\nu_{-(e+1)}$ categorifies the embedding $\wedge^\nu U_e \subset \wedge^\nu U_{e+1}$ (see also Lemma 3.1).

3E. ***Reduction of the number of idempotents.*** In this section we show that it is possible to reduce the number of idempotents in the quotient in Definition 2.6. This is necessary to generalize Theorem 3.5. Here we assume the quivers $\Gamma = (I, H)$ and $\overline{\Gamma} = (\bar{I}, \overline{H})$ are as in Section 2B.

We fix $\alpha \in Q_I^+$ and put $\bar{\alpha} = \phi(\alpha)$. We say that the sequence $\boldsymbol{i} \in \bar{I}^{\bar{\alpha}}$ is *almost ordered* if there exists a well-ordered sequence $\boldsymbol{j} \in \bar{I}^{\bar{\alpha}}$ such that there exists an index r such that $j_r \in \bar{I}_1$ and $\boldsymbol{i} = s_r(\boldsymbol{j})$. It is clear from the definition that each almost ordered sequence is unordered because the subsequence $(i_1, i_2, \dots, i_r)$ of $\boldsymbol{i}$ contains more elements from $\bar{I}_2$ than from $\bar{I}_1$. The following lemma reduces the number of generators of the kernel of $\boldsymbol{e}R_{\bar{\alpha},\boldsymbol{k}}\boldsymbol{e} \to S_{\bar{\alpha},\boldsymbol{k}}$ (see Definition 2.6).

Lemma 3.7. *The kernel of the homomorphism $\boldsymbol{e}R_{\bar{\alpha},\boldsymbol{k}}\boldsymbol{e} \to S_{\bar{\alpha},\boldsymbol{k}}$ is equal to $\sum_{\boldsymbol{i}} \boldsymbol{e}R_{\bar{\alpha},\boldsymbol{k}}e(\boldsymbol{i})R_{\bar{\alpha},\boldsymbol{k}}\boldsymbol{e}$, where $\boldsymbol{i}$ runs over the set of all almost ordered sequences in $\bar{I}^{\bar{\alpha}}$.*

Proof. Denote by J the ideal $\sum_{\boldsymbol{i}} \boldsymbol{e}R_{\bar{\alpha},\boldsymbol{k}}e(\boldsymbol{i})R_{\bar{\alpha},\boldsymbol{k}}\boldsymbol{e}$ of $\boldsymbol{e}R_{\bar{\alpha},\boldsymbol{k}}\boldsymbol{e}$, where $\boldsymbol{i}$ runs over the set of all almost ordered sequences in $\bar{I}^{\bar{\alpha}}$.

By definition, each element of the kernel of $\boldsymbol{e}R_{\bar{\alpha},\boldsymbol{k}}\boldsymbol{e} \to S_{\bar{\alpha},\boldsymbol{k}}$ is a linear combination of elements of the form $\boldsymbol{e}ae(\boldsymbol{j})b\boldsymbol{e}$, where a and b are in $R_{\bar{\alpha},\boldsymbol{k}}$ and the sequence $\boldsymbol{j}$ is unordered. By Remark 2.4, it is enough to prove that for each $\boldsymbol{i} \in \bar{I}^{\bar{\alpha}}_{\mathrm{ord}}$, $\boldsymbol{j} \in \bar{I}^{\bar{\alpha}}_{\mathrm{un}}$, $b \in R_{\bar{\alpha},\boldsymbol{k}}$ and indices $p_1, \dots, p_k$ the element $e(\boldsymbol{i})\tau_{p_1} \cdots \tau_{p_k}e(\boldsymbol{j})b\boldsymbol{e}$ is in J. We will prove this statement by induction on k.

Assume that $k = 1$. Write $p = p_1$. The element $e(\boldsymbol{i})\tau_p e(\boldsymbol{j})b\boldsymbol{e}$ may be nonzero only if $\boldsymbol{i} = s_p(\boldsymbol{j})$. This is possible only if the sequence $\boldsymbol{j}$ is almost ordered. Thus the element $e(\boldsymbol{i})\tau_p e(\boldsymbol{j})b\boldsymbol{e}$ is in J.

Now, assume that $k > 1$ and that the statement is true for each value $< k$. Set $w = s_{p_1} \cdots s_{p_k}$. We may assume that $\boldsymbol{i} = w(\boldsymbol{j})$, otherwise the element $e(\boldsymbol{i})\tau_{p_1} \cdots \tau_{p_k}e(\boldsymbol{j})b\boldsymbol{e}$ is zero. By assumptions on $\boldsymbol{i}$ and $\boldsymbol{j}$ there is an index $r \in [1, d]$ such that $i_r \in \bar{I}_1$ and $w^{-1}(r+1) < w^{-1}(r)$. Thus w has a reduced expression of the form $w = s_r s_{r_1} \cdots s_{r_h}$. This implies that $\tau_{p_1} \cdots \tau_{p_k}e(\boldsymbol{j})$ is equal to a monomial of the form $\tau_r\tau_{r_1} \cdots \tau_{r_h}e(\boldsymbol{j})$ modulo monomials of the form $\tau_{q_1} \cdots \tau_{q_t}x_1^{b_1} \cdots x_d^{b_d}e(\boldsymbol{j})$ with $t < k$, see Remark 2.4. Thus the element $e(\boldsymbol{i})\tau_1 \cdots \tau_k e(\boldsymbol{j})b\boldsymbol{e}$ is equal to $e(\boldsymbol{i})\tau_r\tau_{r_1} \cdots \tau_{r_h}e(\boldsymbol{j})b\boldsymbol{e}$ modulo the elements of the same form $e(\boldsymbol{i})\tau_{p_1} \cdots \tau_{p_k}e(\boldsymbol{j})b\boldsymbol{e}$ with smaller k. The element $e(\boldsymbol{i})\tau_r\tau_{r_1} \cdots \tau_{r_h}e(\boldsymbol{j})b\boldsymbol{e}$ is in J because the sequence $s_r(\boldsymbol{i})$ is almost ordered and the additional terms are in J by the induction assumption. □

3F. *Generalization of Theorem 3.5.* In this section we modify slightly the definition of a categorical representation given in Definition 3.2. The only difference is that we use the lattice Q_I instead of X_I. This new definition is not equivalent to Definition 3.2. In this section we work with an arbitrary quiver $\Gamma = (I, H)$ without 1-loops.

Let $\boldsymbol{k}$ be a field. Let $\mathcal{C}$ be a $\boldsymbol{k}$-linear Hom-finite category.

Definition 3.8. A $\mathfrak{g}_I$-quasicategorical representation (E, F, x, τ) in $\mathcal{C}$ is the following data

(1) a decomposition $\mathcal{C} = \bigoplus_{\alpha \in Q_I} \mathcal{C}_\alpha$,

(2) a pair of biadjoint exact endofunctors (E, F) of $\mathcal{C}$,

(3) morphisms of functors $x\colon F \to F$, $\tau\colon F^2 \to F^2$,

(4) decompositions $E = \bigoplus_{i \in I} E_i$, $F = \bigoplus_{i \in I} F_i$,

satisfying the following conditions.

(a) We have $E_i(\mathcal{C}_\alpha) \subset \mathcal{C}_{\alpha-\alpha_i}$, $F_i(\mathcal{C}_\alpha) \subset \mathcal{C}_{\alpha+\alpha_i}$.

(b) For each $d \in \mathbb{N}$ there is an algebra homomorphism $\psi_d \colon R_{d,\boldsymbol{k}} \to \operatorname{End}(F^d)^{\mathrm{op}}$ such that $\psi_d(e(\boldsymbol{i}))$ is the projector to $F_{i_d} \cdots F_{i_1}$, where $\boldsymbol{i} = (i_1, \ldots, i_d)$ and
$$\psi_d(x_r) = F^{d-r} x F^{r-1} \quad \text{and} \quad \psi_d(\tau_r) = F^{d-r-1} \tau F^{r-1}.$$

(c) For each $M \in \mathcal{C}$ the endomorphism of $F(M)$ induced by x is nilpotent.

If the quiver Γ is infinite, condition (4) should be understood in the same way as in Remark 3.3(b).

Now, fix a decomposition $I = I_0 \sqcup I_1$ as in Section 2B. We consider the quiver $\overline{\Gamma} = (\bar{I}, \overline{H})$ and the map ϕ as in Section 2B. To distinguish the elements of Q_I and $Q_{\bar{I}}$, we write $Q_{\bar{I}} = \bigoplus_{i \in \bar{I}} \mathbb{Z}\bar{\alpha}_i$. For each $\alpha \in Q_I$ we set $\bar{\alpha} = \phi(\alpha) \in Q_{\bar{I}}$. (See Section 2B for the notation.) However we can sometimes use the symbol $\bar{\alpha}$ for an arbitrary element of $Q_{\bar{I}}$ that is not associated with some α in Q_I. Let $\overline{\mathcal{C}}$ be a Hom-finite abelian $\boldsymbol{k}$-linear category. Let $\overline{E} = \bigoplus_{i \in \bar{I}} \overline{E}_i$ and $\overline{F} = \bigoplus_{i \in \bar{I}} \overline{F}_i$ be endofunctors defining a $\mathfrak{g}_{\bar{I}}$-quasicategorical representation in $\overline{\mathcal{C}}$. Let $\overline{\psi}_d \colon R_{d,\boldsymbol{k}}(\overline{\Gamma}) \to \operatorname{End}(\overline{F}^d)^{\mathrm{op}}$ be the corresponding algebra homomorphism. We set $\overline{F}_{\boldsymbol{i}} = \overline{F}_{i_d} \cdots \overline{F}_{i_1}$ for any tuple $\boldsymbol{i} = (i_1, \ldots, i_d) \in \bar{I}^d$ and $\overline{F}_{\bar{\alpha}} = \bigoplus_{\boldsymbol{i} \in \bar{I}^{\bar{\alpha}}} \overline{F}_{\boldsymbol{i}}$ for any element $\bar{\alpha} \in Q_{\bar{I}}^+$. If $|\bar{\alpha}| = d$, let $\overline{\psi}_{\bar{\alpha}} \colon R_{\bar{\alpha},\boldsymbol{k}} \to \operatorname{End}(\overline{F}_{\bar{\alpha}})^{\mathrm{op}}$ be the $\bar{\alpha}$-component of $\overline{\psi}_d$.

Assume that $\mathcal{C}$ is an abelian subcategory of $\overline{\mathcal{C}}$ satisfying the following conditions:

(a) $\mathcal{C}$ is stable by $\overline{F}_i$ and $\overline{E}_i$ for each $i \in I_0$.

(b) $\mathcal{C}$ is stable by $\overline{F}_{i^2}\overline{F}_{i^1}$ and $\overline{E}_{i^1}\overline{E}_{i^2}$ for each $i \in I_1$.

(c) We have $\overline{F}_{i^2}(\mathcal{C}) = 0$ for each $i \in I_1$.

(d) We have $\mathcal{C} = \bigoplus_{\alpha \in Q_I} \mathcal{C} \cap \overline{\mathcal{C}}_{\bar{\alpha}}$.

By (d), we get a decomposition $\mathcal{C} = \bigoplus_{\alpha \in Q_I} \mathcal{C}_\alpha$, where $\mathcal{C}_\alpha = \mathcal{C} \cap \overline{\mathcal{C}}_{\bar{\alpha}}$. For each $i \in I$ we consider the following endofunctors E_i and F_i of $\mathcal{C}$:
$$F_i = \begin{cases} \overline{F}_i|_{\mathcal{C}} & \text{if } i \in I_0, \\ \overline{F}_{i^2}\overline{F}_{i^1}|_{\mathcal{C}} & \text{if } i \in I_1, \end{cases} \quad \text{and} \quad E_i = \begin{cases} \overline{E}_i|_{\mathcal{C}} & \text{if } i \in I_0, \\ \overline{E}_{i^1}\overline{E}_{i^2}|_{\mathcal{C}} & \text{if } i \in I_1. \end{cases}$$

As in the notations above we set $F_{\boldsymbol{i}} = F_{i_d} \cdots F_{i_1}$ for any tuple $\boldsymbol{i} = (i_1, \ldots, i_d) \in I^d$ and $F_\alpha = \bigoplus_{\boldsymbol{i} \in I^\alpha} F_{\boldsymbol{i}}$ for any element $\alpha \in Q_I^+$. Note that we have $F_{\boldsymbol{i}} = \overline{F}_{\phi(\boldsymbol{i})}|_{\mathcal{C}}$ for each $\boldsymbol{i} \in I^\alpha$.

Let $\alpha \in Q_I^+$. We have
$$F_\alpha = \bigoplus_{\boldsymbol{i} \in \bar{I}^{\bar{\alpha}}_{\mathrm{ord}}} \overline{F}_{\boldsymbol{i}}|_{\mathcal{C}}.$$

The homomorphism $\overline{\psi}_{\bar{\alpha}}$ yields a homomorphism $\boldsymbol{e}R_{\bar{\alpha},\boldsymbol{k}}\boldsymbol{e} \to \operatorname{End}(F_\alpha)^{\mathrm{op}}$, where $\boldsymbol{e} = \sum_{\boldsymbol{i} \in \bar{I}^{\bar{\alpha}}_{\mathrm{ord}}} e(\boldsymbol{i})$.

Since the category $\mathcal{C}$ satisfies (a), (b) and (c), for each almost ordered sequence $\boldsymbol{i} = (i_1, \ldots, i_d) \in I^\alpha$ we have $\overline{F}_{i_d} \cdots \overline{F}_{i_1}(\mathcal{C}) = 0$. By Lemma 3.7, this implies that the homomorphism $\boldsymbol{e}R_{\bar{\alpha},\boldsymbol{k}}\boldsymbol{e} \to \operatorname{End}(F_\alpha)^{\mathrm{op}}$ factors through a homomorphism $S_{\bar{\alpha},\boldsymbol{k}} \to \operatorname{End}(F_\alpha)^{\mathrm{op}}$. Let us call it $\overline{\psi}'_{\bar{\alpha}}$. Then we can define an algebra homomorphism $\psi_\alpha \colon R_{\alpha,\boldsymbol{k}} \to \operatorname{End}(F_\alpha)^{\mathrm{op}}$ by setting $\psi_\alpha = \overline{\psi}'_{\bar{\alpha}} \circ \Phi_{\alpha,\boldsymbol{k}}$.

Now, Theorem 2.12 implies the following result.

Theorem 3.9. *For each abelian subcategory $\mathcal{C} \subset \bar{\mathcal{C}}$ as above, that satisfies (a)–(d), we have a $\mathfrak{g}_I$-quasicategorical representation in $\mathcal{C}$ given by functors F_i and E_i and the algebra homomorphisms* $\psi_\alpha \colon R_{\alpha,\boldsymbol{k}} \to \mathrm{End}(F_\alpha)^{\mathrm{op}}$. □

Remark 3.10. Assume that the category $\bar{\mathcal{C}}$ is such that we have $\bar{\mathcal{C}}_{\bar{\alpha}} = 0$ whenever $\bar{\alpha} = \sum_{i \in \bar{I}} d_i \bar{\alpha}_i \in Q_{\bar{I}}$ is such that $d_{i^1} < d_{i^2}$ for some $i \in I_1$. In this case the subcategory $\mathcal{C} \subset \bar{\mathcal{C}}$ defined by $\mathcal{C} = \bigoplus_{\alpha \in Q_I} \bar{\mathcal{C}}_{\bar{\alpha}}$ satisfies conditions (a)–(d).

Appendix A: The geometric construction of the isomorphism Φ

The goal of this section is to give a geometric construction of the isomorphism Φ in Theorem 2.12.

A1. ***The geometric construction of the KLR algebra.*** Let $\boldsymbol{k}$ be a field. Let $\Gamma = (I, H)$ be a quiver without 1-loops. See Section 2A for the notations related to quivers. For an arrow $h \in H$ we will write h' and h'' for its source and target respectively. Fix $\alpha = \sum_{i \in I} d_i \alpha_i \in Q_I^+$ and set $d = |\alpha|$. Set also

$$E_\alpha = \bigoplus_{h \in H} \mathrm{Hom}(V_{h'}, V_{h''}), \qquad V_i = \mathbb{C}^{d_i}, \qquad V = \bigoplus_{i \in I} V_i.$$

The group $G_\alpha = \prod_{i \in I} \mathrm{GL}(V_i)$ acts on E_α by base changes.

Set

$$I^\alpha = \left\{ \boldsymbol{i} = (i_1, \dots, i_d) \in I^d : \sum_{r=1}^{d} \alpha_{i_r} = \alpha \right\}.$$

We denote by $F_{\boldsymbol{i}}$ the variety of all flags

$$\phi = (V = V^0 \supset V^1 \supset \cdots \supset V^d = \{0\})$$

in V that are homogeneous with respect to the decomposition $V = \bigoplus_{i \in I} V_i$ and such that the I-graded vector space V^{r-1}/V^r has graded dimension i_r for $r \in [1, d]$. We denote by $\tilde{F}_{\boldsymbol{i}}$ the variety of pairs $(x, \phi) \in E_\alpha \times F_{\boldsymbol{i}}$ such that x preserves ϕ, i.e., we have $x(V^r) \subset V^r$ for $r \in \{0, 1, \dots, m\}$. Let $\pi_{\boldsymbol{i}}$ be the natural projection from $\tilde{F}_{\boldsymbol{i}}$ to E_α, i.e., $\pi_{\boldsymbol{i}} : \tilde{F}_{\boldsymbol{i}} \to E_\alpha, (x, \phi) \mapsto x$. For $\boldsymbol{i}, \boldsymbol{j} \in I^\alpha$ we denote by $Z_{\boldsymbol{i},\boldsymbol{j}}$ the variety of triples $(x, \phi_1, \phi_2) \in E_\alpha \times F_{\boldsymbol{i}} \times F_{\boldsymbol{j}}$ such that x preserves ϕ_1 and ϕ_2 (i.e., we have $Z_{\boldsymbol{i},\boldsymbol{j}} = \tilde{F}_{\boldsymbol{i}} \times_{E_\alpha} \tilde{F}_{\boldsymbol{j}}$). Set

$$Z_\alpha = \coprod_{\boldsymbol{i},\boldsymbol{j} \in I^\alpha} Z_{\boldsymbol{i},\boldsymbol{j}} \quad \text{and} \quad \tilde{F}_\alpha = \coprod_{\boldsymbol{i} \in I^\alpha} \tilde{F}_{\boldsymbol{i}}.$$

We have an algebra structure on $H_*^{G_\alpha}(Z_\alpha, \boldsymbol{k})$ such that the multiplication is the convolution product with respect to the inclusion $Z_\alpha \subset \tilde{F}_\alpha \times \tilde{F}_\alpha$. Here $H_*^{G_\alpha}(\bullet, \boldsymbol{k})$ denotes the G_α-equivariant Borel–Moore homology with coefficients in $\boldsymbol{k}$. See [Chriss and Ginzburg 1997, §2.7] for the definition of the convolution product.

The following result is proved by Rouquier [2008] and by Varagnolo and Vasserot [2011] in the situation $\mathrm{char}\, \boldsymbol{k} = 0$. See [Maksimau 2015a] for the proof over an arbitrary field.

Proposition A.1. *There is an algebra isomorphism $R_{\alpha,\boldsymbol{k}}(\Gamma) \simeq H_*^{G_\alpha}(Z_\alpha, \boldsymbol{k})$. Moreover, for each $\boldsymbol{i}, \boldsymbol{j} \in I^\alpha$, the vector subspace $e(\boldsymbol{i})R_{\alpha,\boldsymbol{k}}(\Gamma)e(\boldsymbol{j}) \subset R_{\alpha,\boldsymbol{k}}(\Gamma)$ corresponds to the vector subspace $H_*^{G_\alpha}(Z_{\boldsymbol{i},\boldsymbol{j}}, \boldsymbol{k}) \subset H_*^{G_\alpha}(Z_\alpha, \boldsymbol{k})$.* □

A2. ***The geometric construction of the isomorphism* Φ.** As in Section 2B, fix a decomposition $I = I_0 \sqcup I_1$ and consider the quiver $\overline{\Gamma} = (\bar{I}, \overline{H})$; also fix $\alpha \in Q_I^+$ and consider $\bar{\alpha} = \phi(\alpha) \in Q_{\bar{I}}^+$.

We start from the variety $Z_{\bar{\alpha}}$ defined with respect to the quiver $\overline{\Gamma}$. By Proposition A.1, we have an algebra isomorphism $R_{\bar{\alpha},\boldsymbol{k}}(\overline{\Gamma}) \simeq H_*^{G_{\bar{\alpha}}}(Z_{\bar{\alpha}}, \boldsymbol{k})$. We have an obvious projection $p\colon Z_{\bar{\alpha}} \to E_{\bar{\alpha}}$ defined by $(x, \phi_1, \phi_2) \mapsto x$. For each $i \in I_1$ denote by h_i the unique arrow in $\overline{\Gamma}$ that goes from i^1 to i^2. Consider the following open subset of $E_{\bar{\alpha}}$: $E_{\bar{\alpha}}^0 = \{x \in E_{\bar{\alpha}} : x_{h_i} \text{ is invertible } \forall i \in I_1\}$. Set $Z_{\bar{\alpha}}^0 = p^{-1}(E_{\bar{\alpha}}^0)$. The pullback with respect to the inclusion $Z_{\bar{\alpha}}^0 \subset Z_{\bar{\alpha}}$ yields an algebra homomorphism $H_*^{G_{\bar{\alpha}}}(Z_{\bar{\alpha}}, \boldsymbol{k}) \to H_*^{G_{\bar{\alpha}}}(Z_{\bar{\alpha}}^0, \boldsymbol{k})$ (see [Chriss and Ginzburg 1997, Lemma 2.7.46]).

Remark A.2. If the sequence $\boldsymbol{i} \in \bar{I}^{\bar{\alpha}}$ is unordered, then a flag from $F_{\boldsymbol{i}}$ is never preserved by an element from $E_{\bar{\alpha}}^0$. This implies that $Z_{\boldsymbol{i},\boldsymbol{j}} \cap Z_{\bar{\alpha}}^0 = \varnothing$ if $\boldsymbol{i}$ or $\boldsymbol{j}$ is unordered. Thus for each $\boldsymbol{i} \in \bar{I}_{\mathrm{un}}^{\bar{\alpha}}$, the idempotent $e(\boldsymbol{i})$ is in the kernel of the homomorphism $H_*^{G_{\bar{\alpha}}}(Z_{\bar{\alpha}}, \boldsymbol{k}) \to H_*^{G_{\bar{\alpha}}}(Z_{\bar{\alpha}}^0, \boldsymbol{k})$.

Let $\boldsymbol{e}$ be the idempotent as in Definition 2.6. Consider the following subset of $Z_{\bar{\alpha}}$:

$$Z'_{\bar{\alpha}} = \coprod_{\boldsymbol{i},\boldsymbol{j} \in \bar{I}_{\mathrm{ord}}^{\bar{\alpha}}} Z_{\boldsymbol{i},\boldsymbol{j}}.$$

The algebra isomorphism $R_{\bar{\alpha},\boldsymbol{k}}(\overline{\Gamma}) \simeq H_*^{G_{\bar{\alpha}}}(Z_{\bar{\alpha}}, \boldsymbol{k})$ above restricts to an algebra isomorphism $\boldsymbol{e}R_{\bar{\alpha}}(\overline{\Gamma})\boldsymbol{e} \simeq H_*^{G_{\bar{\alpha}}}(Z'_{\bar{\alpha}}, \boldsymbol{k})$.

Now, set $Z'^0_\alpha = Z'_\alpha \cap Z^0_\alpha$. Similarly to the construction above, we have an algebra homomorphism $H_*^{G_{\bar{\alpha}}}(Z'_{\bar{\alpha}}, \boldsymbol{k}) \to H_*^{G_{\bar{\alpha}}}(Z'^0_{\bar{\alpha}}, \boldsymbol{k})$. By Remark A.2, the kernel of this homomorphism contains the kernel of $\boldsymbol{e}R_{\bar{\alpha},\boldsymbol{k}}(\overline{\Gamma})\boldsymbol{e} \to R_{\alpha,\boldsymbol{k}}(\Gamma)$ (see Theorem 2.12). The following result implies that these kernels are the same.

Lemma A.3. *We have the following algebra isomorphism $R_{\alpha,\boldsymbol{k}}(\Gamma) \simeq H_*^{G_{\bar{\alpha}}}(Z'^0_{\bar{\alpha}}, \boldsymbol{k})$.*

Proof. For each $i \in I_0$ we identify $V_i \simeq V_{i^0}$. For each $i \in I_1$ we identify $V_i \simeq V_{i^1} \simeq V_{i^2}$. We have a diagonal inclusion $G_\alpha \subset G_{\bar{\alpha}}$, i.e., the component $\mathrm{GL}(V_i)$ of G_α with $i \in I_0$ goes to $\mathrm{GL}(V_{i^0})$ and the component $\mathrm{GL}(V_i)$ with $i \in I_1$ goes diagonally to $\mathrm{GL}(V_{i^1}) \times \mathrm{GL}(V_{i^2})$.

Set $G_\alpha^{\mathrm{bis}} = \prod_{i \in I_1} \mathrm{GL}(V_{i^2}) \subset G_{\bar{\alpha}}$. We have an obvious group isomorphism $G_{\bar{\alpha}}/G_\alpha^{\mathrm{bis}} \simeq G_\alpha$.

Let us denote by X the choice of isomorphisms $V_{i^1} \simeq V_{i^2}$ mentioned above. Let $E_{\bar{\alpha}}^X$ be the subset of $E_{\bar{\alpha}}$ that contains only $x \in E_{\bar{\alpha}}$ such that for each $i \in I_1$ the component x_{h_i} is the isomorphism chosen in X.

The group G_α^{bis} acts freely on $E_{\bar{\alpha}}^0$ such that each orbit intersects $E_{\bar{\alpha}}^X$ once. This implies that we have an isomorphism of algebraic varieties $E_{\bar{\alpha}}^0/G_\alpha^{\mathrm{bis}} \simeq E_{\bar{\alpha}}^X$. Now, set $Z'^X_{\bar{\alpha}} = p^{-1}(E_{\bar{\alpha}}^X)$. The same argument as above yields $Z'^0_{\bar{\alpha}}/G_\alpha^{\mathrm{bis}} \simeq Z'^X_{\bar{\alpha}}$. We get the following chain of algebra isomorphisms

$$H_*^{G_{\bar{\alpha}}}(Z'^0_{\bar{\alpha}}, \boldsymbol{k}) \simeq H_*^{G_{\bar{\alpha}}/G_\alpha^{\mathrm{bis}}}(Z'^0_{\bar{\alpha}}/G_\alpha^{\mathrm{bis}}, \boldsymbol{k}) \simeq H_*^{G_\alpha}(Z'^X_{\bar{\alpha}}, \boldsymbol{k}).$$

To complete the proof we have to show that the G_α-variety $Z'^X_{\bar\alpha}$ is isomorphic to Z_α. Each element of $I^{\bar\alpha}_{\mathrm{ord}}$ is of the form $\phi(\boldsymbol{i})$ for a unique $\boldsymbol{i}\in I^\alpha$, where ϕ is as in Section 2B. Let us abbreviate $\boldsymbol{i}'=\phi(\boldsymbol{i})$. By definition we have

$$Z'_{\bar\alpha}=\coprod_{\boldsymbol{i},\boldsymbol{j}\in I^\alpha} Z_{\boldsymbol{i}',\boldsymbol{j}'}.$$

Set $Z^X_{\boldsymbol{i}',\boldsymbol{j}'}=Z_{\boldsymbol{i}',\boldsymbol{j}'}\cap Z'^X_{\bar\alpha}$. We have an obvious isomorphism of G_α-varieties $Z^X_{\boldsymbol{i}',\boldsymbol{j}'}\simeq Z_{\boldsymbol{i},\boldsymbol{j}}$. (Beware, the variety $Z_{\boldsymbol{i},\boldsymbol{j}}$ is defined with respect to the quiver Γ and the variety $Z_{\boldsymbol{i}',\boldsymbol{j}'}$ is defined with respect to the quiver $\overline{\Gamma}$.) Taking the union for all $\boldsymbol{i},\boldsymbol{j}\in I^\alpha$ yields an isomorphism of G_α-varieties $Z'^X_{\bar\alpha}\simeq Z_\alpha$. □

Corollary A.4. *We have the following commutative diagram.*

$$\begin{tikzcd}
\boldsymbol{e}R_{\bar\alpha,\boldsymbol{k}}(\overline{\Gamma})\boldsymbol{e} \arrow[r] \arrow[d] & R_{\alpha,\boldsymbol{k}}(\Gamma) \arrow[d] \\
H^{G_{\bar\alpha}}_*(Z'_{\bar\alpha},\boldsymbol{k}) \arrow[r] & H^{G_{\bar\alpha}}_*(Z'^0_{\bar\alpha},\boldsymbol{k}).
\end{tikzcd}$$

Here the left vertical map is the isomorphism from Proposition A.1, the right vertical map is the isomorphism from Lemma A.3, the top horizontal map is obtained from Theorem 2.12 and the bottom horizontal map is the pullback with respect to the inclusion $Z'^0_{\bar\alpha}\subset Z'_{\bar\alpha}$.

Proof. The result follows directly from Lemma A.3. The commutativity of the diagram is easy to see on the generators of $R_{\bar\alpha,\boldsymbol{k}}(\overline{\Gamma})$.

Indeed, the isomorphism $R_{\alpha,\boldsymbol{k}}\simeq H^{G_\alpha}_*(Z_\alpha,\boldsymbol{k})$ is defined in the following way (see [Maksimau 2015a, §2.9, Theorem 2.4] for more details). The element $e(\boldsymbol{i})$ corresponds to the fundamental class $[Z_{\boldsymbol{i},\boldsymbol{i}}]$. The element $x_re(\boldsymbol{i})$ corresponds to the first Chern class of some line bundle on $Z_{\boldsymbol{i},\boldsymbol{i}}$. The element $\psi_re(\boldsymbol{i})$ corresponds to the fundamental class of some correspondence in $Z_{s_r(i),\boldsymbol{i}}$. The commutativity of the diagram in the statement follows from standard properties of Chern classes and fundamental classes. □

Appendix B: A local ring version in type A

In this appendix we give some versions of the main results of the paper (Theorems 2.12 and 3.5) over a local ring. These ring versions are interesting because the study of the category $\mathcal{O}$ for $\widehat{\mathfrak{gl}}_N$ in [Maksimau 2015b] uses a deformation argument. For this we need a version of Theorem 1.2 over a local ring.

It is known that the affine Hecke algebra over a field is related with the KLR algebra (see Propositions B.5, B.6). This allows to reformulate the definition of a categorical representation (see Definition 3.2) that is given in term of KLR algebras in an equivalent way in terms of Hecke algebras (see Definition B.14). The main difficulty is that there is no known relation between Hecke and KLR algebras over a ring. Over a local ring, we can give a definition of a categorical representation using the Hecke algebra (see Definition B.17). But we have no equivalent definition in terms of KLR algebras. That is why, Proposition B.12, that is a ring analogue of Theorem 2.12, is formulated in terms of Hecke algebras and not in terms of KLR algebras.

B1. ***Intertwining operators.*** The center of the algebra $R_{\alpha,\boldsymbol{k}}$ is the ring of symmetric polynomials $\boldsymbol{k}_d[x]^{\mathfrak{S}_d}$, see [Rouquier 2008, Proposition 3.9]. Thus $S_{\bar{\alpha},\boldsymbol{k}}$ is a $\boldsymbol{k}_d[x]^{\mathfrak{S}_d}$-algebra under the isomorphism $\Phi_{\alpha,\boldsymbol{k}}$ in Section 2G. Let Σ be the polynomial $\prod_{a<b}(x_a-x_b)^2 \in \boldsymbol{k}_d[x]^{\mathfrak{S}_d}$. Let $R_{\alpha,\boldsymbol{k}}[\Sigma^{-1}]$ and $S_{\bar{\alpha},\boldsymbol{k}}[\Sigma^{-1}]$ be the rings of quotients of $R_{\alpha,\boldsymbol{k}}$ and $S_{\bar{\alpha},\boldsymbol{k}}$ obtained by inverting Σ. We can extend the isomorphism $\Phi_{\alpha,\boldsymbol{k}}$ from Theorem 2.12 to an algebra isomorphism

$$\Phi_{\alpha,\boldsymbol{k}}\colon R_{\alpha,\boldsymbol{k}}[\Sigma^{-1}] \to S_{\bar{\alpha},\boldsymbol{k}}[\Sigma^{-1}].$$

Assume that the connected components of the quiver Γ are of the form Γ_a for $a \in \mathbb{N}$, $a > 1$ or $a = \infty$. (The quiver Γ_a is defined in Section 3B.)

Note that there is an action of the symmetric group $\mathfrak{S}_d$ on $\boldsymbol{k}_d^{(I)}$ permuting the variables and the components of $\boldsymbol{i}$. Consider the following element in $R_{\alpha,\boldsymbol{k}}[\Sigma^{-1}]$:

$$\Psi_r e(\boldsymbol{i}) = \begin{cases} ((x_r - x_{r+1})\tau_r + 1)e(\boldsymbol{i}) & \text{if } i_{r+1} = i_r, \\ -(x_r - x_{r+1})^{-1}\tau_r e(\boldsymbol{i}) & \text{if } i_{r+1} = i_r - 1, \\ \tau_r e(\boldsymbol{i}) & \text{else.} \end{cases}$$

The element $\Psi_r e(\boldsymbol{i})$ is called *intertwining operator*. Using the formulas (3) we can check that $\Psi_r e(\boldsymbol{i})$ still acts on the polynomial representation and the corresponding operator is equal to $s_r e(\boldsymbol{i})$. Note also that $\tilde{\Psi}_r = (x_r - x_{r+1})\Psi_r$ is an element of $R_{\alpha,\boldsymbol{k}}$.

Lemma B.1. *The images of intertwining operators by $\Phi_{\alpha,\boldsymbol{k}}\colon R_{\alpha,\boldsymbol{k}} \to S_{\bar{\alpha},\boldsymbol{k}}$ can be described in the following way. For $\boldsymbol{i} \in I^\alpha$ such that $i_r - 1 \neq i_{r+1}$ we have*

$$\Phi_{\alpha,\boldsymbol{k}}(\Psi_r e(\boldsymbol{i})) = \begin{cases} \Psi_{r'} e(\phi(\boldsymbol{i})) & \text{if } i_r, i_{r+1} \in I_0, \\ \Psi_{r'}\Psi_{r'+1} e(\phi(\boldsymbol{i})) & \text{if } i_r \in I_1, i_{r+1} \in I_0, \\ \Psi_{r'+1}\Psi_{r'} e(\phi(\boldsymbol{i})) & \text{if } i_r \in I_0, i_{r+1} \in I_1, \\ \Psi_{r'+1}\Psi_{r'+2}\Psi_{r'}\Psi_{r'+1} e(\phi(\boldsymbol{i})) & \text{if } i_r, i_{r+1} \in I_1. \end{cases}$$

For $\boldsymbol{i} \in I^\alpha$ such that $i_r - 1 = i_{r+1}$ we have

$$\Phi_{\alpha,\boldsymbol{k}}(\tilde{\Psi}_r e(\boldsymbol{i})) = \begin{cases} \tilde{\Psi}_{r'} e(\phi(\boldsymbol{i})) & \text{if } i_r, i_{r+1} \in I_0, \\ \tilde{\Psi}_{r'}\Psi_{r'+1} e(\phi(\boldsymbol{i})) & \text{if } i_r \in I_1, i_{r+1} \in I_0, \\ \Psi_{r'+1}\tilde{\Psi}_{r'} e(\phi(\boldsymbol{i})) & \text{if } i_r \in I_0, i_{r+1} \in I_1. \end{cases}$$

Here $r' = r'_{\boldsymbol{i}}$ is as in Section 2F.

Proof. By construction of $\Phi_{\alpha,\boldsymbol{k}}$, the elements $\Phi_{\alpha,\boldsymbol{k}}(\Psi_r e(\boldsymbol{i}))$ and $\Phi_{\alpha,\boldsymbol{k}}(\tilde{\Psi}_r e(\boldsymbol{i}))$ are the unique elements of $S_{\bar{\alpha},\boldsymbol{k}}$ that acts on the polynomial representation by the same operator as $\Psi_r e(\boldsymbol{i})$ and $\tilde{\Psi}_r e(\boldsymbol{i})$, respectively.

The right hand side in the formulas for $\Phi_{\alpha,\boldsymbol{k}}(\Psi_r e(\boldsymbol{i}))$ or $\Phi_{\alpha,\boldsymbol{k}}(\tilde{\Psi}_r e(\boldsymbol{i}))$ in the statement is an element X in $S_{\bar{\alpha},\boldsymbol{k}}[\Sigma^{-1}]$. To complete the proof we have to show that:

(1) X acts by the same operator as $\Psi_r e(\boldsymbol{i})$ or $\tilde{\Psi}_r e(\boldsymbol{i})$, respectively, on the polynomial representation.

(2) X is in $S_{\bar{\alpha},\boldsymbol{k}}$.

Part (1) is obvious. Part (2) follows from part (1) and from the faithfulness of the polynomial representation of $S_{\bar{\alpha},k}[\Sigma^{-1}]$ (see Lemma 2.10). (In fact, part (2) is not obvious only in the case $i_r = i_{r+1} \in I_1$.) □

B2. ***Special quivers.*** From now on we will be interested only in some special types of quivers.

First, consider the quiver $\Gamma = \Gamma_e$, where e is an integer > 1. In particular, from now on we fix $I = \mathbb{Z}/e\mathbb{Z}$. Fix $k \in [0, e-1]$ and set $I_1 = \{k\}$ and $I_0 = I\backslash\{k\}$. In this case the quiver $\overline{\Gamma}$ is isomorphic to Γ_{e+1}. More precisely, the decomposition $\bar{I} = \bar{I}_0 \sqcup \bar{I}_1 \sqcup \bar{I}_2$ is such that $\bar{I}_1 = \{k\}$ and $\bar{I}_2 = \{k+1\}$. To avoid confusion, for $i \in \bar{I}$ we will write $\bar{\alpha}_i$ and $\bar{\varepsilon}_i$ for α_i and ε_i respectively.

Remark B.2. If Γ is as above, a sequence $\boldsymbol{i} = (i_1, \dots, i_d) \in \bar{I}^d$ is well ordered if for each index a such that $i_a = k$ we have $a < d$ and $i_{a+1} = k+1$. The sequence $\boldsymbol{i}$ is unordered if there is $r \leqslant d$ such that the subsequence $(i_1, \dots, i_r)$ contains more elements equal to $k+1$ than elements equal to k.

Let $\Upsilon\colon \mathbb{Z} \to \mathbb{Z}$ be the map given for $a \in \mathbb{Z}$ and $b \in [0, e-1]$ by

$$\Upsilon(ae+b) = \begin{cases} a(e+1)+b & \text{if } b \in [0,k], \\ a(e+1)+b+1 & \text{if } b \in [k+1, e-1]. \end{cases} \tag{8}$$

Now, consider the quiver $\tilde{\Gamma} = (\Gamma_\infty)^{\sqcup l}$ (i.e., $\tilde{\Gamma}$ is a disjoint union of l copies of Γ_∞). Set $\tilde{\Gamma} = (\tilde{I}, \tilde{H})$ and write $\tilde{\alpha}_i$ and $\tilde{\varepsilon}_i$ and for α_i and ε_i respectively for each $i \in \tilde{I}$. We identify an element of $\tilde{I}$ with an element $(a,b) \in \mathbb{Z} \times [1,l]$ in the obvious way. Consider the decomposition $\tilde{I} = \tilde{I}_0 \sqcup \tilde{I}_1$ such that $(a,b) \in \tilde{I}_1$ if and only if $a \equiv k \mod e$. In this case the quiver $\overline{\tilde{\Gamma}}$ is isomorphic to $\tilde{\Gamma}$. We will often write $\tilde{\Gamma}$ instead of $\overline{\tilde{\Gamma}}$ (but sometimes, if confusion is possible, we will use the notation $\overline{\tilde{\Gamma}}$ to stress that we work with the doubled quiver). More precisely, in this case we have

$$\begin{aligned} (a,b)^0 &= (\Upsilon(a), b), \\ (a,b)^1 &= (\Upsilon(a), b), \\ (a,b)^2 &= (\Upsilon(a)+1, b). \end{aligned}$$

To distinguish notations, we will always write $\tilde{\phi}$ for any of the maps $\tilde{\phi}\colon \tilde{I}^\infty \to \tilde{I}^\infty$, $Q_{\tilde{I}} \to Q_{\tilde{I}}$, $X_{\tilde{I}} \to X_{\tilde{I}}$ in Section 2B.

From now on we write $\Gamma = \Gamma_e$, $\overline{\Gamma} = \Gamma_{e+1}$ and $\tilde{\Gamma} = (\Gamma_\infty)^{\sqcup l}$. Recall that

$$I = I_e = \mathbb{Z}/e\mathbb{Z}, \quad \bar{I} = I_{e+1} = \mathbb{Z}/(e+1)\mathbb{Z}, \quad \tilde{I} = (I_\infty)^{\sqcup l} = \mathbb{Z} \times [1,l].$$

Consider the quiver homomorphism $\pi_e\colon \tilde{\Gamma} \to \Gamma$ such that

$$\pi_e\colon \tilde{I} \to I, \quad (a,b) \mapsto a \mod e.$$

Then π_{e+1} is a quiver homomorphism $\pi_{e+1}\colon \tilde{\Gamma} \to \overline{\Gamma}$. They yield $\mathbb{Z}$-linear maps

$$\pi_e\colon Q_{\tilde{I}} \to Q_I, \quad \pi_e\colon X_{\tilde{I}} \to X_I, \quad \pi_{e+1}\colon Q_{\tilde{I}} \to Q_{\bar{I}}, \quad \pi_{e+1}\colon X_{\tilde{I}} \to X_{\bar{I}}.$$

The following diagrams are commutative for $\alpha \in Q_I^+$ and $\tilde{\alpha} \in Q_{\tilde{I}}^+$ such that $\pi_e(\tilde{\alpha}) = \alpha$,

$$\begin{tikzcd} Q_{\tilde{I}} \arrow[r, "\tilde{\phi}"] \arrow[d, "\pi_e"'] & Q_{\tilde{I}} \arrow[d, "\pi_{e+1}"'] \\ Q_I \arrow[r, "\phi"] & Q_{\bar{I}} \end{tikzcd} \qquad \begin{tikzcd} X_{\tilde{I}} \arrow[r, "\tilde{\phi}"] \arrow[d, "\pi_e"'] & X_{\tilde{I}} \arrow[d, "\pi_{e+1}"'] \\ X_I \arrow[r, "\phi"] & X_{\bar{I}} \end{tikzcd} \qquad \begin{tikzcd} \tilde{I}^{\tilde{\alpha}} \arrow[r, "\tilde{\phi}"] \arrow[d, "\pi_e"'] & \tilde{I}^{\tilde{\phi}(\tilde{\alpha})} \arrow[d, "\pi_{e+1}"'] \\ I^{\alpha} \arrow[r, "\phi"] & \bar{I}^{\phi(\alpha)} \end{tikzcd}$$

The quiver $\tilde{\Gamma}$ is infinite. We will sometimes use its truncated version. Fix a positive integer N. Denote by $\tilde{\Gamma}^{\leqslant N}$ the full subquiver (i.e., a quiver with a smaller set of vertices and the same arrows between these vertices) of $\tilde{\Gamma}$ that contains only vertices (a, b) such that $|a| \leqslant eN$. Let $\bar{\tilde{\Gamma}}^{\leqslant N}$ be the doubled quiver associated with $\tilde{\Gamma}^{\leqslant N}$. We can see the quiver $\bar{\tilde{\Gamma}}^{\leqslant N}$ as a full subquiver of $\bar{\tilde{\Gamma}}$ that contains only vertices (a, b) such that we have

$$\begin{cases} -(e+1)N \leqslant a \leqslant (e+1)N & \text{if } k \neq 0, \\ -(e+1)N \leqslant a \leqslant (e+1)N+1 & \text{else.} \end{cases}$$

(Attention, it is not true that the isomorphism of quivers $\tilde{\Gamma} \simeq \bar{\tilde{\Gamma}}$ takes $\tilde{\Gamma}^{\leqslant N}$ to $\bar{\tilde{\Gamma}}^{\leqslant N}$.)

B3. ***Hecke algebras.*** Let R be a commutative ring with 1. Fix an element $q \in R$.

Definition B.3. The *affine Hecke algebra* $H_{R,d}(q)$ is the R-algebra generated by $T_1, \dots, T_{d-1}$ and the invertible elements $X_1, \dots, X_d$ modulo the following defining relations

$$\begin{aligned} X_r X_s &= X_s X_r, & \\ T_r X_r &= X_r T_r & \text{if } |r-s| > 1, \\ T_r T_s &= T_s T_r & \text{if } |r-s| > 1, \\ T_r T_{r+1} T_r &= T_{r+1} T_r T_{r+1}, & \\ T_r X_{r+1} &= X_r T_r + (q-1) X_{r+1}, & \\ T_r X_r &= X_{r+1} T_r - (q-1) X_{r+1}, & \\ 0 &= (T_r - q)(T_r + 1). & \end{aligned}$$

Assume that $R = \boldsymbol{k}$ is a field and $q \neq 0, 1$. The algebra $H_{d,\boldsymbol{k}}(q)$ has a faithful representation (see [Miemietz and Stroppel 2016, Proposition 3.11]) in the vector space $\boldsymbol{k}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]$ such that $X_r^{\pm 1}$ acts by multiplication by $X_r^{\pm 1}$ and T_r by

$$T_r(P) = q s_r(P) + (q-1) X_{r+1} (X_r - X_{r+1})^{-1} (s_r(P) - P).$$

The following operator acts on $\boldsymbol{k}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]$ as the reflection s_r

$$\Psi_r = \frac{X_r - X_{r+1}}{q X_r - X_{r+1}} (T_r - q) + 1 = (T_r + 1) \frac{X_r - X_{r+1}}{X_r - q X_{r+1}} - 1.$$

For a future use, consider the element $\tilde{\Psi}_r \in H_{d,\boldsymbol{k}}$ given by

$$\tilde{\Psi}_r = (q X_r - X_{r+1}) \Psi_r = (X_r - X_{r+1}) T_r + (q-1) X_{r+1}.$$

B4. ***The isomorphism between Hecke and KLR algebras.*** First, we define some localized versions of Hecke algebras and KLR algebras. Let $\mathscr{F}$ be a finite subset of $\boldsymbol{k}^\times$. We view $\mathscr{F}$ as the vertex set of a quiver with an arrow $i \to j$ if and only if $j = qi$. Consider the algebra

$$A_1 = \bigoplus_{\boldsymbol{i} \in \mathscr{F}^d} \boldsymbol{k}[X_1^{\pm 1}, \dots, X_d^{\pm 1}][(X_r - X_t)^{-1}, (qX_r - X_t)^{-1} : r \neq t]e(\boldsymbol{i}),$$

where $e(\boldsymbol{i})$ are orthogonal idempotents and X_r commutes with $e(\boldsymbol{i})$. Let $H^{\mathrm{loc}}_{d,\boldsymbol{k}}(q)$ be the A_1-module given by the extension of scalars from the $\boldsymbol{k}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]$-module $H_{d,\boldsymbol{k}}(q)$. It has a $\boldsymbol{k}$-algebra structure such that

$$T_r e(\boldsymbol{i}) - e(s_r(\boldsymbol{i}))T_r = (1-q)X_{r+1}(X_r - X_{r+1})^{-1}(e(\boldsymbol{i}) - e(s_r(\boldsymbol{i})))$$

and

$$Z^{-1}T_r = T_r Z^{-1}, \qquad \text{where } Z = \prod_{r<t}(X_r - X_t)^2 \prod_{r \neq t}(qX_r - X_t)^2.$$

In this section the KLR algebras are always defined with respect to the quiver $\mathscr{F}$. We consider the algebra

$$A_2 = \bigoplus_{\boldsymbol{i} \in \mathscr{F}^d} \boldsymbol{k}[x_1, \dots, x_d][S_{\boldsymbol{i}}^{-1}]e(\boldsymbol{i}),$$

where

$$S_{\boldsymbol{i}} = \{(x_r + 1), (i_r(x_r+1) - i_t(x_t+1)), (qi_r(x_r+1) - i_t(x_t+1) : r \neq t)\}.$$

Consider the following central element in $R_{d,\boldsymbol{k}}$

$$z = \prod_r (x_r + 1) \prod_{i,j \in \mathscr{F}, r \neq t} (i(x_r+1) - j(x_t+1)).$$

The A_2-module $R^{\mathrm{loc}}_{d,\boldsymbol{k}} = A_2 \otimes_{\boldsymbol{k}_d^{(\mathscr{F})}} R_{d,\boldsymbol{k}}$ has a $\boldsymbol{k}$-algebra structure because it is a subalgebra in $R_{d,\boldsymbol{k}}[z^{-1}]$, where $\boldsymbol{k}_d^{(\mathscr{F})}$ is as in (2).

Remark B.4. We assumed above that the set $\mathcal{F}$ is finite. This assumption is important because it implies that A_1 contains $\boldsymbol{k}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]$ and A_2 contains $\boldsymbol{k}[x_1, \dots, x_d]$. However, it is possible to define the algebras above (A_1, A_2, $H^{\mathrm{loc}}_{d,\boldsymbol{k}}(q)$ and $R^{\mathrm{loc}}_{d,\boldsymbol{k}}$) for arbitrary $\mathcal{F} \subset \boldsymbol{k}^\times$. Indeed, if $\mathcal{F}_1 \subset \mathcal{F}_2$ are finite, then the algebra defined with respect to $\mathcal{F}_1$ is obviously a nonunitary subalgebra of the algebra defined with respect to $\mathcal{F}_2$. Then we can define the algebras A_1, A_2, $H^{\mathrm{loc}}_{d,\boldsymbol{k}}(q)$ and $R^{\mathrm{loc}}_{d,\boldsymbol{k}}$ with respect to any arbitrary $\mathcal{F}$. For example, we define the algebra $R^{\mathrm{loc}}_{d,\boldsymbol{k}}$ associated with $\mathcal{F}$ as

$$R^{\mathrm{loc}}_{d,\boldsymbol{k}}(\mathcal{F}) = \varinjlim_{\mathcal{F}_0 \subset \mathcal{F}} R^{\mathrm{loc}}_{d,\boldsymbol{k}}(\mathcal{F}_0),$$

where the direct limit is taken over all finite subsets $\mathcal{F}_0$ of $\mathcal{F}$. Note that if the set $\mathcal{F}$ is infinite, then the algebras A_1, A_2, $H^{\mathrm{loc}}_{d,\boldsymbol{k}}(q)$ and $R^{\mathrm{loc}}_{d,\boldsymbol{k}}$ are not unitary.

From now on we assume that $\mathcal{F}$ is an arbitrary subset of $\boldsymbol{k}^\times$.

Proposition B.5. *There is an isomorphism of $\boldsymbol{k}$-algebras $R^{\mathrm{loc}}_{d,\boldsymbol{k}} \simeq H^{\mathrm{loc}}_{d,\boldsymbol{k}}(q)$ such that*

$$\begin{aligned} e(\boldsymbol{i}) &\mapsto e(\boldsymbol{i}), \\ x_r e(\boldsymbol{i}) &\mapsto (i_r^{-1} X_r - 1) e(\boldsymbol{i}), \\ \Psi_r e(\boldsymbol{i}) &\mapsto \Psi_r e(\boldsymbol{i}). \end{aligned}$$

Proof. The polynomial representations of $H_{d,\boldsymbol{k}}(q)$ and $R_{d,\boldsymbol{k}}$ yield faithful representations of $H^{\mathrm{loc}}_{d,\boldsymbol{k}}(q)$ and $R^{\mathrm{loc}}_{d,\boldsymbol{k}}$ on A_1 and A_2 respectively. Moreover, there is an isomorphism of $\boldsymbol{k}$-algebras $A_2 \simeq A_1$ given by $x_r e(\boldsymbol{i}) \mapsto (i_r^{-1} X_r - 1) e(\boldsymbol{i})$.

This implies the statement. Indeed, the elements $e(\boldsymbol{i}) \in R^{\mathrm{loc}}_{d,\boldsymbol{k}}$ and $e(\boldsymbol{i}) \in H^{\mathrm{loc}}_{d,\boldsymbol{k}}(q)$ act on $A_2 \simeq A_1$ by the same operators. The elements $x_r e(\boldsymbol{i}) \in R^{\mathrm{loc}}_{d,\boldsymbol{k}}$ and $(i_r^{-1} X_r - 1) e(\boldsymbol{i}) \in H^{\mathrm{loc}}_{d,\boldsymbol{k}}(q)$ act on $A_2 \simeq A_1$ by the same operators. Finally, the elements $\Psi_r e(\boldsymbol{i}) \in R^{\mathrm{loc}}_{d,\boldsymbol{k}}$ and $\Psi_r e(\boldsymbol{i}) \in H^{\mathrm{loc}}_{d,\boldsymbol{k}}(q)$ also act on $A_2 \simeq A_1$ by the same operators. The elements above generate the algebras $R^{\mathrm{loc}}_{d,\boldsymbol{k}}$ and $H^{\mathrm{loc}}_{d,\boldsymbol{k}}(q)$. □

Now, we consider the subalgebra $\hat{R}_{d,\boldsymbol{k}}$ of $R^{\mathrm{loc}}_{d,\boldsymbol{k}}$ generated by

- the elements of $R_{d,\boldsymbol{k}}$,
- the elements $(x_r + 1)^{-1}$,
- the elements of the form $(i_r(x_r + 1) - i_t(x_t + 1))^{-1} e(\boldsymbol{i})$ such that $r \neq t$ and $i_r \neq i_t$,
- the elements of the form $(q i_r(x_r + 1) - i_t(x_t + 1))^{-1} e(\boldsymbol{i})$ such that $r \neq t$ and $q i_r \neq i_t$.

Similarly, consider the subalgebra $\hat{H}_{d,\boldsymbol{k}}(q)$ of $H^{\mathrm{loc}}_{d,\boldsymbol{k}}(q)$ generated by

- the elements of $H_{d,\boldsymbol{k}}(q)$,
- the elements of the form $(X_r - X_t)^{-1} e(\boldsymbol{i})$ such that $r \neq t$ and $i_r \neq i_t$,
- the elements of the form $(q X_r - X_t)^{-1} e(\boldsymbol{i})$ such that $r \neq t$ and $q i_r \neq i_t$.

Note that the element $\Psi_r e(\boldsymbol{i}) \in H^{\mathrm{loc}}_{d,\boldsymbol{k}}(q)$ belongs to $\hat{H}_{d,\boldsymbol{k}}(q)$ if $i_r \neq q i_{r+1}$. We have the following proposition, see also [Rouquier 2008, §3.2].

Proposition B.6. *The isomorphism $R^{\mathrm{loc}}_{d,\boldsymbol{k}} \simeq H^{\mathrm{loc}}_{d,\boldsymbol{k}}(q)$ from Proposition B.5 restricts to an isomorphism $\hat{R}_{d,\boldsymbol{k}} \simeq \hat{H}_{d,\boldsymbol{k}}(q)$.* □

B5. ***Deformation rings.*** In this section we introduce some general definitions from [Rouquier et al. 2016] for a later use.

We call the *deformation ring* $(R, \kappa, \kappa_1, \dots, \kappa_l)$ a regular commutative noetherian $\mathbb{C}$-algebra R with 1 equipped with a homomorphism $\mathbb{C}[\kappa^{\pm 1}, \kappa_1, \dots, \kappa_l] \to R$. Let $\kappa, \kappa_1, \dots, \kappa_l$ also denote the images of $\kappa, \kappa_1, \dots, \kappa_l$ in R. A deformation ring is *in general position* if any two elements of the set

$$\{\kappa_u - \kappa_v + a\kappa + b,\ \kappa - c : a, b \in \mathbb{Z},\ c \in \mathbb{Q},\ u \neq v\}$$

have no common nontrivial divisors. A *local deformation ring* is a deformation ring which is a local ring such that $\kappa_1, \dots, \kappa_l, \kappa - e$ belong to the maximal ideal of R. Note that each $\mathbb{C}$-algebra that is a field has a *trivial* local deformation ring structure, i.e., such that $\kappa_1 = \cdots = \kappa_l = 0$ and $\kappa = e$. We always consider $\mathbb{C}$ as a local deformation ring with a trivial deformation ring structure.

We will write $\bar{\kappa} = \kappa(e+1)/e$ and $\bar{\kappa}_r = \kappa_r(e+1)/e$. We will abbreviate R for $(R, \kappa, \kappa_1, \dots, \kappa_l)$ and $\bar{R}$ for $(R, \bar{\kappa}, \bar{\kappa}_1, \dots, \bar{\kappa}_l)$.

Let R be a complete local deformation ring with residue field $\boldsymbol{k}$. Consider the elements $q_e = \exp(2\pi\sqrt{-1}/\kappa)$ and $q_{e+1} = \exp(2\pi\sqrt{-1}/\bar{\kappa})$ in R. These elements specialize to $\zeta_e = \exp(2\pi\sqrt{-1}/e)$ and $\zeta_{e+1} = \exp(2\pi\sqrt{-1}/(e+1))$ in $\boldsymbol{k}$.

B6. ***The choice of $\mathcal{F}$.*** From now on we assume that R is a complete local deformation ring in general position with residue field $\boldsymbol{k}$ and field of fractions K. In this section we define some special choice of the set $\mathcal{F}$. This choice of parameters is particularly interesting because it is related with the categorical action on the category $\mathcal{O}$ for $\widehat{\mathfrak{gl}}_N$, see [Rouquier et al. 2016].

Fix a tuple $\nu = (\nu_1, \dots, \nu_l) \in \mathbb{Z}^l$. Put $Q_r = \exp(2\pi\sqrt{-1}(\nu_r + \kappa_r)/\kappa)$ for $r \in [1, l]$. The canonical homomorphism $R \to \boldsymbol{k}$ maps q_e to ζ_e and Q_r to $\zeta_e^{\nu_r}$.

Now, consider the subset $\mathscr{F}$ of R given by

$$\mathscr{F} = \bigcup_{r \in \mathbb{Z}, t \in [1,l]} \{q_e^r Q_t\}.$$

Denote by $\mathcal{F}_{\boldsymbol{k}}$ the image of $\mathcal{F}$ in $\boldsymbol{k}$ with respect to the surjection $R \to \boldsymbol{k}$. Recall from Section B4 that we consider $\mathcal{F}$ (and $\mathcal{F}_{\boldsymbol{k}}$) as a vertex set of a quiver. The set $\mathcal{F}$ is a vertex set of a quiver that is a disjoint union if l infinite linear quivers. The set $\mathcal{F}_{\boldsymbol{k}}$ is a vertex set of a cyclic quiver of length e.

Fix $k \in [0, e-1]$. To this k we associate a map $\Upsilon \colon \mathbb{Z} \to \mathbb{Z}$ as in (8). Now, consider the tuple

$$\bar{\nu} = (\bar{\nu}_1, \dots, \bar{\nu}_l) \in \mathbb{Z}^l, \quad \bar{\nu}_r = \Upsilon(\nu_r)\ \forall r \in [1, l].$$

Let $\bar{R}$ be as in the previous section. Let $\bar{\boldsymbol{k}}$ and $\bar{K}$ be the residue field and the field of fractions of $\bar{R}$ respectively. Now, consider $\bar{Q} = (\bar{Q}_1, \dots, \bar{Q}_l)$, where $\bar{Q}_r = \exp(2\pi\sqrt{-1}(\bar{\nu}_r + \bar{\kappa}_r)/\bar{\kappa})$ and $\bar{\kappa}$ and $\bar{\kappa}_r$ are defined in Section B5. Consider the subset $\bar{\mathscr{F}}$ of $\bar{R}$ given by

$$\bar{\mathscr{F}} = \bigcup_{r \in \mathbb{Z}, t \in [1,l]} \{q_{e+1}^r \bar{Q}_t\}.$$

Denote by $\bar{\mathcal{F}}_{\bar{\boldsymbol{k}}}$ the image of $\bar{\mathcal{F}}$ in $\bar{\boldsymbol{k}}$ with respect to the surjection $\bar{R} \to \bar{\boldsymbol{k}}$. The set $\bar{\mathcal{F}}$ is a vertex set of a quiver that is a disjoint union of l infinite linear quivers. The set $\bar{\mathcal{F}}_{\bar{\boldsymbol{k}}}$ is a vertex set of a cyclic quiver of length $e+1$.

B7. ***Algebras $\hat{H}$, $\widehat{SH}$, $\hat{R}$ and $\hat{S}$.*** Let $\Gamma = (I, H)$, $\bar{\Gamma} = (\bar{I}, \bar{H})$ and $\tilde{\Gamma} = (\tilde{I}, \tilde{H})$ be as in Section B2.

We will use the notation $\mathcal{F}$, $\mathcal{F}_{\boldsymbol{k}}$, $\bar{\mathcal{F}}$ and $\bar{\mathcal{F}}_{\bar{\boldsymbol{k}}}$ as in previous section. (In particular, we fix some $\nu = (\nu_1, \dots, \nu_l)$.)

We have the following isomorphisms of quivers

$$\begin{aligned} &\tilde{I} \simeq \mathscr{F}, \quad && i=(a,b) \mapsto p_i := \exp(2\pi\sqrt{-1}(a+\kappa_b)/\kappa),\\ &\tilde{I} \simeq \overline{\mathscr{F}}, \quad && i=(a,b) \mapsto \bar{p}_i := \exp(2\pi\sqrt{-1}(a+\bar{\kappa}_b)/\bar{\kappa}),\\ &I \simeq \mathscr{F}_{\boldsymbol{k}}, && \qquad\qquad i \mapsto p_i := \zeta_e^i,\\ &\bar{I} \simeq \overline{\mathscr{F}}_{\bar{\boldsymbol{k}}}, && \qquad\qquad i \mapsto \bar{p}_i := \zeta_{e+1}^i. \end{aligned}$$

These isomorphisms yield the following commutative diagrams

$$\begin{tikzcd} \tilde{I} \arrow[r, "\sim"] \arrow[d, "\pi_e"'] & \mathscr{F} \arrow[d] \\ I \arrow[r, "\sim"] & \mathscr{F}_{\boldsymbol{k}}, \end{tikzcd} \qquad\qquad \begin{tikzcd} \tilde{I} \arrow[r, "\sim"] \arrow[d, "\pi_{e+1}"'] & \overline{\mathscr{F}} \arrow[d] \\ \bar{I} \arrow[r, "\sim"] & \overline{\mathscr{F}}_{\bar{\boldsymbol{k}}}. \end{tikzcd}$$

We will identify

$$I \simeq \mathscr{F}_{\boldsymbol{k}}, \quad \bar{I} \simeq \overline{\mathscr{F}}_{\bar{\boldsymbol{k}}}, \quad \tilde{I} \simeq \mathscr{F}, \quad \tilde{I} \simeq \overline{\mathscr{F}}$$

as above.

Our goal is to obtain an analogue of Theorem 2.12 over the ring R. First, consider the algebras $\hat{H}_{d,\boldsymbol{k}}(\zeta_e)$ and $\hat{H}_{d,K}(q_e)$ defined in the same way as in Section B4 with respect to the sets $\mathscr{F}_{\boldsymbol{k}} \subset \boldsymbol{k}$ and $\mathscr{F} \subset K$. We can consider the R-algebra $\hat{H}_{d,R}(q_e)$ defined in a similar way with respect to the same set of idempotents as $\hat{H}_{d,\boldsymbol{k}}(\zeta_e)$ (i.e., with respect to the set $\mathcal{F}_{\boldsymbol{k}}$, not $\mathcal{F}$).

The algebra $\hat{H}_{d,K}(q_e)$ is not unitary because the quiver $\tilde{\Gamma}$ is infinite. To avoid this problem we consider the truncated version of this algebra. Let $\hat{H}_{d,K}^{\leqslant N}(q_e)$ be the quotient of $\hat{H}_{d,K}(q_e)$ by the two-sided ideal generated by the idempotents $e(\boldsymbol{j}) \in \tilde{I}^d$ such that $\boldsymbol{j}$ contains a component that is not a vertex of the truncated quiver $\tilde{\Gamma}^{\leqslant N}$ (see Section B2). (In fact, the algebra $\hat{H}_{d,K}^{\leqslant N}(q_e)$ is isomorphic to a direct summand of $\hat{H}_{d,K}(q_e)$).

Similarly, we define the algebras $\hat{H}_{d,\bar{\boldsymbol{k}}}(\zeta_{e+1})$, $\hat{H}_{d,\bar{K}}(q_{e+1})$ and $\hat{H}_{d,\bar{R}}(q_{e+1})$ using the sets $\overline{\mathcal{F}}$ and $\overline{\mathcal{F}}_{\bar{\boldsymbol{k}}}$ instead of $\mathcal{F}$ and $\mathcal{F}_{\boldsymbol{k}}$. We define a truncation $\hat{H}_{d,\bar{K}}^{\leqslant N}(q_{e+1})$ of $\hat{H}_{d,\bar{K}}(q_{e+1})$ using the quiver $\bar{\tilde{\Gamma}}^{\leqslant N}$.

For each $\boldsymbol{i} \in I^d$ we consider the following idempotent in $\hat{H}_{d,K}^{\leqslant N}(q_e)$:

$$e(\boldsymbol{i}) = \sum_{\boldsymbol{j} \in \tilde{I}^d, \pi_e(\boldsymbol{j}) = \boldsymbol{i}} e(\boldsymbol{j}).$$

Here we mean that $e(\boldsymbol{j})$ is zero if $\boldsymbol{j}$ contains a vertex that is not in the truncated quiver $\tilde{\Gamma}^{\leqslant N}$. The idempotent $e(\boldsymbol{i})$ is well defined because only a finite number of terms in the sum are nonzero. For each $\boldsymbol{i} \in \bar{I}^d$ we can define an idempotent $e(\boldsymbol{i}) \in \hat{H}_{d,\bar{K}}^{\leqslant N}(q_{e+1})$ in a similar way.

Lemma B.7. *There is an injective algebra homomorphism $\hat{H}_{d,R}(q_e) \to \hat{H}_{d,K}^{\leqslant N}(q_e)$ such that $e(\boldsymbol{i}) \mapsto e(\boldsymbol{i})$, $X_r e(\boldsymbol{i}) \mapsto X_r e(\boldsymbol{i})$ and $T_r e(\boldsymbol{i}) \mapsto T_r e(\boldsymbol{i})$.*

Proof. It is clear that we have an algebra homomorphism $\hat{H}_{d,R}(q_e) \to \hat{H}_{d,K}^{\leqslant N}(q_e)$ as in the statement. We only have to check the injectivity.

For each $w\in\mathfrak{S}_d$ we have an element $T_w\in H_{d,R}(q)$ defined in the following way. We have $T_w=T_{i_1}\cdots T_{i_r}$, where $w=s_{i_1}\cdots s_{i_r}$ is a reduced expression. It is well-known that T_w is independent of the choice of the reduced expression. Moreover, the algebra $H_{d,R}(q)$ is free over $R[X_1^{\pm1},\dots,X_d^{\pm1}]$ with a basis $\{T_w:w\in\mathfrak{S}_d\}$.

Set

$$B=\bigoplus_{\boldsymbol{i}\in\mathscr{F}_{\boldsymbol{k}}^d} R[X_1^{\pm1},\dots,X_d^{\pm1}][(X_r-X_t)^{-1},(q_eX_r-X_t)^{-1}:r\neq t]e(\boldsymbol{i}),$$

where we invert (X_r-X_t) only if $i_r\neq i_t$ and we invert $(q_eX_r-X_t)$ only if $\zeta_e i_r\neq i_t$. We have $\hat{H}_{d,R}(q_e)=B\otimes_{R[X_1^{\pm1},\dots,X_d^{\pm1}]}H_{d,R}(q_e)$. This implies that the B-module $\hat{H}_{d,R}(q_e)$ is free with a basis $\{T_w:w\in\mathfrak{S}_d\}$.

Similarly, we can show that the algebra $\hat{H}^{\leqslant N}_{d,K}(q_e)$ is free (with a basis $\{T_w:w\in\mathfrak{S}_d\}$) over

$$B'=\bigoplus_{\boldsymbol{j}\in\mathscr{F}^d} K[X_1^{\pm1},\dots,X_d^{\pm1}][(X_r-X_t)^{-1},(q_eX_r-X_t)^{-1}:r\neq t]e(\boldsymbol{j}),$$

where we invert (X_r-X_t) only if $j_r\neq j_t$ and we invert $(q_eX_r-X_t)$ only if $q_e j_r\neq j_t$, and we take only $\boldsymbol{j}$ that are supported on the vertices of the truncated quiver $\Gamma^{\leqslant N}$.

Now, the injectivity of the homomorphism follows from the fact that it takes a B-basis of $\hat{H}_{d,R}(q_e)$ to a B'-linearly independent set in $\hat{H}^{\leqslant N}_{d,K}(q_e)$. □

Now we define the algebra $\widehat{SH}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1})$ that is a Hecke analogue of a localization of the balanced KLR algebra $S_{\bar{\alpha},\boldsymbol{k}}$. To do so, consider the idempotent $\boldsymbol{e}=\sum_{\boldsymbol{i}\in\bar{I}^{\bar{\alpha}}_{\mathrm{ord}}}e(\boldsymbol{i})$ in $\hat{H}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1})$. We set

$$\widehat{SH}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1})=\boldsymbol{e}\hat{H}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1})\boldsymbol{e}/\sum_{\boldsymbol{j}\in\bar{I}^{\bar{\alpha}}_{\mathrm{un}}}\boldsymbol{e}\hat{H}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1})e(\boldsymbol{j})\hat{H}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1})\boldsymbol{e}.$$

Now, we define a similar algebra over K. To do this, we need to introduce some additional notation. Denote by $Q^+_{\tilde{I},\mathrm{eq}}$ the subset of $Q^+_{\tilde{I}}$ that contains only $\tilde{\alpha}$ such that for each $k\in\tilde{I}_1$, the dimension vector $\tilde{\alpha}$ has the same dimensions at vertices k^1 and k^2.

Set

$$\hat{H}^{\leqslant N}_{\bar{\alpha},\overline{K}}(q_{e+1})=\bigoplus_{\pi_{e+1}(\tilde{\alpha})=\bar{\alpha}}\hat{H}_{\tilde{\alpha},\overline{K}}(q_{e+1}),\quad\text{and}\quad \widehat{SH}^{\leqslant N}_{\bar{\alpha},\overline{K}}(q_{e+1})=\bigoplus_{\pi_{e+1}(\tilde{\alpha})=\bar{\alpha}}\widehat{SH}_{\tilde{\alpha},\overline{K}}(q_{e+1}),$$

where in the sums we take only $\tilde{\alpha}\in Q^+_{\tilde{I},\mathrm{eq}}$ that are supported on the vertices of the truncated quiver $\bar{\bar{\Gamma}}^{\leqslant N}$ and $\widehat{SH}_{\tilde{\alpha},\overline{K}}(q_{e+1})$ is defined similarly to $\widehat{SH}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1})$. More precisely, we have

$$\widehat{SH}_{\tilde{\alpha},\overline{K}}(q_{e+1})=\tilde{\boldsymbol{e}}_{\tilde{\alpha}}H_{\tilde{\alpha},\overline{K}}(q_{e+1})\tilde{\boldsymbol{e}}_{\tilde{\alpha}}/\sum_{\boldsymbol{j}\in\tilde{I}^{\tilde{\alpha}}_{\mathrm{un}}}\tilde{\boldsymbol{e}}_{\tilde{\alpha}}H_{\tilde{\alpha},\overline{K}}(q_{e+1})e(\boldsymbol{j})H_{\tilde{\alpha},\overline{K}}(q_{e+1})\tilde{\boldsymbol{e}}_{\tilde{\alpha}},$$

where $\tilde{\boldsymbol{e}}_{\tilde{\alpha}}=\sum_{\boldsymbol{j}\in\tilde{I}^{\tilde{\alpha}}_{\mathrm{ord}}}e(\boldsymbol{j})$.

Remark B.8. Consider the following idempotents in $\hat H^{\leqslant N}_{\bar\alpha,\bar K}(q_{e+1})$:

$$\tilde{\boldsymbol{e}} = \sum_{\pi_{e+1}(\tilde\alpha)=\bar\alpha} \tilde{\boldsymbol{e}}_{\tilde\alpha} \quad\text{and}\quad \boldsymbol{e} = \sum_{\boldsymbol{i}\in \bar I^{\bar\alpha}_{\mathrm{ord}}} e(\boldsymbol{i}),$$

where the first sum is taken only by $\tilde\alpha \in Q^+_{\tilde I,\mathrm{eq}}$. (Note that $\hat H^{\leqslant N}_{\bar\alpha,\bar K}(q_{e+1})$ was defined as a quotient of $\hat H_{\bar\alpha,\bar K}(q_{e+1})$. So, if $\tilde\alpha$ is not supported on $\bar{\tilde\Gamma}^{\leqslant N}$, then the idempotent $\tilde{\boldsymbol{e}}_{\tilde\alpha}$ is zero by definition. In particular, the sum has a finite number of nonzero terms.) Set also $\tilde I^{\bar\alpha} = \coprod_{\pi_{e+1}(\tilde\alpha)=\bar\alpha} \tilde I^{\tilde\alpha}$, where the sum is taken only by $\tilde\alpha \in Q^+_{\tilde I,\mathrm{eq}}$. By definition, the algebra $\widehat{SH}^{\leqslant N}_{\bar\alpha,\bar K}(q_{e+1})$ is a quotient of $\tilde{\boldsymbol{e}}\hat H^{\leqslant N}_{\bar\alpha,\bar K}(q_{e+1})\tilde{\boldsymbol{e}}$. But we can see this algebra as the same quotient of $\boldsymbol{e}\hat H^{\leqslant N}_{\bar\alpha,\bar K}(q_{e+1})\boldsymbol{e}$ (we do the quotient with respect to the same idempotents). Indeed, the idempotent $\boldsymbol{e}$ is a sum of a bigger number of standard idempotents $e(\boldsymbol{j})$, $\boldsymbol{j} \in \tilde I^{\bar\alpha}$ than the idempotent $\tilde{\boldsymbol{e}}$. More precisely, the idempotent $\tilde{\boldsymbol{e}}$ is the sum all $e(\boldsymbol{j})$ such that $\boldsymbol{j}$ is well-ordered while $\boldsymbol{e}$ is the sum of all $e(\boldsymbol{j})$ such that $\pi_{e+1}(\boldsymbol{j})$ is well-ordered. But each $\boldsymbol{j} \in \tilde I^{\bar\alpha}$ such that $\pi_{e+1}(\boldsymbol{j})$ is well-ordered and $\boldsymbol{j}$ is not well-ordered must be unordered. Then such $e(\boldsymbol{j})$ becomes zero after taking the quotient.

Finally, we define the R-algebra $\widehat{SH}^{N}_{\bar\alpha,\bar R}(q_{e+1})$ as the image in $\widehat{SH}^{\leqslant N}_{\bar\alpha,\bar K}(q_{e+1})$ of the following composition of homomorphisms

$$\boldsymbol{e}\hat H_{\bar\alpha,\bar R}(q_{e+1})\boldsymbol{e} \to \boldsymbol{e}\hat H^{\leqslant N}_{\bar\alpha,\bar K}(q_{e+1})\boldsymbol{e} \to \widehat{SH}^{\leqslant N}_{\bar\alpha,\bar K}(q_{e+1}).$$

The lemma below shows that the algebra $\widehat{SH}^{N}_{\bar\alpha,\bar R}(q_{e+1})$ is independent of N for N large enough. So, we can write simply $\widehat{SH}_{\bar\alpha,\bar R}(q_{e+1})$ instead of $\widehat{SH}^{N}_{\bar\alpha,\bar R}(q_{e+1})$ for N large enough.

Lemma B.9. *Assume $N \geqslant 2d$. Then the algebra $\widehat{SH}^{N}_{\bar\alpha,\bar R}(q_{e+1})$ is independent of N.*

Proof. Denote by J_N the kernel of $\boldsymbol{e}\hat H_{\bar\alpha,\bar R}(q_{e+1})\boldsymbol{e} \to \widehat{SH}^{\leqslant N}_{\bar\alpha,\bar K}(q_{e+1})$. Take $M > N$. It is clear that we have $J_M \subset J_N$.

Let us show that we also have an opposite inclusion if $N \geqslant 2d$. We want to show that each element $x \in J_N$ is also in J_M. It is enough to show this for x of the form $x = Xe(\boldsymbol{i})$, where $\boldsymbol{i} \in I^{\bar\alpha}_{\mathrm{ord}}$ and X is composed of the elements of the form T_r and X_r. Then $Xe(\boldsymbol{i}) \in J_N$ means that the element $Xe(\boldsymbol{j}) \in \widehat{SH}^{\leqslant N}_{\bar\alpha,\bar K}(q_{e+1})$ is zero for each $\boldsymbol{j} \in \tilde I^{\bar\alpha}$ supported on $\bar{\tilde\Gamma}^{\leqslant N}$ such that $\pi_{e+1}(\boldsymbol{j}) = \boldsymbol{i}$. To show that we have $Xe(\boldsymbol{i}) \in J_M$ we must check that the element $Xe(\boldsymbol{j}) \in \widehat{SH}^{\leqslant M}_{\bar\alpha,\bar K}(q_{e+1})$ is zero for each $\boldsymbol{j} \in \tilde I^{\bar\alpha}$ supported on $\bar{\tilde\Gamma}^{\leqslant M}$ such that $\pi_{e+1}(\boldsymbol{j}) = \boldsymbol{i}$.

Let $\tilde\alpha \in Q^+_{\tilde I,\mathrm{eq}}$ be such that $\boldsymbol{j} \in \tilde I^{\tilde\alpha}$. It is clear that we can find an $\tilde\alpha' \in Q^+_{\tilde I,\mathrm{eq}}$ supported on $\bar{\tilde\Gamma}^{\leqslant 2d}$ such that we have an isomorphism $\hat H_{\tilde\alpha,\bar K}(q_{e+1}) \simeq \hat H_{\tilde\alpha,\bar K}(q_{e+1})$ that induces an isomorphism $\widehat{SH}_{\tilde\alpha,\bar K}(q_{e+1}) \simeq \widehat{SH}_{\tilde\alpha,\bar K}(q_{e+1})$ and such that this isomorphism preserves the generators X_r and T_r and sends the idempotent $e(\boldsymbol{j})$ to some idempotent $e(\boldsymbol{j}')$ such that $\boldsymbol{j}'$ is supported on $\bar{\tilde\Gamma}^{\leqslant 2d}$ and $\pi_{e+1}(\boldsymbol{j}) = \pi_{e+1}(\boldsymbol{j}')$. Then the element $Xe(\boldsymbol{j}) \in \widehat{SH}^{\leqslant M}_{\bar\alpha,\bar K}(q_{e+1})$ is zero because $Xe(\boldsymbol{j}') \in \widehat{SH}^{\leqslant M}_{\bar\alpha,\bar K}(q_{e+1})$ is zero. This implies $x \in J_M$. □

Now we define the KLR versions of the algebras $\widehat{SH}_{\bar\alpha,\bar{\boldsymbol{k}}}(\zeta_{e+1})$ and $\widehat{SH}^{\leqslant N}_{\bar\alpha,\overline{K}}(q_{e+1})$. As for the Hecke version, we denote by $\boldsymbol{e}$ the idempotent $\sum_{\boldsymbol{i}\in\bar I^{\bar\alpha}_{\mathrm{ord}}} e(\boldsymbol{i})$ in $\hat R_{\bar\alpha,\boldsymbol{k}}(\overline\Gamma)$. Set

$$\hat S_{\bar\alpha,\boldsymbol{k}}(\overline\Gamma)=\boldsymbol{e}\hat R_{\bar\alpha,\boldsymbol{k}}(\overline\Gamma)\boldsymbol{e}/\sum_{\boldsymbol{i}\in\bar I^{\bar\alpha}_{\mathrm{un}}}\boldsymbol{e}\hat R_{\bar\alpha,\boldsymbol{k}}(\overline\Gamma)e(\boldsymbol{i})R_{\bar\alpha,\boldsymbol{k}}(\overline\Gamma)\boldsymbol{e}.$$

For each $\tilde\alpha\in Q^+_{\tilde I,\mathrm{eq}}$ we consider the idempotent $\tilde{\boldsymbol{e}}_{\tilde\alpha}=\sum_{\boldsymbol{j}\in\tilde I^{\tilde\alpha}_{\mathrm{ord}}} e(\boldsymbol{j})$ in $\hat R_{\tilde\alpha,K}(\overline{\tilde\Gamma})$. Set

$$\hat S_{\bar\alpha,K}(\overline{\tilde\Gamma}^{\leqslant N})=\bigoplus_{\pi_{e+1}(\tilde\alpha)=\bar\alpha}\hat S_{\tilde\alpha,K}(\overline{\tilde\Gamma}),$$

where we take only $\tilde\alpha\in Q^+_{\tilde I,\mathrm{eq}}$ that are supported on the vertices of the truncated quiver $\overline{\tilde\Gamma}^{\leqslant N}$ and

$$\hat S_{\tilde\alpha,K}(\overline{\tilde\Gamma})=\tilde{\boldsymbol{e}}_{\tilde\alpha}\hat R_{\tilde\alpha,K}(\overline{\tilde\Gamma})\tilde{\boldsymbol{e}}_{\tilde\alpha}/\sum_{\boldsymbol{j}\in\tilde I^{\tilde\alpha}_{\mathrm{un}}}\tilde{\boldsymbol{e}}_{\tilde\alpha}\hat R_{\tilde\alpha,K}(\overline{\tilde\Gamma})e(\boldsymbol{j})R_{\tilde\alpha,K}(\overline{\tilde\Gamma})\tilde{\boldsymbol{e}}_{\tilde\alpha}.$$

Remark B.10. By Proposition B.6 we have algebra isomorphisms

$$\begin{aligned}\hat R_{\alpha,\boldsymbol{k}}(\Gamma)&\simeq\hat H_{\alpha,\boldsymbol{k}}(\zeta_e), &\quad \hat R_{\alpha,K}(\tilde\Gamma^{\leqslant N})&\simeq\hat H^{\leqslant N}_{\alpha,K}(q_e),\\ \hat R_{\bar\alpha,\boldsymbol{k}}(\overline\Gamma)&\simeq\hat H_{\bar\alpha,\bar{\boldsymbol{k}}}(\zeta_{e+1}), &\quad \hat R_{\bar\alpha,K}(\overline{\tilde\Gamma}^{\leqslant N})&\simeq\hat H^{\leqslant N}_{\bar\alpha,\overline{K}}(q_{e+1}),\end{aligned}$$

from which we deduce the isomorphisms

$$\hat S_{\bar\alpha,\boldsymbol{k}}(\overline\Gamma)\simeq\widehat{SH}_{\bar\alpha,\bar{\boldsymbol{k}}}(\zeta_{e+1})\quad\text{and}\quad\hat S_{\bar\alpha,K}(\overline{\tilde\Gamma}^{\leqslant N})\simeq\widehat{SH}^{\leqslant N}_{\bar\alpha,\overline{K}}(q_{e+1}).$$

We may use these isomorphisms without mentioning them explicitly. Using the identifications above between KLR algebras and Hecke algebras, a localization of the isomorphism in Theorem 2.12 yields an isomorphism

$$\Phi_{\alpha,\boldsymbol{k}}\colon\hat H_{\alpha,\boldsymbol{k}}(\zeta_e)\to\widehat{SH}_{\bar\alpha,\bar{\boldsymbol{k}}}(\zeta_{e+1}).$$

In the same way we also obtain an algebra isomorphism

$$\Phi_{\tilde\alpha,K}\colon\hat H_{\tilde\alpha,K}(q_e)\to\widehat{SH}_{\tilde\phi(\tilde\alpha),\overline{K}}(q_{e+1})$$

for each $\tilde\alpha\in Q^+_{\tilde I}$. Taking the sum over all $\tilde\alpha\in Q^+_{\tilde I}$ such that $\pi_e(\tilde\alpha)=\alpha$ and such that $\tilde\alpha$ is supported on the vertices of the truncated quiver $\tilde\Gamma^{\leqslant N}$ yields an isomorphism

$$\Phi_{\alpha,K}\colon\hat H^{\leqslant N}_{\alpha,K}(q_e)\to\widehat{SH}^{\leqslant N}_{\bar\alpha,\overline{K}}(q_{e+1}).$$

Lemma B.11. *The homomorphism $\boldsymbol{e}\hat H_{\bar\alpha,\overline{R}}(q_{e+1})\boldsymbol{e}\to\boldsymbol{e}\hat H_{\bar\alpha,\bar{\boldsymbol{k}}}(\zeta_{e+1})\boldsymbol{e}$ factors through a homomorphism $\widehat{SH}_{\bar\alpha,\overline{R}}(q_{e+1})\to\widehat{SH}_{\bar\alpha,\bar{\boldsymbol{k}}}(\zeta_{e+1})$.*

Proof. In Section 2E we constructed a faithful polynomial representation of $S_{\bar\alpha,\boldsymbol{k}}$. Let us call it $\mathcal{P}ol_{\boldsymbol{k}}$. It is constructed as a quotient of the standard polynomial representation of $\boldsymbol{e}R_{\bar\alpha,\boldsymbol{k}}\boldsymbol{e}$. After localization we get a faithful representation $\widehat{\mathcal{P}ol}_{\boldsymbol{k}}$ of $\hat S_{\bar\alpha,\boldsymbol{k}}$. Thus the kernel of the algebra homomorphism $\boldsymbol{e}\hat R_{\bar\alpha,\boldsymbol{k}}\boldsymbol{e}\to\hat S_{\bar\alpha,\boldsymbol{k}}$ is the annihilator of the representation $\widehat{\mathcal{P}ol}_{\boldsymbol{k}}$. We can transfer this to the Hecke side (because the isomorphism in Proposition B.6 comes from the identification of the polynomial representations) and we obtain that the kernel of the algebra homomorphism $\boldsymbol{e}\hat{H}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1})\boldsymbol{e} \to \widehat{SH}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1})$ is the annihilator of the representation $\widehat{\mathcal{P}ol}_{\boldsymbol{k}}$. Similarly, we can characterize the kernel of the algebra homomorphism $\boldsymbol{e}\hat{H}^{\leqslant N}_{\bar{\alpha},\bar{K}}(q_{e+1})\boldsymbol{e} \to \widehat{SH}^{\leqslant N}_{\bar{\alpha},\bar{K}}(q_{e+1})$ as the annihilator of a similar representation $\widehat{\mathcal{P}ol}^{\leqslant N}_{K}$.

The K-vector space $\widehat{\mathcal{P}ol}^{\leqslant N}_{K}$ has an R-submodule $\widehat{\mathcal{P}ol}_{R}$ stable by the action of $\boldsymbol{e}\hat{H}_{\bar{\alpha},\bar{R}}(q_{e+1})\boldsymbol{e}$ such that $\boldsymbol{k} \otimes_R \widehat{\mathcal{P}ol}_{R} = \widehat{\mathcal{P}ol}_{\boldsymbol{k}}$ and it is compatible with the algebra homomorphism $\boldsymbol{e}\hat{H}_{\bar{\alpha},\bar{R}}(q_{e+1})\boldsymbol{e} \to \boldsymbol{e}\hat{H}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1})\boldsymbol{e}$. By definition of $\widehat{SH}_{\bar{\alpha},\bar{R}}(q_{e+1})$ and the discussion above, the kernel of the algebra homomorphism $\boldsymbol{e}\hat{H}_{\bar{\alpha},\bar{R}}(q_{e+1})\boldsymbol{e} \to \widehat{SH}_{\bar{\alpha},\bar{R}}(q_{e+1})$ is formed by the elements that act by zero on $\widehat{\mathcal{P}ol}^{\leqslant N}_{K}$ (we assume that N is big enough). Thus each element of this kernel acts by zero on $\widehat{\mathcal{P}ol}_{R}$. This implies, that an element of the kernel of $\boldsymbol{e}\hat{H}_{\bar{\alpha},\bar{R}}(q_{e+1})\boldsymbol{e} \to \widehat{SH}_{\bar{\alpha},\bar{R}}(q_{e+1})$ specializes to an element of the kernel of $\boldsymbol{e}\hat{H}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1})\boldsymbol{e} \to \widehat{SH}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1})$. This proves the statement. □

B8. *The deformation of the isomorphism* Φ.

Proposition B.12. *There is a unique algebra homomorphism* $\Phi_{\alpha,R}\colon \hat{H}_{\alpha,R}(q_e) \to \widehat{SH}_{\bar{\alpha},R}(q_{e+1})$ *such that the following diagram is commutative*:

$$\begin{tikzcd}
\hat{H}_{\alpha,\boldsymbol{k}}(\zeta_e) \arrow[r, "\Phi_{\alpha,\boldsymbol{k}}"] & \widehat{SH}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1}) \\
\hat{H}_{\alpha,R}(q_e) \arrow[r, "\Phi_{\alpha,R}"] \arrow[u] \arrow[d] & \widehat{SH}_{\bar{\alpha},\bar{R}}(q_{e+1}) \arrow[u] \arrow[d] \\
\hat{H}^{\leqslant N}_{\alpha,K}(q_e) \arrow[r, "\Phi_{\alpha,K}"] & \widehat{SH}^{\leqslant N}_{\bar{\alpha},\bar{K}}(q_{e+1}).
\end{tikzcd}$$

Proof. First we consider the algebras $H^{\mathrm{loc}}_{\alpha,\boldsymbol{k}}(\zeta_e)$, $H^{\mathrm{loc}}_{\alpha,R}(q_e)$ and $H^{\mathrm{loc},\leqslant N}_{\alpha,K}(q_e)$ obtained from $\hat{H}_{\alpha,\boldsymbol{k}}(\zeta_e)$, $\hat{H}_{\alpha,R}(q_e)$ and $\hat{H}^{\leqslant N}_{\alpha,K}(q_e)$ by inverting

- $(X_r - X_t)$ and $(\zeta_e X_r - X_t)$ with $r \neq t$,
- $(X_r - X_t)$ and $(q_e X_r - X_t)$ with $r \neq t$,
- $(X_r - X_t)$ and $(q_e X_r - X_t)$ with $r \neq t$

respectively. Let $SH^{\mathrm{loc}}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1})$ and $SH^{\mathrm{loc},\leqslant N}_{\bar{\alpha},\bar{K}}(q_{e+1})$ be the localizations of $\widehat{SH}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1})$ and $\widehat{SH}^{\leqslant N}_{\bar{\alpha},\bar{K}}(q_{e+1})$ such that the isomorphisms $\Phi_{\alpha,\boldsymbol{k}}$ and $\Phi_{\alpha,K}$ above induce isomorphisms

$$\Phi_{\alpha,\boldsymbol{k}}\colon H^{\mathrm{loc}}_{\alpha,\boldsymbol{k}}(\zeta_e) \to SH^{\mathrm{loc}}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1}) \quad \text{and} \quad \Phi_{\alpha,K}\colon H^{\mathrm{loc},\leqslant N}_{\alpha,K}(q_e) \to SH^{\mathrm{loc},\leqslant N}_{\bar{\alpha},\bar{K}}(q_{e+1}).$$

Let $SH^{\mathrm{loc}}_{\bar{\alpha},\bar{R}}(q_{e+1})$ be the image in $SH^{\mathrm{loc},\leqslant N}_{\bar{\alpha},\bar{K}}(q_{e+1})$ of the following composition of homomorphisms

$$\boldsymbol{e}H^{\mathrm{loc}}_{\bar{\alpha},\bar{R}}(q_{e+1})\boldsymbol{e} \to \boldsymbol{e}H^{\mathrm{loc},\leqslant N}_{\bar{\alpha},\bar{K}}(q_{e+1})\boldsymbol{e} \to SH^{\mathrm{loc},\leqslant N}_{\bar{\alpha},\bar{K}}(q_{e+1}).$$

(We assume $N \geqslant 2d$. Then, similarly to Lemma B.9, the algebra $SH^{\mathrm{loc}}_{\bar{\alpha},\bar{R}}$ is independent of N under this assumption.)

Next, we want to prove that there exists an algebra homomorphism $\Phi_{\alpha,R}\colon H^{\mathrm{loc}}_{\alpha,R}(q_e) \to SH^{\mathrm{loc}}_{\bar{\alpha},\bar{R}}(q_{e+1})$ such that the following diagram is commutative:

$$\begin{tikzcd}
H^{\mathrm{loc}}_{\alpha,\boldsymbol{k}}(\zeta_e) \arrow[r, "\Phi_{\alpha,\boldsymbol{k}}"] & SH^{\mathrm{loc}}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1}) \\
H^{\mathrm{loc}}_{\alpha,R}(q_e) \arrow[r, "\Phi_{\alpha,R}"] \arrow[u] \arrow[d] & SH^{\mathrm{loc}}_{\bar{\alpha},\bar{R}}(q_{e+1}) \arrow[u] \arrow[d] \\
H^{\mathrm{loc},\leqslant N}_{\alpha,K}(q_e) \arrow[r, "\Phi_{\alpha,K}"] & SH^{\mathrm{loc},\leqslant N}_{\bar{\alpha},\bar{K}}(q_{e+1}).
\end{tikzcd} \tag{9}$$

We just need to check that the map $\Phi_{\alpha,K}$ takes an element of $H^{\mathrm{loc}}_{\alpha,R}(q_e)$ to an element of $SH^{\mathrm{loc}}_{\bar{\alpha},\bar{R}}(q_{e+1})$ and that it specializes to the map $\Phi_{\alpha,\boldsymbol{k}}\colon H^{\mathrm{loc}}_{\alpha,\boldsymbol{k}}(\zeta_e) \to SH^{\mathrm{loc}}_{\bar{\alpha},\bar{\boldsymbol{k}}}(\zeta_{e+1})$. We will check this on the generators $e(\boldsymbol{i})$, $X_r e(\boldsymbol{i})$ and $\Psi_r e(\boldsymbol{i})$ of $H^{\mathrm{loc}}_{\alpha,R}(q_e)$.

This is obvious for the idempotents $e(\boldsymbol{i})$.

Let us check this for $X_r e(\boldsymbol{i})$. Assume that $\boldsymbol{i} \in I^\alpha$ and $\boldsymbol{j} \in \tilde{I}^{|\alpha|}$ are such that we have $\pi_e(\boldsymbol{j}) = \boldsymbol{i}$. Write $\boldsymbol{i}' = \phi(\boldsymbol{i})$ and $\boldsymbol{j}' = \tilde{\phi}(\boldsymbol{j})$. Set $r' = r'_{\boldsymbol{j}} = r'_{\boldsymbol{i}}$, see the notation in Section 2F. By Theorem 2.12 and Proposition B.5 we have

$$\Phi_{\alpha,K}(X_r e(\boldsymbol{j})) = \bar{p}^{-1}_{j'_{r'}} p_{j_r} X_{r'} e(\boldsymbol{j}').$$

Since, $\bar{p}^{-1}_{j'_{r'}} p_{j_r}$ depends only on $\boldsymbol{i}$ and r and $e(\boldsymbol{i}) = \sum_{\pi_e(\boldsymbol{j})=\boldsymbol{i}} e(\boldsymbol{j})$, we deduce that

$$\Phi_{\alpha,K}(X_r e(\boldsymbol{i})) = \bar{p}^{-1}_{j'_{r'}} p_{j_r} X_{r'} e(\boldsymbol{i}').$$

Thus the element $\Phi_{\alpha,K}(X_r e(\boldsymbol{i}))$ is in $SH^{\mathrm{loc}}_{\bar{\alpha},R}$ and its image in $SH^{\mathrm{loc}}_{\bar{\alpha},\boldsymbol{k}}$ is $\bar{p}^{-1}_{i'_{r'}} p_{i_r} X_{r'} e(\boldsymbol{i}') = \Phi_{\alpha,\boldsymbol{k}}(X_r e(\boldsymbol{i}))$.

Next, we consider the generators $\Psi_r e(\boldsymbol{i})$. We must prove that for each $\boldsymbol{j}$ such that $\pi_e(\boldsymbol{j}) = \boldsymbol{i}$ and for each r we have

- $\Phi_{\alpha,K}(\Psi_r e(\boldsymbol{j})) = \Xi e(\boldsymbol{j}')$, for some element $\Xi \in H^{\mathrm{loc}}_{\alpha,R}(q_e)$ that depends only on r and $\boldsymbol{i}$,
- the image of $\Xi e(\boldsymbol{i}')$ in $SH^{\mathrm{loc}}_{\bar{\alpha},\bar{\boldsymbol{k}}}(q_{e+1})$ under the specialization $R \to \boldsymbol{k}$ is $\Phi_{\alpha,\boldsymbol{k}}(\Psi_r e(\boldsymbol{i}))$.

This follows from Lemma B.1.

Now we obtain the diagram from the claim of Proposition B.12 as the restriction of the diagram (9). □

B9. ***Alternative definition of a categorical representation.*** There is an alternative definition of a categorical representation, where the KLR algebra is replaced by the affine Hecke algebra.

Let R be a $\mathbb{C}$-algebra. Fix an invertible element $q \in R$, $q \neq 1$. Let $\mathcal{C}$ be an R-linear exact category.

Definition B.13. A *representation datum* in $\mathcal{C}$ is a tuple (E, F, X, T) where (E, F) is a pair of exact biadjoint functors $\mathcal{C} \to \mathcal{C}$ and $X \in \mathrm{End}(F)^{\mathrm{op}}$ and $T \in \mathrm{End}(F^2)^{\mathrm{op}}$ are endomorphisms of functors such that for each $d \in \mathbb{N}$, there is an R-algebra homomorphism $\psi_d \colon H_{d,R}(q) \to \mathrm{End}(F^d)^{\mathrm{op}}$ given by

$$\begin{aligned} X_r &\mapsto F^{d-r} X F^{r-1} && \forall r \in [1, d], \\ T_r &\mapsto F^{d-r-1} T F^{r-1} && \forall r \in [1, d-1]. \end{aligned}$$

Now, assume that $R = \boldsymbol{k}$ is a field. Assume that $\mathcal{C}$ is a Hom-finite $\boldsymbol{k}$-linear abelian category. Let $\mathscr{F}$ be a subset of $\boldsymbol{k}^\times$ (possibly infinite). As in Section B4, we view $\mathscr{F}$ as the vertex set of a quiver with an arrow $i \to j$ if and only if $j = qi$.

Definition B.14. A $\mathfrak{g}_{\mathscr{F}}$-categorical representation in $\mathcal{C}$ is the datum of a representation datum (E, F, X, T) and a decomposition $\mathcal{C} = \bigoplus_{\mu \in X_{\mathscr{F}}} \mathcal{C}_\mu$ satisfying the conditions (a) and (b) below. For $i \in \mathscr{F}$, let E_i and F_i be endofunctors of $\mathcal{C}$ such that for each $M \in \mathcal{C}$ the objects $E_i(M)$ and $F_i(M)$ are the generalized i-eigenspaces of X acting on $E(M)$ and $F(M)$ respectively, see also Remark 3.3 (a). We assume

(a) $F = \bigoplus_{i \in \mathscr{F}} F_i$ and $E = \bigoplus_{i \in \mathscr{F}} E_i$,

(b) $E_i(\mathcal{C}_\mu) \subset \mathcal{C}_{\mu+\alpha_i}$ and $F_i(\mathcal{C}_\mu) \subset \mathcal{C}_{\mu-\alpha_i}$.

If the set $\mathscr{F}$ is infinite, condition (a) should be understood in the same way as in Remark 3.3 (b).

Remark B.15. (a) By definition, for each object $M \in \mathcal{C}$ and each $d \in \mathbb{Z}_{\geqslant 0}$, we have $F_{i_d} \cdots F_{i_1}(M) \neq 0$ only for a finite number of sequences $(i_1, \dots, i_d) \in \mathcal{F}^d$. (Else, the endomorphism algebra of $F^d(M)$ is infinite-dimensional.) Then the homomorphism $H_{d,\boldsymbol{k}}(q) \to \mathrm{End}(F^d(M))^{\mathrm{op}}$ extends to a homomorphism $\hat{H}_{d,\boldsymbol{k}}(q) \to \mathrm{End}(F^d(M))^{\mathrm{op}}$ such that only a finite number of idempotents $e(\boldsymbol{j})$ has a nonzero image. (We define the action of $e(\boldsymbol{i})$ as the projection from F^d to $F_{i_d} \cdots F_{i_1}$. Note that the action of $(X_r - X_t)^{-1} e(\boldsymbol{i})$ such that $i_r \neq i_t$ is well defined because X_r and X_t have different eigenvalues. Similarly, the action of $(qX_r - X_t)^{-1} e(\boldsymbol{i})$ such that $r \neq t$ and $qi_r \neq i_t$ is well defined.) In particular, we obtain a homomorphism $\hat{H}_{d,\boldsymbol{k}}(q) \to \mathrm{End}(F^d)^{\mathrm{op}}$.

(b) As in part (a), if we have a categorical representation in the sense of Definition 3.2, then the homomorphism $R_{d,\boldsymbol{k}} \to \mathrm{End}(F^d)^{\mathrm{op}}$ extends to a homomorphism $\hat{R}_{d,\boldsymbol{k}} \to \mathrm{End}(F^d)^{\mathrm{op}}$. Then Proposition B.6 implies that the two definitions of a categorical representation of $\mathfrak{g}_{\mathcal{F}}$ (Definitions 3.2 and B.14) are equivalent.

B10. ***Categorical representations over R.*** We assume that the ring R is as in Section B6. We are going to obtain an analogue of Theorem 3.5 over R.

Let $\mathcal{C}_R$, $\mathcal{C}_{\boldsymbol{k}}$ and $\mathcal{C}_K$ be R-, $\boldsymbol{k}$- and K-linear categories, respectively. Assume that $\mathcal{C}_{\boldsymbol{k}}$ and $\mathcal{C}_K$ are Hom-finite $\boldsymbol{k}$-linear and K-linear abelian categories, respectively. Assume that the category $\mathcal{C}_R$ is exact. Fix R-linear functors $\Omega_{\boldsymbol{k}} \colon \mathcal{C}_R \to \mathcal{C}_{\boldsymbol{k}}$ and $\Omega_K \colon \mathcal{C}_R \to \mathcal{C}_K$.

Remark B.16. The first example of a situation as above that we should imagine is the following. Let A be an R-algebra that is finitely generated as an R-module. We set $\mathcal{C}_R =$ mod (A), $\mathcal{C}_{\boldsymbol{k}} =$ mod $(\boldsymbol{k} \otimes_R A)$, $\mathcal{C}_K =$ mod $(K \otimes_R A)$, $\Omega_{\boldsymbol{k}} = \boldsymbol{k} \otimes \bullet$ and $\Omega_K = K \otimes \bullet$.

Another interesting situation (that in fact motivated the result of this section) is when $\mathcal{C}_B$, for $B \in \{R, \boldsymbol{k}, H\}$, is the category $\mathcal{O}$ for $\widehat{\mathfrak{gl}}_N$ over B at a negative level. We do not want to assume in this section that the category $\mathcal{C}_R$ is abelian because [Rouquier et al. 2016] constructs a categorical representation only in the Δ-filtered category $\mathcal{O}$ over R (and not in the whole abelian category $\mathcal{O}$ over R).

Definition B.17. A categorical representation of $(\widetilde{\mathfrak{sl}}_e, \mathfrak{sl}_\infty^{\oplus l})$ in $(\mathcal{C}_R, \mathcal{C}_{\boldsymbol{k}}, \mathcal{C}_K)$ is the following data

(1) a categorical representation of $\mathfrak{g}_I = \widetilde{\mathfrak{sl}}_e$ in $\mathcal{C}_{\boldsymbol{k}}$,

(2) a categorical representation of $\mathfrak{g}_{\tilde{I}} = \mathfrak{sl}_\infty^{\oplus l}$ in $\mathcal{C}_K$,

(3) a representation datum (E, F) in $\mathcal{C}_R$ (with respect to the Hecke algebra $H_{d,R}(q_e)$) such that the functors E and F commute with $\Omega_{\boldsymbol{k}}$ and Ω_K,

(4) lifts (with respect to $\Omega_{\boldsymbol{k}}$) of decompositions $E = \bigoplus_{i\in I} E_i$, $F = \bigoplus_{i\in I} F_i$ and $\mathcal{C}_{\boldsymbol{k}} = \bigoplus_{X_I} \mathcal{C}_{\boldsymbol{k},\mu}$ from $\mathcal{C}_{\boldsymbol{k}}$ to $\mathcal{C}_R$

such that the following compatibility conditions are satisfied:

- The decomposition $\mathcal{C}_R = \bigoplus_{\mu\in X_e} \mathcal{C}_{R,\mu}$ is compatible with the decomposition $\mathcal{C}_K = \bigoplus_{\tilde{\mu}\in X_{\tilde{I}}} \mathcal{C}_{K,\tilde{\mu}}$ (i.e., we have $\Omega_K(\mathcal{C}_{R,\mu}) \subset \bigoplus_{\pi_e(\tilde{\mu})=\mu} \mathcal{C}_{K,\tilde{\mu}}$).
- The decompositions $E = \bigoplus_{i\in I} E_i$ and $F = \bigoplus_{i\in I} F_i$ in $\mathcal{C}_R$ are compatible with the decompositions $E = \bigoplus_{j\in\tilde{I}} E_j$ and $F = \bigoplus_{j\in\tilde{I}} F_j$ in $\mathcal{C}_K$ with respect to Ω_K (i.e., the functors $E_i = \bigoplus_{j\in\tilde{I},\pi_e(j)=i} E_j$ and $F_i = \bigoplus_{j\in\tilde{I},\pi_e(j)=i} F_j$ for $\mathcal{C}_K$ correspond to the functors E_i, F_i for $\mathcal{C}_R$).
- The actions of the Hecke algebras $H_{d,R}(q_e)$, $H_{d,\boldsymbol{k}}(\zeta_e)$ and $H_{d,K}(q_e)$ on $\mathrm{End}(F^d)^{\mathrm{op}}$ for $\mathcal{C}_R$, $\mathcal{C}_{\boldsymbol{k}}$ and $\mathcal{C}_K$ are compatible with $\Omega_{\boldsymbol{k}}$ and Ω_K.

Proposition B.12 yields the following version of Theorem 3.5 over R.

Let $(\overline{\mathcal{C}}_R, \overline{\mathcal{C}}_{\boldsymbol{k}}, \overline{\mathcal{C}}_K)$ be a categorical representation of $(\widetilde{\mathfrak{sl}}_{e+1}, \mathfrak{sl}_\infty^{\oplus l})$. Assume that for each $\mu \in X_{\bar{I}}\backslash X_{\bar{I}}^+$ we have $\overline{\mathcal{C}}_{\boldsymbol{k},\mu} = \overline{\mathcal{C}}_{R,\mu} = 0$ and the for each $\tilde{\mu} \in X_{\tilde{I}}\backslash X_{\tilde{I}}^+$ we have $\overline{\mathcal{C}}_{K,\tilde{\mu}} = 0$. Let $\mathcal{C}_R$, $\mathcal{C}_{\boldsymbol{k}}$ and $\mathcal{C}_K$ be the subcategory of $\overline{\mathcal{C}}_R$, $\overline{\mathcal{C}}_{\boldsymbol{k}}$ and $\overline{\mathcal{C}}_K$, respectively, defined in the same way as in Section 3D. Then we have the following.

Theorem B.18. *There is a categorical representation of* $(\widetilde{\mathfrak{sl}}_e, \mathfrak{sl}_\infty^{\oplus l})$ *in* $(\mathcal{C}_R, \mathcal{C}_{\boldsymbol{k}}, \mathcal{C}_K)$.

Proof. We obtain a categorical representation of $\widetilde{\mathfrak{sl}}_e$ in $\mathcal{C}_{\boldsymbol{k}}$ by Theorem 3.5. A similar argument as in the proof of Theorem 3.5 yields a categorical representation of $\mathfrak{sl}_\infty^{\oplus l}$ in $\mathcal{C}_K$ (we just have to replace the isomorphism Φ from Section 2G associated with the quiver Γ_e by a similar isomorphism associated with the quiver $\tilde{\Gamma}$). To construct a representation datum in $\mathcal{C}_R$, we use the homomorphism $\Phi_{\alpha,R}$ from Proposition B.12. All axioms of a $(\widetilde{\mathfrak{sl}}_e, \mathfrak{sl}_\infty^{\oplus l})$-categorical representation in $(\mathcal{C}_R, \mathcal{C}_{\boldsymbol{k}}, \mathcal{C}_K)$ follow automatically from the axioms of a categorical representation of $(\widetilde{\mathfrak{sl}}_{e+1}, \mathfrak{sl}_\infty^{\oplus l})$ in $(\overline{\mathcal{C}}_R, \overline{\mathcal{C}}_{\boldsymbol{k}}, \overline{\mathcal{C}}_K)$. □

Acknowledgements

I am grateful for the hospitality of the Max-Planck-Institut für Mathematik in Bonn, where a big part of this work is done. I would like to thank Éric Vasserot for his guidance and helpful discussions during my work on this paper. I would like to thank Alexander Kleshchev for useful discussions about KLR algebras. I would like to thank Cédric Bonnafé for his comments on an earlier version of this paper. I would also like to thank the anonymous reviewer for the careful reading and constructive comments.

References

[Chriss and Ginzburg 1997] N. Chriss and V. Ginzburg, *Representation theory and complex geometry*, Birkhäuser, Boston, 1997. MR Zbl

[Chuang and Rouquier 2008] J. Chuang and R. Rouquier, "Derived equivalences for symmetric groups and $\mathfrak{sl}_2$-categorification", *Ann. of Math.* (2) **167**:1 (2008), 245–298. MR Zbl

[Khovanov and Lauda 2009] M. Khovanov and A. D. Lauda, "A diagrammatic approach to categorification of quantum groups, I", *Represent. Theory* **13** (2009), 309–347. MR Zbl

[Maksimau 2015a] R. Maksimau, "Canonical basis, KLR algebras and parity sheaves", *J. Algebra* **422** (2015), 563–610. MR Zbl

[Maksimau 2015b] R. Maksimau, "Categorical representations, KLR algebras and Koszul duality", preprint, 2015. arXiv

[Miemietz and Stroppel 2016] V. Miemietz and C. Stroppel, "Affine quiver Schur algebras and p-adic GL_n", preprint, 2016. arXiv

[Riche and Williamson 2018] S. Riche and G. Williamson, *Tilting modules and the p-canonical basis*, Astérisque **397**, Société Mathématique de France, Paris, 2018. MR Zbl

[Rouquier 2008] R. Rouquier, "2-Kac–Moody algebras", preprint, 2008. arXiv

[Rouquier et al. 2016] R. Rouquier, P. Shan, M. Varagnolo, and E. Vasserot, "Categorifications and cyclotomic rational double affine Hecke algebras", *Invent. Math.* **204**:3 (2016), 671–786. MR Zbl

[Varagnolo and Vasserot 2011] M. Varagnolo and E. Vasserot, "Canonical bases and KLR-algebras", *J. Reine Angew. Math.* **659** (2011), 67–100. MR Zbl

Communicated by Georgia Benkart
Received 2016-05-02 Revised 2018-02-14 Accepted 2018-06-02

ruslmax@gmail.com
Institut Montpelliérain Alexander Grothendieck, Université de Montpellier, Montpellier, France

ALGEBRA AND NUMBER THEORY 12:8 (2018)
dx.doi.org/10.2140/ant.2018.12.1923

On nonprimitive Weierstrass points

Nathan Pflueger

We give an upper bound for the codimension in $\mathcal{M}_{g,1}$ of the variety $\mathcal{M}^S_{G,1}$ of marked curves (C, p) with a given Weierstrass semigroup. The bound is a combinatorial quantity which we call the effective weight of the semigroup; it is a refinement of the weight of the semigroup, and differs from the weight precisely when the semigroup is not primitive. We prove that whenever the effective weight is less than g, the variety $\mathcal{M}^S_{G,1}$ is nonempty and has a component of the predicted codimension. These results extend previous results of Eisenbud, Harris, and Komeda to the case of nonprimitive semigroups. We also survey other cases where the codimension of $\mathcal{M}^S_{G,1}$ is known, as evidence that the effective weight estimate is correct in wider circumstances.

1. Introduction

Given a point p on a smooth curve C of genus g, there is an associated numerical semigroup

$$S(C, p) = \{-\mathrm{val}_p(f) : f \in \Gamma(C \setminus \{p\}, \mathcal{O}_C)\},$$

given by the pole orders of rational functions with no poles away from p. Weierstrass's *Lückensatz* (now an easy consequence of the Riemann–Roch formula) states that there are exactly g gaps in $S(C, p)$.[1]

In reverse, any numerical semigroup S with g gaps defines a (not necessarily closed) subvariety $\mathcal{M}^S_{g,1} \subseteq \mathcal{M}_{g,1}$ of the moduli space of curves with a marked point. These loci stratify $\mathcal{M}_{g,1}$, with the locus defined by the *ordinary* semigroup $H_g = \{0, g+1, g+2, \dots\}$ dense and open, and the value of the i-th gap ($i = 1, 2, \dots, g$) an upper semicontinuous function.

The link between the combinatorics of these numerical semigroups and the geometry of curves and their moduli is a wide and fascinating story that remains largely mysterious, though many intriguing special cases (specific types of semigroups) are well understood. The core of the difficulty (and excitement) in this story lies in the fact that $S(C, p)$ is not an arbitrary sequence of integers, but a semigroup; this combinatorial restriction reflects itself in the geometry of the stratification.

Our objective is to propose a partial answer to a basic question: given a semigroup S, what is the codimension of $\mathcal{M}^S_{g,1}$ in $\mathcal{M}_{g,1}$?

MSC2010: 14H55.
Keywords: Weierstrass points, numerical semigroups, algebraic curves, limit linear series.

[1] The author has heard conflicting stories about whether the number of gaps in a numerical semigroup is called the "genus" due to this fact from geometry, or as a joking reference to the "number of holes" in the semigroup.

Definition 1.1. The *effective weight* of a numerical semigroup S is

$$\operatorname{ewt}(S) = \sum_{\text{gaps } b} (\#\text{ generators } a < b).$$

Alternatively, $\operatorname{ewt}(S)$ is the number of pairs (a, b), where $0 < a < b$, a is a generator, and b is a gap.

In almost every situation where $\operatorname{codim} \mathcal{M}^S_{g,1}$ is known for an explicit family of semigroups (as well as for all semigroups of genus up to 6), it is equal to $\operatorname{ewt}(S)$; we summarize a number of these cases in Section 2. The first genus in which the author is aware of a semigroup with $\operatorname{codim} \mathcal{M}^S_{g,1} < \operatorname{ewt}(S)$ is $g = 9$ (the example is discussed in Section 2F).

Our main results are the following, which give much stronger evidence for the utility of $\operatorname{ewt}(S)$ in the study of this stratification of $\mathcal{M}_{g,1}$.

Theorem 1.2. *If $\mathcal{M}^S_{g,1}$ is nonempty and X is any irreducible component of it, then*

$$\dim X \geq \dim \mathcal{M}_{g,1} - \operatorname{ewt}(S).$$

We call a point or irreducible component of $\mathcal{M}^S_{g,1}$ *effectively proper* if the local dimension of $\mathcal{M}^S_{g,1}$ is exactly $\dim \mathcal{M}_{g,1} - \operatorname{ewt}(S)$.

Theorem 1.3. *If S is a genus g numerical semigroup with* $\operatorname{ewt}(S) \leq g - 2$, *then $\mathcal{M}^S_{g,1}$ has an effectively proper component. If* $\operatorname{char} k = 0$, *then the same is true for all numerical semigroups with* $\operatorname{ewt}(S) \leq g - 1$.

The effective weight is a refinement of a more naive quantity, the *weight* of a semigroup, and the two quantities are equal for S if and only if S is *primitive*, meaning that the sum of any two nonzero elements is greater than the largest gap (see Section 2A).

Theorems 1.2 and 1.3 were originally proved, with a characteristic 0 hypothesis, for primitive semigroups (using the weight) by Eisenbud and Harris [1987] and Komeda [1991]. Our proofs are based on theirs, using the theory of limit linear series as the central technical tool. Our primary innovation is to apply the machinery of limit linear series to produce *incomplete* linear series with specified vanishing data on smooth curves. The basic technique is the same: curves with Weierstrass semigroups of genus g are constructed by choosing a suitable genus $g-1$ semigroup and a marked curve realizing it, attaching an elliptic curve at the Weierstrass point, marking a second point on the elliptic curve differing by torsion, and deforming the resulting nodal curve.

Remark 1.4. The choice of terminology "effective weight" was made in reference to terminology from the numerical semigroup literature. The set of all numerical semigroups can be arranged in a rooted tree, with each level corresponding to a different genus, where the parent of a semigroup S is given by adding the largest gap back into S. The children of a given semigroup S correspond to the "effective generators" of S, which are defined to be the generators that are larger than the largest gap. For details, and a study of the structure of this tree, see [Bras-Amorós and Bulygin 2009]. The effective weight of S is determined by examining, in the path from the root (genus 0 semigroup) to S, the index of the effective generator removed at each step (when the effective generators are listed in increasing order). If a similar procedure were followed, arranging all cofinite subsets of $\mathbb{N}$ into a tree (not just semigroups), then the quantity constructed in the same way would be the weight, rather than the effective weight.

1A. ***Speculation and conjectures.*** While there are semigroups S for which $\operatorname{codim} \mathcal{M}^S_{g,1} < \operatorname{ewt}(S)$, to the author's knowledge all such examples fall in the range $g \le \operatorname{codim} \mathcal{M}^S_{g,1} \le 2g$. Therefore, we (somewhat speculatively) conjecture that no such semigroups exist in codimension less than g.

Conjecture 1.5. *If $\mathcal{M}^S_{g,1}$ has a component of codimension less than g in $\mathcal{M}_{g,1}$, then all components of $\mathcal{M}^S_{g,1}$ have codimension exactly* $\operatorname{ewt}(S)$.

Curiously, we are not aware of any numerical semigroups of any genus for which $\operatorname{codim} \mathcal{M}^S_{g,1} > 2g$. Therefore we also make the following (equally speculative) conjecture.

Conjecture 1.6. *For any numerical semigroup such that $\mathcal{M}^S_{g,1} \neq \varnothing$, all components of $\mathcal{M}^S_{g,1}$ have codimension at most* $2g$.

Note that in the above conjectures, g and $2g$ perhaps ought to be replaced with $g + C_1$ and $2g + C_2$ for some constants C_1 and C_2, the value of which we have no strong beliefs about. We have stated the conjectures as above merely to make them specific.

Although not relevant to the present paper, we also mention a purely combinatorial conjecture about the effective weight that arose during this work. We have verified this conjecture by a computer search[2] up to genus 50.

Conjecture 1.7. *For any numerical semigroup of genus g,*

$$\operatorname{ewt}(S) \le \left\lfloor \frac{(g+1)^2}{8} \right\rfloor.$$

Remark 1.8. If true, this conjecture is sharp. For $g \le 5$, this follows from case analysis. For $g \ge 6$, this follows from a general construction. Let η be an integer between -2 and 2 inclusive such that $\eta \equiv g + 1 \pmod 4$ (there are two choices if $g \equiv 1 \pmod 4$, and one otherwise). Let $c = \frac{1}{4}(3g + 3 + \eta)$ and $d = \frac{1}{4}(5g + 1 + 3\eta)$. Then the semigroup

$$S = \langle c, c+1, \dots, d-1, d \rangle = \mathbb{N} \setminus \{1, 2, \dots, c-1, d+1, d+2, \dots, 2c-1\}$$

has genus g and effective weight $\frac{1}{8}(g+1)^2 - \frac{1}{8}\eta^2 = \left\lfloor \frac{1}{8}(g+1)^2 \right\rfloor$. For $10 \le g \le 50$, a computer search shows that these are the only semigroups of this effective weight, while for $g \le 9$ there are some additional sporadic examples achieving the same maximum.

The semigroups above (that appear, empirically, to maximize $\operatorname{ewt}(S)$ in a given genus) also provide examples where $\operatorname{codim} \mathcal{M}^S_{g,1} < \operatorname{ewt}(S)$ [Pflueger 2016].

[2] C++ source code is available on the author's website. The search took approximately 17 hours on a 3.4Ghz Intel i7-3770 CPU.

Outline of the paper

We summarize in Section 2 several cases where $\operatorname{codim} \mathcal{M}^S_{g,1}$ is known in the literature, including the simplest case where strict inequality $\operatorname{codim} \mathcal{M}^S_{g,1} < \operatorname{ewt}(S)$ occurs. Section 3 summarizes background on linear series and limit linear series needed for the proofs of the main theorems. Theorem 1.2 is proved in Section 4. Section 5 is purely combinatorial, and provides some preliminary results on the structure of numerical semigroups of low effective weight. Section 6 gives the proof of Theorem 1.3.

Conventions

Throughout this paper, we work over an algebraically closed field k. In Section 2, we assume $\operatorname{char} k = 0$. A *point* of a scheme will always refer to a closed point, and when we say that a *general point* of a scheme satisfies a property, we mean that there exists a dense open subset in which all points satisfy the property. A *curve* is always reduced, connected, and complete. A *marked curve* is a pair (C, p) of a curve C and a point $p \in C$.

We denote by $\mathbb{N}$ the set of nonnegative integers; a *numerical semigroup* is a cofinite subset $S \subseteq \mathbb{N}$ containing 0 and closed under addition. The elements of $\mathbb{N} \setminus S$ are called the *gaps* of S, and the number of gaps is called the *genus*. A positive element of S that is not equal to the sum of two positive elements of S is called a *generator*, and a sum of two positive elements is called *composite*.

We denote the set $\{0, g+1, g+2, \ldots\}$ by H_g, which we call the *ordinary semigroup of genus g*.

2. Background

The classification question of Weierstrass points can be asked on various levels. Hurwitz [1892] first raised the simple existence question, while we are concerned with the more geometric dimension question.

Question 2.1. For which S is $\mathcal{M}^S_{g,1} \neq \varnothing$?

Question 2.2. Given S, how many irreducible components does $\mathcal{M}^S_{g,1}$ have? What are their codimensions?

Question 2.3. Given S, what is the maximum codimension of an irreducible component of $\mathcal{M}^S_{g,1}$?

The number of semigroups of genus g grows exponentially with g with limiting ratio $\frac{1+\sqrt{5}}{2}$ [Zhai 2013], and present knowledge, even about Question 2.1, becomes quite sparse for large genus if all semigroups are considered (see [Kaplan and Ye 2013]). This is one reason we prefer to focus on Questions 2.2 and 2.3: if one hopes for general results, matters become much more tractable upon restricting to the more plentiful sorts of semigroups, i.e., those for which $\operatorname{codim} \mathcal{M}^S_{g,1}$ is small compared to g.

This restriction, to semigroups expected to appear in low codimension, is what allowed Eisenbud and Harris to prove their rather strong results, later extended by Komeda. The downside of their results is that they needed to impose not just a quantitative restriction (weight being less than g) but a qualitative one: that the semigroup is primitive.

In the remainder of this section, we summarize some known results and simple cases of answers to these questions, in order to highlight the extent to which the effective weight brings many known cases under one umbrella. We conclude in Section 2F, however, with the first example we know in which the effective weight does not give the correct codimension.

Throughout this background section, we will assume that $\operatorname{char} k = 0$, as much of the literature makes this assumption.

2A. ***The work of Eisenbud, Harris and Komeda.*** The *weight* of a numerical semigroup is most simply defined as the sum of the gaps minus $\binom{g+1}{2}$. An alternate description, more suggestive of the link to the effective weight, is that $\operatorname{wt}(S)$ is the number of pairs (a, b) where $0 < a < b$, $a \in S$, and $b \notin S$. This description shows that $\operatorname{wt}(S) - \operatorname{ewt}(S)$ is equal to the number of pairs (a, b) where $a < b$, a is *composite*, and b is a gap; hence $\operatorname{wt}(S) = \operatorname{ewt}(S)$ if and only if S is primitive.

All semigroups satisfy $\operatorname{codim} \mathcal{M}^S_{g,1} \leq \operatorname{wt}(S)$ (see Remark 4.3 for one argument), and a point of $\mathcal{M}^S_{g,1}$ at which equality holds locally is called *dimensionally proper*. Eisenbud and Harris [1987] proved that if S is primitive and $\operatorname{wt}(S) \leq g - 2$, then $\mathcal{M}^S_{g,1}$ has dimensionally proper points. Their proof made a characteristic 0 assumption, since this assumption was built into their theory of limit linear series developed in [Eisenbud and Harris 1986], but modern treatments of limit linear series (e.g., [Osserman 2006; 2013]) make no such assumption. The proofs of Eisenbud and Harris [1987] can therefore be carried to characteristic p with no modification.

The argument of Eisenbud and Harris [1987] proceeds by induction on g. It nearly succeeds in proving the same result for primitive semigroups with $\operatorname{wt}(S) \leq g - 1$ (rather than $g - 2$), except that the inductive step fails for one very specific class of semigroups of weight $g - 1$. Komeda's contribution [1991] is to prove the theorem in this special case by a different argument, without limit linear series (and with a characteristic 0 hypothesis), thus extending the results of Eisenbud and Harris [1987] to the case $\operatorname{wt}(S) = g - 1$. A second argument for this special case appears in [Coppens and Kato 1994].

Eisenbud and Harris observe [1987, Corollary on p. 497] that the primitivity hypothesis is necessary, i.e., $\operatorname{codim} \mathcal{M}^S_{g,1} < \operatorname{wt}(S)$ if S is nonprimitive. This fact of course also now follows from our Theorem 1.2. This is no minor difficulty, as many semigroups, including those that appear with low codimension in the Weierstrass stratification of $\mathcal{M}_{g,1}$, are not primitive. The main example, which provided substantial motivation regarding how to refine $\operatorname{wt}(S)$, is the following.

Example 2.4. A hyperelliptic curve of genus g has $2g + 2$ points with semigroup $\{2, 4, 6, \ldots, 2g - 2\} \cup H_{2g} = \langle 2, 2g + 1 \rangle$ (the ramification points of the double cover of $\mathbb{P}^1$), while the rest of the points have the ordinary semigroup (see e.g., [Arbarello et al. 1985, Exercise I.E-3]). Furthermore, if $2 \in S(C, p)$ then C is necessarily hyperelliptic. Hence $S = \langle 2, 2g + 1 \rangle$ is called the *hyperelliptic semigroup*, and $\operatorname{codim} \mathcal{M}^S_{g,1} = g - 1$ (the codimension of the hyperelliptic locus in $\mathcal{M}_g$ plus 1).

The hyperelliptic semigroup has the distinction of having the maximum weight of all genus g semigroups, namely $\binom{g}{2}$. So the weight bound is spectacularly off in this case. However $\operatorname{ewt}(S) = g - 1$.

Remark 2.5. Since the semigroups of maximum weight provide a nice example where the weight bound fails to be exact (and suggested the definition of the effective weight), it seems reasonable to try to find cases where the effective weight bound fails to be exact in the semigroups of maximum effective weight.

Indeed, these semigroups provide such examples; see [Pflueger 2016]. See also Conjecture 1.7 and the remark following it.

One notable extension of Eisenbud and Harris's result, and method of proof, was given by Bullock [2013]. Using a variation on Eisenbud and Harris's inductive argument, Bullock proves that for the nonprimitive semigroup

$$S = \{0, g-1, g+1, g+2, \dots, 2g-2\} \cup H_{2g}$$

of weight g, the locus $\mathcal{M}^S_{g,1}$ is irreducible of codimension $g-1$. The manner in which Bullock treated a nonprimitive semigroup with Eisenbud and Harris's basic method provided inspiration for our method in proving the more general Theorem 1.3. Note that S is "barely nonprimitive," as there is only one gap exceeding one composite element. See Remark 2.9 for more about Bullock's work.

2B. ***The Deligne bound and negatively graded semigroups.*** The best general-purpose *lower* bound on $\operatorname{codim} \mathcal{M}^S_{g,1}$ is the Deligne bound, defined below.

Definition 2.6. For any numerical semigroup S, let $\lambda(S)$ be the number of gaps $b \notin S$ such that $b+a \in S$ for all positive elements $a \in S$.

Proposition 2.7. *Let S be a numerical semigroup of genus g. If $\mathcal{M}^S_{g,1}$ is nonempty, then*

$$\operatorname{codim} \mathcal{M}^S_{g,1} \geq g - \lambda(S).$$

Proof. This bound follows from results of Deligne [SGA 7_{II} 1973], first applied to the moduli of Weierstrass points by Pinkham [1974, Theorems 10.3 and 13.9]. For a discussion of the bound in this form, see [Rim and Vitulli 1977, Corollary 6.3]. □

In most cases, the Deligne bound and the effective weight bound do not coincide. Interestingly, the cases where they do coincide are semigroups of a particular structure: they are the "negatively graded semigroups" studied by Rim and Vitulli. Rim and Vitulli [1977, Theorem 4.7] prove that a semigroup is negatively graded (a deformation-theoretic condition) if and only if it is one of the following.

Definition 2.8. Let $g \geq 2$ be a positive integer. For each integer e between 1 and $g-1$ inclusive, define:

$$\begin{aligned} \mathrm{NG}^1_{g,e} &= (g-e+1)\cdot\mathbb{Z} \cup H_c, \quad \text{where } c = g + \lfloor g/(g-e) \rfloor, \\ \mathrm{NG}^2_{g,e} &= \{0, g, g+1, \dots, g+e-1\} \cup H_{g+e}, \end{aligned}$$

and also define, for $g \geq 3$,

$$\mathrm{NG}^3_g = \{0, g-1, g+1, g+2, \dots, 2g-2\} \cup H_{2g-1}.$$

Observe that $\operatorname{ewt}(\mathrm{NG}^1_{g,e}) = \operatorname{ewt}(\mathrm{NG}^2_{g,e}) = e$, and $\operatorname{ewt}(\mathrm{NG}^3_g) = g-1$. Note that $\mathrm{NG}^1_{g,1} = \mathrm{NG}^2_{g,1}$ and $\mathrm{NG}^1_{3,2} = \mathrm{NG}^3_3$, but in all other cases the semigroups described above are distinct. Therefore for $g \geq 4$, there is one negatively graded semigroup of effective weight 1, two negatively graded semigroups of effective weight e for $2 \leq e \leq g-2$, and three negatively graded semigroups of effective weight $g-1$.

Half of the semigroups $\mathrm{NG}^1_{g,e}$ $\left(\text{those for which } e \le \frac{g}{2}\right)$, and all of the semigroups $\mathrm{NG}^2_{g,e}$ are primitive, while NG^3_g and the other half of the $\mathrm{NG}^1_{g,e}$ are not.

Remark 2.9. The three semigroups $\mathrm{NG}^1_{g,g-1}$, $\mathrm{NG}^2_{g,g-1}$, and NG^3_g of effective weight $g-1$ were studied by Bullock [2013]; they correspond to the three irreducible components of the locus of "subcanonical points" $\{(C,p) \in \mathcal{M}_{g,1} : (2g-2)p \sim K_C\}$; see [Kontsevich and Zorich 2003] for further background about this locus. All three are called *symmetric* semigroups, since a positive integer n is a gap if and only if $2g-1-n$ is not a gap; this condition is equivalent to the condition that $2g-1$ is a gap.

Remark 2.10. The semigroups $\mathrm{NG}^2_{g,e}$ are among the first semigroups for which $\mathcal{M}^S_{g,1}$ was studied in detail; see [Pinkham 1974, Theorem 14.7].

In fact, the negatively graded semigroups are precisely the semigroups for which the Deligne lower bound (on codimension) and the effective weight upper bound coincide.

Proposition 2.11. *For any numerical semigroup S of genus g,*

$$\mathrm{ewt}(S) \ge g - \lambda(S),$$

with equality if and only if S is either ordinary or one of the semigroups $\mathrm{NG}^1_{g,e}$, $\mathrm{NG}^2_{g,e}$, or NG^3_g.

Proof. Let E denote the set of pairs $(a,b) \in \mathbb{N}^2$ such that $a < b$, a is a generator of S, and b is a gap. Let Λ denote the set of gaps b such that $b+a \in S$ for all positive elements $a \in S$. By definition, $\mathrm{ewt}(S) = |E|$ and $\lambda(S) = |\Lambda|$.

For all $(a,b) \in E$, $b-a$ is necessarily a gap that is not in Λ. Conversely, any gap b' that is not an element of Λ must be equal to $b-a$ for some $(a,b) \in E$. This shows that the complement of Λ in $\mathbb{N} \setminus S$ has at most $|E|$ elements, hence $\mathrm{ewt}(S) \ge g - \lambda(S)$. Furthermore, this argument shows that equality holds if and only if each $(a,b) \in E$ gives a *distinct* difference $b-a$. Assume now that S satisfies $\mathrm{ewt}(S) = g - \lambda(S)$; we will show that S is of one of the three forms stated. The case where S is ordinary is immediate, so assume that S is nonordinary. Denote by m, n the first two generators of S.

Case 1: Suppose there are no gaps above n. In this case $S = \mathrm{NG}^1_{g,g-m+1}$.

Case 2: Suppose that $n = m+1$. There can be no two consecutive gaps b and $b+1$ of S greater than $m+1$, since otherwise (m,b) and $(m+1,b+1)$ both lie in E. Similarly, there is at most one gap b such that $b-1$ is a generator. Since all elements of S less than $2m$ are generators, these two facts show that there is at most one gap b of S between m and $2m$. If there are no gaps between m and $2m$, then S is ordinary. If there is one gap b between m and $2m$, then S contains $\{m, m+1, \dots, b-1, b+1, b+2, \dots, 2m-1\}$, which generate all integers greater than $2m$ (recall that $b > n = m+1$ by assumption), so in fact b is the only gap greater than m. Hence $S = \mathrm{NG}^2_{g,b-g}$.

Case 3: Suppose that $n \ge m+2$ and there is some gap $b > n$. Assume that b is the smallest such gap. The gap b is less than $m+n$, since otherwise $b-m$ would be an element of S and b could not be a gap. Since $(n,b) \in E$ and $1 \le b-n \le m-1$, it follows that not all of $(m,m+1), \dots, (m,2m-1)$ can lie in E; this implies that $n \le 2m-1$, hence $m+3 \le b \le 3m-1$. The pair $(m,m+1)$ lies in E, so $(b-1,b)$ cannot lie in E, hence $b-1$ is a composite element of S. The only possibility is that $b=2m+1$. Therefore $(2m-1,2m+1)\in E$. This shows that $(m,m+2)\notin E$, so $n=m+2$. Therefore $m+2,m+3,\dots,2m\in S$ and $2m+1\notin S$. The numbers $m,m+2,m+3,\dots,2m$ generate all integers greater than $2m+1$, so $2m+1$ is the largest gap of S. Therefore $m=g-1$ and $S=\mathrm{NG}_g^3$ in this case. □

Remark 2.12. It would be interesting to find a more direct connection between negative grading and the equality of the Deligne and effective weight bounds. It seems improbable that the fact that the same list of semigroups is found in both contexts is merely a combinatorial coincidence.

2C. *Semigroups of low genus.*

The exact codimension of $\mathcal{M}_{g,1}^S$ is known for all semigroups of genus less than or equal to 6; in all of these cases, $\operatorname{ewt}(S)=\operatorname{codim}\mathcal{M}_{g,1}^S$. We now summarize where these results can be found in the literature.

Most of these loci $\mathcal{M}_{g,1}^S$ have been described by Nakano; see Table 2 of [Nakano 2008].[3] Of the rows in Nakano's table where $\dim\mathcal{M}_{g,1}^S$ is not known, all but one are in fact one of the negatively graded semigroups discussed in Section 2B, hence $\operatorname{codim}\mathcal{M}_{g,1}^S$ is equal to $\operatorname{ewt}(S)$ in those cases. The remaining semigroup is $S=\langle 5,7,9,11,13\rangle$ ($N(6)_{12}$, in the naming system of [Nakano 2008]). The discussion in [Bullock 2014, Section 2.2] shows that for this semigroup, $\mathcal{M}_{g,1}^S$ has a component of codimension $\operatorname{ewt}(S)$ (equal to $\operatorname{wt}(S)$ in this case since S is primitive), and the main theorem of [Bullock 2014] shows that this is the only component.

2D. *Two-generator semigroups.*

We now show that any numerical semigroup S with only two generators exists as a Weierstrass semigroup, and that $\mathcal{M}_{g,1}^S$ is irreducible of codimension $\operatorname{ewt}(S)$ in $\mathcal{M}_{g,1}$. This furnishes an infinite family of nonprimitive semigroups of effective weight larger than g for which the effective weight gives the correct codimension.

Let $1<e<d$ be relatively prime integers and let $S=\langle e,d\rangle$. The genus of S is $\frac{1}{2}(e-1)(d-1)$, as a short combinatorial argument shows.

The effective weight is the number of gaps greater than e plus the number of gaps greater that d, which can be expressed as:

$$\operatorname{ewt}(S)=2g-d-e+\left\lfloor\frac{d}{e}\right\rfloor+2.$$

To analyze $\mathcal{M}_{g,1}^S$, we use the following description.

Proposition 2.13. *Let S,g,d,e be as above, and let P denote the convex lattice polygon $\{(i,j)\in\mathbb{R}^2 : i,j\geq 0, ei+dj\leq ed\}$. Let $(c_{i,j})_{(i,j)\in P}$ be coefficients such that the affine curve $\tilde{C}$ defined by*

$$0=\sum_{(i,j)\in P\cap\mathbb{Z}^2} c_{i,j}x^iy^j$$

[3] There is a typographical error in that table: the semigroup $\langle 5,6,7\rangle$ is stated in one column to be 11-dimensional, while the following column indicates that $\mathcal{M}_{g,1}^S$ is an open subset of a 10-dimensional weighted projective space. The second column is correct.

is smooth, and such that the coefficients $c_{d,0}$ and $c_{0,e}$ are nonzero. Then the completion C of $\tilde{C}$ has only one additional point p, which has Weierstrass semigroup S. Viewing the coordinates x and y as rational functions on C regular on $\tilde{C}$, the pole orders of x and y at p are d and e, respectively.

Conversely, given any $(C, p) \in \mathcal{M}^S_{g,1}$ and rational functions f and g of pole orders e and d at p and regular elsewhere, the map (f, g) embeds $\tilde{C} = C \setminus \{p\}$ as an affine curve of the form above.

Proof. Embed the affine plane as the set $U = \{(x, y, 1)\}$ in the weighted projective plane $\mathbb{P}(e, d, 1)$, and let C denote the closure of $\tilde{C}$ in $\mathbb{P}(e, d, 1)$. Denote by X, Y and Z the weighted homogeneous coordinates on $\mathbb{P}(e, d, 1)$. The equation of C is

$$0 = \sum_{(i,j)\in P\cap\mathbb{Z}^2} c_{i,j} X^i Y^j Z^{de-ei-dj}. \tag{1}$$

Note that the nonvanishing of $c_{d,0}$ and $c_{0,e}$ ensures that the scheme cut out by this homogeneous equation has no components supported on the complement of U, so since this scheme matches $\tilde{C}$ on U it is indeed equal to the closure of $\tilde{C}$.

Neither of the points $(1, 0, 0)$, $(0, 1, 0)$ lie on C since $c_{d,0}$ and $c_{0,e}$ are nonzero. Therefore any points of $C \setminus \tilde{C}$ lie on $\{(x, y, 0) : x, y \neq 0\} \cong \operatorname{Spec} k[u, u^{-1}]$, where $u = X^d Y^{-e}$. The scheme-theoretic intersection of C with this curve is given by the equation $c_{d,0}u + c_{0,e} = 0$. Hence C meets the boundary transversely in a single point; it follows that C is smooth, hence it is the completion of $\tilde{C}$, and has exactly one additional point on the boundary; denote this point by p.

The rational functions x and y are regular on $\tilde{C}$ and their divisors of zeros are degree e and d, respectively, hence they have poles of orders e and d at p. It follows that the Weierstrass semigroup of p contains e and d, hence it contains all of S. It suffices to verify that the genus of C is equal to the genus of S, which is $\frac{1}{2}(d-1)(e-1)$. This can be deduced from standard results in the geometry of toric surfaces; we summarize an argument using results from [Cox et al. 2011]. To the convex lattice polygon P, we may associate, as described in [loc. cit.], a toric variety X_P together with a projective embedding. The variety X_P is isomorphic to $\mathbb{P}(e, d, 1)$ [loc. cit., Exercise 10.2.6(a)]. The hyperplane sections in this embedding are subschemes cut out by equations of the form of (1), so the curve C is one such hyperplane section. By [loc. cit., Proposition 10.5.8], the arithmetic genus of the subscheme cut out by (1) is equal to the number of interior lattice points of P. The area of P is $\frac{1}{2}de$, and the number of boundary vertices of P is $d + e + 1$ (since d and e are relatively prime, there are no lattice points interior to the edge from $(d, 0)$ to $(0, e)$, so we need only count the points on the other two edges). It follows from Pick's theorem that the number of interior lattice points of P is $\frac{1}{2}(d-1)(e-1)$, as desired.

For the converse, suppose that f and g are rational functions on C as in the proposition statement, and let $\tilde{C}$ be $C\setminus\{p\}$. Then (f, g) defines a map from $\tilde{C}$ to the affine plane. The ring generated by f and g includes functions of every possible pole order at p, hence this ring includes all regular functions on $\tilde{C}$, and (f, g) is an embedding. Both f^d and g^e have pole order de at p, so some linear combination of them has a strictly smaller pole order, hence is expressible as a linear combination of functions $f^i g^j$, where $ei + dj \leq de$.

In other words, $\tilde{C}$ satisfies a relation of the form $0 = \sum_{(i,j)\in P\cap\mathbb{Z}^2} c_{i,j}x^i y^j$. Since there can be no relations of smaller degree, this must be the generator of the ideal of (the image in the affine plane of) $\tilde{C}$. □

We can use this description to determine the dimension of $\mathcal{M}^S_{g,1}$. The dimension of the space of curves in the affine plane of the form in the Proposition is $|P \cap \mathbb{Z}^2| - 1$. Using Pick's theorem and the fact that there are $d + e + 1$ vertices on the boundary of P (as in the proof of the proposition), this dimension is $\frac{1}{2}(d+1)(e+1)$. This exceeds $\dim \mathcal{M}^S_{g,1}$ by the dimension of the set of ways to embed a given $(C, p) \in \mathcal{M}^S_{g,1}$ in this manner, which is equal to $h^0(\mathcal{O}_C(e\cdot p)) + h^0(\mathcal{O}_C(d\cdot p))$, which in turn is equal to $4 + \lfloor \frac{d}{e} \rfloor$. Therefore

$$\dim \mathcal{M}^S_{g,1} = \tfrac{1}{2}(d+1)(e+1) - 4 - \left\lfloor \tfrac{d}{e} \right\rfloor = g + d + e - 4 - \left\lfloor \tfrac{d}{e} \right\rfloor.$$

Combining with the earlier calculation of $\operatorname{ewt}(S)$, we have proved:

Proposition 2.14. *Let $S = \langle e, d\rangle$ be a numerical semigroup with two generators. Then $\mathcal{M}^S_{g,1}$ is irreducible of codimension* $\operatorname{ewt}(S)$ *in $\mathcal{M}_{g,1}$.*

2E. ***Total inflection points of nodal plane curves.*** Another naturally arising class of semigroups for which the effective weight bound is exact are those arising from nodal plane curves. These have been investigated by Coppens and Kato [1994]. Although they do not explicitly analyze the dimension of $\mathcal{M}^S_{g,1}$, their results readily give its value, which coincides with the value that the effective weight would predict.

Definition 2.15. Let $d \geq 3$ be an integer, and δ a nonnegative integer less than $\binom{d-1}{2}$. Let

$$N_{d,\delta} = \langle d-1, d\rangle \cup H_c,$$

where $g = \binom{d-1}{2} - \delta$ and c is the gth gap in $\langle d-1, d\rangle$.

Remark 2.16. The genus of the semigroup $\langle d-1, d\rangle$ is $\binom{d-1}{2}$, so this is well defined. The semigroup $N_{d,\delta}$ can be thought of as the "simplest" (e.g., the lowest-effective-weight) semigroup of genus g containing both d and $d-1$. It can also be described as the genus g ancestor of $\langle d-1, d\rangle$ in the semigroup tree (see Remark 1.4).

Theorem 2.17 [Coppens and Kato 1994, Theorem 2.3]. *Let L be a fixed line in $\mathbb{P}^2$. Let X denote the variety of degree d plane curves C with δ simple nodes and smooth at all other points, such that C intersects L at a smooth point of C to multiplicity d. Then for a general point $[C] \in X$, the Weierstrass semigroup of (C, p) is $N_{d,\delta}$.*

Proposition 2.18. *For $S = N_{d,\delta}$, with d and δ as in Definition 2.15, $\mathcal{M}^S_{g,1}$ is irreducible of codimension* $\operatorname{ewt}(S)$ *in $\mathcal{M}_{g,1}$.*

Proof. The genus of S is $g = \binom{d-1}{2} - \delta$ by definition, and its only generators that are below any gaps are $d-1$ and d, which lie below all gaps of S except $1, 2, \dots, d-2$. Therefore

$$\operatorname{ewt}(S) = 2g - 2d + 4.$$

Let X be the variety in the statement of Theorem 2.17. It has a dense open subset U consisting of curves C such that the normalization of (C, p) lies in $\mathcal{M}^S_{g,1}$, and the induced map $U \to \mathcal{M}^S_{g,1}$ has irreducible fibers of dimension 6, since there is a 6-dimensional space of automorphisms of $\mathbb{P}^2$ fixing a line. The map is dominant since any (C, p) with Weierstrass semigroup S may be given a morphism to $\mathbb{P}^2$ using two rational functions of pole orders $d-1$ and d at p; the image curve will be smooth at the image of p, and the image of p will be a total inflection point since the divisor $d \cdot p$ must be the pullback of some hyperplane section. Therefore $\dim X = \dim \mathcal{M}^S_{g,1} + 6$. It suffices to show that X is irreducible of dimension $g + 2d$. The dimension of X is equal to $g + 2d$ by [Harris 1986, Lemma 2.4], and the irreducibility of X follows from [Ran 1989, Irreducibility Theorem (bis)]. □

2F. ***A case where*** $\mathbf{codim}\, \mathcal{M}^S_{g,1} \neq \mathbf{ewt}(S)$**.** The smallest genus in which we are aware of a semigroup S for which $\operatorname{codim} \mathcal{M}^S_{g,1} \neq \operatorname{ewt}(S)$ is $g = 9$.

The example is

$$S = \langle 6, 7, 8 \rangle = \mathbb{N} \setminus \{1, 2, 3, 4, 5, 9, 10, 11, 17\}.$$

For this semigroup, $\operatorname{ewt}(S) = 12$, but we claim that $\operatorname{codim} \mathcal{M}^S_{g,1} = 11$. We will sketch a proof of this fact, omitting the full details. In [Pflueger 2016], we describe $\mathcal{M}^S_{g,1}$ for all semigroups of the form $\langle d-r+1, d-r+2, \dots, d \rangle$ in complete detail. These semigroups furnish a large collection of cases where $\operatorname{codim} \mathcal{M}^S_{g,1} < \operatorname{ewt}(S)$.

If $(C, p) \in \mathcal{M}^S_{g,1}$, then one can show that the complete linear series $|8p|$ embeds C in $\mathbb{P}^3$ as the complete intersection of a quadric Q and a quartic R, and in this embedding the osculating plane H at p meets C at p only. Hence we can study $\mathcal{M}^S_{g,1}$ via the variety of triples (C, H, p) of a smooth complete intersections C of a quadric and quartic, a hyperplane H, and a point p such that C and H meet at p only. One can calculate that the dimension of this variety is 29, and verify that for a general point of this variety, (C, p) does indeed have Weierstrass semigroup S. Since a point $(C, p) \in \mathcal{M}^S_{g,1}$ determines the triple (C, H, p) up to automorphisms of $\mathbb{P}^3$, this shows that $\dim \mathcal{M}^S_{g,1} = 29 - \dim \operatorname{Aut} \mathbb{P}^3 = 14$, hence $\operatorname{codim} \mathcal{M}^S_{g,1} = 25 - 14 = 11$.

3. Dimensionally proper linear series

This section collects several key facts and definitions about families of linear series on marked algebraic curves, including a "regeneration lemma" from the theory of limit linear series. The regeneration lemma is the basic inductive tool in the proof of Theorem 1.3.

Our discussion will be brief, and a number of proofs and precise definitions are omitted where they are not necessary for the application in this paper. A complete discussion of these matters can be found in [Osserman 2013, Chapter 4]; other useful references are [Arbarello et al. 1985, Chapter IV], [Harris and Morrison 1998, Chapter 5] and [Arbarello et al. 2011, Chapter XXI].

3A. *Varieties of linear series with specified ramification.*

Definition 3.1. Let C be a smooth curve. A *linear series of rank r and degree d* on C, or "a g^r_d," is a pair (L, V) consisting of a degree d line bundle on C and an $(r+1)$-dimensional vector space V of global sections of L. We will sometimes refer to the linear series simply as V.

Let p be a point of C. The *vanishing sequence* $a_0^V(p), \dots, a_r^V(p)$ of V consists of the $r+1$ distinct orders of vanishing of elements $s \in V$ at the point p, in (strictly) increasing order.

We will often use the phrase *vanishing sequence* to refer to a set of $r+1$ nonnegative integers between 0 and d inclusive (when the values of r and d are clear from context). Vanishing sequences will be denoted by capital roman letters, while the individual elements of a vanishing sequence will be denoted by the corresponding lowercase letter, with a subscript. For example, the elements of a vanishing sequence A will be denoted $a_0, a_1, \dots, a_r$, in increasing order.

In the following two definitions, we describe the set of closed points of a scheme without specifying the scheme structure. We hope the reader will forgive this, as the scheme structure is not relevant to our application. Full details, including the functors that these schemes represent, can be found in [Osserman 2013, Section 4.1]. Although we only need the following two definitions in the cases $n=1$ and $n=2$, we state them in fuller generality.

Definition 3.2. Let C be a smooth curve, $p_1, \dots, p_n$ be distinct points of C, and $A^1, \dots, A^n$ be vanishing sequences. Denote by

$$G^r_d(C; (p_1, A^1), \dots, (p_n, A^n))$$

a scheme whose closed points correspond to the g^r_ds (L, V) on C such that for $i=1, \dots, n$ and $j=0, \dots, r$, the inequality $a_j^V(p_i) \geq a_j^i$ holds (recall that we write a_j^i to denote the j-th element of the set A^i). Denote by

$$\tilde{G}^r_d(C; (p_1, A^1), \dots, (p_n, A^n))$$

the open subscheme where equality $a_j^V(p_i) = a_j^i$ holds for all i and j.

Remark 3.3. In this definition and those that follow, our notation differs slightly from that of, for example, [Osserman 2013]. In particular, we specify the *vanishing* sequence at each marked point, whereas most authors specify the *ramification* sequence, defined by $\alpha_i(p) = a_i(p) - i$. We have chosen to work exclusively with vanishing orders, as it significantly reduces clutter in several parts of the present paper.

Definition 3.4. Let $\mathcal{C} \to B$ be a smooth, proper family of curves and $s_1, \dots, s_n$ be disjoint sections. In the case $n=0$, assume that the family has at least one section. Denote by

$$G^r_d(\mathcal{C}/B; (s_1, A^1), \dots, (s_n, A^n)) \to B$$

a scheme whose fiber over $b \in B$ is $G^r_d(\mathcal{C}_b; (s_1(b), A^1), \dots, (s_n(b), A^n))$.

Denote by $\mathcal{G}^r_{g,d}(A^1, \dots, A^n) \to \mathcal{M}_{g,n}$ the scheme formed by gluing these schemes together (or, more precisely, gluing these schemes together over a versal family of n-marked curves, and then taking the quotient by a finite group action).

The notation $\tilde{G}^r_d$ or $\tilde{\mathcal{G}}^r_{g,d}$ will refer to the open subscheme where the vanishing sequences match the prescribed sequences exactly.

Here, $\mathcal{M}_{g,n}$ denotes the coarse moduli space of smooth curves with n distinct marked points. We omit the details of the gluing process; it suffices for our purposes that a scheme $\mathcal{G}^r_{g,d}(A^1, \dots, A^n)$ exists, whose fibers over $\mathcal{M}_{g,n}$ are isomorphic to the varieties $G^r_d(C; (p_1, A^1), \dots, (p_n, A^n))$.

3B. *Dimensionally proper points.*

Definition 3.5. For integers g, r, d and vanishing sequences $A^1, \dots, A^n$, define

$$\rho_g(r, d; A^1, \dots, A^n) = (r+1)(d-r) - rg - \sum_{i=1}^{n} \sum_{j=0}^{r} (a^i_j - j).$$

When $g, r, d, A^1, \dots, A^n$ are clear from context, we will denote this number simply by ρ.

Lemma 3.6. *If $G^r_d(\mathcal{C}/B; (s_1, A^1), \dots, (s_n, A^n))$ is nonempty, its local dimension at any point is greater than or equal to* $\dim B + \rho$.

Proof. See, for example, [Osserman 2013, Theorem 4.1.3] for full details; what follows is a brief summary. First, describe $G^r_d(\mathcal{C}/B)$ (where we must assume that $\mathcal{C} \to B$ has a section) as a degeneracy locus of a map of vector bundles over the relative Picard scheme $\mathrm{Pic}^d(\mathcal{C}/B)$, and bound its dimension with this description. Note that the assumption that $\mathcal{C} \to B$ has a section is needed to construct the relative Picard scheme $\mathrm{Pic}^d(\mathcal{C}/B)$. Then impose the vanishing conditions by intersecting the pullback of n Schubert cells under n maps of Grassmannian bundles; this imposes at most a number of conditions equal to the double summation in the formula for ρ. □

Definition 3.7. A linear series $(L, V) \in G^r_d(C; (p_1, A^1), \dots, (p_n, A^n))$ is called *dimensionally proper* (with respect to the choice of $A^1, \dots, A^n$) if there exists a deformation $(\mathcal{C}/B, s_1, \dots, s_n)$ of $(C, p_1, \dots, p_n)$ such that

$$\dim G^r_d(\mathcal{C}/B; (s_1, A^1), \dots, (s_n, A^n)) = \dim B + \rho,$$

locally at (L, V).

Equivalently, (L, V) is dimensionally proper if the local dimension of $\mathcal{G}^r_{g,d}(A^1, \dots, A^n)$ at (L, V) is equal to $3g + n - 3 + \rho$.

3C. *Regeneration.*

We will reduce the proof of Theorem 1.3 to the existence of dimensionally proper points of a suitable variety of linear series. The existence results will come from an induction on genus, made possible by the following "regeneration lemma."

Lemma 3.8. *Fix positive integers g_1, g_2, d, r and two vanishing sequences A and A'. Denote by $d - A$ the vanishing sequence $\{d - a_r, d - a_{r-1}, \dots, d - a_0\}$.*

If $\tilde{\mathcal{G}}^r_{g_1,d}(A)$ and $\tilde{\mathcal{G}}^r_{g_2,d}(d - A, A')$ both have dimensionally proper points, then $\tilde{\mathcal{G}}^r_{g_1+g_2,d}(A')$ also has dimensionally proper points.

This lemma is a standard application of the theory of *limit linear series*, pioneered by Eisenbud and Harris [1986]. It is essentially a special case of the "smoothing theorem" [Eisenbud and Harris 1986, Theorem 3.4], which is referred to as the "regeneration theorem" in the expository account [Harris and Morrison

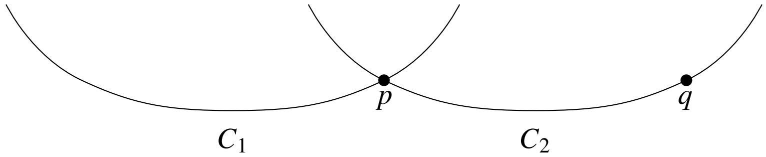


Figure 1. The nodal curve X of Situation 3.9.

1998, Theorem 5.41]. Both of these sources work over the complex numbers and work locally in the complex-analytic setting. We give a proof of Lemma 3.8 below based on the more recent [Osserman 2006], which is therefore valid in characteristic p. The theory of limit linear series has subsequently been expanded (for example, to include curves not of compact type) in various ways (e.g., [Osserman 2014a; Osserman 2014b; Amini and Baker 2015]), but for our purposes the theory developed in [Osserman 2006] is sufficient.

Limit linear series, as their name suggests, provide a way to construct, from a family of smooth algebraic curves degenerating to a nodal curve and a family of linear series on the smooth curves, an object over the nodal curve that serves as a well-defined limit of the linear series on smooth curves. For the purpose of Lemma 3.8, we need only consider particularly simple nodal curves.

Situation 3.9. Fix positive integers g_1 and g_2. Let C_1 and C_2 be smooth curves of genus g_1 and g_2 respectively, let p_1 be a point of C_1 and let p_2 and q be distinct points of C_2. Denote by X the nodal curve obtained by gluing p_1 to p_2, and denote the attachment point by $p \in X$. See Figure 1.

We will only require a specific type of limit linear series, namely *refined* series. In general, the refined series form an open subset of all limit linear series. We do not require nonrefined series (called *coarse* series in the Eisenbud and Harris theory) for our application, so we will not discuss them.

Definition 3.10. In Situation 3.9, a *refined limit linear series* of rank r and degree d on X (or a *limit g^r_d on X*), is a pair $((L_1, V_1), (L_2, V_2))$ of g^r_ds on C_1 and C_2 respectively, such that for $i = 0, 1, \dots, r$,

$$a_i^{V_1}(p_1) = d - a_{r-i}^{V_2}(p_2). \tag{2}$$

Equation (2) is called the compatibility condition. The linear series (L_i, V_i) is called the *C_i-aspect* of the limit linear series.

Another way to view a refined limit linear series on X is that it consists of a choice of vanishing sequence A (with respect to the data r, d) and a point in

$$\tilde{G}^r_d(C_1; (p_1, A)) \times \tilde{G}^r_d(C_2; (p_2, d - A)).$$

Therefore a natural way to define a scheme structure for the set of refined limit linear series is as follows.

Definition 3.11. In Situation 3.9, the scheme of refined limit g^r_ds on X is

$$G^{r,\mathrm{ref}}_d(X) = \bigcup_A \tilde{G}^r_d(C_1; (p_1, A)) \times \tilde{G}^r_d(C_2; (p_2, d - A)),$$

where the union is taken within the scheme $G^r_d(C_1) \times G^r_d(C_2)$.

Definition 3.11 extends in an obvious way to families $\mathcal{X} \to B$ of two-component curves. What is less obvious is that it can also be extended to certain families of curves in which some members are smooth and some are singular. For our purposes, we require the following facts:

(1) There is a special type of family $\mathcal{X} \to B$ of nodal curves, called a *smoothing family* [Osserman 2006, Definition 3.1]. For every flat, proper family $\mathcal{X} \to B$ of genus g curves, all are either smooth curves or two-component curves with one node, with $\mathcal{X}$ regular and B regular and connected, and for every choice of point $b \in B$, there is an étale neighborhood $B' \to B$ of b such that the fiber product $\mathcal{X}' \to B'$ is a smoothing family [loc. cit., Lemma 3.3].

(2) If $\mathcal{X} \to B$ is a smoothing family of curves, all either smooth or two-component, there is a scheme $G_d^{r,\mathrm{ref}}(\mathcal{X}/B) \to B$, whose fiber over any $b \in B$ is either $G_d^r(\mathcal{X}_b)$ (if $\mathcal{X}_b$ is smooth) or $G_d^{r,\mathrm{ref}}(\mathcal{X}_b)$ (if $\mathcal{X}_b$ is a two-component curve) [loc. cit., Proposition 6.6].

(3) For such a smoothing family, the dimension bound

$$\dim G_d^{r,\mathrm{ref}}(\mathcal{X}/B) \geq \dim B + \rho_g(r, d)$$

holds locally at every point [loc. cit., Theorem 5.3].

(4) With a family $\mathcal{X} \to B$ as above, given a section s whose image lies in the smooth locus of every fiber, and a vanishing sequence A, there also exists a scheme

$$\tilde{G}_d^{r,\mathrm{ref}}(\mathcal{X}/B; (s, A)) \to B,$$

whose fiber over a point $b \in B$ such that $\mathcal{X}_b$ is smooth is isomorphic to $\tilde{G}_d^r(\mathcal{X}_b; (s(b), A))$, and whose fiber over a point $b \in B$ such that $\mathcal{X}_b$ is singular consists (set-theoretically) of those refined limit linear series such that the aspect of the component on which $s(b)$ lies has vanishing sequence equal to A at $s(b)$ [loc. cit., Corollary 6.10].

(5) In the previous situation, the dimension bound

$$\dim G_d^{r,\mathrm{ref}}(\mathcal{X}/B; (s, A)) \geq \dim B + \rho_g(r, d; A)$$

holds locally at every point [loc. cit., Theorem 4.4.10].

With this machinery in place, we can prove the regeneration lemma.

Proof of Lemma 3.8. Suppose that there are two dimensionally proper linear series

$$(L_1, V_1) \in \tilde{G}_d^r(C_1; (p_1, A)) \quad \text{and} \quad (L_2, V_2) \in \tilde{G}_d^r(C_2; (p_2, d - A), (q, A')),$$

where C_1 and C_2 are smooth curves of genus g_1 and g_2, respectively. Form from (C_1, p_1) and (C_2, p_2, q) a nodal two-component marked curve (X, q) as in Situation 3.9. Let $(\mathcal{X}/B, s)$ be a versal deformation of (X, q), and let $\Delta \subset B$ denote the locus of singular curves, which is of codimension 1 in B. We may assume (perhaps after taking a base change to an étale neighborhood) that $\mathcal{X}/B$ is a smoothing family, and hence form the scheme $\tilde{G}_d^{r,\mathrm{ref}}(\mathcal{X}/B; (s, A'))$ of refined limit linear series. The two linear series (L_1, V_1)

and (L_2, V_2) constitute the aspects of a refined limit linear series on X. Since both of these aspects are dimensionally proper, the local dimension, at this point, of the preimage of Δ in $\tilde{G}^{r,\mathrm{ref}}_d(\mathcal{X}/B; (s, A'))$ must be equal to exactly $\dim \Delta + \rho_{g_1}(r, d; A) + \rho_{g_2}(r, d; d - A, A')$. A bit of algebra shows that this is equal to $\dim \Delta + \rho_g(r, d; A')$ (this bit of algebra is sometimes referred to as "the additivity of the Brill–Noether number," e.g., in [Eisenbud and Harris 1986, Lemma 3.6]). On the other hand, the local dimension, at this same point, of the entire space $\tilde{G}^{r,\mathrm{ref}}_d(\mathcal{X}/B; (s, A'))$ is *at least* $\dim B + \rho_g(r, d; A')$; since Δ has codimension one, it follows that the local dimension of $\tilde{G}^{r,\mathrm{ref}}_d(\mathcal{X}/B; (s, A'))$ is in fact exactly equal to $\dim B + \rho_g(r, d; A')$, and that no irreducible component containing this point lies entirely over Δ. Taking any irreducible component and restricting it to the complement of Δ in B, we obtain a dimensionally proper family of g^r_ds on *smooth* marked curves of genus g, with imposed vanishing sequence A' at the marked point. Hence $\tilde{\mathcal{G}}^r_{g,d}(A')$ has dimensionally proper points. □

4. The effective weight bound

We will prove Theorem 1.2 in this section. The proof comes from the dimension bound of Lemma 3.6, applied to carefully chosen vanishing data at the marked point. Our main point of departure from previous work on this subject (e.g., [Eisenbud and Harris 1987; Bullock 2013]) is that we consider *incomplete* linear series (that is, (L, V) where V is a strict subspace of the space of global sections of L), which nonetheless determine the Weierstrass semigroup. This innovation allows the weight bound to be improved to the effective weight bound.

Definition 4.1. Let $S \subset \mathbb{N}$ be a numerical semigroup of genus g. An *effective subsequence for S* is a finite subset $T \subset S$ such that

(1) T contains 0,

(2) T contains all generators of S, and

(3) T does not contain any composite elements of S that are less than the largest gap of S.

In the statement below and elsewhere, we will write $d - T$ to denote the set $\{d - t : t \in T\}$.

Lemma 4.2. *Let T be an effective subsequence for a numerical semigroup S of genus g, and $d \geq \max T$ an integer. Let $r = |T| - 1$. For any smooth marked curve (C, p) of genus g:*

(1) *If the Weierstrass semigroup of (C, p) is not S, then*

$$\tilde{G}^r_d(C; (p, d - T)) = \varnothing.$$

(2) *If the Weierstrass semigroup of (C, p) is S, then the reduced structure of $\tilde{G}^r_d(C; (p, d - T))$ is isomorphic to the affine space of dimension*

$$\rho_g(r, d; d - T) + \mathrm{ewt}(S).$$

Proof. Suppose that $(L, V) \in \tilde{G}^r_d(C; (p, d - T))$. Since $0 \in T$, one of the vanishing orders of V must be d itself. Therefore L must be $\mathcal{O}_C(d \cdot p)$, and V may be regarded as a vector space of rational functions on C, regular away from p, including functions of pole orders $t \in T$ and no others. In particular, T is a subset of the Weierstrass semigroup of p. Hence $S(C, p)$ contains all of the generators of S, and hence is precisely equal to S since S and $S(C, p)$ have the same genus. This proves part (1).

Now suppose that the Weierstrass semigroup of (C, p) is S. Let W be the vector space of global sections of $\mathcal{O}_C(d \cdot p)$; regard the elements of W as rational functions on C. This space has a complete flag $\{0\} = W_0 \subset W_1 \subset \cdots \subset W_\ell = W$, where W_i consists of those rational functions of pole order less than s_i, where $S = \{0 = s_0, s_1, s_2, \ldots\}$ (written in increasing order). Then the reduced structure of $\tilde{G}^r_d(C; (p, d - T))$ may be identified with an open Schubert cell in the Grassmannian of $(r + 1)$-dimensional subspaces of W with respect to this flag, hence it is isomorphic to an affine space. If we write $T = \{s_{j_i} : i = 0, \ldots, r\}$ (j_i increasing with i), then the dimension of this Schubert cell is equal to $\sum_{i=0}^r (j_i - i)$. For $i = 0, 1, 2, \ldots, r$, $s_{j_i} - j_i$ is equal to the number of gaps below s_{j_i}, and therefore

$$j_i - i = (s_{j_i} - i) - g + (\#\text{gaps of } S \text{ greater than } s_{j_i}).$$

Summing over all i and performing some algebra, we obtain

$$\dim \tilde{G}^r_d(C; (p, d - T)) = \rho_g(r, d; d - T) - g + \sum_{t \in T} (\#\text{gaps of } S \text{ greater than } t).$$

Now, the value $0 \in T$ contributes g to the sum on the right side of this equation, the set of generators of S contribute $\operatorname{ewt}(S)$ total to the sum, and all elements of T that are composite in S have no gaps of S above them, thus contribute 0. Therefore $\dim \tilde{G}^r_d(C; (p, d - T)) = \rho_g(r, d; d - T) + \operatorname{ewt}(S)$. □

Remark 4.3. If T were selected to be $S \cap \{n \in \mathbb{N} : n \leq 2g - 1\}$ (that is, if we include many composite elements), then the same proof would show that $\tilde{G}^r_d(C; (p, d - T))$ is either empty or a single point, and the following corollary would prove the ordinary weight bound $\operatorname{codim} \mathcal{M}^S_{g,1} \leq \operatorname{wt}(S)$. Omitting the composite elements is precisely what strengthens the bound from $\operatorname{wt}(S)$ to $\operatorname{ewt}(S)$.

Corollary 4.4. *Let $(\mathcal{C}/B, s)$ be a smooth, proper family of genus g curves with a section, and consider the subvariety*

$$B^S = \{b \in B : (\mathcal{C}_b, s(b)) \in \mathcal{M}^S_{g,1}\}$$

of marked curves with Weierstrass semigroup S. If B^S is nonempty, then $\dim B^S \geq \dim B - \operatorname{ewt}(S)$.

Proof. The morphism $\tilde{G}^r_d(\mathcal{C}/B; (s, d - T)) \to B$ has image equal to B^S, and all fibers of dimension $\rho_g(T) + \operatorname{ewt}(S)$. Hence

$$\dim B^S = \dim \tilde{G}^r_d(\mathcal{C}/B; (s, d - T)) - \rho_g(r, d; d - T) - \operatorname{ewt}(S).$$

Lemma 3.6 now gives the result. □

Proposition 4.5. *Let S be a numerical semigroup and let T be an effective subsequence for S. Let d be any integer greater than or equal to* $\max T$. *Also let g be the genus of S and $r = |T| - 1$. The map $\tilde{\mathcal{G}}^r_{g,d}(d - T) \to \mathcal{M}_{g,1}$ gives a bijection between the irreducible components of $\tilde{\mathcal{G}}^r_{g,d}(d - T)$ and $\mathcal{M}^S_{g,1}$.*

Under this bijection, the effectively proper components of $\mathcal{M}^S_{g,1}$ *correspond to the dimensionally proper components of* $\tilde{\mathcal{G}}^r_{g,d}(d-T)$.

In particular, $\mathcal{M}^S_{g,1}$ *has effectively proper points if and only if* $\tilde{\mathcal{G}}^r_{g,d}(d-T)$ *has dimensionally proper points.*

Proof. The fiber of this morphism over the point corresponding to a marked curve (C, p) is equal to $\tilde{G}^r_d(C; (p, d-T))$. By Lemma 4.2, this fiber is either irreducible of dimension $\rho_g(r, d; d-T) + \operatorname{ewt}(S)$ (if $(C, p) \in \mathcal{M}^S_{g,1}$), or empty (otherwise). From this it follows that the irreducible components are in bijection, and that a component of $\mathcal{M}^S_{g,1}$ has dimension $\dim \mathcal{M}_{g,1} - \operatorname{ewt}(S)$ if and only if the corresponding component of $\tilde{\mathcal{G}}^r_{g,d}(d-T)$ has dimension $\dim \mathcal{M}_{g,1} + \rho_g(r, d; d-T)$. □

We can now prove Theorem 1.2.

Proof of Theorem 1.2. Let (C, p) be any marked smooth curve with Weierstrass semigroup S. Let $(\mathcal{C}/B, s)$ be a versal deformation of (C, p). Corollary 4.4, applied to $(\mathcal{C}/B, s)$, implies that the local dimension of $\mathcal{M}^S_{g,1}$ at (C, p) is at least $\dim \mathcal{M}_{g,1} - \operatorname{ewt}(S)$. For any irreducible component X of $\mathcal{M}^S_{g,1}$, a general point of X lies on no other irreducible components, hence $\dim X \geq \dim \mathcal{M}_{g,1} - \operatorname{ewt}(S)$. □

5. Secundive semigroups

This section collects several purely combinatorial ingredients needed to perform the inductive proof of Theorem 1.3.

Definition 5.1. A numerical semigroup S is called *secundive* if the largest gap is smaller than the sum of the two smallest generators.

Remark 5.2. The author has chosen "secundive" as a weaker form of "primitive" ("primus" and "secundus" meaning, respectively, "first" and "second" in Latin).

Lemma 5.3. *If S is a semigroup with* $\operatorname{ewt}(S) \leq g - 1$, *then S is secundive.*

Proof. Let S be a semigroup that is not secundive; we will show that $\operatorname{ewt}(S) \geq g$. Let m and n be the smallest and second-smallest generators of S, and let f be the largest gap of S. Since S is not secundive, $f > m + n$.

Consider the following three subsets of $\mathbb{N} \times \mathbb{N}$:

(1) $\{(m, a) : m < a \text{ and } a \notin S\}$.

(2) $\{(n, a) : n \leq a < m + n \text{ and } a \notin S\}$.

(3) $\{(a, f) : n \leq a < m + n, m \nmid a, \text{ and } a \in S\}$.

These three sets are disjoint, and every pair (x, y) in one of the three sets consists of a generator x and a gap y, with $x < y$. Therefore the sum of the sizes of the three sets is less than or equal to $\operatorname{ewt}(S)$.

The size of the first set is $g - m + 1$. The sum of the sizes of the second and third sets is equal to the number of integers $a \in \{n, n+1, \dots, m+n-1\}$ that are not divisible by m. There is exactly one a such that $m \mid a$ and $n \le a < m+n$, hence the sum of the sizes of the second and third sets is equal to $m - 1$. It follows that $\mathrm{ewt}(S) \ge g$. □

Remark 5.4. The method of the proof above, with slight modification, shows that the inequality $\mathrm{ewt}(S) \ge g$ is sharp (for nonsecundive semigroups), and provides a method to enumerate the equality cases. In fact, there exist nonsecundive semigroups with $\mathrm{ewt}(S) = g$ for all $g \ge 6$. On the other hand, all semigroups of genus $g \le 5$ are secundive.

Our inductive argument requires reducing the study of one secundive semigroup to another, which must be smaller both in genus and in effective weight. This is accomplished with the following operation.

Definition 5.5. For two integers s and k, with $k \ge 2$, define

$$\mathrm{slide}_k(s) = \begin{cases} s & \text{if } s \equiv 0 \bmod k, \\ s-2 & \text{if } s \equiv 1 \bmod k, \\ s-1 & \text{otherwise.} \end{cases}$$

For a set S of integers and an integer $k \ge 2$, define

$$\mathrm{slide}_k(S) = \{\mathrm{slide}_k(s) : s \in S\}.$$

In other words, slide_k fixes all multiples of k in place, and replaces each nonmultiple with the preceding nonmultiple. In particular, this function is order-preserving when restricted to nonmultiples of k; this is the feature which makes it interact well with the effective weight.

Definition 5.6. Let S be a secundive numerical semigroup of genus g. Call an element $k \in S$ a *good slider* if the following three conditions are met:

(a) $S' = \mathrm{slide}_k(S)$ is a secundive numerical semigroup of genus $g - 1$.

(b) $\mathrm{ewt}(S') = \mathrm{ewt}(S) - 1$.

(c) There exists an effective subsequence (Definition 4.1) T for S such that $\mathrm{slide}_k(T)$ is an effective subsequence for S'.

Lemma 5.7. *Let S be a secundive numerical semigroup and let m be the smallest generator of S*:

(1) *If $m + 1 \notin S$, then m is a good slider.*

(2) *If the largest gap of S is less than $2m - 1$, then any $k \in S$ such that $k + 1 \notin S$ is a good slider.*

(3) *If $m \ge 3$, $2m - 2 \in S$ and $2m - 1$ is the largest gap of S, then $2m - 2$ is a good slider.*

Proof. Part (1). Suppose that $m + 1 \notin S$, and let $S' = \mathrm{slide}_m(S)$. Let n be the second-smallest generator of S, and let f be the largest gap of S; note that neither is divisible by m. Then m is the smallest positive element of S' (this is where we use the hypothesis that $m + 1 \notin S$), the smallest element of S' that isn't a multiple of m is $n' = \mathrm{slide}_m(n)$, and the largest integer that is not in S' is $f' = \mathrm{slide}_m(f)$.

Since S is secundive, $f - n \leq m - 1$. Equivalently, there are fewer than $m - 1$ nonmultiples of m between n and f inclusive. The same is true of n' and f' since sliding preserves order among nonmultiples of m, hence $f' - n' \leq m - 1$ as well.

The sum of any two elements of S' is either a multiple of m or exceeds $m + n'$, which exceeds f', hence this sum lies in S'. So S' is indeed a numerical semigroup. Since $m + n' > f'$, S' is secundive. The gaps of S' are precisely $\{\mathrm{slide}_m(a) : a \notin S, a \geq 2\}$, so the genus of S' is $g - 1$.

To compare $\mathrm{ewt}(S)$ and $\mathrm{ewt}(S')$, observe first that in a secundive semigroup, an element a smaller than the largest gap is a generator if and only if it is either equal to m or not divisible by m. All other generators (those larger than the largest gap) do not contribute to the effective weight. Next observe that m has one fewer gap above it in S' than in S. Finally, note that slide_m establishes a bijection between the generators of S between m and f exclusive and the generators of S' between m and f' exclusive, and that the number of gaps above a given generator is preserved by this bijection. This shows that $\mathrm{ewt}(S') = \mathrm{ewt}(S) - 1$.

For part (c) of Definition 5.6, let T consist of 0 and also the smallest positive element of S in each congruence class modulo m. This set necessarily includes all generators of S, and the fact that S is secundive implies that any composite elements of S in T exceeds the largest gap, hence T is an effective subsequence of S. The set $T' = \mathrm{slide}_m(T)$ is precisely equal to the set containing 0 and the smallest positive element of S' in each congruence class modulo m, so since S' is also a secundive semigroup, T' is an effective subsequence of S' by the same reasoning. This completes the proof that m is a good slider when $m + 1 \notin S$.

Part (2). Now assume that the largest gap of S is less than $2m - 1$, and that $k \in S$ is an element with $k + 1 \notin S$. Then S is primitive. The smallest positive element of S' is either $m - 1$ or m, and the largest gap of S' is less than $2m - 2$, hence S' is in fact a *primitive* semigroup as well. The operation slide_k preserves the number of gaps above every element of S, except in one case: the number of gaps of S' above $\mathrm{slide}_k(k) = k$ is one less than the number of gaps above k in S. So $\mathrm{ewt}(S') = \mathrm{ewt}(S) - 1$. Finally, the set T can be constructed in a manner similar to Part (1): let T consist of 0 and the smallest positive element of S in each congruence class modulo k. Then $T' = \mathrm{slide}_k(T)$ is the result of an identical construction applied to S', and a set constructed this way contains all generators of the semigroup. Since S and S' are primitive, the sets T and T' are effective subsequences, since the condition of containing no composite elements less than the largest gap is vacuous.

Part (3). Now assume that $2m - 2 \in S$, $2m - 1 \notin S$, and all integers larger than $2m$ are in S. Then again, S is primitive. The largest gap of $S' = \mathrm{slide}_{2m-2}(S)$ is $2m - 3$, and the smallest element of S' is $m - 1$ (note that $m \neq 2m - 2$ since we are assuming $m \geq 3$), so S' is also a primitive semigroup. The rest of the argument is now analogous to the proof of Part (2). □

Lemma 5.8. *Let S be a secundive numerical semigroup and let m be the smallest generator of S. If $m + 1 \in S$, $2m - 2 \notin S$, and $2m - 1 \notin S$, then* $\mathrm{ewt}(S) \geq g - 1$. *Furthermore, equality* $\mathrm{ewt}(S) = g - 1$ *occurs if and only if $S = \{0, m, m + 1\} \cup H_{2m}$.*

Proof. Note that the hypotheses imply that $m \geq 4$, so $m + 1 < 2m - 2$. Also note that since S is secundive and contains $m + 1$, in fact S is primitive and the largest gap is $2m - 1$.

Since all elements of S less than the largest gap are generators, the effective weight (which is equal to the weight) is equal to the size of the set $E = \{(a, b) \in \mathbb{N}^2 : 0 < a < b, a \in S, b \notin S\}$.

The elements of E can be partitioned into four types:

(1) The pairs $(m, 2m-2)$, $(m, 2m-1)$, $(m+1, 2m-2)$, and $(m+1, 2m-1)$.

(2) Pairs of the form (m, a) or $(m+1, a)$, where $m+2 \le a \le 2m-3$ and $a \notin S$.

(3) Pairs of the form $(a, 2m-2)$ or $(a, 2m-1)$, where $m+2 \le a \le 2m-3$ and $a \in S$.

(4) Pairs of the form (a, b), where $m+2 \le a < b \le 2m-3$, $a \in S$, and $b \notin S$.

There are four pairs of the first type. Since every element a between $m+2$ and $2m-3$ inclusive appears in either two pairs of the second type or two pairs of the third type (depending on whether or not $a \in S$), the total number of pairs of either the second or third type is exactly $2(m-4)$. Therefore, adding the four pairs of the first type, $\mathrm{ewt}(S)$ is equal to $2m-4$ plus the number of pairs of the fourth type. On the other hand, the genus of S is at most $(m-1)+(m-2) = 2m-3$, with equality if and only if S contains no elements between $m+2$ and $2m-3$ inclusive. Hence $g-1 \le 2m-4 \le \mathrm{ewt}(S)$, with equality throughout if and only if S consists precisely of $0, m, m+1$ and all integers greater than or equal to $2m$. □

Corollary 5.9. *If S is a numerical semigroup with $1 \le \mathrm{ewt}(S) \le g-2$, then S has a good slider. If $\mathrm{ewt}(S) = g-1$, then S has a good slider unless $S = \{0, m, m+1\} \cup H_{2m}$ for some $m \ge 4$.*

Proof. Suppose that $1 \le \mathrm{ewt}(S) \le g-1$ and that S does not have a good slider. By Lemma 5.3, S is secundive. Let m be the smallest generator of S. By Lemma 5.7(1), $m+1 \in S$; this must be the second-smallest generator. Since S is secundive, the largest gap of S is less than $m+(m+1)$, so it is at most $2m-1$. Since $\mathrm{ewt}(S) > 0$, the largest gap is greater than m, hence at least $m+2$. Therefore $m+2 \le 2m-1$, so $m \ge 3$. By Lemma 5.7(2), the largest gap is in fact equal to $2m-1$, and by Lemma 5.7(3), $2m-2 \notin S$. This implies that $m+1 < 2m-2$, hence $m \ge 4$. By Lemma 5.8, it follows that $\mathrm{ewt}(S)$ is equal to $g-1$ and $S = \{0, m, m+1\} \cup H_{2m}$. □

6. Existence of effectively proper points

We can now prove Theorem 1.3 by assembling the ingredients of the previous sections and the following statement about elliptic curves. This lemma is similar to [Eisenbud and Harris 1987, Proposition 5.2], and plays an analogous role in our argument.

Lemma 6.1. *Fix integers d and r, and let T and T' be two vanishing sequences. As usual, denote the elements of these, in increasing order, by t_i and t'_i. Suppose that there exists an integer k, $1 \le k \le r$, such that:*

(1) *$t_0 = t'_0 = 0$ and $t_k = t'_k$.*

(2) *for all $i \notin \{0, k\}$, neither t_i nor t'_i is divisible by t_k.*

(3) *for all $i \notin \{0, k\}$, the inequalities $t_{i-1} \le t'_i < t_i$ hold.*

Then $\tilde{\mathcal{G}}^r_{1,d}(T', d-T)$ has dimensionally proper points.

Proof. Denote the number $t_k = t'_k$ by m. Fix an elliptic curve E with a point p. Consider the trivial family $E \times E \to E$ given by projection to the second coordinate, with two sections $s_1(q) = (p, q)$ and $s_2(q) = (q, q)$. Fix a point $q_0 \in E$ differing from p by torsion of order exactly m. Let B be the open subset of E given by removing p and all points q' differing from p by torsion of order dividing m, except the point q_0 itself. We can now regard $(E \times B, s_1, s_2)$ as a family of twice-marked elliptic curves $\{(E, p, q)\}_{q \in B}$, with the property that exactly one member (E, p, q_0) of this family has $p - q$ of torsion order m, and all others have $p - q$ either nontorsion or torsion of order not dividing m. We will show that $\tilde{G}^r_d(E \times B/B; (s_1, T'), (s_2, d - T))$ is nonempty of dimension $\rho_1(r, d; T', d - T) + 1$, which will prove the result.

More specifically, we will show that if p and q are points not differing by torsion of order dividing m (which is the case for all but one member of the family), $\tilde{G}^r_d(E; (p, T'), (q, d - T))$ is empty, while if p and q differ by torsion of order exactly m (the case for one member of the family), then $\tilde{G}^r_d(E; (p, T'), (q, d - T))$ is nonempty of dimension $\rho_1(r, d; T', d - T) + 1$.

Suppose that (E, p, q) is a twice-marked elliptic curve and (L, V) is some linear series with vanishing orders exactly T' at p and $d - T$ at q.

The key observation is that for any $i \in \{0, 1, \dots, r\}$, the subspace

$$V_i = V(-t'_i p - (d - t_i)q)$$

of V consisting of sections vanishing to order at least t'_i at p and order at least $d - t_i$ at q must be at least 1-dimensional.

In particular, for $i = 0$ and $i = k$ it follows that the divisors $d \cdot q$ and $m \cdot p + (d - m) \cdot q$ are both in the divisor class defined by the line bundle L. Hence L must be the line bundle $\mathcal{O}_E(d \cdot q)$, and the points p and q must differ by an element of $\operatorname{Pic}^0(E)$ of order dividing m. This shows that $G^r_d(E; (p, T'), (q, d - T))$ is indeed empty whenever p, q do not differ by torsion of order dividing m.

We will now assume that p and q differ by torsion of order exactly m.

Claim 1. *For $i = 0, 1, \dots, r - 1$, there are no sections $s \in V$ whose divisor of zeros contains $t'_{i+1} p + (d - t_i)q$.*

Proof of claim 1. If $i = 0, k - 1$, or k, then the divisor $t'_{i+1} p + (d - t_i)q$ has degree greater than d, so it certainly cannot be contained in the divisor of zeros of s. Otherwise, $t'_{i+1} + (d - t_i) \geq d$, so the only way for the divisor of s to contain such a divisor is if $t'_{i+1} = t_i$ and the divisor of s is exactly $t_i p + (d - t_i)q$. But this implies that $p - q$ is t_i-torsion, which is impossible since $m \nmid t_i$ when $i \neq 0, k$. □

Claim 2. *The space V_i is exactly 1-dimensional, and a nonzero section of V_i vanishes to order exactly t'_i at p and $d - t_i$ at q.*

Proof of claim 2. The second statement follows from claim 1. The first part follows from the second: in a 2-dimensional space of sections, there must be 2 distinct orders of vanishing at any given point. □

Therefore we see that (L, V) has a very simple form: $L = \mathcal{O}_E(d \cdot q)$ and V is the span of $r + 1$ disjoint 1-dimensional subspaces V_i, each of which is spanned by a section of L vanishing along the divisor

$t_i'p+(d-t_i)q$. Conversely, it is clear that any choice of these $r+1$ spaces V_i gives rise to a point of $\tilde{G}^r_d(E;\,(p,T'),(q,d-T))$. From this description, we can calculate from Riemann–Roch:

$$\begin{aligned}\dim \tilde{G}^r_d(E;\,(p,T'),(q,d-T)) &= \sum_{i=0}^{r} \dim \mathbb{P}H^0(L(-t_i'p-(d-t_i)q))\\ &= \sum_{i=0}^{r} \dim \mathbb{P}H^0(\mathcal{O}_E(-t_i'p+t_iq))\\ &= 2+\sum_{i=0}^{r}(t_i-t_i'-1).\end{aligned}$$

In the third line, we use the fact that for all $i\notin\{0,k\}$, $t_i-t_i'>0$, hence $h^1(E,\mathcal{O}_E(-t_i'p+t_iq))=0$, while for $i=0$ and $i=k$, $h^1(\mathcal{O}_E(-t_i'p+t_iq))=h^1(\mathcal{O}_E)=1$.

On the other hand, a bit of algebra shows that

$$\rho_1(r,d;\,T',d-T)=1+\sum_{i=0}^{r}(t_i-t_i'-1).$$

So we have established that, in the case where $p-q$ differ by torsion of order m,

$$\dim G^r_d(E;\,(p,T'),(q,d-T))=1+\rho_1(r,d;\,T',d-T).$$

By the remarks in the first paragraph, this proves that $\tilde{\mathcal{G}}^r_{1,d}(T',d-T)$ has dimensionally proper points. □

Corollary 6.2. *If S is a secundive numerical semigroup, k is a good slider for S, and $\mathcal{M}^{\mathrm{slide}_k(S)}_{g-1,1}$ has effectively proper points, then $\mathcal{M}^S_{g,1}$ has effectively proper points.*

Proof. Let T be an effective subsequence of S such that $T'=\mathrm{slide}_k(T)$ is an effective subsequence of $S'=\mathrm{slide}_k(S)$. Removing some elements if necessary, we may assume that T and T' contain no multiples of k other than 0 and k. Let $r=|T|-1$ and let $d=\max T$. By Proposition 4.5, $\tilde{\mathcal{G}}^r_{g-1,d}(d-T')$ has dimensionally proper points. By Lemma 6.1, $\tilde{\mathcal{G}}^r_{1,d}(T',d-T)$ also has dimensionally proper points. The regeneration lemma, Lemma 3.8, implies that $\tilde{\mathcal{G}}^r_{g,d}(d-T)$ also has dimensionally proper points; Proposition 4.5 now implies that $\mathcal{M}^S_{g,1}$ has an effectively proper component. □

Remark 6.3. In fact, the proof of Corollary 6.2 shows that the existence of effectively proper points of $\mathcal{M}^S_{g,1}$ can be deduced from the existence of effectively proper points of $\mathcal{M}^{S'}_{g-1,1}$ whenever S' and S are semigroups of genus $g-1$ and g respectively possessing effective subsequences T' and T that satisfy the hypotheses of Lemma 6.1. It is possible that more $\mathcal{M}^S_{g,1}$ can be shown to have effective proper points by constructing S' in a different way from the slide construction.

Proof of Theorem 1.3. Let S be a numerical semigroup of genus g, such that $\mathrm{ewt}(S)\le g-2$. We will prove that $\mathcal{M}^S_{g,1}$ has effectively proper components by induction on g. For $g=1$ the only semigroup is H_1, and there is nothing to prove.

Suppose that $g \geq 2$ and the result holds for genus $g-1$. If $\mathrm{ewt}(S) = 0$, then $S = H_g$ and $\mathcal{M}^S_{g,1}$ is a dense open subset of $\mathcal{M}_{g,1}$, so the result follows. Otherwise, Corollaries 5.9 and 6.2 show that the existence of an effectively proper point of $\mathcal{M}^S_{g,1}$ follows from the existence of an effectively proper point of $\mathcal{M}^{S'}_{g-1,1}$ for some semigroup S' of genus $g-1$ and effective weight $\mathrm{ewt}(S) - 1 \leq g-3$. This completes the induction.

Now suppose that $\operatorname{char} k = 0$ and S is a numerical semigroup of genus g such that $\mathrm{ewt}(S) = g-1$. The argument above works without modification, except in one case: g is odd and $S = \{0,\ \frac{1}{2}g+\frac{3}{2},\ \frac{1}{2}g+\frac{5}{2}\} \cup H_{g+3}$ (this is the exception in Corollary 5.9). The main theorem of [Komeda 1991] is that for this specific semigroup, in characteristic 0, $\mathcal{M}^S_{g,1}$ has dimensionally proper points (which are the same as effectively proper points, since S is primitive). With this possibility accounted for, the induction is complete in the case $\mathrm{ewt}(S) = g-1$ as well. □

Acknowledgments

Most of this paper was developed from part of my thesis at Harvard University, and I owe Joe Harris a great debt for suggesting the questions and providing several key insights. I also thank Nathan Kaplan for a number of helpful conversations on numerical semigroups, and the anonymous referee for detailed comments on the manuscript.

References

[Amini and Baker 2015] O. Amini and M. Baker, “Linear series on metrized complexes of algebraic curves”, *Math. Ann.* **362**:1-2 (2015), 55–106. MR Zbl

[Arbarello et al. 1985] E. Arbarello, M. Cornalba, P. A. Griffiths, and J. Harris, *Geometry of algebraic curves, I*, Grundlehren der Math. Wissenschaften **267**, Springer, 1985. MR Zbl

[Arbarello et al. 2011] E. Arbarello, M. Cornalba, and P. A. Griffiths, *Geometry of algebraic curves, II*, Grundlehren der Math. Wissenschaften **268**, Springer, 2011. MR Zbl

[Bras-Amorós and Bulygin 2009] M. Bras-Amorós and S. Bulygin, “Towards a better understanding of the semigroup tree”, *Semigroup Forum* **79**:3 (2009), 561–574. MR Zbl

[Bullock 2013] E. M. Bullock, “Subcanonical points on algebraic curves”, *Trans. Amer. Math. Soc.* **365**:1 (2013), 99–122. MR Zbl

[Bullock 2014] E. M. Bullock, “Irreducibility and stable rationality of the loci of curves of genus at most six with a marked Weierstrass point”, *Proc. Amer. Math. Soc.* **142**:4 (2014), 1121–1132. MR Zbl

[Coppens and Kato 1994] M. Coppens and T. Kato, “Weierstrass gap sequence at total inflection points of nodal plane curves”, *Tsukuba J. Math.* **18**:1 (1994), 119–129. MR Zbl

[Cox et al. 2011] D. A. Cox, J. B. Little, and H. K. Schenck, *Toric varieties*, Graduate Studies in Math. **124**, Amer. Math. Soc., Providence, RI, 2011. MR Zbl

[Eisenbud and Harris 1986] D. Eisenbud and J. Harris, “Limit linear series: basic theory”, *Invent. Math.* **85**:2 (1986), 337–371. MR Zbl

[Eisenbud and Harris 1987] D. Eisenbud and J. Harris, “Existence, decomposition, and limits of certain Weierstrass points”, *Invent. Math.* **87**:3 (1987), 495–515. MR Zbl

[Harris 1986] J. Harris, “On the Severi problem”, *Invent. Math.* **84**:3 (1986), 445–461. MR Zbl

[Harris and Morrison 1998] J. Harris and I. Morrison, *Moduli of curves*, Graduate Texts in Math. **187**, Springer, 1998. MR Zbl

[Hurwitz 1892] A. Hurwitz, "Ueber algebraische Gebilde mit eindeutigen Transformationen in sich", *Math. Ann.* **41**:3 (1892), 403–442. MR JFM

[Kaplan and Ye 2013] N. Kaplan and L. Ye, "The proportion of Weierstrass semigroups", *J. Algebra* **373** (2013), 377–391. MR Zbl

[Komeda 1991] J. Komeda, "On primitive Schubert indices of genus g and weight $g-1$", *J. Math. Soc. Japan* **43**:3 (1991), 437–445. MR Zbl

[Kontsevich and Zorich 2003] M. Kontsevich and A. Zorich, "Connected components of the moduli spaces of abelian differentials with prescribed singularities", *Invent. Math.* **153**:3 (2003), 631–678. MR Zbl

[Nakano 2008] T. Nakano, "On the moduli space of pointed algebraic curves of low genus, II: Rationality", *Tokyo J. Math.* **31**:1 (2008), 147–160. MR Zbl

[Osserman 2006] B. Osserman, "A limit linear series moduli scheme", *Ann. Inst. Fourier* (*Grenoble*) **56**:4 (2006), 1165–1205. MR Zbl

[Osserman 2013] B. Osserman, "Limit linear series", draft monograph, 2013, Available at https://tinyurl.com/ossdraft.

[Osserman 2014a] B. Osserman, "Limit linear series for curves not of compact type", preprint, 2014. To appear in *J. Reine Angew. Math.* arXiv

[Osserman 2014b] B. Osserman, "Limit linear series moduli stacks in higher rank", preprint, 2014. arXiv

[Pflueger 2016] N. Pflueger, "Weierstrass semigroups on Castelnuovo curves", preprint, 2016. arXiv

[Pinkham 1974] H. C. Pinkham, *Deformations of algebraic varieties with G_m action*, Astérisque **20**, Soc. Math. France, Paris, 1974. MR Zbl

[Ran 1989] Z. Ran, "Families of plane curves and their limits: Enriques' conjecture and beyond", *Ann. of Math.* (2) **130**:1 (1989), 121–157. MR Zbl

[Rim and Vitulli 1977] D. S. Rim and M. A. Vitulli, "Weierstrass points and monomial curves", *J. Algebra* **48**:2 (1977), 454–476. MR Zbl

[SGA 7_{II} 1973] P. Deligne, "Quadriques", pp. 62–81 in *Groupes de monodromie en géométrie algébrique, II: Exposés X–XXII* (Séminaire de Géométrie Algébrique du Bois Marie 1967–1969), edited by P. Deligne and N. Katz, Lecture Notes in Math. **340**, Springer, 1973. MR Zbl

[Zhai 2013] A. Zhai, "Fibonacci-like growth of numerical semigroups of a given genus", *Semigroup Forum* **86**:3 (2013), 634–662. MR Zbl

Communicated by Gavril Farkas
Received 2016-09-02 Revised 2018-01-05 Accepted 2018-03-10

Department of Mathematics and Statistics, Amherst College, Amherst, MA, United States
npflueger@amherst.edu

ALGEBRA AND NUMBER THEORY 12:8 (2018)
dx.doi.org/10.2140/ant.2018.12.1949

Bounded generation of SL_2 over rings of S-integers with infinitely many units

Aleksander V. Morgan, Andrei S. Rapinchuk and Balasubramanian Sury

To Alex Lubotzky on his 60th birthday

Let $\mathcal{O}$ be the ring of S-integers in a number field k. We prove that if the group of units $\mathcal{O}^\times$ is infinite then every matrix in $\Gamma = \mathrm{SL}_2(\mathcal{O})$ is a product of at most 9 elementary matrices. This essentially completes a long line of research in this direction. As a consequence, we obtain a new proof of the fact that Γ is boundedly generated as an abstract group that uses only standard results from algebraic number theory.

1. Introduction

Let k be a number field. Given a finite subset S of the set V^k of valuations of k containing the set V^k_∞ of archimedean valuations, we let $\mathcal{O}_{k,S}$ denote the ring of S-integers in k, i.e.,

$$\mathcal{O}_{k,S} = \{a \in k^\times \mid v(a) \geq 0 \text{ for all } v \in V^k \setminus S\} \cup \{0\}.$$

As usual, for any commutative ring R, we let $\mathrm{SL}_2(R)$ denote the group of unimodular 2×2-matrices over R and refer to the $\mathrm{SL}_2(R)$-matrices

$$E_{12}(a) = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad E_{21}(b) = \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix} \quad (a, b \in R)$$

as *elementary* (over R).

It was established in [Vasershtein 1972] (see also [Liehl 1981]) that if the ring of S-integers $\mathcal{O} = \mathcal{O}_{k,S}$ has infinitely many units, the group $\Gamma = \mathrm{SL}_2(\mathcal{O})$ is generated by elementary matrices. The goal of this paper is to prove that in this case Γ is actually *boundedly* generated by elementaries. More precisely, we prove the following.

Theorem 1.1. *Let $\mathcal{O} = \mathcal{O}_{k,S}$ be the ring of S-integers in a number field k, and assume that the group of units $\mathcal{O}^\times$ is infinite. Then every matrix in $\mathrm{SL}_2(\mathcal{O})$ is a product of at most 9 elementary matrices.*

The quest to validate the property that every element of $\mathrm{SL}_2(\mathcal{O})$ is a product of a bounded number of elementary matrices has a considerable history. First, G. Cooke and P. J. Weinberger [1975] established it (with the same bound as in Theorem 1.1) assuming the truth of a suitable form of the generalized Riemann hypothesis, which still remains unproven. Later, it was shown in [Loukanidis and Murty 1994] (see also

MSC2010: primary 11F06; secondary 11R37, 20H05.
Keywords: bounded generation, arithmetic groups, congruence subgroup problem.

[Murty 1995]) by analytic tools that the argument can be made unconditional if $|S| \geq \max(5, 2[k:\mathbb{Q}]-3)$. On the other hand, B. Liehl [1984] proved the result by algebraic methods for some special fields k. The first unconditional proof in full generality was given by D. Carter, G. Keller and E. Paige in an unpublished preprint; their argument was streamlined and made available to the public by D. W. Morris [2007]. This argument is based on model theory and provides no explicit bound on the number of elementaries required; besides, it uses difficult results from additive number theory.

M. Vsemirnov [2014] proved Theorem 1.1 for $\mathcal{O} = \mathbb{Z}[1/p]$ using the results of D. R. Heath-Brown [1986] on Artin's primitive root conjecture (thus, in a broad sense, this proof develops the initial approach of Cooke and Weinberger [1975]); his bound on the number of elementaries required is ≤ 5. Subsequently, the third-named author reworked the argument from [Vsemirnov 2014] to avoid the use of [Heath-Brown 1986] in an unpublished note. These notes were the beginning of the work of the first two authors that eventually led to a proof of Theorem 1.1 in the general case. It should be noted that our proof uses only standard results from number theory such as Artin reciprocity and Chebotarev's density theorem, and is relatively short and constructive with an explicit bound which is independent of the field k and the set S. This, in particular, implies that Theorem 1.1 remains valid for any *infinite* S.

The problem of bounded generation (particularly by elementaries) has been considered for S-arithmetic subgroups of algebraic groups other than SL_2. A few years after [Cooke and Weinberger 1975], Carter and Keller [1983] showed that $\mathrm{SL}_n(\mathcal{O})$ for $n \geq 3$ is boundedly generated by elementaries for any ring $\mathcal{O}$ of algebraic integers (see [Tavgen 1990] for other Chevalley groups of rank > 1, and [Erovenko and Rapinchuk 2006] for isotropic, but nonsplit (or quasisplit), orthogonal groups). The upper bound on the number of factors required to write every matrix in $\mathrm{SL}_n(\mathcal{O})$ as a product of elementaries given in [Carter and Keller 1983] is $\frac{1}{2}(3n^2-n)+68\Delta-1$, where Δ is the number of prime divisors of the discriminant of k; in particular, this estimate depends on the field k. Using our Theorem 1.1, one shows in all cases where the group of units $\mathcal{O}^\times$ is infinite, this estimate can be improved to $\frac{1}{2}(3n^2-n)+4$, hence made independent of k — see Corollary 4.6. The situation not covered by this result are when $\mathcal{O}$ is either $\mathbb{Z}$ or the ring of integers in an imaginary quadratic field — see below. The former case was treated in [Carter and Keller 1984] with an estimate $\frac{1}{2}(3n^2-n)+36$, so only in the case of imaginary quadratic fields the question of the existence of a bound on the number of elementaries independent of the k remains open.

From a more general perspective, Theorem 1.1 should be viewed as a contribution to the sustained effort aimed at proving that all higher rank lattices are boundedly generated as abstract groups. We recall that a group Γ is said to have *bounded generation* (BG) if there exist elements $\gamma_1, \dots, \gamma_d \in \Gamma$ such that

$$\Gamma = \langle \gamma_1 \rangle \cdots \langle \gamma_d \rangle,$$

where $\langle \gamma_i \rangle$ denotes the cyclic subgroup generated by γ_i. The interest in this property stems from the fact that while being purely combinatorial in nature, it is known to have a number of far-reaching consequences for the structure and representations of a group, particularly if the latter is S-arithmetic. For example, under one additional (necessary) technical assumption, (BG) implies the rigidity of completely reducible complex representations of Γ (known as SS-rigidity) — see [Rapinchuk 1990; Platonov and Rapinchuk

1994, Appendix A]. Furthermore, if Γ is an S-arithmetic subgroup of an absolutely simple simply connected algebraic group G over a number field k, then assuming the truth of the Margulis–Platonov conjecture for the group $G(k)$ of k-rational points [Platonov and Rapinchuk 1994, §9.1], (BG) implies the *congruence subgroup property* (i.e., the finiteness of the corresponding congruence kernel — see [Lubotzky 1995; Platonov and Rapinchuk 1992]). For applications of (BG) to the Margulis–Zimmer conjecture, see [Shalom and Willis 2013]. Given these and other implications of (BG), we would like to point out the following consequence of Theorem 1.1.

Corollary 1.2. *Let $\mathcal{O} = \mathcal{O}_{k,S}$ be the ring of S-integers, in a number field k. If the group of units $\mathcal{O}^\times$ is infinite, then the group $\Gamma = \mathrm{SL}_2(\mathcal{O})$ has bounded generation.*

We note that combining this fact with the results of [Lubotzky 1995; Platonov and Rapinchuk 1992], one obtains an alternative proof of the centrality of the congruence kernel for $\mathrm{SL}_2(\mathcal{O})$ (provided that $\mathcal{O}^\times$ is infinite), originally established by J.-P. Serre [1970]. We also note that (BG) of $\mathrm{SL}_2(\mathcal{O})$ is needed to prove (BG) for some other groups [Tavgen 1990; Erovenko and Rapinchuk 2006].

Next, it should be pointed out that the assumption that the unit group $\mathcal{O}^\times$ is infinite is *necessary* for the bounded generation of $\mathrm{SL}_2(\mathcal{O})$, hence cannot be omitted. Indeed, it follows from Dirichlet's unit theorem [Cassels and Fröhlich 1967, §2.18] that $\mathcal{O}^\times$ is finite only when $|S| = 1$ which happens precisely when S is the set of archimedian valuations in the following two cases:

(1) $k = \mathbb{Q}$ and $\mathcal{O} = \mathbb{Z}$. In this case, the group $\mathrm{SL}_2(\mathbb{Z})$ is generated by the elementaries, but has a nonabelian free subgroup of finite index, which prevents it from having bounded generation.

(2) $k = \mathbb{Q}(\sqrt{-d})$ for some square-free integer $d \geq 1$, and $\mathcal{O}_d$ is the ring of algebraic integers in k. According to [Grunewald and Schwermer 1981], the group $\Gamma = \mathrm{SL}_2(\mathcal{O}_d)$ has a finite index subgroup that admits an epimorphism onto a nonabelian free group, hence again cannot possibly be boundedly generated. Moreover, P. M. Cohn [1966] shows that if $d \notin \{1, 2, 3, 7, 11\}$ then Γ is not even generated by elementary matrices.

The structure of the paper is the following. In Section 2 we prove an algebraic result about abelian subextensions of radical extensions of general field — see Proposition 2.1. This statement, which may be of independent interest, is used in the paper to prove Theorem 3.7. This theorem is one of the number-theoretic results needed in the proof of Theorem 1.1, and it is established in Section 3 along with some other facts from algebraic number theory. One of the key notions in the paper is that of a $\mathbb{Q}$-split prime: we say that a prime $\mathfrak{p}$ of a number field k is $\mathbb{Q}$-split if it is nondyadic and its local degree over the corresponding rational prime is 1. In Section 3, we establish some relevant properties of such primes (see Section 3A) and prove in Section 3B the following (known — see the remark in Section 3) refinement of Dirichlet's theorem from [Bass et al. 1967].

Theorem 3.3. *Let $\mathcal{O}$ be the ring of S-integers in a number field k for some finite $S \subset V^k$ containing V^k_∞. If nonzero $a, b \in \mathcal{O}$ are relatively prime (i.e., $a\mathcal{O} + b\mathcal{O} = \mathcal{O}$) then there exist infinitely many principal $\mathbb{Q}$-split prime ideals $\mathfrak{p}$ of $\mathcal{O}$ with a generator π such that $\pi \equiv a \pmod{b\mathcal{O}}$ and $\pi > 0$ in all real completions of k.*

Section 3C is devoted to the statement and proof of Theorem 3.7, which is another key number-theoretic result needed in the proof of Theorem 1.1. In Section 4, we prove Theorem 1.1 and Corollary 1.2. Finally, in Section 5 we correct the faulty example from [Vsemirnov 2014] of a matrix in $\mathrm{SL}_2(\mathbb{Z}[1/p])$, where p is a prime $\equiv 1 \pmod{29}$, that is not a product of four elementary matrices — see Proposition 5.1, confirming thereby that the bound of 5 in [Vsemirnov 2014] is optimal.

Notations and conventions. For a field k, we let k^{ab} denote the maximal abelian extension of k. Furthermore, $\mu(k)$ will denote the group of all roots of unity in k; if $\mu(k)$ is finite, we let μ denote its order. For $n \geq 1$ prime to $\operatorname{char} k$, we let ζ_n denote a primitive n-th root of unity.

In this paper, with the exception of Section 2, the field k will be a field of algebraic numbers (i.e., a finite extension of $\mathbb{Q}$), in which case $\mu(k)$ is automatically finite. We let $\mathcal{O}_k$ denote the ring of algebraic integers in k. Furthermore, we let V^k denote the set of (the equivalence classes of) nontrivial valuations of k, and let V^k_∞ and V^k_f denote the subsets of archimedean and nonarchimedean valuations, respectively. For any $v \in V^k$, we let k_v denote the corresponding completion; if $v \in V^k_f$ then $\mathcal{O}_v$ will denote the valuation ring in k_v with the valuation ideal $\hat{\mathfrak{p}}_v$ and the group of units $U_v = \mathcal{O}_v^\times$.

Throughout the paper, S will denote a fixed finite subset of V^k containing V^k_∞, and $\mathcal{O} = \mathcal{O}_{k,S}$ the corresponding ring of S-integers (see above). Then the nonzero prime ideals of $\mathcal{O}$ are in a natural bijective correspondence with the valuations in $V^k \setminus S$. So, for a nonzero prime ideal $\mathfrak{p} \subset \mathcal{O}$ we let $v_{\mathfrak{p}} \in V^k \setminus S$ denote the corresponding valuation, and conversely, for a valuation $v \in V^k \setminus S$ we let $\mathfrak{p}_v \subset \mathcal{O}$ denote the corresponding prime ideal (note that $\mathfrak{p}_v = \mathcal{O} \cap \hat{\mathfrak{p}}_v$). Generalizing Euler's φ-function, for a nonzero ideal $\mathfrak{a}$ of $\mathcal{O}$, we set

$$\phi(\mathfrak{a}) = |(\mathcal{O}/\mathfrak{a})^\times|.$$

For simplicity of notation, for an element $a \in \mathcal{O}$, $\phi(a)$ will always mean $\phi(a\mathcal{O})$. Finally, for $a \in k^\times$, we let $V(a) = \{v \in V^k_f \mid v(a) \neq 0\}$.

Given a prime number p, one can write any integer n in the form $n = p^e \cdot m$, for some nonnegative integer e, where $p \nmid m$. We then call p^e the *p-primary component* of n.

2. Abelian subextensions of radical extensions

In this section, k is an arbitrary field. For a prime $p \neq \operatorname{char} k$, we let $\mu(k)_p$ denote the subgroup of $\mu(k)$, consisting of elements satisfying $x^{p^d} = 1$ for some $d \geq 0$. If this subgroup is finite, we set $\lambda(k)_p$ to be the nonnegative integer satisfying $|\mu(k)_p| = p^{\lambda(k)_p}$; otherwise, set $\lambda(k)_p = \infty$. Clearly if $\mu(k)$ is finite, then $\mu = \prod_p p^{\lambda(k)_p}$. For $a \in k^\times$, we write $\sqrt[n]{a}$ to denote an arbitrary root of the polynomial $x^n - a$.

The goal of this section is to prove the following.

Proposition 2.1. *Let $n \geq 1$ be an integer prime to* $\operatorname{char} k$*, and let $u \in k^\times$ be such that $u \notin \mu(k)_p k^{\times p}$ for all $p \mid n$. Then the polynomial $x^n - u$ is irreducible over k, and for $t = \sqrt[n]{u}$ we have*

$$k(t) \cap k^{\mathrm{ab}} = k(t^m) \quad \textit{where } m = \frac{n}{\prod_{p \mid n} \gcd(n, p^{\lambda(k)_p})},$$

with the convention that $\gcd(n, p^\infty)$ *is simply the p-primary component of n.*

We first treat the case $n = p^d$ where p is a prime.

Proposition 2.2. *Let p be a prime number $\neq \operatorname{char} k$, and let $u \in k^\times \setminus \mu(k)_p (k^\times)^p$. Fix an integer $d \geq 1$, set $t = \sqrt[p^d]{u}$. Then*

$$k(t) \cap k^{\mathrm{ab}} = k(t^{p^\gamma}) \quad \textit{where } \gamma = \max(0, d - \lambda(k)_p).$$

We begin with the following lemma.

Lemma 2.3. *Let p be a prime number $\neq \operatorname{char} k$, and let $u \in k^\times \setminus \mu(k)_p (k^\times)^p$. Set $k_1 = k(\sqrt[p]{u})$. Then*:

(i) $[k_1 : k] = p$.

(ii) $\mu(k_1)_p = \mu(k)_p$.

(iii) *None of the $\sqrt[p]{u}$ are in $\mu(k_1)_p (k_1^\times)^p$.*

Proof. (i) follows from [Lang 2002, Chapter VI, Theorem 9.1], as $u \notin (k^\times)^p$.

(ii) If $\lambda(k)_p = \infty$, then there is nothing to prove. Otherwise, we need to show that for $\lambda = \lambda(k)_p$, we have $\zeta_{p^{\lambda+1}} \notin k_1$. Assume the contrary. Then, first, $\lambda > 0$. Indeed, we have a tower of inclusions $k \subseteq k(\zeta_p) \subseteq k_1$. Since $[k_1 : k] = p$ by (i), and $[k(\zeta_p) : k] \leq p - 1$, we conclude that $[k(\zeta_p) : k] = 1$, i.e., $\zeta_p \in k$.

Now, since $\zeta_{p^{\lambda+1}} \notin k$, we have

$$k_1 = k(\zeta_{p^{\lambda+1}}) = k\big(\sqrt[p]{\zeta_{p^\lambda}}\big). \tag{1}$$

But according to Kummer's theory (which applies because $\zeta_p \in k$), the fact that $k(\sqrt[p]{a}) = k(\sqrt[p]{b})$ for $a, b \in k^\times$ implies that the images of a and b in $k^\times/(k^\times)^p$ generate the same subgroup. So, it follows from (1) that $u\zeta_p^i \in (k^\times)^p$ for some i, and therefore $u \in \mu(k)_p (k^\times)^p$, contradicting our choice of u.

(iii) Assume the contrary, i.e., some p-th root $\sqrt[p]{u}$ can be written in the form $\sqrt[p]{u} = \zeta a^p$ for some $a \in k_1^\times$ and $\zeta \in \mu(k_1)_p$. Let $N = N_{k_1/k} \colon k_1^\times \to k^\times$ be the norm map. Then

$$N(\sqrt[p]{u}) = N(\zeta) N(a)^p.$$

Clearly, $N(\zeta) \in \mu(k)_p$, so $N(\sqrt[p]{u}) \in \mu(k)_p (k^\times)^p$. On the other hand, $N(\sqrt[p]{u}) = u$ for p odd, and $-u$ for $p = 2$. In all cases, we obtain that $u \in \mu(k)_p (k^\times)^p$. A contradiction. □

A simple induction now yields the following:

Corollary 2.4. *Let p be a prime number $\neq \operatorname{char} k$, and let $u \in k^\times \setminus \mu(k)_p (k^\times)^p$. For a fixed integer $d \geq 1$, set $k_d = k(\sqrt[p^d]{u})$. Then*:

(i) $[k_d : k] = p^d$.

(ii) $\mu(k_d)_p = \mu(k)_p$, *hence* $\lambda(k_d)_p = \lambda(k)_p$.

Of course, assertion (i) is well known and follows, for example, from [Lang 2002, Chapter VI, §9].

Lemma 2.5. *Let p be a prime number $\neq \operatorname{char} k$, and let $u \in k^\times \setminus \mu(k)_p (k^\times)^p$. Fix an integer $d \geq 1$, and set $t = \sqrt[p^d]{u}$ and $k_d = k(t)$. Furthermore, for an integer j between 0 and d define $\ell_j = k(t^{p^{d-j}}) \simeq k(\sqrt[p^j]{u})$. Then any intermediate subfield $k \subseteq \ell \subseteq k_d$ is of the form $\ell = \ell_j$ for some $j \in \{0, \dots, d\}$.*

Proof. Given such an ℓ, it follows from Corollary 2.4(i) that $[k_d : \ell] = p^j$ for some $0 \le j \le d$. Since any conjugate of t is of the form $\zeta \cdot t$ where $\zeta^{p^d} = 1$, we see that the norm $N_{k_d/\ell}(t)$ is of the form $\zeta_0 t^{p^j}$, where again $\zeta_0^{p^d} = 1$. Then $\zeta_0 \in \mu(k_d)_p$, and using Corollary 2.4(ii), we conclude that $\zeta_0 \in k \subseteq \ell$. So, $t^{p^j} \in \ell$, implying the inclusion $\ell_{d-j} \subseteq \ell$. Now, the fact that $[k_d : \ell_{d-j}] = p^j$ implies that $\ell = \ell_{d-j}$, yielding our claim. □

Proof of Proposition 2.2. Set $\lambda = \lambda(k)_p$. Then for any $d \le \lambda$ the extension $k(\sqrt[p^d]{u})/k$ is abelian, and our assertion is trivial. So, we may assume that $\lambda < \infty$ and $d > \lambda$. It follows from Lemma 2.5 that $\ell := k(t) \cap k^{\mathrm{ab}}$ is of the form $\ell_{d-j} = k(t^{p^j})$ for some $j \in \{0, \dots, d\}$. On the other hand, ℓ_{d-j}/k is a Galois extension of degree p^{d-j}, so must contain the conjugate $\zeta_{p^{d-j}} t^{p^{d-j}}$ of $t^{p^{d-j}}$, implying that $\zeta_{p^{d-j}} \in \ell_{d-j}$. Since $\ell_{d-j} \simeq k(\sqrt[p^{d-j}]{u})$, we conclude from Corollary 2.4(ii) that $d - j \le \lambda$, i.e., $j \ge d - \lambda$. This proves the inclusion $\ell \subseteq k(t^{p^\gamma})$; the opposite inclusion is obvious. □

Proof of Proposition 2.1. Let $n = p_1^{\alpha_1} \cdots p_s^{\alpha_s}$ be the prime factorization of n, and for $i = 1, \dots, s$ set $n_i = n/p_i^{\alpha_i}$. Let $t = \sqrt[n]{u}$ and $t_i = t^{n_i}$ (so, t_i is a $p_i^{\alpha_i}$-th root of u). Using again [Lang 2002, Chapter VI, Theorem 9.1] we conclude that $[k(t) : k] = n$, which implies that

$$[k(t) : k(t_i)] = n_i \quad \text{for all } i = 1, \dots, r. \tag{2}$$

Since for $K := k(t) \cap k^{\mathrm{ab}}$ the degree $[K : k]$ divides n, we can write $K = K_1 \cdots K_s$ where K_i is an abelian extension of k of degree $p_i^{\beta_i}$ for some $\beta_i \le \alpha_i$. Then the degree $[K_i(t_i) : k(t_i)]$ must be a power of p_i. Comparing with (2), we conclude that $K_i \subseteq k(t_i)$. Applying Proposition 2.2 with $d = \alpha_i$, we obtain the inclusion

$$K_i \subseteq k(t_i^{p_i^{\gamma_i}}) = k(t^{n_i p_i^{\gamma_i}}) \quad \text{where } \gamma_i = \max(0, \alpha_i - \lambda(k)_{p_i}). \tag{3}$$

It is easy to see that the gcd of the numbers $n_i p_i^{\gamma_i}$ for $i = 1, \dots, s$ is

$$m = \frac{n}{\prod_{p \mid n} \gcd(n, p^{\lambda(k)_p})}.$$

Furthermore, the subgroup of $k(t)^\times$ generated by $t^{n_1 p_1^{\gamma_1}}, \dots, t^{n_s p_s^{\gamma_s}}$ coincides with the cyclic subgroup with generator t^m. Then (3) yields the following inclusion

$$K = K_1 \cdots K_s \subseteq k(t^m).$$

Since the opposite inclusion is obvious, our claim follows. □

Corollary 2.6. *Assume that $\mu = |\mu(k)| < \infty$. Let P be a finite set of rational primes $\neq \operatorname{char} k$, and define*

$$\mu' = \mu \cdot \prod_{p \in P} p.$$

Given $u \in k^\times$ such that

$$u \notin \mu(k)_p (k^\times)^p \quad \textit{for all } p \in P,$$

for any abelian extension F of k the intersection

$$E := F \cap k(\sqrt[\mu']{u}, \zeta_{\mu'})$$

is contained in $k(\sqrt[\mu]{u}, \zeta_{\mu'})$.

Proof. Without loss of generality we may assume that $\zeta_{\mu'} \in F$, and then we have the following tower of field extensions

$$k(\sqrt[\mu]{u}, \zeta_{\mu'}) \subset E(\sqrt[\mu]{u}) \subset k(\sqrt[\mu']{u}, \zeta_{\mu'}).$$

We note that the degree $[k(\sqrt[\mu']{u}, \zeta_{\mu'}) : k(\sqrt[\mu]{u}, \zeta_{\mu'})]$ divides $\prod_{p \in P} p$. So, if we assume that the assertion of the lemma is false, then we should be able to find to find a prime $p \in P$ that divides the degree $[E(\sqrt[\mu]{u}) : k(\sqrt[\mu]{u}, \zeta_{\mu'})]$, and therefore does *not* divide the degree $[k(\sqrt[\mu']{u}, \zeta_{\mu'}) : E(\sqrt[\mu]{u})]$. The latter implies that $\sqrt[p\mu]{u} \in E(\sqrt[\mu]{u})$. But this contradicts Proposition 2.1 since $E(\sqrt[\mu]{u}) = E \cdot k(\sqrt[\mu]{u})$ is an abelian extension of k. □

3. Results from algebraic number theory

3A. $\mathbb{Q}$-*split primes.* Our proof of Theorem 1.1 heavily relies on properties of so-called $\mathbb{Q}$-split primes in $\mathcal{O}$.

Definition. Let $\mathfrak{p}$ be a nonzero prime ideal of $\mathcal{O}$, and let p be the corresponding rational prime. We say that $\mathfrak{p}$ is $\mathbb{Q}$-*split* if $p > 2$, and for the valuation $v = v_{\mathfrak{p}}$ we have $k_v = \mathbb{Q}_p$.

For the convenience of further references, we list some simple properties of $\mathbb{Q}$-split primes.

Lemma 3.1. *Let $\mathfrak{p}$ be a $\mathbb{Q}$-split prime in $\mathcal{O}$, and for $n \geq 1$ let $\rho_n \colon \mathcal{O} \to \mathcal{O}/\mathfrak{p}^n$ be the corresponding quotient map. Then*:

(a) *The group of invertible elements $(\mathcal{O}/\mathfrak{p}^n)^\times$ is cyclic for any n.*

(b) *If $c \in \mathcal{O}$ is such that $\rho_2(c)$ generates $(\mathcal{O}/\mathfrak{p}^2)^\times$ then $\rho_n(c)$ generates $(\mathcal{O}/\mathfrak{p}^n)^\times$ for any $n \geq 2$.*

Proof. Let $p > 2$ be the rational prime corresponding to $\mathfrak{p}$, and $v = v_{\mathfrak{p}}$ be the associated valuation of k. By definition, $k_v = \mathbb{Q}_p$, hence $\mathcal{O}_v = \mathbb{Z}_p$. So, for any $n \geq 1$ we will have canonical ring isomorphisms

$$\mathcal{O}/\mathfrak{p}^n \simeq \mathcal{O}_v/\hat{\mathfrak{p}}_v^n = \mathbb{Z}_p/p^n\mathbb{Z}_p \simeq \mathbb{Z}/p^n\mathbb{Z}. \tag{4}$$

Then (a) follows from the well-known fact that the group $(\mathbb{Z}/p^n\mathbb{Z})^\times$ is cyclic. Furthermore, the isomorphisms in (4) are compatible for different n. Since the kernel of the group homomorphism $(\mathbb{Z}/p^n\mathbb{Z})^\times \to (\mathbb{Z}/p^2\mathbb{Z})^\times$ is contained in the Frattini subgroup of $(\mathbb{Z}/p^n\mathbb{Z})^\times$ for $n \geq 2$, the same is true for the homomorphism $(\mathcal{O}/\mathfrak{p}^n)^\times \to (\mathcal{O}/\mathfrak{p}^2)^\times$. This easily implies (b). □

Let $\mathfrak{p}$ be a $\mathbb{Q}$-split prime, let $v = v_{\mathfrak{p}}$ be the corresponding valuation. We will now define the *level* $\ell_{\mathfrak{p}}(u)$ of an element $u \in \mathcal{O}_v^\times$ and establish some properties of this notion that we will need later.

Let $p > 2$ be the corresponding rational prime. The group of p-adic units $\mathbb{U}_p = \mathbb{Z}_p^\times$ has the natural filtration by the congruence subgroups

$$\mathbb{U}_p^{(i)} = 1 + p^i\mathbb{Z}_p \quad \text{for } i \in \mathbb{N}.$$

It is well-known that

$$\mathbb{U}_p = C \times \mathbb{U}_p^{(1)}$$

where C is the cyclic group of order $(p-1)$ consisting of all roots of unity in $\mathbb{Q}_p$. Furthermore, the logarithmic map yields a continuous isomorphism $\mathbb{U}_p^{(i)} \to p^i\mathbb{Z}_p$, which implies that for any $u \in \mathbb{U}_p \setminus C$, the closure of the cyclic group generated by u has a decomposition of the form

$$\overline{\langle u \rangle} = C' \times \mathbb{U}_p^{(\ell)}$$

for some subgroup $C' \subset C$ and some integer $\ell = \ell_p(u) \geq 1$ which we will refer to as the *p-level* of u. We also set $\ell_p(u) = \infty$ for $u \in C$.

Returning now to a $\mathbb{Q}$-split prime $\mathfrak{p}$ of k and keeping the above notations, we define the *$\mathfrak{p}$-level* $\ell_{\mathfrak{p}}(u)$ of $u \in \mathcal{O}_v^\times$ as the p-level of the element in $\mathbb{U}_p$ that corresponds to u under the natural identification $\mathcal{O}_v = \mathbb{Z}_p$. We will need the following.

Lemma 3.2. *Let $\mathfrak{p}$ be a $\mathbb{Q}$-split prime in $\mathcal{O}$, let p be the corresponding rational prime, and $v = v_{\mathfrak{p}}$ the corresponding valuation. Suppose we are given an integer $d \geq 1$ not divisible by p, a unit $u \in \mathcal{O}_v^\times$ of infinite order having $\mathfrak{p}$-level $s = \ell_{\mathfrak{p}}(u)$, an integer n_s, and an element $c \in \mathcal{O}_v$ such that $u^{n_s} \equiv c \pmod{\mathfrak{p}^s}$. Then for any $t \geq s$ there exists an integer $n_t \equiv n_s \pmod{d}$ for which $u^{n_t} \equiv c \pmod{\mathfrak{p}^t}$.*

Proof. In view of the identification $\mathcal{O}_v = \mathbb{Z}_p$, it is enough to prove the corresponding statement for $\mathbb{Z}_p$. More precisely, we need to show the following: *Let $u \in \mathbb{U}_p$ be a unit of infinite order and p-level $s = \ell_p(u)$. If $c \in \mathbb{U}_p$ and $n_s \in \mathbb{Z}$ are such that $u^{n_s} \equiv c \pmod{p^s}$, then for any $t \geq s$ there exists $n_t \equiv n_s \pmod{d}$ such that $u^{n_t} \equiv c \pmod{p^t}$.* Thus, we have that $u^{n_s} \in c\mathbb{U}_p^{(s)}$, and we wish to show that

$$u^{n_s} \cdot \langle u^d \rangle \cap c\mathbb{U}_p^{(t)} \neq \varnothing.$$

Since $c\mathbb{U}_p^{(t)}$ is open, it is enough to show that

$$u^{n_s} \cdot \overline{\langle u^d \rangle} \cap c\mathbb{U}_p^{(t)} \neq \varnothing. \tag{5}$$

But since $\ell_p(u) = s$ and d is prime to p, we have the inclusion $\overline{\langle u^d \rangle} \supset \mathbb{U}_p^{(s)}$, and (5) is obvious. □

3B. *Dirichlet's theorem for $\mathbb{Q}$-split primes.* The following known (see the remark below) result gives the existence of $\mathbb{Q}$-split primes in arithmetic progressions.

Theorem 3.3. *Let $\mathcal{O}$ be the ring of S-integers in a number field k for some finite $S \subset V^k$ containing V_∞^k. If nonzero $a, b \in \mathcal{O}$ are relatively prime (i.e., $a\mathcal{O} + b\mathcal{O} = \mathcal{O}$) then there exist infinitely many principal $\mathbb{Q}$-split prime ideals $\mathfrak{p}$ of $\mathcal{O}$ with a generator π such that $\pi \equiv a \pmod{b\mathcal{O}}$ and $\pi > 0$ in all real completions of k.*

The proof follows the same general strategy as the proof of Dirichlet's theorem in [Bass et al. 1967] — see Theorem A.10 in the appendix on number theory. First, we will quickly review some basic facts from global class field theory (see, for example, [Cassels and Fröhlich 1967, Chapter VII]) and fix some notations. Let J_k denote the *group of ideles* of k with the natural topology; as usual, we identify $k^\times$ with the (discrete) *subgroup of principal ideles* in J_k. Then for every open subgroup $\mathcal{U} \subset J_k$ of finite index containing $k^\times$ there exists a finite abelian Galois extension L/k and a continuous surjective homomorphism $\alpha_{L/k} \colon J_k \to \mathrm{Gal}(L/k)$ (known as the *norm residue map*) such that:

- $\mathcal{U} = \operatorname{Ker} \alpha_{L/k} = N_{L/k}(J_L)k^\times$.
- For every nonarchimedean $v \in V^k$ which is unramified in L we let $\mathrm{Fr}_{L/k}(v)$ denote the Frobenius automorphism of L/k at v (i.e., the Frobenius automorphism $\mathrm{Fr}_{L/k}(w|v)$ associated to some (equivalently, any) extension $w|v$) and let $\boldsymbol{i}(v) \in J_k$ be an idele with the components
$$\boldsymbol{i}(v)_{v'} = \begin{cases} 1 & \text{if } v' \neq v, \\ \pi_v & \text{if } v' = v, \end{cases}$$
where $\pi_v \in k_v$ is a uniformizer; then $\alpha_{L/k}(\boldsymbol{i}(v)) = \mathrm{Fr}_{L/k}(v)$.

For our fixed finite subset $S \subset V^k$ containing V^k_∞, we define the following open subgroup of J_k:

$$U_S := \prod_{v \in S} k_v^\times \times \prod_{v \in V^k \setminus S} U_v.$$

Then the abelian extension of k corresponding to the subgroup $\mathcal{U}_S := U_S k^\times$ will be called the *Hilbert S-class field* of k and denoted K throughout the rest of the paper.

Next, we will introduce the idelic S-analogs of *ray groups*. Let $\mathfrak{b}$ be a nonzero ideal of $\mathcal{O} = \mathcal{O}_{k,S}$ with the prime factorization

$$\mathfrak{b} = \mathfrak{p}_1^{n_1} \cdots \mathfrak{p}_t^{n_t}, \tag{6}$$

let $v_i = v_{\mathfrak{p}_i}$ be the valuation in $V^k \setminus S$ associated with $\mathfrak{p}_i$, and let $V(\mathfrak{b}) = \{v_1, \ldots, v_t\}$. We then define an open subgroup

$$R_S(\mathfrak{b}) = \prod_{v \in V^k} R_v$$

where the open subgroups $R_v \subseteq k_v^\times$ are defined as follows. For v real, we let R_v be the subgroup of positive elements, letting $R_v = k_v^\times$ for all other $v \in S$, and setting $R_v = U_v$ for all $v \notin S \cup V(\mathfrak{b})$. It remains to define R_v for $v = v_i \in V(\mathfrak{b})$, in which case we set it to be the congruence subgroup $U_{v_i}^{(n_i)}$ of U_{v_i} modulo $\hat{\mathfrak{p}}_{v_i}^{n_i}$. We then let $K(\mathfrak{b})$ denote the abelian extension of k corresponding to $\boldsymbol{R}_S(\mathfrak{b}) := R_S(\mathfrak{b})k^\times$ ("ray class field"). (Obviously, $K(\mathfrak{b})$ contains K for any nonzero ideal $\mathfrak{b}$ of $\mathcal{O}$.) Furthermore, given $c \in k^\times$, we let $\boldsymbol{j}_{\mathfrak{b}}(c)$ denote the idele with the following components:

$$\boldsymbol{j}_{\mathfrak{b}}(c)_v = \begin{cases} c & \text{if } v \in V(\mathfrak{b}), \\ 1 & \text{if } v \notin V(\mathfrak{b}). \end{cases}$$

Then $\theta_{\mathfrak{b}} \colon k^\times \to \mathrm{Gal}(K(\mathfrak{b})/k)$ defined by $c \mapsto \alpha_{K(\mathfrak{b})/k}(\boldsymbol{j}_{\mathfrak{b}}(c))^{-1}$ is a group homomorphism.

The following lemma summarizes some simple properties of these definitions.

Lemma 3.4. *Let $\mathfrak{b} \subset \mathcal{O}$ be a nonzero ideal.*

(a) *If a nonzero $c \in \mathcal{O}$ is relatively prime to $\mathfrak{b}$ (i.e., $c\mathcal{O} + \mathfrak{b} = \mathcal{O}$) then $\theta_{\mathfrak{b}}(c)$ restricts to the Hilbert S-class field K trivially.*

(b) *If nonzero $c_1, c_2 \in \mathcal{O}$ are both relatively prime to $\mathfrak{b}$ then $c_1 \equiv c_2 \pmod{\mathfrak{b}}$ is equivalent to*

$$\mathrm{pr}_{\mathfrak{b}}(\boldsymbol{j}_{\mathfrak{b}}(c_1)R_S(\mathfrak{b})) = \mathrm{pr}_{\mathfrak{b}}(\boldsymbol{j}_{\mathfrak{b}}(c_2)R_S(\mathfrak{b})) \tag{7}$$

where $\mathrm{pr}_{\mathfrak{b}} \colon J_k \to \prod_{v \in V(\mathfrak{b})} k_v^{\times}$ is the natural projection.

Proof. (a) Since c is relatively prime to $\mathfrak{b}$, we have $\boldsymbol{j}_{\mathfrak{b}}(c) \in U_S$. So, using the functoriality properties of the norm residue map, we obtain

$$\theta_{\mathfrak{b}}(c)|K = \alpha_{K(\mathfrak{b})/k}(\boldsymbol{j}_{\mathfrak{b}}(c))^{-1}|K = \alpha_{K/k}(\boldsymbol{j}_{\mathfrak{b}}(c))^{-1} = \mathrm{id}_K$$

because $\boldsymbol{j}_{\mathfrak{b}}(c) \in U_S \subset \mathcal{U}_S = \operatorname{Ker} \alpha_{K/k}$, as required.

(b) As above, let (6) be the prime factorization of $\mathfrak{b}$, let $v_i = v_{\mathfrak{p}_i} \in V^k \setminus S$ be the valuation associated with $\mathfrak{p}_i$. Then for any $c_1, c_2 \in \mathcal{O}$, the congruence $c_1 \equiv c_2 \pmod{\mathfrak{b}}$ is equivalent to

$$c_1 \equiv c_2 \pmod{\hat{\mathfrak{p}}_{v_i}^{n_i}} \quad \text{for all } i = 1, \dots, t. \tag{8}$$

On the other hand, for any $v \in V_f^k$ and any $u_1, u_2 \in U_v$, the congruence $u_1 \equiv u_2 \pmod{\hat{\mathfrak{p}}_v^n}$ for $n \geq 1$ is equivalent to

$$u_1 U_v^{(n)} = u_2 U_v^{(n)},$$

where $U_v^{(n)}$ is the congruence subgroup of U_v modulo $\hat{\mathfrak{p}}_v^n$. Thus, for (nonzero) $c_1, c_2 \in \mathcal{O}$ prime to $\mathfrak{b}$, the conditions (7) and (8) are equivalent, and our assertion follows. □

We will now establish a result needed for the proof of Theorem 3.3 and its refinements.

Proposition 3.5. *Let $\mathfrak{b}$ be a nonzero ideal of $\mathcal{O}$, let $a \in \mathcal{O}$ be relatively prime to $\mathfrak{b}$, and let F be a finite Galois extension of $\mathbb{Q}$ that contains $K(\mathfrak{b})$. Assume that a rational prime p is unramified in F and there exists an extension w of the p-adic valuation v_p to F such that $\mathrm{Fr}_{F/\mathbb{Q}}(w|v_p)|K(\mathfrak{b}) = \theta_{\mathfrak{b}}(a)$. If the restriction v of w to k does not belong to $S \cup V(\mathfrak{b})$ then:*

(a) *$k_v = \mathbb{Q}_p$.*

(b) *The prime ideal $\mathfrak{p} = \mathfrak{p}_v$ of $\mathcal{O}$ corresponding to v is principal with a generator π satisfying $\pi \equiv a \pmod{\mathfrak{b}}$ and $\pi > 0$ in every real completion of k.*

We note since v is unramified in F which contains $K(\mathfrak{b})$, we in fact *automatically* have that $v \notin V(\mathfrak{b})$.

Proof. (a) Since the Frobenius $\mathrm{Fr}(w|v_p)$ generates $\mathrm{Gal}(F_w/\mathbb{Q}_p)$, our claim immediately follows from the fact that it acts trivially on k.

(b) According to (a), the local degree $[k_v : \mathbb{Q}_p]$ is 1, hence the residual degree $f(v|v_p)$ is also 1, and therefore

$$\mathrm{Fr}(w|v) = \mathrm{Fr}(w|v_p)^{f(v|v_p)} = \mathrm{Fr}(w|v_p).$$

Thus,

$$\alpha_{K(\mathfrak{b})/k}(\boldsymbol{i}(v)) = \mathrm{Fr}(w|v)|K(\mathfrak{b}) = \theta_{\mathfrak{b}}(a) = \alpha_{K(\mathfrak{b})/k}(\boldsymbol{j}_{\mathfrak{b}}(a))^{-1},$$

and therefore

$$\boldsymbol{i}(v)\boldsymbol{j}_{\mathfrak{b}}(a) \in \operatorname{Ker} \alpha_{K(\mathfrak{b})/K} = \boldsymbol{R}_S(\mathfrak{b}) = R_S(\mathfrak{b})k^{\times}.$$

So, we can write

$$\boldsymbol{i}(v)\boldsymbol{j}_{\mathfrak{b}}(a) = \boldsymbol{r}\pi \quad \text{with } \boldsymbol{r} \in R_S(\mathfrak{b}), \pi \in k^{\times}. \tag{9}$$

Then

$$\pi = \boldsymbol{i}(v)(\boldsymbol{j}_{\mathfrak{b}}(a)\boldsymbol{r}^{-1}).$$

Since a is prime to $\mathfrak{b}$, the idele $\boldsymbol{j}_{\mathfrak{b}}(a) \in U_S$, and then $\boldsymbol{j}_{\mathfrak{b}}(a)\boldsymbol{r}^{-1} \in U_S$. For any $v' \in V^k \setminus (S \cup \{v\})$, the v'-component of $\boldsymbol{i}(v)$ is trivial, so we obtain that $\pi \in U_{v'}$. On the other hand, the v-component of $\boldsymbol{i}(v)$ is a uniformizer π_v of k_v implying that π is also a uniformizer. Thus, $\mathfrak{p} = \pi\mathcal{O}$ is precisely the prime ideal associated with v. For any real v', the v'-components of $\boldsymbol{i}(v)$ and $\boldsymbol{j}_{\mathfrak{b}}(a)$ are trivial, so π equals the inverse of the v'-component of $\boldsymbol{r}$, hence positive in $k_{v'}$. Finally, it follows from (9) that

$$\mathrm{pr}_{\mathfrak{b}}(\boldsymbol{j}_{\mathfrak{b}}(a)) = \mathrm{pr}_{\mathfrak{b}}(\boldsymbol{j}_{\mathfrak{b}}(\pi)\boldsymbol{r}),$$

so $\pi \equiv a \pmod{\mathfrak{b}}$ by Lemma 3.4(b), as required. □

Proof of Theorem 3.3. Set $\mathfrak{b} = b\mathcal{O}$ and $\sigma = \theta_{\mathfrak{b}}(a) \in \mathrm{Gal}(K(\mathfrak{b})/k)$. Let F be the Galois closure of $K(\mathfrak{b})$ over $\mathbb{Q}$, and let $\tau \in \mathrm{Gal}(F/\mathbb{Q})$ be such that $\tau|K(\mathfrak{b}) = \sigma$. Applying Chebotarev's density theorem (see [Cassels and Fröhlich 1967, Chapter VII, 2.4] or [Bass et al. 1967, A.6]) we find infinitely many rational primes $p > 2$ for which the p-adic valuation v_p is unramified in F, does not lie below any valuations in $S \cup V(\mathfrak{b})$, and has an extension w to F such that $\mathrm{Fr}_{F/\mathbb{Q}}(w|v_p) = \tau$. Let $v = w|k$, and let $\mathfrak{p} = \mathfrak{p}_v$ be the corresponding prime ideal of $\mathcal{O}$. Since $p > 2$, Proposition 3.5(a) implies that $\mathfrak{p}$ is $\mathbb{Q}$-split. Furthermore, Proposition 3.5(b) asserts that $\mathfrak{p}$ has a generator π such that $\pi \equiv a \pmod{\mathfrak{b}}$ and $\pi > 0$ in every real completion of k, as required. □

Remark. Dong Quan Ngoc Nguyen pointed out to us that Theorem 3.3, hence the essential part of Dirichlet's theorem from [Bass et al. 1967] (in particular, (A.11)), was known already to Hasse [1926, Satz 13]. In the current paper, however, we use the approach described in [Bass et al. 1967] to establish the key Theorem 3.7; the outline of the constructions from [loc. cit.] as well as the technical Lemma 3.4 and Proposition 3.5 are included for this purpose. We note that in contrast to the argument in [loc. cit.], our proofs of Theorems 3.3 and 3.7 involve the application of Chebotarev's density theorem to *noncommutative* Galois extensions.

We will now prove a statement from Galois theory that we will need in the next subsection.

Lemma 3.6. *Let $F/\mathbb{Q}$ be a finite Galois extension, and let κ be an integer for which $F \cap \mathbb{Q}^{\mathrm{ab}} \subseteq \mathbb{Q}(\zeta_\kappa)$. Then $F(\zeta_\kappa) \cap \mathbb{Q}^{\mathrm{ab}} = \mathbb{Q}(\zeta_\kappa)$.*

Proof. We need to show that

$$[F(\zeta_\kappa) : F(\zeta_\kappa) \cap \mathbb{Q}^{\mathrm{ab}}] = [F(\zeta_\kappa) : \mathbb{Q}(\zeta_\kappa)]. \tag{10}$$

Let

$$G = \mathrm{Gal}(F(\zeta_\kappa)/\mathbb{Q}) \quad \text{and} \quad H = \mathrm{Gal}(F/\mathbb{Q}).$$

Then the left-hand side of (10) is equal to the order of the commutator subgroup $[G, G]$, while the right-hand side equals

$$[F : F \cap \mathbb{Q}(\zeta_\kappa)] = [F : F \cap \mathbb{Q}^{\mathrm{ab}}] = \big|[H, H]\big|.$$

Now, the restriction gives an *injective* group homomorphism

$$\psi \colon G \to H \times \mathrm{Gal}(\mathbb{Q}(\zeta_\kappa)/\mathbb{Q}).$$

Since the restriction $G \to H$ is surjective, we obtain that ψ implements an isomorphism between $[G, G]$ and $[H, H] \times \{1\}$. Thus, $[G, G]$ and $[H, H]$ have the same order, and (10) follows. □

3C. ***Key statement.*** In this subsection we will establish another number-theoretic statement which plays a crucial role in the proof of Theorem 1.1. To formulate it, we need to introduce some additional notations. As above, let $\mu = |\mu(k)|$ be the number of roots of unity in k, let K be the Hilbert S-class field of k, and let $\tilde{K}$ be the Galois closure of K over $\mathbb{Q}$. Suppose we are given two finite sets P and Q of rational primes. Let

$$\mu' = \mu \cdot \prod_{p \in P} p,$$

pick an integer $\lambda \geq 1$ which is divisible by μ and for which $\tilde{K} \cap \mathbb{Q}^{\mathrm{ab}} \subseteq \mathbb{Q}(\zeta_\lambda)$, and set

$$\lambda' = \lambda \cdot \prod_{q \in Q} q.$$

Theorem 3.7. *Let $u \in \mathcal{O}^\times$ be a unit of infinite order such that $u \notin \mu(k)_p (k^\times)^p$ for every prime $p \in P$, and let $\mathfrak{q}$ be a $\mathbb{Q}$-split prime of $\mathcal{O}$ which is relatively prime to λ'. Then there exist infinitely many principal $\mathbb{Q}$-split primes $\mathfrak{p} = \pi\mathcal{O}$ of $\mathcal{O}$ with a generator π such that*:

(1) *For each $p \in P$, the p-primary component of $\phi(\mathfrak{p})/\mu$ divides the p-primary component of the order of u* $\pmod{\mathfrak{p}}$.

(2) $\pi \pmod{\mathfrak{q}^2}$ *generates* $(\mathcal{O}/\mathfrak{q}^2)^\times$.

(3) $\gcd(\phi(\mathfrak{p}), \lambda') = \lambda$.

Proof. As in the proof of Theorem 3.3, we will derive the required assertion by applying Chebotarev's density theorem to a specific automorphism of an appropriate finite Galois extension.

Let $K(\mathfrak{q}^2)$ be the abelian extension $K(\mathfrak{b})$ of k introduced in Section 3B for the ideal $\mathfrak{b} = \mathfrak{q}^2$. Set

$$L_1 = K(\mathfrak{q}^2)(\zeta_{\lambda'}), \quad L_2 = k(\zeta_{\mu'}, \sqrt[\mu']{u}), \quad L = L_1 L_2 \quad \text{and} \quad \ell = L_1 \cap L_2.$$

Then

$$\mathrm{Gal}(L/k) = \{\sigma = (\sigma_1, \sigma_2) \in \mathrm{Gal}(L_1/k) \times \mathrm{Gal}(L_2/k) : \sigma_1 \mid \ell = \sigma_2 \mid \ell\}. \tag{11}$$

So, to construct $\sigma \in \mathrm{Gal}(L/k)$ that we will need in the argument it is enough to construct appropriate $\sigma_i \in \mathrm{Gal}(L_i/k)$ for $i = 1, 2$ that have the same restriction to ℓ.

Lemma 3.8. *The restriction maps define the following isomorphisms*:

(1) $\mathrm{Gal}(L_1/K) \simeq \mathrm{Gal}(K(\mathfrak{q}^2)/K) \times \mathrm{Gal}(K(\zeta_{\lambda'})/K)$.

(2) $\mathrm{Gal}(K(\zeta_{\lambda'})/K(\zeta_\lambda)) \simeq \mathrm{Gal}(\mathbb{Q}(\zeta_{\lambda'})/\mathbb{Q}(\zeta_\lambda)) \simeq \prod_{q \in Q} \mathrm{Gal}(\mathbb{Q}(\zeta_{q\lambda})/\mathbb{Q}(\zeta_\lambda))$.

Proof. (1) We need to show that $K(\mathfrak{q}^2) \cap K(\zeta_\lambda) = K$. But the Galois extensions $K(\mathfrak{q}^2)/K$ and $K(\zeta_\lambda)/K$ are respectively totally and unramified at the extensions of $v_{\mathfrak{q}}$ to K (since $\mathfrak{q}$ is prime to λ), so the required fact is immediate.

(2) Since $K(\zeta_{\lambda'}) = K(\zeta_\lambda) \cdot \mathbb{Q}(\zeta_{\lambda'})$, we only need to show that

$$K(\zeta_\lambda) \cap \mathbb{Q}(\zeta_{\lambda'}) = \mathbb{Q}(\zeta_\lambda). \tag{12}$$

We have

$$K(\zeta_\lambda) \cap \mathbb{Q}(\zeta_{\lambda'}) \subseteq \tilde{K}(\zeta_\lambda) \cap \mathbb{Q}^{\mathrm{ab}} = \mathbb{Q}(\zeta_\lambda)$$

by Lemma 3.6. This proves one inclusion in (12); the other inclusion is obvious. □

Since $\mathfrak{q}$ is $\mathbb{Q}$-split, the group $(\mathcal{O}/\mathfrak{q}^2)^\times$ is cyclic (Lemma 3.1(a)), and we pick $c \in \mathcal{O}$ so that $c \pmod{\mathfrak{q}^2}$ is a generator of this group. We then set

$$\sigma_1' = \theta_{\mathfrak{q}^2}(c) \in \mathrm{Gal}(K(\mathfrak{q}^2)/K)$$

in the notations of Section 3B (see Lemma 3.4(a)). Next, for $q \in Q$, we let $q^{e(q)}$ be the q-primary component of λ. Then using the isomorphism from Lemma 3.8(2), we can find $\sigma_1'' \in \mathrm{Gal}(K(\zeta_{\lambda'})/K)$ such that

$$\sigma_1''(\zeta_\lambda) = \zeta_\lambda \quad \text{but} \quad \sigma_1''(\zeta_{q^{e(q)+1}}) \neq \zeta_{q^{e(q)+1}} \quad \text{for all } q \in Q. \tag{13}$$

We then define $\sigma_1 \in \mathrm{Gal}(L_1/K)$ to be the automorphism corresponding to the pair (σ_1', σ_1'') in terms of the isomorphism from Lemma 3.8(1) (in other words, the restrictions of σ_1 to $K(\mathfrak{q}^2)$ and $K(\zeta_{\lambda'})$ are σ_1' and σ_1'', respectively).

We fix a μ'-th root $\sqrt[\mu']{u}$, and for $\nu|\mu'$ set $\sqrt[\nu]{u} = (\sqrt[\mu']{u})^{\mu'/\nu}$ (also denoted $u^{\nu^{-1}}$). To construct $\sigma_2 \in \mathrm{Gal}(L_2/k)$, we need the following.

Lemma 3.9. *Let* $\sigma_0 \in \mathrm{Gal}(\ell/k)$. *Then there exists* $\sigma_2 \in \mathrm{Gal}(L_2/k)$ *such that*

(1) $\sigma_2|\ell = \sigma_0$.

(2) *For any $p \in P$, if $p^{d(p)}$ is the p-primary component of μ then*

$$\sigma_2(u^{p^{-(d(p)+1)}}) \neq u^{p^{-(d(p)+1)}}.$$

Consequently either $\sigma_2(\zeta_{p^{d(p)+1}}) \neq \zeta_{p^{d(p)+1}}$ or σ_2 acts nontrivially on all *$p^{d(p)+1}$-th roots of u.*

Proof. Since L_1/k is an abelian extension, we conclude from Corollary 2.6 that

$$\ell \subseteq k(\sqrt[\mu]{u}, \zeta_{\mu'}) \subseteq k^{\mathrm{ab}}. \tag{14}$$

On the other hand, according to Proposition 2.1, none of the roots $\sqrt[p\mu]{u}$ for $p \in P$ lies in k^{ab}, and the restriction maps yield an isomorphism

$$\mathrm{Gal}\big(k(\sqrt[\mu']{u}, \zeta_{\mu'})/k(\sqrt[\mu]{u}, \zeta_{\mu'})\big) \to \prod_{p \in P} \mathrm{Gal}\big(k(\sqrt[p\mu]{u}, \zeta_{\mu'})/k(\sqrt[\mu]{u}, \zeta_{\mu'})\big).$$

It follows that for each $p \in P$ we can find $\tau_p \in \mathrm{Gal}\big(k(\sqrt[\mu']{u}, \zeta_{\mu'})/k(\sqrt[\mu]{u}, \zeta_{\mu'})\big)$ such that

$$\tau_p(u^{p^{-(d(p)+1)}}) = \zeta_p \cdot u^{p^{-(d(p)+1)}} \quad \text{and} \quad \tau_p(u^{q^{-(d(q)+1)}}) = u^{q^{-(d(q)+1)}} \quad \text{for all } q \in P \setminus \{p\}.$$

Now, let $\tilde{\sigma}_0$ be any extension of σ_0 to L_2. For $p \in P$, define

$$\chi(p) = \begin{cases} 1 & \text{if } \tilde{\sigma}_0(u^{p^{-(d(p)+1)}}) = u^{p^{-(d(p)+1)}}, \\ 0 & \text{if } \tilde{\sigma}_0(u^{p^{-(d(p)+1)}}) \neq u^{p^{-(d(p)+1)}} \end{cases}$$

Set

$$\sigma_2 = \tilde{\sigma}_0 \cdot \prod_{p \in P} \tau_p^{\chi(p)}.$$

In view of (14), all τ_p's act trivially on ℓ, so $\sigma_2 \mid \ell = \tilde{\sigma}_0|\ell = \sigma_0$ and (1) holds. Furthermore, the choice of the τ_p's and the $\chi(p)$'s implies that (2) also holds. □

Continuing the proof of Theorem 3.7, we now use $\sigma_1 \in \mathrm{Gal}(L_1/k)$ constructed above, set $\sigma_0 = \sigma_1|\ell$, and using Lemma 3.9 construct $\sigma_2 \in \mathrm{Gal}(L_2/k)$ with the properties described therein. In particular, part (1) of this lemma in conjunction with (11) implies that the pair (σ_1, σ_2) corresponds to an automorphism $\sigma \in \mathrm{Gal}(L/k)$. As in the proof of Theorem 3.3, we let F denote the Galois closure of L over $\mathbb{Q}$, and let $\tilde{\sigma} \in \mathrm{Gal}(F/\mathbb{Q})$ be such that $\tilde{\sigma}|L = \sigma$. By Chebotarev's density theorem, there exist infinitely many rational primes $\pi > 2$ that are relatively prime to $\lambda' \cdot \mu'$ and for which the π-adic valuation v_π is unramified in F, does not lie below any valuation in $S \cup \{v_{\mathfrak{q}}\}$, and has an extension w to F such that $\mathrm{Fr}_{F/\mathbb{Q}}(w|v_\pi) = \tilde{\sigma}$. Let $v = w|k$, and let $\mathfrak{p} = \mathfrak{p}_v$ be the corresponding prime ideal of $\mathcal{O}$. As in the proof of Theorem 3.3, we see that $\mathfrak{p}$ is $\mathbb{Q}$-split. Furthermore, since $\sigma|K(\mathfrak{q}^2) = \theta_{\mathfrak{q}^2}(c)$, we conclude that $\mathfrak{p}$ has a generator π such that $\pi \equiv c \pmod{\mathfrak{q}^2}$ (see Proposition 3.5(b)). Then by construction $\pi \pmod{\mathfrak{q}^2}$ generates $(\mathcal{O}/\mathfrak{q}^2)^\times$, verifying condition (2) of Theorem 3.7.

To verify condition (1), we fix $p \in P$ and consider two cases. First, suppose $\sigma(\zeta_{p^{d(p)+1}}) \neq \zeta_{p^{d(p)+1}}$. Since p is prime to $\mathfrak{p}$, this means that the residue field $\mathcal{O}/\mathfrak{p}$ does not contain an element of order $p^{d(p)+1}$ (although, since μ is prime to $\mathfrak{p}$, it does contain an element of order μ, hence of order $p^{d(p)}$). So, in this case $\phi(\mathfrak{p})/\mu$ is prime to p, and there is nothing to prove. Now, suppose that $\sigma(\zeta_{p^{d(p)+1}}) = \zeta_{p^{d(p)+1}}$. Then by construction σ acts nontrivially on every $p^{d(p)+1}$-th root of u, and therefore the polynomial $X^{p^{d(p)+1}} - u$ has no roots in k_v. Again, since p is prime to $\mathfrak{p}$, we see from Hensel's lemma that $u \pmod{\mathfrak{p}}$ is not a $p^{d(p)+1}$-th power in the residue field. It follows that the p-primary component of the order of $u \pmod{\mathfrak{p}}$ is not less than the p-primary component of $\phi(\mathfrak{p})/p^{d(p)}$, and (1) follows.

Finally, by construction σ acts trivially on ζ_λ but nontrivially on $\zeta_{q\lambda}$ for any $q \in Q$. Since $\mathfrak{p}$ is prime to λ', we see that the residue field $\mathcal{O}/\mathfrak{p}$ contains an element of order λ, but does not contain an element of order $q\lambda$ for any $q \in Q$. This means that $\lambda \mid \phi(\mathfrak{p})$ but $\phi(\mathfrak{p})/\lambda$ is relatively prime to each $q \in Q$, which is equivalent to condition (3) of Theorem 3.7. □

4. Proof of Theorem 1.1

First, we will introduce some additional notation needed to convert the task of factoring a given matrix $A \in \mathrm{SL}_2(\mathcal{O})$ as a product of elementary matrices into the task of reducing the first row of A to $(1, 0)$. Let

$$\mathcal{R}(\mathcal{O}) = \{(a, b) \in \mathcal{O}^2 \mid a\mathcal{O} + b\mathcal{O} = \mathcal{O}\}$$

(note that $\mathcal{R}(\mathcal{O})$ is precisely the set of all first rows of matrices $A \in \mathrm{SL}_2(\mathcal{O})$). For $\lambda \in \mathcal{O}$, one defines two permutations, $e_+(\lambda)$ and $e_-(\lambda)$, of $\mathcal{R}(\mathcal{O})$ given respectively by

$$(a, b) \mapsto (a, b + \lambda a) \quad \text{and} \quad (a, b) \mapsto (a + \lambda b, b).$$

These permutations will be called *elementary transformations* of $\mathcal{R}(\mathcal{O})$. For (a, b), $(c, d) \in \mathcal{R}(\mathcal{O})$ we will write $(a, b) \overset{n}{\Longrightarrow} (c, d)$ to indicate the fact that (c, d) can be obtained from (a, b) by a sequence of n (equivalently, $\leq n$) elementary transformations. For the convenience of further reference, we will record some simple properties of this relation.

Lemma 4.1. *Let $(a, b) \in \mathcal{R}(\mathcal{O})$.*

(1a) *If $(c, d) \in \mathcal{R}(\mathcal{O})$ and $(a, b) \overset{n}{\Longrightarrow} (c, d)$, then $(c, d) \overset{n}{\Longrightarrow} (a, b)$.*

(1b) *If (c, d), $(e, f) \in \mathcal{R}(\mathcal{O})$ are such that $(a, b) \overset{m}{\Longrightarrow} (c, d)$ and $(c, d) \overset{n}{\Longrightarrow} (e, f)$, then $(a, b) \overset{m+n}{\Longrightarrow} (e, f)$.*

(2a) *If $c \in \mathcal{O}$ such that $c \equiv a \pmod{b\mathcal{O}}$, then $(c, b) \in \mathcal{R}(\mathcal{O})$, and $(a, b) \overset{1}{\Longrightarrow} (c, b)$.*

(2b) *If $d \in \mathcal{O}$ such that $d \equiv b \pmod{a\mathcal{O}}$, then $(a, d) \in \mathcal{R}(\mathcal{O})$, and $(a, b) \overset{1}{\Longrightarrow} (a, d)$.*

(3a) *If $(a, b) \overset{n}{\Longrightarrow} (1, 0)$ then any matrix $A \in \mathrm{SL}_2(\mathcal{O})$ with the first row (a, b) is a product of $\leq n + 1$ elementary matrices.*

(3b) *If $(a, b) \overset{n}{\Longrightarrow} (0, 1)$ then any matrix $A \in \mathrm{SL}_2(\mathcal{O})$ with the second row (a, b) is a product of $\leq n + 1$ elementary matrices.*

(4a) *If $a \in \mathcal{O}^\times$ then $(a, b) \overset{2}{\Longrightarrow} (0, 1)$.*

(4b) *If $b \in \mathcal{O}^\times$ then $(a, b) \overset{2}{\Longrightarrow} (1, 0)$.*

Proof. (1a) We observe that the inverse of an elementary transformation is again an elementary transformation given by $[e_{\pm}(\lambda)]^{-1} = e_{\pm}(-\lambda)$, so the required fact follows. Part (1b) is obvious.

(Note that (1) implies that the relation between (a, b) and $(c, d) \in \mathcal{R}(O)$ defined by $(a, b) \xRightarrow{n} (c, d)$ for *some* $n \in \mathbb{N}$ is an equivalence relation.)

(2a) We have $c = a + \lambda b$ with $\lambda \in \mathcal{O}$. Then

$$c\mathcal{O} + b\mathcal{O} = a\mathcal{O} + b\mathcal{O} = \mathcal{O},$$

so $(c, a) \in \mathcal{R}(\mathcal{O})$, and $e_+(\lambda)$ takes (a, b) to (c, b). The argument for (2b) is similar.

(3a) Suppose $A \in \mathrm{SL}_2(\mathcal{O})$ has the first row (a, b). Then for $\lambda \in \mathcal{O}$, the first row of the product $AE_{12}(\lambda)$ is $(a, b + \lambda a) = e_+(\lambda)(a, b)$, and similarly the first row of $AE_{21}(\lambda)$ is $e_-(\lambda)(a, b)$. So, the fact that $(a, b) \xRightarrow{n} (1, 0)$ implies that there exists a matrix $U \in \mathrm{SL}_2(\mathcal{O})$ which is a product of n elementary matrices and is such that AU has the first row $(1, 0)$. This means that $AU = E_{21}(z)$ for some $z \in \mathcal{O}$, and then $A = E_{21}(z)U^{-1}$ is a product of $\leq n + 1$ elementary matrices. The argument for (3b) is similar.

(4a) This follows since $e_-(-a)e_+(a^{-1}(1 - b))(a, b) = (0, 1)$. The proof of (4b) is similar. □

Remark. All assertions of Lemma 4.1 are valid over any commutative ring $\mathcal{O}$.

Corollary 4.2. *Let $\mathfrak{q}$ be a principal $\mathbb{Q}$-split prime ideal of $\mathcal{O}$ with generator q, and let $z \in \mathcal{O}$ be such that $z \pmod{\mathfrak{q}^2}$ generates $(\mathcal{O}/\mathfrak{q}^2)^\times$. Given an element of $\mathcal{R}(\mathcal{O})$ of the form (b, q^n) with $n \geq 2$, and an integer t_0, there exists an integer $t \geq t_0$ such that $(b, q^n) \xRightarrow{1} (z^t, q^n)$.*

Proof. By Lemma 3.1(b), the element $z \pmod{\mathfrak{q}^n}$ generates $(\mathcal{O}/\mathfrak{q}^n)^\times$. Since b is prime to $\mathfrak{q}$, one can find $t \in \mathbb{Z}$ such that $b \equiv z^t \pmod{\mathfrak{q}^n}$. Adding to t a suitable multiple of $\phi(\mathfrak{q}^n)$ if necessary, we can assume that $t \geq t_0$. Our assertion then follows from Lemma 4.1(2a). □

Lemma 4.3. *Suppose we are given $(a, b) \in \mathcal{R}(\mathcal{O})$, a finite subset $T \subseteq V_f^k$, and an integer $n \neq 0$. Then there exists $\alpha \in \mathcal{O}_k$ and $r \in \mathcal{O}^\times$ such that $V(\alpha) \cap T = \varnothing$, and $(a, b) \xRightarrow{1} (\alpha r^n, b)$.*

Proof. Let h_k be the class number of k. If for each $v \in S \setminus V_\infty^k$ we let $\mathfrak{m}_v$ denote the maximal ideal of $\mathcal{O}_k$ corresponding to v, then the ideal $(\mathfrak{m}_v)^{h_k}$ is principal, and its generator π_v satisfies $v(\pi_v) = h_k$ and $w(\pi_v) = 0$ for all $w \in V_f^k \setminus \{v\}$. Let R be the subgroup of $k^\times$ generated by π_v for $v \in S \setminus V_\infty^k$; note that $R \subset \mathcal{O}^\times$. We can pick $r \in R$ so that $a' := ar^{-n} \in \mathcal{O}_k$. We note that since a and b are relatively prime in $\mathcal{O}$, we have $V(a') \cap V(b) \subset S$.

Now, it follows from the strong approximation theorem that there exists $\gamma \in \mathcal{O}_k$ such that

$$v(\gamma b) \geq 0 \quad \text{and} \quad v(\gamma b) \equiv 0 \pmod{nh_k} \quad \text{for all } v \in S \setminus V_\infty^k,$$

and

$$v(\gamma b) = 0 \quad \text{for all } v \in V(a') \setminus S.$$

Then, in particular, we can find $s \in R$ so that $v(\gamma bs^{-1}) = 0$ for all $v \in S \setminus V_\infty^k$. Set

$$\gamma' := \gamma s^{-1} \in \mathcal{O} \quad \text{and} \quad b' := \gamma' b \in \mathcal{O}_k.$$

By construction,

$$v(b') = 0 \quad \text{for all } v \in V(a') \cup (S \setminus V_\infty^k), \tag{15}$$

implying that $V(a') \cap V(b') = \varnothing$, which means that a' and b' are relatively prime in $\mathcal{O}_k$.

Again, by the strong approximation theorem we can find $t \in \mathcal{O}_k$ such that

$$v(t) = 0 \;\text{ for } v \in T \cap V(a') \quad \text{and} \quad v(t) > 0 \;\text{ for } v \in T \setminus V(a').$$

Set $\alpha = a' + tb' \in \mathcal{O}_k$. Then for $v \in T \cap V(a')$ we have $v(a') > 0$ and $v(tb') = 0$ (in view of (15)), while for $v \in T \setminus V(a')$ we have $v(a') = 0$ and $v(tb') > 0$. In either case,

$$v(\alpha) = v(a' + tb') = 0 \quad \text{for all } v \in T,$$

i.e., $V(\alpha) \cap T = \varnothing$. On the other hand,

$$a + r^n t\gamma' b = r^n(a' + tb') = r^n\alpha,$$

which means that $(a, b) \overset{1}{\Longrightarrow} (\alpha r^n, b)$, as required. □

Recall that we let μ denote the number of roots of unity in k.

Lemma 4.4. *Let $(a, b) \in \mathcal{R}(\mathcal{O})$ be such that $a = \alpha \cdot r^\mu$ for some $\alpha \in \mathcal{O}_k$ and $r \in \mathcal{O}^\times$ where $V(\alpha)$ is disjoint from $S \cup V(\mu)$. Then there exist $a' \in \mathcal{O}$ and infinitely many $\mathbb{Q}$-split prime principal ideals $\mathfrak{q}$ of $\mathcal{O}$ with a generator q such that for any $m \equiv 1 \pmod{\phi(a'\mathcal{O})}$ we have $(a, b) \overset{3}{\Longrightarrow} (a', q^{\mu m})$.*

Proof. The argument below is adapted from the proof of Lemma 3 in [Carter and Keller 1983]. It relies on the properties of the power residue symbol (in particular, the power reciprocity law) described in the appendix on number theory in [Bass et al. 1967]. We will work with all $v \in V^k$ (and not only $v \in V^k \setminus S$), so to each such v we associate a symbol ("modulus") $\mathfrak{m}_v$. For $v \in V_f^k$ we will identify $\mathfrak{m}_v$ with the corresponding maximal ideal of $\mathcal{O}_k$ (obviously, $\mathfrak{p}_v = \mathfrak{m}_v\mathcal{O}$ for $v \in V^k \setminus S$); the valuation ideal and the group of units in the valuation ring $\mathcal{O}_v$ (or $\mathcal{O}_{\mathfrak{m}_v}$) in the completion k_v will be denoted $\hat{\mathfrak{m}}_v$ and U_v respectively. For any divisor $\kappa \mid \mu$, we let

$$\left(\frac{*\,,\,*}{\mathfrak{m}_v}\right)_\kappa$$

be the (bimultiplicative, skew-symmetric) power residue symbol of degree κ on $k_v^\times$ [Bass et al. 1967, p.85]. We recall that $\left(\frac{x,y}{\mathfrak{m}_v}\right)_\kappa = 1$ if one of the elements x, y is a κ-th power in $k_v^\times$ (in particular, if either v is complex or v is real and one of the elements x, y is positive in k_v) or if v is nonarchimedean $\notin V(\kappa)$ and $x, y \in U_v$. It follows that for any $x, y \in k^\times$, we have $\left(\frac{x,y}{\mathfrak{m}_v}\right)_\kappa = 1$ for almost all $v \in V^k$. Furthermore, we have the *reciprocity law*:

$$\prod_{v \in V^k} \left(\frac{x,\, y}{\mathfrak{m}_v}\right)_\kappa = 1. \tag{16}$$

Now, let $\mu = p_1^{e_1} \cdots p_n^{e_n}$ be a prime factorization of μ. For each $i = 1, \dots, n$, pick $v_i \in V(p_i)$. According to [Bass et al. 1967, A.17], the values

$$\left(\frac{x, y}{\mathfrak{m}_{v_i}}\right)_{p_i^{e_i}} \quad \text{for } x, y \in U_{v_i}$$

cover all $p_i^{e_i}$-th roots of unity. Thus, we can pick units $u_i, u_i' \in U_{v_i}$ for $i = 1, \dots, n$ so that $\left(\frac{u_i, u_i'}{\mathfrak{m}_{v_i}}\right)_{p_i^{e_i}}$ is a primitive $p_i^{e_i}$-th root of unity. On the other hand, since $u_i, u_i' \in U_{v_i}$ and $v_i(\mu/p_i^{e_i}) = 0$, we have

$$\left(\frac{u_i, u_i'}{\mathfrak{m}_{v_i}}\right)_{\mu}^{p_i^{e_i}} = \left(\frac{u_i, u_i'}{\mathfrak{m}_{v_i}}\right)_{\mu/p_i^{e_i}} = 1.$$

Thus,

$$\zeta_{p_i^{e_i}} := \left(\frac{u_i, u_i'}{\mathfrak{m}_{v_i}}\right)_{\mu}$$

is a primitive $p_i^{e_i}$-th root of unity for each $i = 1, \dots, n$, making

$$\zeta_\mu := \prod_{i=1}^{n} \left(\frac{u_i, u_i'}{\mathfrak{m}_{v_i}}\right)_{\mu} \tag{17}$$

a primitive μ-th root of unity. Furthermore, it follows from the inverse function theorem or Hensel's lemma that we can find an integer $N > 0$ such that

$$1 + \hat{\mathfrak{m}}_v^N \subset k_v^{\times\mu} \quad \text{for all } v \in V(\mu). \tag{18}$$

We now write $b = \beta t^\mu$ with $\beta \in \mathcal{O}_k$ and $t \in \mathcal{O}^\times$. Since a, b are relatively prime in $\mathcal{O}$, so are α, β, hence $V(\alpha) \cap V(\beta) \subset S$. On the other hand, by our assumption $V(\alpha)$ is disjoint from $S \cup V(\mu)$, so we conclude that $V(\alpha)$ is disjoint from $V(\beta) \cup V(\mu)$. Applying Theorem 3.3 to the ring $\mathcal{O}_k$ we obtain that there exists $\beta' \in \mathcal{O}_k$ having the following properties:

(1_1) $\mathfrak{b} := \beta'\mathcal{O}_k$ is a prime ideal of $\mathcal{O}_k$ and the corresponding valuation $v_{\mathfrak{b}} \notin S \cup V(\mu)$.

(2_1) $\beta' > 0$ in every real completion of k.

(3_1) $\beta' \equiv \beta \pmod{\alpha\mathcal{O}_k}$.

(4_1) For each $i = 1, \dots, n$, we have

$$\beta' \equiv u_i' \pmod{\hat{\mathfrak{m}}_{v_i}^N} \quad \text{and} \quad \beta' \equiv 1 \pmod{\hat{\mathfrak{m}}_v^N}$$

for all $v \in V(p_i) \setminus \{v_i\}$.

Set $b' = \beta' t^\mu$. It is a consequence of $(3)_1$ that $b \equiv b' \pmod{a\mathcal{O}}$, so by Lemma 4.1(2) we have $(a, b) \xRightarrow{1} (a, b')$. Furthermore, it follows from $(4)_1$ and (18) that $\beta'/u_i' \in k_{v_i}^{\times\mu}$, so

$$\left(\frac{u_i, \beta'}{\mathfrak{m}_{v_i}}\right)_{\mu} = \left(\frac{u_i, u_i'}{\mathfrak{m}_{v_i}}\right)_{\mu} = \zeta_{p_i^{e_i}}.$$

Since ζ_μ defined by (17) is a primitive μ-th root of unity, we can find an integer $d > 0$ such that

$$1 = \left(\frac{\alpha, \beta'}{\mathfrak{b}}\right)_\mu \cdot \zeta_\mu^d = \left(\frac{\alpha, \beta'}{\mathfrak{b}}\right)_\mu \cdot \prod_{i=1}^{n} \left(\frac{u_i^d, \beta'}{\mathfrak{m}_{v_i}}\right)_\mu. \tag{19}$$

By construction, $v_{\mathfrak{b}} \notin V(\alpha) \cup V(\mu)$, so applying Theorem 3.3 one more time, we find $\alpha' \in \mathcal{O}_k$ such that:

(1_2) $\mathfrak{a} := \alpha' \mathcal{O}_k$ is a prime ideal of $\mathcal{O}_k$ and the corresponding valuation $v_{\mathfrak{a}} \notin S \cup V(\mu)$.

(2_2) $\alpha' \equiv \alpha \pmod{\mathfrak{b}}$.

(3_2) $\alpha' \equiv u_i^d \pmod{\hat{\mathfrak{m}}_{v_i}^N}$ for $i = 1, \dots, n$.

Set $a' = \alpha' r^\mu$. Then $a'\mathcal{O} = \alpha'\mathcal{O}$ is a prime ideal of $\mathcal{O}$ and $a' \equiv a \pmod{b'\mathcal{O}}$, so $(a, b') \overset{1}{\Longrightarrow} (a', b')$.

Now, we note that $\left(\frac{\alpha', \beta'}{\mathfrak{m}_v}\right)_\mu = 1$ if either $v \in V_\infty^k$ (since $\beta' > 0$ in all real completions of k) or $v \in V_f^k \setminus (V(\alpha') \cup V(\beta') \cup V(\mu))$. Since the ideals $\mathfrak{a} = \alpha'\mathcal{O}_k$ and $\mathfrak{b} = \beta'\mathcal{O}_k$ are prime by construction, we have $V(\alpha') = \{v_{\mathfrak{a}}\}$ and $V(\beta') = \{v_{\mathfrak{b}}\}$. Besides, it follows from (18) and $(4)_1$ that for $v \in V(p_i) \setminus \{v_i\}$ we have $\beta' \in k_v^{\times \mu}$, and therefore again $\left(\frac{\alpha', \beta'}{\mathfrak{m}_v}\right)_\mu = 1$. Thus, the reciprocity law (16) for α', β' reduces to the relation

$$\left(\frac{\alpha', \beta'}{\mathfrak{a}}\right)_\mu \cdot \left(\frac{\alpha', \beta'}{\mathfrak{b}}\right)_\mu \cdot \prod_{i=1}^{n} \left(\frac{\alpha', \beta'}{\mathfrak{m}_{v_i}}\right)_\mu = 1. \tag{20}$$

It follows from $(2)_2$ and $(3)_2$ that

$$\left(\frac{\alpha', \beta'}{\mathfrak{b}}\right)_\mu = \left(\frac{\alpha, \beta'}{\mathfrak{b}}\right)_\mu \quad \text{and} \quad \left(\frac{\alpha', \beta'}{\mathfrak{m}_{v_i}}\right)_\mu = \left(\frac{u_i^d, \beta'}{\mathfrak{m}_{v_i}}\right)_\mu \quad \text{for all } i = 1, \dots, n.$$

Comparing now (19) with (20), we find that

$$\left(\frac{\beta', \alpha'}{\mathfrak{a}}\right)_\mu = \left(\frac{\alpha', \beta'}{\mathfrak{a}}\right)_\mu^{-1} = 1.$$

This implies [Bass et al. 1967, A.16] that β' is a μ-th power modulo $\mathfrak{a}$, i.e., $\beta' \equiv \gamma^\mu \pmod{\mathfrak{a}}$ for some $\gamma \in \mathcal{O}_k$. Clearly, the elements $a' = \alpha' r^\mu$ and γt are relatively prime in $\mathcal{O}$, so applying Theorem 3.3 to this ring, we find infinitely many $\mathbb{Q}$-split principal prime ideals $\mathfrak{q}$ of $\mathcal{O}$ having a generator $q \equiv \gamma t \pmod{a'\mathcal{O}}$. Then for any $m \equiv 1 \pmod{\phi(a'\mathcal{O})}$ we have

$$q^{\mu m} \equiv q^\mu \equiv \beta' t^\mu \equiv b' \pmod{a'\mathcal{O}},$$

so $(a', b') \overset{1}{\Longrightarrow} (a', q^{\mu m})$. Then by Lemma 4.1(1b), we have $(a, b) \overset{3}{\Longrightarrow} (a', q^{\mu m})$, as required. □

The final ingredient that we need for the proof of Theorem 1.1 is the following lemma which uses the notion of the *level* $\ell_{\mathfrak{p}}(u)$ of a unit u of infinite order with respect to a $\mathbb{Q}$-split ideal $\mathfrak{p}$ introduced in Section 3A.

Lemma 4.5. *Let $\mathfrak{p}$ be a principal $\mathbb{Q}$-split ideal of $\mathcal{O}$ with a generator π, and let $u \in \mathcal{O}^\times$ be a unit of infinite order. Set $s = \ell_{\mathfrak{p}}(u)$, and let λ and m be integers satisfying $\lambda | \phi(\mathfrak{p})$ and $m \equiv 0 \pmod{\phi(\mathfrak{p}^s)/\lambda}$.*

Given an integer $\delta > 0$ dividing λ and $b \in \mathcal{O}$ prime to π such that b is a δ-th power $\bmod \mathfrak{p}$ *while $\nu := \lambda/\delta$ divides the order of u* $(\bmod \mathfrak{p})$, *for any integer $t \geq s$ there exists an integer n_t for which*

$$(\pi^t, b^m) \xRightarrow{1} (\pi^t, u^{n_t}).$$

Proof. Let p be the rational prime corresponding to $\mathfrak{p}$. Being a divisor of λ, the integer δ is relatively prime to p. So, the fact that b is a δ-th power $\bmod \mathfrak{p}$ implies that it is also a δ-th power $\bmod \mathfrak{p}^s$. On the other hand, it follows from our assumptions that $\lambda m = \delta \nu m$ is divisible by $\phi(\mathfrak{p}^s)$, and therefore $(b^m)^\nu \equiv 1 \pmod{\mathfrak{p}^s}$. But since ν is prime to p, the subgroup of elements in $(\mathcal{O}/\mathfrak{p}^s)^\times$ of order dividing ν is isomorphic to a subgroup of $(\mathcal{O}/\mathfrak{p})^\times$, hence cyclic. So, the fact that the order of $u \pmod{\mathfrak{p}}$, and consequently the order $u \pmod{\mathfrak{p}^s}$, is divisible by ν implies that every element in $(\mathcal{O}/\mathfrak{p}^s)^\times$ whose order divides ν lies in the subgroup generated by $u \pmod{\mathfrak{p}^s}$. Thus, $b^m \equiv u^{n_s} \pmod{\mathfrak{p}^s}$ for some integer n_s. Since $\mathfrak{p}$ is $\mathbb{Q}$-split, we can apply Lemma 3.2 to conclude that for any $t \geq s$ there exists an integer n_t such that $b^m \equiv u^{n_t} \pmod{\mathfrak{p}^t}$. Then $(\pi^t, b^m) \xRightarrow{1} (\pi^t, u^{n_t})$ by Lemma 4.1(2). □

We will call a unit $u \in \mathcal{O}^\times$ *fundamental* if it has infinite order and the cyclic group $\langle u \rangle$ is a direct factor of $\mathcal{O}^\times$. Since the group $\mathcal{O}^\times$ is finitely generated (Dirichlet's unit theorem, cf. [Cassels and Fröhlich 1967, §2.18]) it always contains a fundamental unit once it is infinite. We note that any fundamental unit has the following property:

$$u \notin \mu(k)_p (k^\times)^p \quad \text{for any prime } p.$$

We are now in a position to give

Proof of Theorem 1.1. We return to the notations of Section 3C: we let K denote the Hilbert S-class field of k, let $\tilde{K}$ be its normal closure over $\mathbb{Q}$, and pick an integer $\lambda \geq 1$ which is divisible by μ and for which $\tilde{K} \cap \mathbb{Q}^{\mathrm{ab}} \subset \mathbb{Q}(\zeta_\lambda)$. Furthermore, since $\mathcal{O}^\times$ is infinite by assumption, we can find a fundamental unit $u \in \mathcal{O}^\times$. By Lemma 4.1(3), it suffices to show that for any $(a, b) \in \mathcal{R}(\mathcal{O})$, we have

$$(a, b) \xRightarrow{8} (1, 0). \tag{21}$$

First, applying Lemma 4.3 with $T = (S \setminus V_\infty^k) \cup V(\mu)$ and $n = \mu$, we see that there exist $\alpha \in \mathcal{O}_k$ and $r \in \mathcal{O}^\times$ such that

$$V(\alpha) \cap (S \cup V(\mu)) = \varnothing \quad \text{and} \quad (a, b) \xRightarrow{1} (\alpha r^\mu, b).$$

Next, applying Lemma 4.4 to the last pair, we find $a' \in \mathcal{O}$ and a $\mathbb{Q}$-split principal prime ideal $\mathfrak{q}$ such that $v_\mathfrak{q} \notin S \cup V(\lambda) \cup V(\phi(a'\mathcal{O}))$ and $(\alpha r^\mu, b) \xRightarrow{3} (a', q^{\mu m})$ for any $m \equiv 1 \pmod{\phi(a'\mathcal{O})}$. Then

$$(a, b) \xRightarrow{4} (a', q^{\mu m}) \quad \text{for any } m \equiv 1 \pmod{\phi(a'\mathcal{O})}. \tag{22}$$

To proceed with the argument we will now specify m. We let P and Q denote the sets of prime divisors of λ/μ and $\phi(a'\mathcal{O})$, respectively, and define λ' and μ' as in Section 3C; we note that by construction $\mathfrak{q}$ is relatively prime to λ'. So, we can apply Theorem 3.7 which yields a $\mathbb{Q}$-split principal prime ideal $\mathfrak{p} = \pi\mathcal{O}$

so that $v_{\mathfrak{p}} \notin V(\phi(a'\mathcal{O}))$ and conditions (1) - (3) are satisfied. Let $s = \ell_{\mathfrak{p}}(u)$ be the $\mathfrak{p}$-level of u. Condition (3) implies that

$$\gcd(\phi(\mathfrak{p})/\lambda, \lambda'/\lambda) = 1 = \gcd(\phi(\mathfrak{p})/\lambda, \phi(a'\mathcal{O}))$$

since λ'/λ is the product of all prime divisors of $\phi(a'\mathcal{O})$. It follows that the numbers $\phi(\mathfrak{p}^s)/\lambda$ and $\phi(a'\mathcal{O})$ are relatively prime, and therefore one can pick a positive integer m so that

$$m \equiv 0 \pmod{\phi(\mathfrak{p}^s)/\lambda} \quad \text{and} \quad m \equiv 1 \pmod{\phi(a'\mathcal{O})}.$$

Fix this m for the rest of the proof.

Condition (2) of Theorem 3.7 enables us to apply Corollary 4.2 with $z = \pi$ and $t_0 = s$ to find $t \geq s$ so that $(a', q^{\mu m}) \overset{1}{\Longrightarrow} (\pi^t, q^{\mu m})$. Since P consists of all prime divisors of λ/μ, condition (1) of Theorem 3.7 implies that λ/μ divides the order of $u \pmod{\mathfrak{p}}$. Now, applying Lemma 4.5 with $\delta = \mu$ and $b = q^{\mu}$, we see that $(\pi^t, q^{\mu m}) \overset{1}{\Longrightarrow} (\pi^t, u^{n_t})$ for some integer n_t. Finally, since u is a unit, we have $(\pi^t, u^{n_t}) \overset{2}{\Longrightarrow} (1, 0)$. Combining these computations with (22), we obtain (21), completing the proof. □

Corollary 4.6. *Assume that the group $\mathcal{O}^\times$ is infinite. Then for $n \geq 2$, any matrix $A \in \mathrm{SL}_n(\mathcal{O})$ is a product of $\leq \frac{1}{2}(3n^2 - n) + 4$ elementary matrices.*

Proof. For $n = 2$, this is equivalent to Theorem 1.1. Now, let $n \geq 3$. Since the ring $\mathcal{O}$ is Dedekind, it is well-known and easy to show that any $A \in \mathrm{SL}_n(\mathcal{O})$ can be reduced to a matrix in $\mathrm{SL}_2(\mathcal{O})$ by at most $\frac{1}{2}(3n^2 - n) - 5$ elementary operations [Carter and Keller 1983, p. 683]. Now, our result immediately follows from Theorem 1.1. □

Proof of Corollary 1.2. Let

$$e_+ : \alpha \mapsto \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad e_- : \alpha \mapsto \begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix}$$

be the standard 1-parameter subgroups. Set $U^{\pm} = e_{\pm}(\mathcal{O})$. In view of Theorem 1.1, it is enough to show that each of the subgroups U^+ and U^- is contained in a product of finitely many cyclic subgroups of $\mathrm{SL}_2(\mathcal{O})$. Let h_k be the class number of k. Then there exists $t \in \mathcal{O}^\times$ such that $v(t) = h_k$ for all $v \in S \setminus V_\infty^k$ and $v(t) = 0$ for all $v \notin S$. Then $\mathcal{O} = \mathcal{O}_k[1/t]$. So, letting $U_0^{\pm} = e_{\pm}(\mathcal{O}_k)$ and $h = \left(\begin{smallmatrix} t & 0 \\ 0 & t^{-1} \end{smallmatrix}\right)$, we will have the inclusion

$$U^{\pm} \subset \langle h \rangle U_0^{\pm} \langle h \rangle.$$

On the other hand, if $w_1, \ldots, w_n$ (where $n = [k : \mathbb{Q}]$) is a $\mathbb{Z}$-basis of $\mathcal{O}_k$ then $U_0^{\pm} = \langle e_{\pm}(w_1) \rangle \cdots \langle e_{\pm}(w_n) \rangle$, hence

$$U^{\pm} \subset \langle h \rangle \langle e_{\pm}(w_1) \rangle \cdots \langle e_{\pm}(w_n) \rangle \langle h \rangle, \tag{23}$$

as required. □

Remarks. (1) Quantitatively, it follows from the proof of Theorem 1.1 that $\mathrm{SL}_2(\mathcal{O}) = U^- U^+ \cdots U^-$ (nine factors), so since the right-hand side of (23) involves $n + 2$ cyclic subgroups, with $\langle h \rangle$ at both ends, we obtain that $\mathrm{SL}_2(\mathcal{O})$ is a product of $9[k:\mathbb{Q}]+10$ cyclic subgroups. Also, it follows from [Vsemirnov 2014] that $\mathrm{SL}_2(\mathbb{Z}[1/p])$ is a product of 11 cyclic subgroups.

(2) If $S = V_\infty^k$, then the proof of Corollary 1.2 yields a factorization of $\mathrm{SL}_2(\mathcal{O})$ as a finite product $\langle\gamma_1\rangle\cdots\langle\gamma_d\rangle$ of cyclic subgroups where all generators γ_i are elementary matrices, hence *unipotent*. On the contrary, when $S \neq V_\infty^k$, the factorization we produce involves some diagonal (*semisimple*) matrices. So, it is worth pointing out in the latter case there is no factorization with all γ_i unipotent. Indeed, let $v \in S \setminus V_\infty^k$ and let $\gamma \in \mathrm{SL}_2(\mathcal{O})$ be unipotent. Then there exists $N = N(\gamma)$ such that for any $a = (a_{ij}) \in \langle\gamma\rangle$ we have $v(a_{ij}) \leq N(\gamma)$ for all $i, j \in \{1, 2\}$. It follows that if $\mathrm{SL}_2(\mathcal{O}) = \langle\gamma_1\rangle\cdots\langle\gamma_d\rangle$ where all γ_i are unipotent, then there exists N_0 such that for any $a = (a_{ij}) \in \mathrm{SL}_2(\mathcal{O})$ we have $v(a_{ij}) \leq N_0$ for $i, j \in \{1, 2\}$, which is absurd.

5. Example

For a ring of S-integers $\mathcal{O}$ in a number field k such that the group of units $\mathcal{O}^\times$ is infinite, we let $\nu(\mathcal{O})$ denote the smallest positive integer with the property that every matrix in $\mathrm{SL}_2(\mathcal{O})$ is a product of $\leq \nu(\mathcal{O})$ elementary matrices. So, the result of [Vsemirnov 2014] implies that $\nu(\mathbb{Z}[1/p]) \leq 5$ for any prime p, and our Theorem 1.1 yields that $\nu(\mathcal{O}) \leq 9$ for any $\mathcal{O}$ as above. It may be of some interest to determine the exact value of $\nu(\mathcal{O})$ in some situations. In Example 2.1 on p.289, Vsemirnov [2014] claims that the matrix

$$M = \begin{pmatrix} 5 & 12 \\ 12 & 29 \end{pmatrix}$$

is not a product of four elementary matrices in $\mathrm{SL}_2(\mathbb{Z}[1/p])$ for any $p \equiv 1 \pmod{29}$, and therefore $\nu(\mathbb{Z}[1/p]) = 5$ in this case. However this example is faulty because for any prime p, in $\mathrm{SL}_2(\mathbb{Z}[1/p])$ we have

$$M = \begin{pmatrix} 5 & 12 \\ 12 & 29 \end{pmatrix} = \left(\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}\right)^2$$

However, it turns out that the assertion that $\nu(\mathbb{Z}[1/p]) = 5$ is valid not only for $p \equiv 1 \pmod{29}$ but in fact for all $p > 7$. More precisely, we have the following.

Proposition 5.1. *Let $\mathcal{O} = \mathbb{Z}[1/p]$, where p is prime > 7. Then not every matrix in $\mathrm{SL}_2(\mathcal{O})$ is a product of four elementary matrices.*

In the remainder of this section, unless stated otherwise, we will work with congruences over the ring $\mathcal{O}$ rather than $\mathbb{Z}$, so the notation $a \equiv b \pmod{n}$ means that elements $a, b \in \mathcal{O}$ are congruent modulo the ideal $n\mathcal{O}$. We begin the proof of the proposition with the following lemma.

Lemma 5.2. *Let $\mathcal{O} = \mathbb{Z}[1/p]$, where p is any prime, and let r be a positive integer satisfying $p \equiv 1 \pmod{r}$. Then any matrix $A \in \mathrm{SL}_2(\mathcal{O})$ of the form*

$$A = \begin{pmatrix} 1-p^\alpha & * \\ * & 1-p^\beta \end{pmatrix}, \quad \alpha, \beta \in \mathbb{Z} \tag{24}$$

which is a product of four elementary matrices, satisfies the congruence

$$A \equiv \pm \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \pmod{r}.$$

Proof. The required congruence is obvious for the diagonal entries, so we only need to establish it for the off-diagonal ones. Since A is a product of four elementary matrices, it admits one of the following presentations:

$$A = E_{12}(a)E_{21}(b)E_{12}(c)E_{21}(d), \tag{25}$$

or

$$A = E_{21}(a)E_{12}(b)E_{21}(c)E_{12}(d), \tag{26}$$

with $a, b, c, d \in \mathcal{O}$.

First, suppose we have (25). Then

$$A = \begin{pmatrix} * & * \\ * & 1+bc \end{pmatrix}.$$

Comparing with (24), we get $bc = -p^{\beta}$, so b and c are powers of p with opposite signs. Thus, A looks as follows:

$$A = E_{12}(a)E_{21}(\pm p^{\gamma})E_{12}(\mp p^{\delta})E_{21}(d) = \begin{pmatrix} * & a(1-p^{\gamma+\delta})\mp p^{\delta} \\ d(1-p^{\gamma+\delta})\pm p^{\gamma} & * \end{pmatrix}.$$

Consequently, the required congruences for the off-diagonal entries immediately follow from the fact that $p \equiv 1 \pmod{r}$, proving the lemma in this case.

Now, suppose we have (26). Then

$$A^{-1} = E_{12}(-d)E_{21}(-c)E_{12}(-b)E_{21}(-a),$$

which means that A^{-1} has a presentation of the form (25). Since the required congruence in this case has already been established, we conclude that

$$A^{-1} \equiv \pm \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \pmod{r}.$$

But then we have

$$A \equiv \pm \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \pmod{r},$$

as required. □

To prove the proposition, we will consider two cases:

CASE 1: *$p-2$ is composite.* Write $p-2 = r_1 \cdot r_2$, where r_1 and r_2 are positive integers > 1, and set $r = p-1$. Then

$$r_i \not\equiv \pm 1 \pmod{r} \quad \text{for } i = 1, 2. \tag{27}$$

Indeed, we can assume that $r_2 \leq \sqrt{p-2}$. If $r_2 \equiv \pm 1 \pmod{r}$ then because r is prime to p, the number $r_2 \mp 1$ would be a nonzero integral multiple of r. Then $r \leq r_2 + 1$, hence

$$p-2 \leq \sqrt{p-2}+1.$$

But this is impossible since $p > 3$. Thus, $r_2 \not\equiv \pm 1 \pmod{r}$. Since $r_1 \cdot r_2 \equiv -1 \pmod{r}$, condition (27) follows.

Now, consider the matrix

$$A = \begin{pmatrix} 1-p & r_1 \cdot p \\ r_2 & 1-p \end{pmatrix}$$

One immediately checks that $A \in \mathrm{SL}_2(\mathcal{O})$. At the same time, A is of the form (24). Then Lemma 5.2 in conjunction with (27) implies that A is not a product of four elementary matrices.

CASE 2. *p and $p-2$ are both primes.* In the beginning of this paragraph we will use congruences in $\mathbb{Z}$. Clearly, a prime > 3 can only be congruent to $\pm 1 \pmod{6\mathbb{Z}}$. Since $p > 5$ and $p-2$ is also prime, in our situation we must have $p \equiv 1 \pmod{6\mathbb{Z}}$. Furthermore, since $p > 7$, the congruence $p \equiv 0$ or $2 \pmod{5\mathbb{Z}}$ is impossible. Thus, in the case at hand we have

$$p \equiv 1, 13, \text{ or } 19 \pmod{30\mathbb{Z}}.$$

If $p \equiv 13 \pmod{30\mathbb{Z}}$, then $p^3 \equiv 7 \pmod{30\mathbb{Z}}$, and therefore $p^3 - 2$ is an integral multiple of 5. Set $r = p-1$ and $s = (p^3-2)/5$, and consider the matrix

$$A = \begin{pmatrix} 1-p^3 & 5p^3 \\ s & 1-p^3 \end{pmatrix}$$

Then A is a matrix in $\mathrm{SL}_2(\mathcal{O})$ having form (24). Note that $5p^3 \equiv 5 \pmod{r}$, which is different from $\pm 1 \pmod{r}$ since $r > 6$. Now, it follows from Lemma 5.2 that A is not a product of four elementary matrices.

It remains to treat the case where $p \equiv 1$ or $19 \pmod{30\mathbb{Z}}$. Consider the following matrix:

$$A = \begin{pmatrix} 900 & 53 \cdot 899 \\ 17 & 900 \end{pmatrix},$$

and note that $A \in \mathrm{SL}_2(\mathbb{Z})$ and

$$A^{-1} = \begin{pmatrix} 900 & -53 \cdot 899 \\ -17 & 900 \end{pmatrix}.$$

It suffices to show that neither A nor A^{-1} can be written in the form

$$E_{12}(a)E_{21}(b)E_{12}(c)E_{21}(d) = \begin{pmatrix} * & c+a(1+bc) \\ b+d(1+bc) & (1+bc) \end{pmatrix}, \quad \text{with } a, b, c, d \in \mathcal{O}. \tag{28}$$

Assume that either A or A^{-1} is written in the form (28). Then $1 + bc = 900$, so

$$b, c \in \{\pm p^n, \pm 29p^n, \pm 31p^n, \pm 899p^n \mid n \in \mathbb{Z}\}.$$

Set

$$t = b + d(1+bc) \quad \text{and} \quad u = c + a(1+bc).$$

We have the following congruences in $\mathcal{O} = \mathbb{Z}[1/p]$:

$$t \equiv b \pmod{30} \quad \text{and} \quad u \equiv c \pmod{30}.$$

Analyzing the above list of possibilities for b and c, we conclude that each of t and u is $\equiv \pm p^n \pmod{30}$ for some integer n. Thus, if $p \equiv 1 \pmod{30}$ then $t, u \equiv \pm 1 \pmod{30}$, and if $p \equiv 19 \pmod{30}$ then $t, u \equiv \pm 1, \pm 19 \pmod{30}$. Since $17 \not\equiv \pm 1, \pm 19 \pmod{30}$, we obtain a contradiction in either case. (We observe that the argument in this last case is inspired by Vsemirnov's argument in his Example 2.1.)

Acknowledgements

During the preparation of the final version of the paper the second author visited Princeton University and the Institute for Advanced Study on a Simons Fellowship. The hospitality of both institutions and the generous support of the Simons Foundation are thankfully acknowledged. The third author gratefully acknowledges fruitful conversations with Maxim Vsemirnov. We thank Dong Quan Ngoc Nguyen for his comments regarding Theorem 3.3. Our most sincere thanks are due to the anonymous referee who has read the paper extremely carefully and offered a number of corrections and valuable suggestions.

References

[Bass et al. 1967] H. Bass, J. Milnor, and J.-P. Serre, "Solution of the congruence subgroup problem for SL_n $(n \geq 3)$ and Sp_{2n} $(n \geq 2)$", *Inst. Hautes Études Sci. Publ. Math.* 33 (1967), 59–137. MR Zbl

[Carter and Keller 1983] D. Carter and G. Keller, "Bounded elementary generation of $\mathrm{SL}_n(\mathcal{O})$", *Amer. J. Math.* **105**:3 (1983), 673–687. MR Zbl

[Carter and Keller 1984] D. Carter and G. Keller, "Elementary expressions for unimodular matrices", *Comm. Algebra* **12**:3-4 (1984), 379–389. MR Zbl

[Cassels and Fröhlich 1967] J. W. S. Cassels and A. Fröhlich (editors), *Algebraic number theory*, Academic Press, Washington D.C., 1967. MR Zbl

[Cohn 1966] P. M. Cohn, "On the structure of the GL_2 of a ring", *Inst. Hautes Études Sci. Publ. Math.* 30 (1966), 5–53. MR Zbl

[Cooke and Weinberger 1975] G. Cooke and P. J. Weinberger, "On the construction of division chains in algebraic number rings, with applications to SL_2", *Comm. Algebra* **3** (1975), 481–524. MR Zbl

[Erovenko and Rapinchuk 2006] I. V. Erovenko and A. S. Rapinchuk, "Bounded generation of S-arithmetic subgroups of isotropic orthogonal groups over number fields", *J. Number Theory* **119**:1 (2006), 28–48. MR Zbl

[Grunewald and Schwermer 1981] F. J. Grunewald and J. Schwermer, "Free nonabelian quotients of SL_2 over orders of imaginary quadratic numberfields", *J. Algebra* **69**:2 (1981), 298–304. MR Zbl

[Hasse 1926] H. Hasse, "Bericht über neuere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper, I: Klassenkörpertheorie", *Jahresber. Dtsch. Math.-Ver.* **35** (1926), 1–55. Zbl

[Heath-Brown 1986] D. R. Heath-Brown, "Artin's conjecture for primitive roots", *Quart. J. Math. Oxford Ser.* (2) **37**:145 (1986), 27–38. MR Zbl

[Lang 2002] S. Lang, *Algebra*, 3rd ed., Graduate Texts in Mathematics **211**, Springer, New York, 2002. MR Zbl

[Liehl 1981] B. Liehl, "On the group SL_2 over orders of arithmetic type", *J. Reine Angew. Math.* **323** (1981), 153–171. MR Zbl

[Liehl 1984] B. Liehl, "Beschränkte Wortlänge in SL_2", *Math. Z.* **186**:4 (1984), 509–524. MR Zbl

[Loukanidis and Murty 1994] D. Loukanidis and V. Murty, “Bounded generation for SL_n ($n \geq 2$) and Sp_{2n} ($n \geq 1$)”, preprint, 1994.

[Lubotzky 1995] A. Lubotzky, “Subgroup growth and congruence subgroups”, *Invent. Math.* **119**:2 (1995), 267–295. MR Zbl

[Morris 2007] D. W. Morris, “Bounded generation of $\mathrm{SL}(n, A)$ (after D. Carter, G. Keller, and E. Paige)”, *New York J. Math.* **13** (2007), 383–421. MR Zbl

[Murty 1995] V. K. Murty, “Bounded and finite generation of arithmetic groups”, pp. 249–261 in *Number theory* (Halifax, NS, 1994), edited by K. Dilcher, CMS Conf. Proc. **15**, Amer. Math. Soc., Providence, RI, 1995. MR Zbl

[Platonov and Rapinchuk 1992] V. P. Platonov and A. S. Rapinchuk, “Abstract properties of S-arithmetic groups and the congruence problem”, *Izv. Ross. Akad. Nauk Ser. Mat.* **56**:3 (1992), 483–508. In Russian; translated in *Russian Acad. Sci. Izv. Math.* **40**:3 (1993), 455–476. MR Zbl

[Platonov and Rapinchuk 1994] V. Platonov and A. Rapinchuk, *Algebraic groups and number theory*, Pure and Applied Mathematics **139**, Academic Press, Boston, 1994. MR Zbl

[Rapinchuk 1990] A. S. Rapinchuk, “Representations of groups of finite width”, *Dokl. Akad. Nauk SSSR* **315**:3 (1990), 536–540. In Russian; translated in *Soviet Math. Dokl.* **42**:3 (1991), 816–820. MR Zbl

[Serre 1970] J.-P. Serre, “Le problème des groupes de congruence pour SL2”, *Ann. of Math.* (2) **92** (1970), 489–527. MR Zbl

[Shalom and Willis 2013] Y. Shalom and G. A. Willis, “Commensurated subgroups of arithmetic groups, totally disconnected groups and adelic rigidity”, *Geom. Funct. Anal.* **23**:5 (2013), 1631–1683. MR Zbl

[Tavgen 1990] O. I. Tavgen, “Bounded generability of Chevalley groups over rings of S-integer algebraic numbers”, *Izv. Akad. Nauk SSSR Ser. Mat.* **54**:1 (1990), 97–122. In Russian; translated in *Math. USSR-Izv.* **36**:1 (1991), 101–128. MR Zbl

[Vasershtein 1972] L. N. Vasershtein, “The group SL_2 over Dedekind rings of arithmetic type”, *Mat. Sb.* (*N.S.*) **89(131)** (1972), 313–322, 351. In Russian; translated in *Mathematics of the USSR-Sbornik* **18**:2 (1972), 321–332. MR

[Vsemirnov 2014] M. Vsemirnov, “Short unitriangular factorizations of $\mathrm{SL}_2(\mathbb{Z}[1/p])$”, *Q. J. Math.* **65**:1 (2014), 279–290. MR Zbl

Communicated by Raman Parimala
Received 2017-09-03 Revised 2018-05-10 Accepted 2018-06-18

avo2t@virginia.edu *Department of Mathematics, University of Virginia, Charlottesville, VA, United States*

asr3x@virginia.edu *Department of Mathematics, University of Virginia, Charlottesville, VA, United States*

surybang@gmail.com *Stat-Math Unit, Indian Statistical Institute, Bangalore, India*

ALGEBRA AND NUMBER THEORY 12:8 (2018)
dx.doi.org/10.2140/ant.2018.12.1975

Tensor triangular geometry of filtered modules

Martin Gallauer

We compute the tensor triangular spectrum of perfect complexes of filtered modules over a commutative ring and deduce a classification of the thick tensor ideals. We give two proofs: one by reducing to perfect complexes of graded modules which have already been studied in the literature by Dell'Ambrogio and Stevenson (2013, 2014) and one more direct for which we develop some useful tools.

Introduction

One of the age-old problems mathematicians engage in is to classify their objects of study, up to an appropriate equivalence relation. In contexts in which the domain is organized in a category with compatible tensor and triangulated structure (we call this a *tt-category*) it is natural to view objects as equivalent when they can be constructed from each other using sums, extensions, translations, tensor product etc., in other words, using the tensor and triangulated structure alone. This can be made precise by saying that the objects generate the same thick tensor ideal (or, *tt-ideal*) in the tt-category. This sort of classification is precisely what tt-geometry, as developed by Balmer, achieves. To a (small) tt-category $\mathcal{T}$ it associates a topological space $\operatorname{Spc}(\mathcal{T})$ called the *tt-spectrum* of $\mathcal{T}$ which, via its Thomason subsets, classifies the (radical) tt-ideals of $\mathcal{T}$. A number of classical mathematical domains have in the meantime been studied through the lens of tt-geometry; we refer to [Balmer 2010b] for an overview of the basic theory, its early successes and applications.

MSC2010: primary 18E30; secondary 18D10, 55U35.
Keywords: tensor triangular geometry, filtered modules, classification.

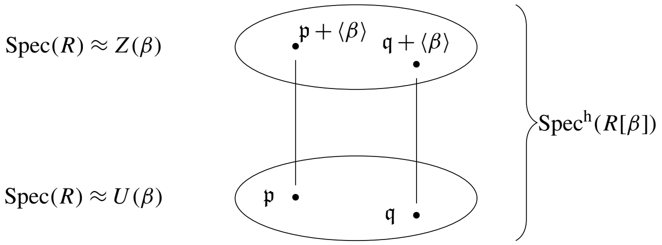
One type of context which does not seem to have received any attention so far arises from filtered objects. Examples pertinent to tt-geometry abound: filtrations by the weight in algebraic geometry induce filtrations on cohomology theories, giving rise to filtered vector spaces, representations or motives; (mixed) Hodge theory involves bifiltered vector spaces; filtrations by the order of a differential operator play an important role in the theory of $\mathcal{D}$-modules.

In this note, we take the first steps in the study of filtered objects through the lens of tt-geometry by focusing on a particularly interesting case whose unfiltered analogue is well-understood. Namely, we give a complete account of the tt-geometry of filtered modules. This is already enough to say something interesting about certain motives, as we explain at the end of this introduction. To describe our results in more detail, let us recall the analogous situation for modules.

Let R be a ring, assumed commutative and with unit. Its derived category $\mathcal{D}(R)$ is a tt-category which moreover is compactly generated, and the compact objects coincide with the rigid (or, strongly dualizable) objects, which are also called perfect complexes. These are (up to isomorphism in the derived category) the bounded complexes of finitely generated projective R-modules. The full subcategory $\mathcal{D}^{\mathrm{perf}}(R)$ of perfect complexes inherits the structure of a (small) tt-category, and the Hopkins–Neeman–Thomason classification of its thick subcategories can be interpreted as the statement that the tt-spectrum $\mathrm{Spc}(\mathcal{D}^{\mathrm{perf}}(R))$ is precisely the Zariski spectrum $\mathrm{Spec}(R)$. In this particular case, thick subcategories are the same as tt-ideals so that this result indeed classifies perfect complexes up to the triangulated and tensor structure available.

In this note we will replicate these results for filtered R-modules. Its derived category $\mathcal{D}(\mathrm{Mod}_{\mathrm{fil}}(R))$ is a tt-category which moreover is compactly generated, and the compact objects coincide with the rigid objects. We characterize these "perfect complexes" as bounded complexes of "finitely generated projective" objects in the category $\mathrm{Mod}_{\mathrm{fil}}(R)$ of filtered R-modules.[1] The full subcategory $\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R)$ of perfect complexes inherits the structure of a (small) tt-category. For a regular ring R this is precisely the filtered derived category of R in the sense first studied by Illusie [1971], and for general rings it is a full subcategory. Our main theorem computes the tt-spectrum of this tt-category.

Theorem 4.1. *The tt-spectrum of* $\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R)$ *is canonically isomorphic to the homogeneous Zariski spectrum* $\mathrm{Spec}^{\mathrm{h}}(R[\beta])$ *of the polynomial ring in one variable. In particular, the underlying topological space contains two copies of* $\mathrm{Spec}(R)$, *connected by specialization. Schematically*:



[1] In the body of the text these are rather called *split finite projective* for reasons which will become apparent when they are introduced.

As a consequence we are able to classify the tt-ideals in $\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R)$. To state it precisely notice that we may associate to any filtered R-module M its underlying R-module $\pi(M)$ as well as the R-module of its graded pieces $\mathrm{gr}(M)$. These induce two tt-functors $\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R) \to \mathcal{D}^{\mathrm{perf}}(R)$. Also, recall that the *support* of an object $M \in \mathcal{D}^{\mathrm{perf}}(R)$, denoted by $\mathrm{supp}(M)$, is the set of primes in $\mathrm{Spc}(\mathcal{D}^{\mathrm{perf}}(R)) = \mathrm{Spec}(R)$ which do not contain M. This is extended to a set $\mathcal{E}$ of objects by taking the union of the supports of its elements: $\mathrm{supp}(\mathcal{E}) := \bigcup_{M \in \mathcal{E}} \mathrm{supp}(M)$. Conversely, starting with a set of primes $Y \subset \mathrm{Spec}(R)$, we define $\mathcal{K}_Y := \{M \in \mathcal{D}^{\mathrm{perf}}(R) \mid \mathrm{supp}(M) \subset Y\}$.

Corollary 4.9. *There is an inclusion preserving bijection*:

$$\begin{aligned} \{\Pi \subset \Gamma \mid \Pi, \Gamma \subset \mathrm{Spec}(R) \textit{ Thomason subsets}\} &\leftrightarrow \{\textit{tt-ideals in } \mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R)\} \\ (\Pi \subset \Gamma) &\mapsto \pi^{-1}(\mathcal{K}_\Pi) \cap \mathrm{gr}^{-1}(\mathcal{K}_\Gamma) \\ (\,\mathrm{supp}(\pi\mathcal{J}) \subset \mathrm{supp}(\mathrm{gr}\,\mathcal{J})) &\leftarrow\!\shortmid \mathcal{J} \end{aligned}$$

Clearly, an important role is played by the element β appearing in the theorem. It can be interpreted as the following morphism of filtered R-modules. Let $R(0)$ be the module R placed in filtration degree 0, while $R(1)$ is R placed in degree 1 (our filtrations are by convention decreasing), and $\beta : R(0) \to R(1)$ is the identity on the underlying modules:[2]

$$\begin{tikzcd} R(1): & \cdots & = & 0 & \subset & R & = & R & = & \cdots \\ R(0): \arrow[u, "\beta"'] & \cdots & = & 0 \arrow[u] & = & 0 \arrow[u] & \subset & R \arrow[u, "\mathrm{id}"'] & = & \cdots \end{tikzcd}$$

Note that β has trivial kernel and cokernel but is not an isomorphism, witnessing the fact that the category of filtered modules is *not* abelian. We will give two proofs of Theorem 4.1, the first of which relies on "abelianizing" the category. It is observed in [Schneiders 1999] that the derived category of filtered modules is canonically identified with the derived category of graded $R[\beta]$-modules. And the tt-geometry of graded modules has been studied in [Dell'Ambrogio and Stevenson 2013; 2014]. Together these two results provide a short proof of Theorem 4.1, but in view of future studies of filtered objects in more general abelian tensor categories we thought it might be worthwhile to study filtered modules in more detail and in their own right. For the second proof we will use the abelianization only minimally to construct the category of perfect complexes of filtered modules (Section 3). The computation of the tt-geometry stays within the realm of filtered modules, as we now proceed to explain.

As mentioned above, forgetting the filtration and taking the associated graded of a filtered R-module gives rise to two tt-functors. It is not difficult to show that $\mathrm{Spc}(\pi)$ and $\mathrm{Spc}(\mathrm{gr})$ are injective with disjoint images (Section 4). The challenge is in proving that they are jointly surjective — more precisely, proving that the images of $\mathrm{Spc}(\pi)$ and $\mathrm{Spc}(\mathrm{gr})$ are exactly the two copies of $\mathrm{Spec}(R)$ in the picture above. As

[2] We call this element β in view of the intended application described at the end of this introduction. In the context of motives considered there, β is the "Bott element" of [Hasemeyer and Hornbostel 2005].

suggested by this then, and as we will prove, inverting β (in a categorical sense) amounts to passing from filtered to unfiltered R-modules, while killing β amounts to passing to the associated graded.

We prove surjectivity first for R a noetherian ring, by reducing to the local case, using some general results we establish on central localization (Section 5), extending the discussion in [Balmer 2010a]. In the local noetherian case, the maximal ideal is "generated by nonzerodivisors and nilpotent elements" (more precisely, it admits a system of parameters); we will study how killing such elements affects the tt-spectrum (Section 6) which allows us to decrease the Krull dimension of R one by one until we reach the case of R a field.

Although the category of filtered modules is not abelian, it has the structure of a *quasi*abelian category, and we will use the results of Schneiders [1999] on the derived category of a quasiabelian category, in particular the existence of two t-structures, to deal with the case of a field (Section 7). In fact, the category of filtered vector spaces can reasonably be called a *semisimple* quasiabelian category, and we will prove in general that the t-structures in that case are hereditary. With this fact it is then possible to deduce the theorem in the case of a field.

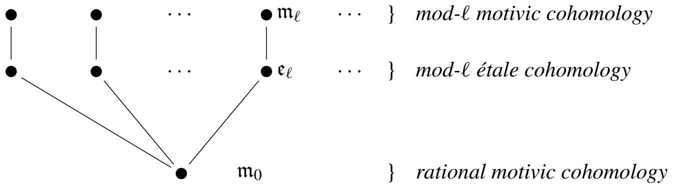
Finally, we will reduce the case of arbitrary rings to noetherian rings (Section 8) by proving in general that tt-spectra are *continuous*, that is for filtered colimits of tt-categories one has a canonical homeomorphism

$$\mathrm{Spc}(\varinjlim_i \mathcal{T}_i) \xrightarrow{\sim} \varprojlim_i \mathrm{Spc}(\mathcal{T}_i).$$

In fact, we will prove a more general statement which we believe will be useful in other studies of tt-geometry as well, because it often allows to reduce the tt-geometry of "infinite objects" to the tt-geometry of "finite objects". For example, it shows immediately that the noetherianity assumption in the results of [Dell'Ambrogio and Stevenson 2013] is superfluous, arguably simplifying the proof given for this observation in [Dell'Ambrogio and Stevenson 2014].

We mentioned above that one of our motivations for studying the questions discussed in this note lies in the theory of motives. Let us therefore give the following application. We are able to describe completely the spectrum of the triangulated category of Tate motives over the algebraic numbers with integer coefficients, $\mathrm{DTM}(\overline{\mathbb{Q}}, \mathbb{Z})$. (Previously, only the rational part $\mathrm{DTM}(\overline{\mathbb{Q}}, \mathbb{Q})$ was known.)

Theorem. *The tt-spectrum of* $\mathrm{DTM}(\overline{\mathbb{Q}}, \mathbb{Z})$ *consists of the following primes, with specialization relations as indicated by the lines going upward.*


$\mathfrak{m}_\ell$
$\cdots$ } *mod-ℓ motivic cohomology*
$\mathfrak{e}_\ell$
$\cdots$ } *mod-ℓ étale cohomology*
$\mathfrak{m}_0$
} *rational motivic cohomology*

Here ℓ runs through all prime numbers and the primes are defined by the vanishing of the cohomology theories as indicated on the right. Moreover, the proper closed subsets are precisely the finite subsets stable under specialization.

As a consequence, we are able to classify the thick tensor ideals of $\mathrm{DTM}(\overline{\mathbb{Q}}, \mathbb{Z})$. This theorem and related results are proved in a separate paper [Gallauer 2017].

1. Conventions

A symmetric, unital monoidal structure on a category is called *tensor structure* if the category is additive and the monoidal product is additive in each variable separately. We also call these data simply a *tensor category*. A *tensor functor* between tensor categories is a strong, unital, symmetric monoidal additive functor.

Our conventions regarding tensor triangular geometry mostly follow those of [Balmer 2010a]. A *tensor triangulated category* (or *tt-category* for short) is a triangulated category with a compatible (symmetric, unital) tensor structure. Typically, one assumes that the category is (essentially) small. If not specified otherwise, the tensor product is denoted by $\otimes$ and the unit by $\mathbb{1}$. A *tt-functor* is an exact tensor functor between tt-categories.

A *tt-ideal* in a tt-category $\mathcal{T}$ is a thick subcategory $\mathcal{I} \subset \mathcal{T}$ such that $\mathcal{T} \otimes \mathcal{I} \subset \mathcal{I}$. If S is a set of objects in $\mathcal{T}$ we denote by $\langle S \rangle$ the tt-ideal generated by S. To a small rigid tt-category $\mathcal{T}$ one associates a locally ringed space $\mathrm{Spec}(\mathcal{T})$, called the *tt-spectrum of* $\mathcal{T}$, whose underlying topological space is denoted by $\mathrm{Spc}(\mathcal{T})$. It consists of *prime ideals* in $\mathcal{T}$, i.e., proper tt-ideals $\mathcal{I}$ such that $a \otimes b \in \mathcal{I}$ implies $a \in \mathcal{I}$ or $b \in \mathcal{I}$. (The underlying topological space $\mathrm{Spc}(\mathcal{T})$ is defined even if $\mathcal{T}$ is not rigid.)

All rings are commutative with unit, and morphisms of rings are unital. For R a ring, we denote by $\mathrm{Spec}(R)$ the Zariski spectrum of R (considered as a locally ringed space) whereas $\mathrm{Spc}(R)$ denotes its underlying topological space (as for the tt-spectrum). We adopt similar conventions regarding graded rings R: they are commutative in a general graded sense [Balmer 2010a, 3.4], and possess a unit. $\mathrm{Spec}^{\mathrm{h}}(R)$ denotes the homogeneous Zariski spectrum with underlying topological space $\mathrm{Spc}^{\mathrm{h}}(R)$.

As a general rule, canonical isomorphisms in categories are typically written as equalities.

2. Category of filtered modules

In this section we describe filtered modules from a slightly nonstandard perspective which will be useful in the sequel. Hereby we follow the treatment in [Schapira and Schneiders 2016]. The idea is to embed the (nonabelian) category of filtered modules into its *abelianization*, the category of presheaves of modules on the poset $\mathbb{Z}$. From this embedding we deduce a number of properties of the category of filtered modules. Much of the discussion in this section applies more generally to filtered objects in suitable abelian tensor categories.

Fix a commutative ring with unit R. Denote by $\mathrm{Mod}(R)$ the abelian category of R-modules, with its canonical tensor structure. We view $\mathbb{Z}$ as a monoidal category where

$$\hom(m, n) = \begin{cases} \{*\} & m \leq n, \\ \varnothing & m > n \end{cases}$$

and $m \otimes n = m + n$. The Day convolution product then induces a tensor structure on the category of presheaves on $\mathbb{Z}$ with values in $\mathrm{Mod}(R)$ which we denote by $\mathbb{Z}^{\mathrm{op}}R$. Explicitly, an object a of $\mathbb{Z}^{\mathrm{op}}R$ is an infinite sequence of morphisms in $\mathrm{Mod}(R)$

$$\cdots \to a_{n+1} \xrightarrow{a_{n,n+1}} a_n \xrightarrow{a_{n-1,n}} a_{n-1} \to \cdots , \tag{2.1}$$

and the tensor product of two such objects a and b is described by

$$(a \otimes_{\mathbb{Z}^{\mathrm{op}}} b)_n = \mathrm{colim}_{p+q \geq n}\, a_p \otimes_R b_q .$$

Let M be an R-module and $n \in \mathbb{Z}$. The associated presheaf $\oplus_{\hom_{\mathbb{Z}}(-,n)} M$ is denoted by $M(n)$. It is the object

$$\cdots \to 0 \to 0 \to M \xrightarrow{\mathrm{id}} M \xrightarrow{\mathrm{id}} M \to \cdots$$

with the first M in degree n. Via the association $\sigma_0 : M \mapsto M(0)$ we view $\mathrm{Mod}(R)$ as a full subcategory of $\mathbb{Z}^{\mathrm{op}}R$. For any object $a \in \mathbb{Z}^{\mathrm{op}}R$ and $n \in \mathbb{Z}$ we denote by $a(n)$ the tensor product $a \otimes R(n)$, and we call it the n-th twist of a. Explicitly, this is the sequence of (2.1) shifted to the left by n places, i.e., $a(n)_m = a_{m-n}$.

The category $\mathbb{Z}^{\mathrm{op}}R$ is R-linear Grothendieck abelian, and the monoidal structure is closed. Explicitly, the internal hom of $a, b \in \mathbb{Z}^{\mathrm{op}}R$ is given by

$$\underline{\hom}(a, b)_n = \hom_{\mathbb{Z}^{\mathrm{op}}R}(a(n), b).$$

Here is another way of thinking about $\mathbb{Z}^{\mathrm{op}}R$. Let $a \in \mathbb{Z}^{\mathrm{op}}R$ be a presheaf of R-modules. Associate to it the graded $R[\beta]$-module $\bigoplus_{n\in\mathbb{Z}} a_n$ with β acting by $\beta : a \to a(1)$, i.e., in degree n by $a_{n-1,n} : a_n \to a_{n-1}$. In particular, β is assumed to have degree -1. Conversely, given a graded $R[\beta]$-module $\bigoplus_{n\in\mathbb{Z}} M_n$, define a presheaf by $n \mapsto M_n$ and transition maps $\cdot\beta : M_n \to M_{n-1}$. This clearly establishes an isomorphism of categories $\mathbb{Z}^{\mathrm{op}}R = \mathrm{Mod}_{\mathrm{gr}}(R[\beta])$, and it is not difficult to see that the isomorphism is compatible with the tensor structures on both sides.

Definition 2.2. (1) A *filtered R-module* is an object $a \in \mathbb{Z}^{\mathrm{op}}R$ such that $a_{n,n+1}$ is a monomorphism for all $n \in \mathbb{Z}$. The full subcategory of filtered R-modules in $\mathbb{Z}^{\mathrm{op}}R$ is denoted by $\mathrm{Mod}_{\mathrm{fil}}(R)$.

(2) A *finitely filtered R-module* is a filtered R-module a such that $a_{n,n+1}$ is an isomorphism for almost all n.

(3) A filtered R-module a is *separated* if $\bigcap_{n\in\mathbb{Z}} a_n = 0$.

For a filtered R-module a we denote the "underlying" R-module $\varinjlim_{n\to-\infty} a_n$ by $\pi(a)$. This clearly defines a functor $\pi : \mathrm{Mod}_{\mathrm{fil}}(R) \to \mathrm{Mod}(R)$ which "forgets the filtration". In this way we recover the more classical perspective on filtrations: an R-module $\pi(a)$ together with a (decreasing, exhaustive) filtration $(a_n)_{n\in\mathbb{Z}}$; a morphism $f : a \to b$ of filtered R-modules a, b is an R-linear morphism $\pi(a) \to \pi(b)$ compatible with the filtration.

To a filtered R-module a one can associate its ($\mathbb{Z}$-)graded R-module whose n-th graded piece is $\mathrm{coker}(a_{n,n+1}) = a_n/a_{n+1}$. This clearly defines a functor $\mathrm{gr}_\bullet : \mathrm{Mod}_{\mathrm{fil}}(R) \to \mathrm{Mod}_{\mathrm{gr}}(R) = \prod_{n\in\mathbb{Z}} \mathrm{Mod}(R)$.

The following observation is simple but very useful.

Lemma 2.3 [Schapira and Schneiders 2016, 3.5]. *The inclusion* $\iota : \mathrm{Mod}_{\mathrm{fil}}(R) \to \mathbb{Z}^{\mathrm{op}}R$ *admits a left adjoint* $\kappa : \mathbb{Z}^{\mathrm{op}}R \to \mathrm{Mod}_{\mathrm{fil}}(R)$ *given by*

$$\kappa(a)_n = \mathrm{im}(a_n \to \varinjlim_{m\to-\infty} a_m)$$

and the canonical transition maps.

It follows from Lemma 2.3 that $\mathrm{Mod}_{\mathrm{fil}}(R)$ is complete and cocomplete. Limits, filtered colimits and direct sums are computed in $\mathbb{Z}^{\mathrm{op}}R$ while pushouts are computed by applying the reflector κ to the pushout in $\mathbb{Z}^{\mathrm{op}}R$. (The statement about limits and pushouts is formal, while the rest stems from the fact that filtered colimits and direct sums are exact in $\mathrm{Mod}(R)$.) In particular, $\mathrm{Mod}_{\mathrm{fil}}(R)$ is additive and has kernels and cokernels. However, it is not an abelian category as witnessed by the morphism

$$\beta : R(0) \to R(1) \tag{2.4}$$

induced by the map $0 \to 1$ in $\mathbb{Z}$ through the Yoneda embedding: both $\ker(\beta)$ and $\mathrm{coker}(\beta)$ are 0 but β is not an isomorphism. It is an example of a *nonstrict* morphism. (A morphism $f : a \to b$ is called *strict* if the canonical morphism $\mathrm{coim}(f) \to \mathrm{im}(f)$ is an isomorphism, or equivalently if $\mathrm{im}(\pi(f)) \cap b_n = \mathrm{im}(f_n)$ for all $n \in \mathbb{Z}$.) However, one can easily check that strict monomorphisms and strict epimorphisms in $\mathrm{Mod}_{\mathrm{fil}}(R)$ are preserved by pushouts and pullbacks, respectively [Schapira and Schneiders 2016, 3.9]. In other words, $\mathrm{Mod}_{\mathrm{fil}}(R)$ is a *quasi*abelian category (we will use [Schneiders 1999] as a reference for the basic theory of quasiabelian categories).

An object a in a quasiabelian category is called *projective* if $\hom(a, -)$ takes strict epimorphisms to surjections. (Note that this convention differs from the categorical notion of a projective object!) For example, for a projective R-module M and $n \in \mathbb{Z}$ the object $M(n)$ is projective since $\hom_{\mathrm{Mod}_{\mathrm{fil}}(R)}(M(n), -) = \hom_R(M, (-)_n)$.

Lemma 2.5 [Schneiders 1999, 3.1.8]. *For any* $a \in \mathrm{Mod}_{\mathrm{fil}}(R)$, *the canonical morphism*

$$\bigoplus_{n\in\mathbb{Z}} \bigoplus_{x\in a_n} R(n) \to a \tag{2.6}$$

is a strict epimorphism with projective domain. In particular, *the quasiabelian category* $\mathrm{Mod}_{\mathrm{fil}}(R)$ *has enough projectives.*

Let us denote by $\sigma : \mathrm{Mod}_{\mathrm{gr}}(R) \to \mathrm{Mod}_{\mathrm{fil}}(R)$ the canonical functor which takes $(M_n)_n$ to $\bigoplus_n M_n(n)$. A filtered R-module is called *split* if it lies in the essential image of σ. Correspondingly we call a filtered R-module *split free*, *split projective* or *split finite projective* if it is (isomorphic to) the image of a free, projective or finite projective graded R-module under σ, respectively. In other words, an object of the form $\bigoplus_n M_n(n)$ with $\bigoplus_n M_n$ free, projective or finite projective, respectively. Lemma 2.5 shows that every object in $\mathrm{Mod}_{\mathrm{fil}}(R)$ admits a canonical split free resolution.

It is clear that split projective objects are projective, and the converse is also true as we now prove (see Lemma 2.8 below).

Lemma 2.7. *The full additive subcategory* $\mathrm{Proj}_{\mathrm{fil}}(R)$ *of split projectives is idempotent complete. The same is true for the full additive subcategory* $\mathrm{proj}_{\mathrm{fil}}(R)$ *of split finite projectives.*

Proof. Let $f : a \xrightarrow{\sim} b \oplus c$ be an isomorphism, with a split projective. Since a is split, there is a canonical isomorphism $g : a \xrightarrow{\sim} \bigoplus_n \mathrm{gr}_n(a)(n)$, and we can define the following composition of isomorphisms:

$$b \oplus c \xrightarrow[\sim]{f^{-1}} a \xrightarrow[\sim]{g} \bigoplus_n \mathrm{gr}_n(a)(n) \xrightarrow[\sim]{f} \bigoplus_n \mathrm{gr}_n(b \oplus c)(n) = \Big(\bigoplus_n \mathrm{gr}_n(b)(n)\Big) \oplus \Big(\bigoplus_n \mathrm{gr}_n(c)(n)\Big).$$

It is easy to see that this induces an isomorphism $b \cong \bigoplus_n \mathrm{gr}_n(b)(n)$, and we also see that $\mathrm{gr}_n(b)$ is a direct summand of $\mathrm{gr}_n(a)$. In other words, b is split projective as required. The same proof applies in the finite case. □

Lemma 2.8. *For a filtered R-module $a \in \mathrm{Mod}_{\mathrm{fil}}(R)$ the following are equivalent*:

(1) *a is projective.*

(2) *a is split projective.*

Proof. Let a be projective. As remarked in Lemma 2.5, there is a canonical strict epimorphism $b \to a$ with b split free. By definition of projectivity, there is a section $a \to b$, and since $\mathrm{Mod}_{\mathrm{fil}}(R)$ has kernels and images, we deduce that a is a direct summand of b. It therefore suffices to prove that every direct summand of a split free is split projective. This follows from Lemma 2.7. □

In general, due to the possibility of the tensor product in $\mathrm{Mod}(R)$ not being exact, the tensor structure on $\mathbb{Z}^{\mathrm{op}}R$ does not restrict to the subcategory $\mathrm{Mod}_{\mathrm{fil}}(R)$. We can use the reflector κ to rectify this: for $a, b \in \mathrm{Mod}_{\mathrm{fil}}(R)$, let

$$a \otimes b = \kappa(\iota(a) \otimes_{\mathbb{Z}^{\mathrm{op}}} \iota(b)).$$

This defines a tensor structure on $\mathrm{Mod}_{\mathrm{fil}}(R)$.[3] It is clear that the internal hom on $\mathbb{Z}^{\mathrm{op}}R$ restricts to a bifunctor on $\mathrm{Mod}_{\mathrm{fil}}(R)$, and it follows formally from Lemma 2.3 that this bifunctor is the internal hom on $\mathrm{Mod}_{\mathrm{fil}}(R)$.

Although we will in the sequel only use the implication (1)⇒(2) of the following result, it is satisfying to see these notions match up as they do in $\mathrm{Mod}(R)$. Recall that an object a in a category with filtered colimits is called *finitely presented* if $\hom(a, -)$ commutes with these filtered colimits.

Lemma 2.9. *For a filtered R-module $a \in \mathrm{Mod}_{\mathrm{fil}}(R)$ the following are equivalent*:

(1) *a is split finite projective.*

(2) *a is rigid* (*or strongly dualizable*).

(3) *a is finitely presented and projective.*

Proof. Since $\sigma : \mathrm{Mod}_{\mathrm{gr}}(R) \to \mathrm{Mod}_{\mathrm{fil}}(R)$ is a tensor functor it preserves rigid objects. This shows the implication (1)⇒(2).

[3] This can be seen as a particular instance of [Day 1972] due to the canonical isomorphisms

$$\kappa(a \otimes_{\mathbb{Z}^{\mathrm{op}}} \kappa(b)) \xleftarrow{\sim} \kappa(a \otimes_{\mathbb{Z}^{\mathrm{op}}} b) \xrightarrow{\sim} \kappa(\kappa(a) \otimes_{\mathbb{Z}^{\mathrm{op}}} b)$$

for any $a, b \in \mathbb{Z}^{\mathrm{op}}$.

For (2)⇒(3) notice that the unit $R(0)$ is both finitely presented and projective. The latter is clear, and the former is true as filtered colimits are computed in $\mathbb{Z}^{\mathrm{op}}R$. The implication is now obtained from the identification

$$\hom(a,-)=\hom(R(0),\underline{\hom}(a,R(0))\otimes -)$$

together with the fact that the tensor product preserves filtered colimits and strict epimorphisms.

Finally for (3)⇒(1), we start with the identification $a=\bigoplus_n \mathrm{gr}_n(a)(n)$ with $\mathrm{gr}_n(a)$ projective R-modules, which exists by Lemma 2.8. Notice that the forgetful functor $\pi:\mathrm{Mod}_{\mathrm{fil}}(R)\to\mathrm{Mod}(R)$ has a right adjoint $\Delta:\mathrm{Mod}(R)\to\mathrm{Mod}_{\mathrm{fil}}(R)$ which takes an R-module to the same R-module with the constant filtration. It is clear that Δ commutes with filtered colimits so that

$$\hom(\pi(a),\varinjlim -)=\hom(a,\varinjlim\Delta-)=\varinjlim\hom(a,\Delta-)=\varinjlim\hom(\pi(a),-),$$

hence $\pi(a)$ is a finitely presented R-module. We conclude that $a=\bigoplus \mathrm{gr}_n(a)(n)$ is split finite projective. □

Corollary 2.10. (1) *If $a\in\mathrm{Mod}_{\mathrm{fil}}(R)$ is projective then $a\otimes -$ preserves kernels of arbitrary morphisms.*

(2) *If $a,b\in\mathrm{Mod}_{\mathrm{fil}}(R)$ are projective then so is $a\otimes b$.*

Proof. Since the tensor product commutes with direct sums both statements follow from Lemma 2.8. □

3. Derived category of filtered modules

Quasiabelian categories are examples of exact categories and can therefore be derived in the same way. However, the theory for quasiabelian categories is more precise and we will exploit this fact starting in the current section. In the case of (separated, finitely) filtered R-modules we obtain what is classically known as the filtered derived category of R. Some of its basic properties are established, a number of which are deduced from the relation with the derived category of $\mathbb{Z}^{\mathrm{op}}R$.

For $*\in\{b,-,+,\varnothing\}$ we denote by $\mathcal{C}^*(\mathrm{Mod}_{\mathrm{fil}}(R))$ the category of bounded (respectively bounded above, bounded below, unbounded) cochain complexes in $\mathrm{Mod}_{\mathrm{fil}}(R)$, and by $\mathcal{K}^*(\mathrm{Mod}_{\mathrm{fil}}(R))$ the associated homotopy category. A complex

$$A:\quad \cdots\to A^{l-1}\xrightarrow{d^{l-1}}A^l\xrightarrow{d^l}A^{l+1}\to\cdots$$

is called *strictly exact* if all differentials d^l are strict, and the canonical morphism $\mathrm{im}(d^{l-1})\to\ker(d^l)$ is an isomorphism for all l. We note the following simple but useful fact.

Lemma 3.1 [Sjödin 1973, 1]. *Let A be a complex in $\mathrm{Mod}_{\mathrm{fil}}(R)$ and consider the following conditions:*

(1) *A is strictly exact.*

(2) *All its differentials d^l are strict and the underlying complex $\pi(A)$ is exact.*

(3) *The associated graded complex $\mathrm{gr}_\bullet(A)$ is exact, i.e., $\mathrm{gr}_n(A)$ is an exact complex for all $n\in\mathbb{Z}$.*

We have (1)⇔(2)⇒(3), and if A^l is finitely filtered and separated for all $l\in\mathbb{Z}$ then all conditions are equivalent.

The class of strictly exact complexes forms a saturated null system $\mathcal{K}^*_{\mathrm{ac}}$ [Schneiders 1999, 1.2.15] and we set $\mathcal{D}^*(\mathrm{Mod}_{\mathrm{fil}}(R)) = \mathcal{K}^*(\mathrm{Mod}_{\mathrm{fil}}(R))/\mathcal{K}^*_{\mathrm{ac}}$. The canonical triangulated structure on $\mathcal{K}^*(\mathrm{Mod}_{\mathrm{fil}}(R))$ induces a triangulated structure on $\mathcal{D}^*(\mathrm{Mod}_{\mathrm{fil}}(R))$. As follows from Lemma 3.1, this definition is an extension of the classical "filtered derived category" considered in [Illusie 1971]. There, complexes are assumed to be (uniformly) finitely filtered separated and the localization is with respect to filtered quasiisomorphisms, i.e., morphisms $f : A \to B$ of complexes such that $\mathrm{gr}_n(f)$ is a quasiisomorphism of complexes of R-modules, for all $n \in \mathbb{Z}$.

The functor $\iota : \mathrm{Mod}_{\mathrm{fil}}(R) \to \mathbb{Z}^{\mathrm{op}}R$ clearly preserves strictly exact complexes (we say that ι is *strictly exact*), hence it derives trivially to an exact functor of triangulated categories $\iota : \mathcal{D}^*(\mathrm{Mod}_{\mathrm{fil}}(R)) \to \mathcal{D}^*(\mathbb{Z}^{\mathrm{op}}R)$.

Proposition 3.2 [Schapira and Schneiders 2016, 3.16]. *The functor $\iota : \mathcal{D}^*(\mathrm{Mod}_{\mathrm{fil}}(R)) \to \mathcal{D}^*(\mathbb{Z}^{\mathrm{op}}R)$ is an equivalence of categories. Its quasiinverse is given by the left derived functor of κ.*

Explicitly, $\mathrm{L}\kappa$ may be computed using the "Rees functor" $\lambda : \mathbb{Z}^{\mathrm{op}}R \to \mathrm{Mod}_{\mathrm{fil}}(R)$ which takes $a \in \mathbb{Z}^{\mathrm{op}}R$ to the filtered R-module $\lambda(a)$ with

$$\lambda(a)_n = \bigoplus_{m \geq n} a_m$$

and the obvious inclusions as transition maps [Schapira and Schneiders 2016, 3.12]. It comes with a canonical epimorphism $\varepsilon_a : \iota\lambda(a) \to a$ and since $\mathrm{Mod}_{\mathrm{fil}}(R)$ is closed under subobjects in $\mathbb{Z}^{\mathrm{op}}R$, objects in $\mathbb{Z}^{\mathrm{op}}R$ admit an additively functorial two-term resolution by objects in $\mathrm{Mod}_{\mathrm{fil}}(R)$. Thus a complex A in $\mathbb{Z}^{\mathrm{op}}R$ is replaced by the cone of $\ker(\varepsilon_A) \to \iota\lambda(A)$ which is a complex in $\mathrm{Mod}_{\mathrm{fil}}(R)$ and computes $\mathrm{L}\kappa(A)$.

The tensor product $\otimes_{\mathbb{Z}^{\mathrm{op}}}$ on $\mathbb{Z}^{\mathrm{op}}R$ can be left derived and yields

$$\otimes^{\mathrm{L}}_{\mathbb{Z}^{\mathrm{op}}} : \mathcal{D}^*(\mathbb{Z}^{\mathrm{op}}R) \times \mathcal{D}^*(\mathbb{Z}^{\mathrm{op}}R) \to \mathcal{D}^*(\mathbb{Z}^{\mathrm{op}}R)$$

for $* \in \{-, \varnothing\}$. This follows for example from [Cisinski and Déglise 2009, 2.3] (where the descent structure is given by $(\mathcal{G} = \{R(n) \mid n \in \mathbb{Z}\}, \mathcal{H} = \{0\})$).

Lemma 3.3. *The tensor product on* $\mathrm{Mod}_{\mathrm{fil}}(R)$ *induces a left-derived tensor product*

$$\otimes^{\mathrm{L}} : \mathcal{D}^*(\mathrm{Mod}_{\mathrm{fil}}(R)) \times \mathcal{D}^*(\mathrm{Mod}_{\mathrm{fil}}(R)) \to \mathcal{D}^*(\mathrm{Mod}_{\mathrm{fil}}(R))$$

where $* \in \{-, \varnothing\}$. Moreover, the equivalence of Proposition 3.2 is compatible with the derived tensor products.

Proof. Recall that the tensor product was defined as $\kappa \circ \otimes_{\mathbb{Z}^{\mathrm{op}}} \circ (\iota \times \iota)$. Therefore the left-derived tensor product is given by

$$\otimes^{\mathrm{L}} = \mathrm{L}\kappa \circ \otimes^{\mathrm{L}}_{\mathbb{Z}^{\mathrm{op}}} \circ (\iota \times \iota).$$

The second statement is then clear. □

Corollary 3.4. *The triangulated category $\mathcal{D}(\mathrm{Mod}_{\mathrm{fil}}(R))$ is compactly generated. For an object $A \in \mathcal{D}(\mathrm{Mod}_{\mathrm{fil}}(R))$ the following are equivalent:*

(1) *A is compact.*

(2) *A is rigid.*

(3) *A is isomorphic to a bounded complex of split finite projectives.*

Proof. It is easy to see [Choudhury and Gallauer 2015, 3.20] that the set $\{R(n) \mid n \in \mathbb{Z}\}$ compactly generates $\mathcal{D}(\mathbb{Z}^{\mathrm{op}}R)$. The first statement therefore follows from Proposition 3.2. As is true in general [Neeman 1992, 2.2], the compact objects span precisely the thick subcategory generated by these generators $R(n)$. From this we see immediately that (3) implies (1). The converse implication follows from Corollary 3.5 below.

That (3) implies (2) is easy to see, using Lemma 2.9. Finally that (2) implies (1) follows formally from the tensor unit being compact (see the proof of Lemma 2.9). □

We denote by $\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R)$ the full subcategory of compact objects in $\mathcal{D}(\mathrm{Mod}_{\mathrm{fil}}(R))$. Its objects are also called *perfect filtered complexes*. Note that this is an idempotent complete, rigid tt-category. We denote the tensor product on $\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R)$ simply by $\otimes$. Recall that $\mathrm{proj}_{\mathrm{fil}}(R)$ denotes the additive category of split finite projective filtered R-modules.

Corollary 3.5. *The canonical functor $\mathcal{K}^b(\mathrm{proj}_{\mathrm{fil}}(R)) \to \mathcal{D}(\mathrm{Mod}_{\mathrm{fil}}(R))$ induces an equivalence of tt-categories*

$$\mathcal{K}^b(\mathrm{proj}_{\mathrm{fil}}(R)) \xrightarrow{\sim} \mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R).$$

Proof. The fact that the image of the functor is contained in $\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R)$ was proved in Corollary 3.4. It therefore makes sense to consider the following square of canonical exact functors

$$\begin{tikzcd}
\mathcal{K}^b(\mathrm{proj}_{\mathrm{fil}}(R)) \arrow[r] \arrow[d] & \mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R) \arrow[d] \\
\mathcal{K}^-(\mathrm{Proj}_{\mathrm{fil}}(R)) \arrow[r] & \mathcal{D}^-(\mathrm{Mod}_{\mathrm{fil}}(R))
\end{tikzcd}$$

The vertical arrows are the inclusions of full subcategories. (For the right vertical arrow this follows from [Keller 1996, 11.7].) Moreover, the bottom horizontal arrow is an equivalence, by [Schneiders 1999, 1.3.22] together with Lemma 2.5. We conclude that the top horizontal arrow is fully faithful as well.

Next, we notice that since $\mathrm{proj}_{\mathrm{fil}}(R)$ is idempotent complete by Lemma 2.7, the same is true of its bounded homotopy category [Balmer and Schlichting 2001, 2.8]. It follows that the image of the top horizontal arrow is a thick subcategory containing $R(n)$, $n \in \mathbb{Z}$. As remarked in the proof of Corollary 3.4, this implies essential surjectivity.

As tensoring with a split finite projective is strictly exact, by Corollary 2.10, the same is true for objects in $\mathcal{K}^b(\mathrm{proj}_{\mathrm{fil}}(R))$. It is then clear that the equivalence just established preserves the tensor product. □

For future reference we record the following simple fact.

Lemma 3.6. *Let $\mathcal{J} \subset \mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R)$ be a thick subcategory. Then the following are equivalent*:

(1) *$\mathcal{J}$ is a tt-ideal.*

(2) *$\mathcal{J}$ is closed under $R(n) \otimes -$, $n \in \mathbb{Z}$.*

Proof. As remarked in the proof of Corollary 3.4, the category of filtered complexes $\mathcal{D}_{\text{fil}}^{\text{perf}}(R)$ is generated as a thick subcategory by $R(n)$, $n \in \mathbb{Z}$. Thus (2) implies (1):

$$\mathcal{J} \otimes \mathcal{D}_{\text{fil}}^{\text{perf}}(R) = \mathcal{J} \otimes \langle R(n) \mid n \in \mathbb{Z}\rangle^{\text{thick}} \subset \mathcal{J}.$$

The converse is trivial. □

Let us discuss the derived analogues of the functors π and $\text{gr}_\bullet$ introduced earlier.

Lemma 3.7. *The functor* $\pi : \text{Mod}_{\text{fil}}(R) \to \text{Mod}(R)$ *is strictly exact and derives trivially to a tt-functor* $\pi : \mathcal{D}(\text{Mod}_{\text{fil}}(R)) \to \mathcal{D}(R)$. *The latter preserves compact objects and restricts to a tt-functor*

$$\pi : \mathcal{D}_{\text{fil}}^{\text{perf}}(R) \to \mathcal{D}^{\text{perf}}(R),$$

where $\mathcal{D}^{\text{perf}}(R)$ *denotes the category of perfect complexes over* R, *i.e., the compact objects in* $\mathcal{D}(R)$.

Proof. The first statement follows from Lemma 3.1. The functor π being tensor, it preserves rigid objects and the second statement follows from Corollary 3.4. □

Lemma 3.8. *The functor* $\text{gr}_\bullet : \text{Mod}_{\text{fil}}(R) \to \text{Mod}_{\text{gr}}(R)$ *is strictly exact and derives trivially to an exact functor* $\text{gr}_\bullet : \mathcal{D}(\text{Mod}_{\text{fil}}(R)) \to \mathcal{D}(\text{Mod}_{\text{gr}}(R))$. *The latter preserves compact objects and induces a conservative tt-functor*

$$\text{gr} := \oplus\, \text{gr}_\bullet : \mathcal{D}_{\text{fil}}^{\text{perf}}(R) \to \mathcal{D}^{\text{perf}}(R).$$

Proof. That $\text{gr}_\bullet$ is strictly exact is Lemma 3.1. It follows that $\text{gr}_\bullet$ derives trivially to give an exact functor $\text{gr}_\bullet : \mathcal{D}(\text{Mod}_{\text{fil}}(R)) \to \mathcal{D}(\text{Mod}_{\text{gr}}(R))$. For each n, gr_n clearly sends perfect filtered complexes to perfect complexes, i.e., $\text{gr}_\bullet$ preserves compact objects (by Corollary 3.4).

The functor $\oplus : \text{Mod}_{\text{gr}}(R) \to \text{Mod}(R)$ is strictly exact (in fact, it preserves arbitrary kernels and cokernels) and hence derives trivially as well to give a tt-functor which preserves compact objects.

There is a canonical natural transformation (on the underived level)

$$\text{gr}_\bullet \otimes \text{gr}_\bullet \to \text{gr}_\bullet \circ \otimes$$

endowing $\text{gr}_\bullet$ with the structure of a unital lax monoidal functor [Sjödin 1973, 3]. This natural transformation is easily seen to be an isomorphism for split finite projective filtered R-modules [Sjödin 1973, 12]. It follows that $\text{gr} : \mathcal{K}^b(\text{proj}_{\text{fil}}(R)) \to \mathcal{K}^b(\text{proj}(R))$ is a tt-functor ($\text{proj}(R)$ is the category of finitely generated projective R-modules). Conservativity of this functor follows from Lemma 3.1. □

Finally, notice that viewing $\text{Mod}(R)$ as a tensor subcategory of $\text{Mod}_{\text{fil}}(R)$ induces a section

$$\sigma_0 : \mathcal{D}^{\text{perf}}(R) \to \mathcal{D}_{\text{fil}}^{\text{perf}}(R)$$

to both gr and π.

4. Main result

The set of endomorphisms of the unit in a tt-category $\mathcal{T}$ is a (unital commutative) ring $\mathcal{R}_\mathcal{T}$, called the central ring of $\mathcal{T}$. There is a canonical morphism of locally ringed spaces

$$\rho_\mathcal{T} : \mathrm{Spec}(\mathcal{T}) \to \mathrm{Spec}(\mathcal{R}_\mathcal{T})$$

comparing the tt-spectrum of $\mathcal{T}$ with the Zariski spectrum of its central ring, as explained in [Balmer 2010a].

There is also a graded version of this construction. Given an invertible object $u \in \mathcal{T}$, it makes sense to consider the graded central ring with respect to u ([Balmer 2010a, 3.2], see also Section 5 for further discussion):

$$\mathcal{R}^\bullet_{\mathcal{T},u} := \hom_\mathcal{T}(\mathbb{1}, u^{\otimes\bullet}), \quad \bullet \in \mathbb{Z}.$$

This is a unital ϵ-commutative graded ring [Balmer 2010a, 3.3] and we can therefore consider its homogeneous spectrum. There is again a canonical morphism of locally ringed spaces

$$\rho^\bullet_{\mathcal{T},u} : \mathrm{Spec}(\mathcal{T}) \to \mathrm{Spec}^\mathrm{h}(\mathcal{R}^\bullet_{\mathcal{T},u}).$$

The inclusion $\mathcal{R}_\mathcal{T} \to \mathcal{R}^\bullet_{\mathcal{T},u}$ as the degree 0 part provides a factorization $\rho_\mathcal{T} = (R_\mathcal{T} \cap -) \circ \rho^\bullet_{\mathcal{T},u}$.

Let us specialize to $\mathcal{T} = \mathcal{D}^\mathrm{perf}_\mathrm{fil}(R)$. The object $R(1) \in \mathcal{D}^\mathrm{perf}_\mathrm{fil}(R)$ is clearly invertible and we define $\mathcal{R}^\bullet_R := \mathcal{R}^\bullet_{\mathcal{D}^\mathrm{perf}_\mathrm{fil}(R),R(1)}$, so that

$$\mathcal{R}^\bullet_R = \hom_{\mathcal{D}^\mathrm{perf}_\mathrm{fil}(R)}(R(0), R(\bullet)), \quad \bullet \in \mathbb{Z}.$$

Also, $\rho^\bullet_R := \rho^\bullet_{\mathcal{D}^\mathrm{perf}_\mathrm{fil}(R),R(1)}$.

We are now in a position to state our main result.

Theorem 4.1. (1) *The graded central ring $\mathcal{R}^\bullet_R$ is canonically isomorphic to the polynomial ring $R[\beta]$ where $\beta : R(0) \to R(1)$ as in* (2.4) *has degree* 1.

(2) *The morphism*

$$\rho^\bullet_R : \mathrm{Spec}(\mathcal{D}^\mathrm{perf}_\mathrm{fil}(R)) \to \mathrm{Spec}^\mathrm{h}(R[\beta])$$

is an isomorphism of locally ringed spaces.

The first part is immediate: by Corollary 3.5, morphisms $R(0) \to R(n)$ in $\mathcal{D}^\mathrm{perf}_\mathrm{fil}(R)$ may be computed in the homotopy category into which $\mathrm{proj}_\mathrm{fil}(R)$ embeds fully faithfully. Using the Yoneda embedding we therefore find

$$\begin{aligned}\hom_{\mathcal{D}^\mathrm{perf}_\mathrm{fil}(R)}(R(0), R(n)) &= \hom_{\mathcal{K}^\mathrm{b}(\mathrm{proj}_\mathrm{fil}(R))}(R(0), R(n))\\ &= \hom_{\mathrm{proj}_\mathrm{fil}(R)}(R(0), R(n))\\ &= \oplus_{\hom_\mathbb{Z}(0,n)} R\\ &= \begin{cases} R \cdot \{0 \to n\} & n \geq 0,\\ 0 & n < 0\end{cases}\end{aligned}$$

and under this identification, $\{0 \to n\}$ corresponds to β^n.

In the remainder of this section we outline two proofs of the second part of Theorem 4.1 and deduce the classification of tt-ideals in $\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)$ in Corollary 4.9. The subsequent sections will provide the missing details.

First proof of Theorem 4.1(2). It is proven in [Dell'Ambrogio and Stevenson 2013, 5.1] (R noetherian); [Dell'Ambrogio and Stevenson 2014, 4.7] (R general) that the comparison map

$$\rho^\bullet : \mathrm{Spc}(\mathcal{D}^{\mathrm{perf}}(R[\beta])_{\mathrm{gr}}) \to \mathrm{Spc}^{\mathrm{h}}(R[\beta])$$

is a homeomorphism, where the thick subcategory of compact objects in $\mathcal{D}(\mathrm{Mod}_{\mathrm{gr}}(R[\beta]))$ is denoted by $\mathcal{D}(R[\beta])_{\mathrm{gr}}^{\mathrm{perf}}$. It then follows from Proposition 3.2 and Lemma 3.3 (as well as the identification $\mathbb{Z}^{\mathrm{op}}R \cong \mathrm{Mod}_{\mathrm{gr}}(R[\beta])$ discussed in Section 2) that the same is true for $\rho_R^\bullet : \mathrm{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)) \to \mathrm{Spc}^{\mathrm{h}}(R[\beta])$. By [Balmer 2010a, 6.11], the morphism of locally ringed spaces $\rho_R^\bullet$ is then automatically an isomorphism. □

Second proof of Theorem 4.1(2). For the second proof of Theorem 4.1.(2) we proceed as follows. By [Balmer 2010a, 6.11], it suffices to show that

$$\rho_R^\bullet : \mathrm{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)) \to \mathrm{Spc}^{\mathrm{h}}(\mathcal{R}_R^\bullet)$$

is a homeomorphism on the underlying topological spaces.

Consider the invertible object $R \in \mathcal{D}^{\mathrm{perf}}(R)$ and the associated graded central ring $R[t, t^{-1}]$ where $t = \mathrm{id} : R \to R$ has degree 1. The morphisms of graded R-algebras induced by gr and π respectively are given by

$$\begin{array}{cc} R[\beta] \xrightarrow{\mathrm{gr}} R[t, t^{-1}] & \qquad R[\beta] \xrightarrow{\pi} R[t, t^{-1}] \\ \beta \longmapsto 0 & \qquad \beta \longmapsto t \end{array} \tag{4.2}$$

Recall (Section 3) the existence of a section σ_0 to gr and π. We therefore obtain for $\xi \in \{\mathrm{gr}, \pi\}$ commutative diagrams of topological spaces and continuous maps

$$\begin{tikzcd}
\mathrm{Spc}(\mathcal{D}^{\mathrm{perf}}(R)) \arrow[r, "\mathrm{Spc}(\xi)"] \arrow[d, "\rho^\bullet"'] \arrow[dd, bend right=60, "\rho"'] & \mathrm{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)) \arrow[r, "\mathrm{Spc}(\sigma_0)"] \arrow[d, "\rho_R^\bullet"'] & \mathrm{Spc}(\mathcal{D}^{\mathrm{perf}}(R)) \arrow[d, "\rho^\bullet"] \arrow[dd, bend left=60, "\rho"] \\
\mathrm{Spc}^{\mathrm{h}}(R[t, t^{-1}]) \arrow[r, "\mathrm{Spc}^{\mathrm{h}}(\xi)"] \arrow[d, "\sim"'] & \mathrm{Spc}^{\mathrm{h}}(R[\beta]) \arrow[r, "\mathrm{Spc}^{\mathrm{h}}(\sigma_0)"] \arrow[d] & \mathrm{Spc}^{\mathrm{h}}(R[t, t^{-1}]) \arrow[d, "\sim"] \\
\mathrm{Spc}(R) \arrow[r, "\mathrm{Spc}(\xi)", "="'] & \mathrm{Spc}(R) \arrow[r, "\mathrm{Spc}(\sigma_0)", "="'] & \mathrm{Spc}(R)
\end{tikzcd}$$

where the outer vertical maps are all homeomorphisms [Balmer 2010a, 8.1] and the composition of the horizontal morphisms in each row is the identity. It follows immediately that both Spc(gr) and Spc(π)

are homeomorphisms onto their respective images which are disjoint by (4.2). More precisely, we have

$$\begin{aligned} \operatorname{im}(\operatorname{Spc}(\mathrm{gr})) &\subseteq (\rho_R^\bullet)^{-1}(Z(\beta)) = \operatorname{supp}(\operatorname{cone}(\beta)), \\ \operatorname{im}(\operatorname{Spc}(\pi)) &\subseteq (\rho_R^\bullet)^{-1}(U(\beta)) = U(\operatorname{cone}(\beta)). \end{aligned} \tag{4.3}$$

It now remains to prove two things:

- $\operatorname{Spc}(\mathrm{gr})$ and $\operatorname{Spc}(\pi)$ are jointly surjective.
- Specializations lift along $\rho_R^\bullet$.

Indeed, since $\rho_R^\bullet$ is a spectral map between spectral spaces [Balmer 2010a, 5.7], it being a homeomorphism is equivalent to it being bijective and lifting specializations [Hochster 1967, 8.16].

The first bullet point is the subject of the subsequent sections. Let us assume it for now and establish the second bullet point. In particular, we now assume that the inclusions in (4.3) are equalities. Let $\rho_R^\bullet(\mathfrak{P}) \rightsquigarrow \rho_R^\bullet(\mathfrak{Q})$ be a specialization in $\operatorname{Spc}^{\mathrm{h}}(R[\beta])$, i.e., $\rho_R^\bullet(\mathfrak{P}) \subset \rho_R^\bullet(\mathfrak{Q})$. If $\beta \notin \rho_R^\bullet(\mathfrak{Q})$ then both primes lie in the image of $\operatorname{Spc}(\pi)$ and we already know that $\mathfrak{P} \rightsquigarrow \mathfrak{Q}$. Similarly, if $\beta \in \rho_R^\bullet(\mathfrak{P})$ then both primes lie in the image of $\operatorname{Spc}(\mathrm{gr})$ and we deduce again that $\mathfrak{P} \rightsquigarrow \mathfrak{Q}$. So we may assume $\beta \in \rho_R^\bullet(\mathfrak{Q})\backslash\rho_R^\bullet(\mathfrak{P})$. Define $\mathfrak{r} = \rho_R^\bullet(\mathfrak{Q}) \cap R \in \operatorname{Spc}(R)$ and notice that

$$\rho_R^\bullet(\mathfrak{P}) \subset \mathfrak{r}[\beta] \subset \mathfrak{r} + \langle\beta\rangle = \rho_R^\bullet(\mathfrak{Q}).$$

Consequently, the preimage of $\mathfrak{r}[\beta]$ under $\rho_R^\bullet$ is the prime

$$\mathfrak{R} = \ker(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R) \xrightarrow{\pi} \mathcal{D}^{\mathrm{perf}}(R) \to \mathcal{D}^{\mathrm{perf}}(R/\mathfrak{r}))$$

which contains the prime

$$\mathfrak{Q} = \ker(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R) \xrightarrow{\mathrm{gr}} \mathcal{D}^{\mathrm{perf}}(R) \to \mathcal{D}^{\mathrm{perf}}(R/\mathfrak{r})).$$

We now obtain specialization relations

$$\mathfrak{P} \rightsquigarrow \mathfrak{R} \rightsquigarrow \mathfrak{Q}$$

and the proof is complete. □

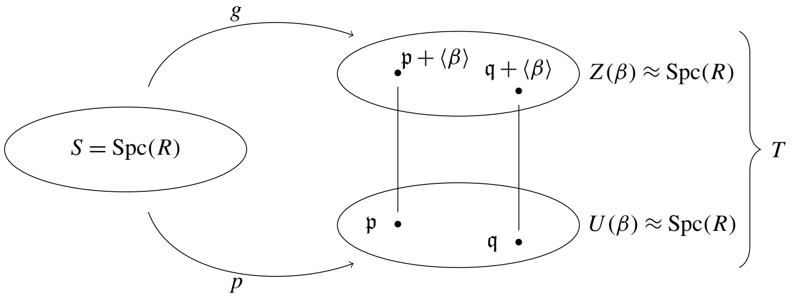
As a consequence of Theorem 4.1 we will classify the tt-ideals in $\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)$.

Lemma 4.4. *The following two maps set up an order preserving bijection*

$$\begin{aligned} \{\Pi \subset \Gamma \mid \Pi, \Gamma \subset \operatorname{Spc}(R) \textit{ Thomason subsets}\} &\leftrightarrow \{\textit{Thomason subsets of } \operatorname{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R))\} \\ (\Pi \subset \Gamma) &\mapsto \operatorname{Spc}(\pi)(\Pi) \sqcup \operatorname{Spc}(\mathrm{gr})(\Gamma) \\ (\operatorname{Spc}(\pi)^{-1}(Y) \subset \operatorname{Spc}(\mathrm{gr})^{-1}(Y)) &\leftarrow\!\shortmid Y \end{aligned}$$

Here, the order relation on the left is given by $(\Pi \subset \Gamma) \le (\Pi' \subset \Gamma')$ if $\Pi \subset \Pi'$ and $\Gamma \subset \Gamma'$.

Proof. To ease the notation, let us denote in this proof by $p : S \to T$ (respectively, $g : S \to T$) the map $\operatorname{Spc}(\pi) : \operatorname{Spc}(R) \to \operatorname{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R))$ (respectively $\operatorname{Spc}(\mathrm{gr})$). It might be helpful to keep the following picture in mind.


g
$\mathfrak{p}+\langle\beta\rangle$
$\mathfrak{q}+\langle\beta\rangle$
$Z(\beta)\approx \mathrm{Spc}(R)$
$S=\mathrm{Spc}(R)$
T
$\mathfrak{p}$
$\mathfrak{q}$
$U(\beta)\approx \mathrm{Spc}(R)$
p

Thus g and p are spectral maps between spectral spaces, homeomorphisms onto disjoint images which jointly make up all of T. Moreover, the image of p is open, and there is a common retraction $r : T \to S$ to both g and p.

First, the preimages of a Thomason subset $Y \subset T$ under the spectral maps g and p are Thomason. Moreover, every Thomason subset is closed under specializations from which one deduces $p^{-1}(Y) \subset g^{-1}(Y)$. This shows that the map from right to left is well defined.

Next, given $\Pi \subset \Gamma \subset \mathrm{Spc}(R)$ two Thomason subsets we claim that $p(\Pi) \sqcup g(\Gamma)$ is Thomason as well. By assumption, we may write $\Pi = \bigcup_i \Pi_i$ and $\Gamma = \bigcup_j \Gamma_j$ with Π_i^c and Γ_j^c quasicompact open subsets, and hence also

$$\Pi = \Pi \cap \Gamma = \left(\bigcup_i \Pi_i\right) \cap \left(\bigcup_j \Gamma_j\right) = \bigcup_{i,j} (\Pi_i \cap \Gamma_j)$$

with $(\Pi_i \cap \Gamma_j)^c = \Pi_i^c \cup \Gamma_j^c$ quasicompact open. Then

$$p(\Pi) \sqcup g(\Gamma) = \left(\bigcup_{i,j} p(\Pi_i \cap \Gamma_j)\right) \sqcup \left(\bigcup_j g(\Gamma_j)\right) = \bigcup_{i,j} (p(\Pi_i \cap \Gamma_j) \sqcup g(\Gamma_j))$$

and we reduce to the case where Π^c and Γ^c are quasicompact open. But in that case,

$$(p(\Pi) \sqcup g(\Gamma))^c = (p(\Gamma^c) \sqcup g(\Gamma^c)) \cup p(\Pi^c) = r^{-1}(\Gamma^c) \cup p(\Pi^c).$$

Again, r is a spectral map and hence the first set is quasicompact open. The second one is quasicompact by assumption, and also open since p is a homeomorphism onto an open subset. This shows that the map from left to right is also well defined.

It is obvious that the two maps are order preserving and inverses to each other. □

To state the classification result more succinctly, let us make the following definition.

Definition 4.5. Let $a \in \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)$. For $\xi \in \{\pi, \mathrm{gr}\}$ set

$$\mathrm{supp}_\xi(a) := \{\mathfrak{p} \in \mathrm{Spc}(R) \mid \xi(a \otimes \kappa(\mathfrak{p})) \neq 0 \in \mathcal{D}^{\mathrm{perf}}(\kappa(\mathfrak{p}))\}.$$

We extend this definition to arbitrary subsets $\mathcal{J} \subset \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)$ by

$$\mathrm{supp}_\xi(\mathcal{J}) := \bigcup_{a \in \mathcal{J}} \mathrm{supp}_\xi(a).$$

Lemma 4.6. *Let* $a \in \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)$. *Then*:

(1) $\mathrm{supp}_{\mathrm{gr}}(a) = \{\mathfrak{p} \in \mathrm{Spc}(R) \mid a \otimes \kappa(\mathfrak{p}) \neq 0 \in \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(\kappa(\mathfrak{p}))\}$.

(2) $\mathrm{supp}_{\pi}(a) \subset \mathrm{supp}_{\mathrm{gr}}(a)$.

(3) $\mathrm{supp}_{\xi}(a) = \mathrm{supp}(\xi(a))$.

Proof.

(1) The functor $\mathrm{gr} : \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(\kappa(\mathfrak{p})) \to \mathcal{D}^{\mathrm{perf}}(\kappa(\mathfrak{p}))$ is conservative by Lemma 3.8, thus the claim.

(2) This follows immediately from the first part.

(3) We have

$$\begin{aligned} \mathrm{supp}_{\xi}(a) &= \{\mathfrak{p} \in \mathrm{Spc}(R) \mid \xi(a \otimes \kappa(\mathfrak{p})) \neq 0 \in \mathcal{D}^{\mathrm{perf}}(\kappa(\mathfrak{p}))\} \\ &= \{\mathfrak{p} \in \mathrm{Spc}(R) \mid \xi(a) \otimes \kappa(\mathfrak{p}) \neq 0 \in \mathcal{D}^{\mathrm{perf}}(\kappa(\mathfrak{p}))\} \\ &= \mathrm{supp}(\xi(a)). \end{aligned}$$ □

The relation to the usual support can be expressed in two (equivalent) ways.

Lemma 4.7. *Let* $a \in \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)$. *Then*

(1) $\mathrm{supp}(a) = \mathrm{Spc}(\pi)(\mathrm{supp}_{\pi}(a)) \sqcup \mathrm{Spc}(\mathrm{gr})(\mathrm{supp}_{\mathrm{gr}}(a))$.

(2) *Under the bijection of Lemma 4.4*, $\mathrm{supp}(a)$ *corresponds to the pair* $\mathrm{supp}_{\pi}(a) \subset \mathrm{supp}_{\mathrm{gr}}(a)$.

Proof. Both statements follow from

$$\mathrm{Spc}(\xi)^{-1}(\mathrm{supp}(a)) = \mathrm{supp}(\xi(a)) = \mathrm{supp}_{\xi}(a),$$

the last equality being true by Lemma 4.6. □

Lemma 4.8. *Let* $Y \subset \mathrm{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R))$ *be a Thomason subset*, *corresponding to* $\Pi \subset \Gamma$ *under the bijection in Lemma 4.4. For* $a \in \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)$ *the following are equivalent*:

(1) $\mathrm{supp}(a) \subset Y$.

(2) $\mathrm{supp}_{\pi}(a) \subset \Pi$ *and* $\mathrm{supp}_{\mathrm{gr}}(a) \subset \Gamma$.

Proof. This follows immediately from the way $\Pi \subset \Gamma$ is associated to Y, together with Lemma 4.7. □

Corollary 4.9. *There is an inclusion preserving bijection*:

$$\begin{aligned} \{\Pi \subset \Gamma \mid \Pi, \Gamma \subset \mathrm{Spc}(R) \textit{ Thomason subsets}\} &\leftrightarrow \{\textit{tt-ideals in } \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)\} \\ (\Pi \subset \Gamma) &\mapsto \{a \mid \mathrm{supp}_{\pi}(a) \subset \Pi, \mathrm{supp}_{\mathrm{gr}}(a) \subset \Gamma\} \\ (\mathrm{supp}_{\pi}(\mathcal{J}) \subset \mathrm{supp}_{\mathrm{gr}}(\mathcal{J})) &\leftarrow\!\shortmid \mathcal{J} \end{aligned}$$

Proof. A bijection between Thomason subsets of $\mathrm{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R))$ and tt-ideals in $\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)$ is described in [Balmer 2010b, 14]. Explicitly, it is given by $Y \mapsto \{a \mid \mathrm{supp}(a) \subset Y\}$ and $\mathrm{supp}(\mathcal{J}) \leftarrow\!\shortmid \mathcal{J}$. The Corollary follows by composing this bijection with the one of Lemma 4.4, using Lemmas 4.7 and 4.8. □

5. Central localization

In this section we study several localizations of $\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R)$ which will allow us to catch primes (points for the tt-spectrum). In order to accommodate the different localizations we are interested in, we want to work in the following setting. Let $\mathcal{A}$ be a tensor category with central ring $R = \hom_{\mathcal{A}}(\mathbb{1}, \mathbb{1})$, and fix an invertible object $u \in \mathcal{A}$. Most of the discussion in [Balmer 2010a, §3] regarding graded homomorphisms and central localization carries over to our setting. Let us recall what we will need from [loc. cit.].

The graded central ring of $\mathcal{A}$ with respect to u is $\mathcal{R}^\bullet = \hom_{\mathcal{A}}(\mathbb{1}, u^{\otimes \bullet})$. This is a $\mathbb{Z}$-graded ε-commutative ring, where $\varepsilon \in R$ is the central switch for u, i.e., the switch $u \otimes u \xrightarrow{\sim} u \otimes u$ is given by multiplication by ε. For any objects $a, b \in \mathcal{A}$, the $\mathbb{Z}$-graded abelian group $\hom^\bullet_{\mathcal{A}}(a, b) = \hom_{\mathcal{A}}(a, b \otimes u^{\otimes \bullet})$ has the structure of a graded $\mathcal{R}^\bullet$-module in a natural way.

Let $S \subset \mathcal{R}^\bullet$ be a multiplicative subset of central homogeneous elements. The central localization $S^{-1}\mathcal{A}$ of $\mathcal{A}$ with respect to S is obtained as follows: it has the same objects as $\mathcal{A}$, and for $a, b \in \mathcal{A}$,

$$\hom_{S^{-1}\mathcal{A}}(a, b) = (S^{-1} \hom^\bullet_{\mathcal{A}}(a, b))^0,$$

the degree 0 elements in the graded localization.

We now prove that this is in fact a categorical localization.

Proposition 5.1. *The canonical functor $Q : \mathcal{A} \to S^{-1}\mathcal{A}$ is the localization with respect to*

$$\Sigma = \{a \xrightarrow{s} a \otimes u^{\otimes n} \mid a \in \mathcal{A}, s \in S, |s| = n\}.$$

Moreover, $S^{-1}\mathcal{A}$ has a canonical structure of a tensor category, and Q is a tensor functor.

Proof. Denote by Q' the localization functor $\mathcal{A} \to \Sigma^{-1}\mathcal{A}$. It is clear by construction that every morphism in Σ is inverted in $S^{-1}\mathcal{A}$ thus Q factors through Q', say via the functor $F : \Sigma^{-1}\mathcal{A} \to S^{-1}\mathcal{A}$. The functor F is clearly essentially surjective. And fully faithfulness follows readily from the fact, easy to verify, that Σ admits a calculus of left (and right) fractions [Gabriel and Zisman 1967, 2.2].

The fact that $\Sigma^{-1}\mathcal{A}$ is an additive category and Q' an additive functor is [Gabriel and Zisman 1967, 3.3], and the analogous statement about the monoidal structure is proven in [Day 1973]. The monoidal product in $\Sigma^{-1}\mathcal{A}$ is automatically additive in each variable. □

Consider the homotopy category $\mathcal{K}^b(\mathcal{A})$ of $\mathcal{A}$. This is a tt-category (large if $\mathcal{A}$ is) with the same graded central ring $\mathcal{R}^\bullet$ (with respect to u considered in degree 0).

Lemma 5.2. *There is a canonical equivalence of tt-categories $S^{-1}\mathcal{K}^b(\mathcal{A}) \simeq \mathcal{K}^b(S^{-1}\mathcal{A})$, and both are equal to the Verdier localization of $\mathcal{K}^b(\mathcal{A})$ with kernel $\langle \mathrm{cone}(s) \mid s \in S \rangle$.*

Proof. The first statement can be shown in two steps. First, consider the category of chain complexes $\mathcal{C}^b(\mathcal{A})$ and the canonical functor $\mathcal{C}^b(\mathcal{A}) \to \mathcal{C}^b(S^{-1}\mathcal{A})$. By Proposition 5.1, it factors through $S^{-1}\mathcal{C}^b(\mathcal{A}) \to \mathcal{C}^b(S^{-1}\mathcal{A})$; fully faithfulness and essential surjectivity of this functor are an easy exercise using the explicit nature of the central localization. (The point is that for bounded complexes there are always only finitely many morphisms involved thus the possibility of finding a "common denominator".)

Next, since $\mathcal{C}^b(-) \to \mathcal{K}^b(-)$ is a categorical localization (with respect to chain homotopy equivalences), Proposition 5.1 easily implies the claim.

Compatibility with the tt-structure is also straightforward. The second statement follows from [Balmer 2010a, 3.6]. □

We want to draw two consequences from this discussion. For the first one, denote by $\operatorname{proj}(R)$ the tensor category of rigid objects in $\operatorname{Mod}(R)$, i.e., the category of finitely generated projective R-modules. We let $\mathcal{A} = \operatorname{proj}_{\mathrm{fil}}(R)$ and as the invertible object u we choose $R(1)$ so that $\mathcal{R}^\bullet = R[\beta]$.

Lemma 5.3. *The functor* $\pi : \operatorname{proj}_{\mathrm{fil}}(R) \to \operatorname{proj}(R)$ *is the central localization at the multiplicative set* $\{\beta^n \mid n \geq 0\} \subset R[\beta]$.

Proof. Consider the set of arrows $\Sigma = \{\beta^n : a \to a(n) \mid a \in \operatorname{proj}_{\mathrm{fil}}(R), n \geq 0\}$. By Proposition 5.1, the central localization in the statement of the Lemma is the localization at Σ. We have $\Sigma^{-1}\operatorname{proj}_{\mathrm{fil}}(R)(a, b) = \varinjlim_n \hom_{\operatorname{proj}_{\mathrm{fil}}(R)}(a(-n), b)$. At each level n, this maps injectively into $\hom_{\operatorname{proj}(R)}(\pi a, \pi b)$, and the transition maps $f \mapsto f \circ \beta$ are injective as well since β is an epimorphism, hence the induced map

$$\varinjlim_n \hom_{\operatorname{proj}_{\mathrm{fil}}(R)}(a(-n), b) \to \hom_{\operatorname{proj}(R)}(\pi a, \pi b)$$

is injective. For surjectivity, we may assume $a, b \in \operatorname{proj}_{\mathrm{fil}}(R)$ are of "weight in $[m, n]$", i.e., $m \leq n$ and $\operatorname{gr}_i(a) = \operatorname{gr}_i(b) = 0$ for all $i \notin [m, n]$. In that case $f : \pi a \to \pi b$ comes from a map $f : a(m-n) \to b$.

We have proved that $\pi : \Sigma^{-1}\operatorname{proj}_{\mathrm{fil}}(R) \to \operatorname{proj}(R)$ is fully faithful. Essential surjectivity is clear. □

Corollary 5.4. *The functor* $\pi : \mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R) \to \mathcal{D}^{\mathrm{perf}}(R)$ *is the Verdier localization at the morphisms* $\beta : A \to A(1)$, *every* $A \in \mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R)$. *In particular*, $\ker(\pi) = \langle \operatorname{cone}(\beta) \rangle$.

Proof. Let $S = \{\beta^n\} \subset R[\beta]$. We know from Lemma 5.3 that $S^{-1}\operatorname{proj}_{\mathrm{fil}}(R) = \operatorname{proj}(R)$ hence also $S^{-1}\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R) = \mathcal{D}^{\mathrm{perf}}(R)$, by Lemma 5.2, and this is the Verdier localization with kernel $\langle \operatorname{cone}(\beta^n) \mid n \geq 0 \rangle$. The latter tt-ideal is equal to $\langle \operatorname{cone}(\beta) \rangle$ by [Balmer 2010a, 2.16] and we conclude. □

Still in the same context let $\mathfrak{p} \subset R$ be a prime ideal. Denote by $q : R \to R_{\mathfrak{p}}$ the canonical localization morphism, and set $S = R \backslash \mathfrak{p}$.

Lemma 5.5. *The morphism q induces an equivalence of tensor categories*

$$S^{-1}\operatorname{proj}_{\mathrm{fil}}(R) \simeq \operatorname{proj}_{\mathrm{fil}}(R_{\mathfrak{p}}).$$

Proof. The functor $S^{-1}\operatorname{proj}_{\mathrm{fil}}(R) \to \operatorname{proj}_{\mathrm{fil}}(R_{\mathfrak{p}})$ is given by $\otimes_R R_{\mathfrak{p}}$. This is clearly a tensor functor. Since $R_{\mathfrak{p}}$ is local every finitely generated projective $R_{\mathfrak{p}}$-module is free thus $\otimes_R R_{\mathfrak{p}}$ is essentially surjective. For full faithfulness notice that $\otimes_R R_{\mathfrak{p}}$ is additive and one therefore reduces to check this for twists of R:

$$\begin{aligned} S^{-1}\hom_{\operatorname{proj}_{\mathrm{fil}}(R)}(R(m), R(n)) &= \begin{cases} S^{-1}R & n \geq m, \\ 0 & n < m \end{cases} \\ &= \begin{cases} R_{\mathfrak{p}} & n \geq m, \\ 0 & n < m \end{cases} \\ &= \hom_{\operatorname{proj}_{\mathrm{fil}}(R_{\mathfrak{p}})}(R_{\mathfrak{p}}(m), R_{\mathfrak{p}}(n)). \end{aligned}$$

□

Corollary 5.6. *The square of topological spaces*

$$\begin{tikzcd}
\mathrm{Spc}(\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R)) \arrow[d, "\rho_R"'] & \mathrm{Spc}(\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R_{\mathfrak{p}})) \arrow[l, "\mathrm{Spc}(q)"'] \arrow[d, "\rho_{R_{\mathfrak{p}}}"] \\
\mathrm{Spc}(R) & \mathrm{Spc}(R_{\mathfrak{p}}) \arrow[l, "\mathrm{Spc}(q)"]
\end{tikzcd}$$

is cartesian.

Proof. By Lemmas 5.2 and 5.5, we know that $\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R_{\mathfrak{p}})$ is the Verdier localization of $\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R)$ with kernel $\langle \mathrm{cone}(s) \mid s \in S\rangle$. The claim now follows from [Balmer 2010a, 5.6]. □

Remark 5.7. Lemma 5.5 is false for general multiplicative subsets $S \subset R$, even without taking into account filtrations. The proof shows that the functor $S^{-1}\,\mathrm{proj}_{\mathrm{fil}}(R) \to \mathrm{proj}_{\mathrm{fil}}(S^{-1}R)$ is always fully faithful but it may fail to be essentially surjective. The correct statement would therefore be that $(S^{-1}\,\mathrm{proj}_{\mathrm{fil}}(R))^{\natural} \simeq \mathrm{proj}_{\mathrm{fil}}(S^{-1}R)$, where $(-)^{\natural}$ denotes the idempotent completion. We then deduce

$$\begin{aligned}
\mathcal{K}^b(\mathrm{proj}_{\mathrm{fil}}(S^{-1}R)) &\simeq \mathcal{K}^b((S^{-1}\,\mathrm{proj}_{\mathrm{fil}}(R))^{\natural}) \\
&\simeq (\mathcal{K}^b(S^{-1}\,\mathrm{proj}_{\mathrm{fil}}(R)))^{\natural} && \text{[Balmer and Schlichting 2001, 2.8]} \\
&\simeq (S^{-1}\mathcal{K}^b(\mathrm{proj}_{\mathrm{fil}}(R)))^{\natural} && \text{Lemma 5.2}
\end{aligned}$$

and since the tt-spectrum is invariant under idempotent completion, we obtain a cartesian square as in Corollary 5.6 for arbitrary multiplicative subsets $S \subset R$.

6. Reduction steps

Let R be a noetherian ring. Recall from Section 4 that we would like to prove that the tt-functors $\pi, \mathrm{gr} : \mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R) \to \mathcal{D}^{\mathrm{perf}}(R)$ induce jointly surjective maps

$$\mathrm{Spc}(\pi), \mathrm{Spc}(\mathrm{gr}) : \mathrm{Spc}(\mathcal{D}^{\mathrm{perf}}(R)) \to \mathrm{Spc}(\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R)).$$

In this section, we will explain how to reduce this statement to R a field. The latter case will be proved in Section 7, and the case of arbitrary (i.e., not necessarily noetherian) rings will be addressed in Section 8.

Proposition 6.1. *If $r \in R$ is nilpotent then the canonical map*

$$\mathrm{Spc}(\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R/r)) \to \mathrm{Spc}(\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R))$$

is surjective.

Proof. Let $F = \otimes_R R/r : \mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R) \to \mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R/r)$. We will use the criterion in [Balmer 2017, 1.3] to establish surjectivity of $\mathrm{Spc}(F)$, i.e., we want to prove that F detects $\otimes$-nilpotent morphisms. Let $f : A \to B \in \mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(R)$ such that $\overline{f} := F(f) = 0$. Equivalently, we may consider f as a morphism $f' : R(0) \to A^{\vee} \otimes B$, where $A^{\vee}$ denotes the dual of A. Then $\overline{f'} = 0$ and if $(f')^{\otimes m} = 0$ then also $f^{\otimes m} = 0$, in other words we reduce to $A = R(0)$.

The morphism f in $\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)$ is then determined by a map $f^0 : R(0) \to B^0$ such that $\delta^0 f^0 = 0$, and $\bar{f} = 0$ means that there is a map $\bar{h} : R/r(0) \to B^{-1}/r$ such that $\overline{f^0} = \overline{\delta^{-1}}\bar{h}$. Choose a lift $h : R(0) \to B^{-1}$ of $\bar{h}$ to $\mathrm{proj}_{\mathrm{fil}}(R)$. There exists $g : R(0) \to B^0$ such that $f^0 - gr = \delta^{-1}h$. The composite gr determines a chain morphism, and we may assume that f^0 is of the form gr for some $g : R(0) \to B^0$. (The map g itself does not necessarily determine a chain morphism.)

Let $m \geq 1$ such that $r^{\circ m} = 0$. Then $f^{\otimes m} : R(0) \to B^{\otimes m}$ is described by the morphism

$$R(0) \xrightarrow{(gr)^{\otimes m}} (B^0)^{\otimes m} \hookrightarrow (B^{\otimes m})^0$$

which factors as

$$R(0) \xrightarrow{r^{\circ m} = 0} R(0) \xrightarrow{g^{\otimes m}} (B^0)^{\otimes m} \hookrightarrow (B^{\otimes m})^0.$$

We conclude that f is $\otimes$-nilpotent as required. □

Proposition 6.2. *Let $r \in R$ be a nonzerodivisor. The image of the canonical map*

$$\mathrm{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R/r)) \to \mathrm{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R))$$

is precisely the support of $\mathrm{cone}(r)$.

Proof. Let $F = \otimes_R R/r : \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R) \to \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R/r)$. The fact that r is a nonzerodivisor means that $R/r(0)$ is an object in $\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)$ hence F admits a right adjoint $G : \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R/r) \to \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)$ (which is simply the forgetful functor). We may therefore invoke [Balmer 2017, 1.7]: the image of $\mathrm{Spc}(F)$ is the support of $G(R/r(0)) = \mathrm{cone}(r)$. □

We can now put these pieces together. Notice that we have, for any ring morphism $R \to R'$ and $\xi \in \{\pi, \mathrm{gr}\}$, commutative squares

$$\begin{tikzcd}
\mathrm{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)) & \mathrm{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R')) \arrow[l, "\mathrm{Spc}(\otimes_R R')"'] \\
\mathrm{Spc}(\mathcal{D}^{\mathrm{perf}}(R)) \arrow[u, "\mathrm{Spc}(\xi)"] & \mathrm{Spc}(\mathcal{D}^{\mathrm{perf}}(R')). \arrow[l, "\mathrm{Spc}(\otimes_R R')"] \arrow[u, "\mathrm{Spc}(\xi)"']
\end{tikzcd} \tag{6.3}$$

Let $\mathfrak{P} \in \mathrm{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R))$ be a prime and set $\mathfrak{p} = \rho_R(\mathfrak{P}) \in \mathrm{Spc}(R)$. From Corollary 5.6 we know that $\mathfrak{P}$ lies in the subspace $\mathrm{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R_{\mathfrak{p}}))$. Using (6.3) we therefore reduce to a local ring $(R, \mathfrak{p})$ (still assuming $\mathfrak{p} = \rho_R(\mathfrak{P})$). We now do induction on the dimension d of R. In each case, repeated application of Proposition 6.1 in conjunction with (6.3) allows us to assume R reduced. If $d = 0$, R is necessarily a field and this case will be dealt with in Corollary 7.9. If $d > 0$ there exists a nonzerodivisor $r \in \mathfrak{p}$. Proposition 6.2 in conjunction with (6.3) reduce us to R/r but this ring has dimension $d - 1$ and we conclude by induction.

7. The case of a field

In this section we will prove Theorem 4.1 in the case of $R = k$ a field. This will follow easily from a more precise description of $\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(k)$.

We begin with a result describing the structure of any morphism in $\mathrm{proj}_{\mathrm{fil}}(k)$. For this, let us agree to call a quasiabelian category *semisimple* if every short strictly exact sequence splits. Equivalently, a quasiabelian category is semisimple if every object is projective.

Lemma 7.1. *The category* $\mathrm{proj}_{\mathrm{fil}}(k)$ *is semisimple quasiabelian.*

Proof. Notice that $\mathrm{proj}_{\mathrm{fil}}(k) \subset \mathrm{Mod}_{\mathrm{fil}}(k)$ is simply the full subcategory of separated filtered vector spaces whose underlying vector space is finite dimensional. This is an additive subcategory and the set of objects is closed under kernels and cokernels in $\mathrm{Mod}_{\mathrm{fil}}(k)$. We deduce that it is a quasiabelian subcategory.

Since every object in $\mathrm{proj}_{\mathrm{fil}}(k)$ is projective (Lemma 2.9), semisimplicity follows. □

Lemma 7.2. *Let $f : a \to b$ be a morphism in a semisimple quasiabelian category. Then f is the composition*

$$f = f_m \circ f_{em} \circ f_e, \tag{7.3}$$

where

- *f_e is the projection onto a direct summand* (*in particular a strict epimorphism*),
- *f_{em} is an epimonomorphism,*
- *f_m is the inclusion of a direct summand* (*in particular a strict monomorphism*).

Proof. As in every quasiabelian category, f factors as

$$a \xrightarrow{f_e} \mathrm{coim}(f) \xrightarrow{f_{em}} \mathrm{im}(f) \xrightarrow{f_m} b,$$

where f_e is a strict epimorphism, f_{em} is an epimonomorphism, and f_m is a strict monomorphism. The lemma now follows from the definition of semisimplicity. □

Remark 7.4. Lemma 7.2 allows to characterize certain properties of morphisms $f : a \to b$ in a particularly simple way:

(1) f is a monomorphism if and only if f_e is an isomorphism.

(2) f is an epimorphism if and only if f_m is an isomorphism.

(3) f is strict if and only if f_{em} is an isomorphism.

Fix a semisimple quasiabelian category $\mathcal{A}$. Its bounded derived category $\mathcal{D}^b(\mathcal{A})$ admits a bounded t-structure whose heart $\mathcal{D}^{\heartsuit}(\mathcal{A})$ is the subcategory of objects represented by complexes

$$0 \to a \xrightarrow{f} b \to 0, \tag{7.5}$$

where b sits in degree 0 and f is a monomorphism in $\mathcal{A}$.[4]

[4] This is [Schneiders 1999, 1.2.18, 1.2.21]. The reader who is puzzled by the asymmetry of this statement should rest assured that there is a dual t-structure for which the objects in the heart are represented by *epi*morphisms [Schneiders 1999, 1.2.23]. Also, the existence of the t-structures does not require $\mathcal{A}$ to be semisimple.

Lemma 7.6. *The t-structure on $\mathcal{D}^b(\mathcal{A})$ is strongly hereditary, i.e., for any $A, B \in \mathcal{D}^\heartsuit(\mathcal{A})$ and $i \geq 2$, we have* $\hom_{\mathcal{D}^b(\mathcal{A})}(A, B[i]) = 0$.

Proof. This follows from the fact that A and B are represented by complexes of the form (7.5), and that homomorphisms can be computed in the homotopy category. Indeed, as every object in $\mathcal{A}$ is projective, the canonical functor $\mathcal{K}^b(\mathcal{A}) \to \mathcal{D}^b(\mathcal{A})$ is an equivalence [Schneiders 1999, 1.3.22]. □

Assume now in addition that $\mathcal{A}$ is a tensor category and every object is a finite sum of invertibles. Clearly, $\mathrm{proj}_{\mathrm{fil}}(k)$ satisfies this condition.

Proposition 7.7. *Every object in $\mathcal{D}^b(\mathcal{A})$ is of the form*

$$\bigoplus_i \mathrm{cone}(g_i)[i] \oplus \bigoplus_j c_j[a_j],$$

where the sums are finite, the c_j are invertible in $\mathcal{A}$, and the g_i are epimonomorphisms in $\mathcal{A}$.

Proof. Let $A \in \mathcal{D}^b(\mathcal{A})$. By Lemma 7.6, the object A is a finite direct sum of shifts of objects in $\mathcal{D}^\heartsuit(\mathcal{A})$. As discussed above, every object in the heart is represented by a complex as in (7.5). We then deduce from Remark 7.4 that f is an epimonomorphism g followed by the inclusion of a direct summand, say with direct complement c. Thus

$$\mathrm{cone}(f) = \mathrm{cone}(g) \oplus c. \qquad \square$$

We now come to the study of tt-ideals in $\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(k) = \mathcal{D}^b(\mathrm{proj}_{\mathrm{fil}}(k))$. Proposition 7.7 tells us that every prime ideal is generated by cones of epimonomorphisms in $\mathrm{proj}_{\mathrm{fil}}(k)$. However, it turns out that all these cones generate the same prime ideal (except if 0, of course).

Proposition 7.8. *There is a unique nontrivial, proper tt-ideal in $\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(k)$ given by*

$$\ker(\pi) = \langle \mathrm{cone}(\beta) \rangle.$$

In particular, $\langle \mathrm{cone}(\beta) \rangle$ is a prime ideal.

Proof. The equality of the two tt-ideals follows from Corollary 5.4. Since π is a tt-functor and $\mathcal{D}^{\mathrm{perf}}(k)$ is local, it is clear that its kernel is a prime ideal.

Let A be a nonzero object in $\mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(k)$ such that $\langle A \rangle \neq \mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(k)$. We would like to show $\langle A \rangle = \langle \mathrm{cone}(\beta) \rangle$. By Proposition 7.7, we may assume $A = \mathrm{cone}(g)$ where g is a nonstrict epimonomorphism in $\mathrm{proj}_{\mathrm{fil}}(k)$. Writing the domain and codomain of g as a sum of invertibles, we may identify g with a square matrix with entries in the polynomial ring $k[\beta]$. Let $p(\beta) \in k[\beta]$ be the determinant. Since g is not an isomorphism neither is $\mathrm{gr}(g) \in \mathcal{D}^{\mathrm{perf}}(k)$ by Lemma 3.8. We deduce that $p(0) = 0$, or in other words $p(\beta) = \beta \cdot p'(\beta)$ for some $p'(\beta) \in k[\beta]$.

Let $\mathcal{T} = \mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(k)/\langle \mathrm{cone}(g) \rangle$ and denote by $\varphi : \mathcal{D}^{\mathrm{perf}}_{\mathrm{fil}}(k) \to \mathcal{T}$ the localization functor. As $\mathcal{T}$ is a tt-category we can consider the graded (automatically commutative) central ring $\mathcal{R}^\bullet_{\mathcal{T}}$ with respect to $\varphi(k(1))$. Since $\varphi(g)$ is invertible, $\varphi(p) \in \mathcal{R}^\bullet_{\mathcal{T}}$ is invertible as well. But then we must have

$$(\varphi(p)^{-1} \cdot \varphi(p')) \cdot \varphi(\beta) = \varphi(p)^{-1} \cdot \varphi(p) = 1$$

so $\varphi(\beta)$ is invertible as well. In other words, $\mathrm{cone}(\beta) \in \ker(\varphi) = \langle \mathrm{cone}(g) \rangle$.

Conversely, $\pi(g)$ is an isomorphism since g is an epimonomorphism. In other words, $\mathrm{cone}(g) \in \langle \mathrm{cone}(\beta)\rangle$. □

Corollary 7.9. *The canonical morphism*

$$\rho_k^\bullet : \mathrm{Spec}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(k)) \to \mathrm{Spec}^{\mathrm{h}}(k[\beta])$$

is an isomorphism of locally ringed spaces. The tt-spectrum $\mathrm{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(k))$ *is the topological space*

$$\begin{tikzcd}
\langle 0\rangle = \ker(\mathrm{gr}) \arrow[d, dash] \\
\langle \mathrm{cone}(\beta)\rangle = \ker(\pi)
\end{tikzcd}$$

where the only nontrivial specialization relation is indicated by the vertical line going upward.

8. Continuity of tt-spectra

Our primary goal in this section is to deduce the veracity of Theorem 4.1 from its veracity for noetherian rings. The idea is to write an arbitrary ring as a filtered colimit of noetherian rings, and since this technique of reducing some statement in tt-geometry to the analogous statement about "more finite" objects can be useful in other contexts we decided to approach the question in greater generality.

Denote by $ttCat$ the category of small tt-categories and tt-functors. For the moment we assume that all structure is strict, e.g., the tt-functors preserve the tensor product and translation functor on the nose.

Lemma 8.1. *The forgetful functor* $ttCat \to Cat$ *creates filtered colimits.*

Proof sketch. The fact that filtered colimits of monoidal categories are created by the forgetful functor is [Johnstone 2002, C1.1.8]. Since filtered colimits commute with finite products, the colimit will be an additive category. It is obvious how to endow the filtered colimit with a translation functor and a class of distinguished triangles. The axioms for the triangulated structure all involve only finitely many objects and morphisms each and therefore are easily seen to hold. The same is true for exactness of the monoidal product.

It remains to check universality. But given a cocone on the diagram there is a unique morphism (a priori not respecting the tt-structure) from the filtered colimit. Hence all one needs to know is that it actually does respect the tt-structure. Again, in each case this only involves finitely many objects and morphisms and is easily seen to hold. □

Let us be given a filtered diagram $(\mathcal{T}_i, f_{ij} : \mathcal{T}_i \to \mathcal{T}_j)_{i\in I}$ in $ttCat$ and denote by $\mathcal{T}$ its colimit, and by $f_i : \mathcal{T}_i \to \mathcal{T}$ the canonical functors.

Proposition 8.2. *The induced map*

$$\varphi := \varprojlim_i f_i^{-1} : \mathrm{Spc}(\mathcal{T}) \to \varprojlim_i \mathrm{Spc}(\mathcal{T}_i)$$

is a homeomorphism.

Proof. This follows from Proposition 8.5. □

Remark 8.3. In practice, of course, tt-categories and tt-functors are rarely strict, and (filtered) diagrams of such things are rarely strictly functorial. Denote by $2\text{-}tt\mathcal{C}at$ the 2-category of small tt-categories, tt-functors, and tt-isotransformations without any strictness assumptions.

Given a pseudofunctor $F : I \to 2\text{-}tt\mathcal{C}at$, where I is a small filtered category, we are going to endow its pseudocolimit $2\text{-}\varinjlim_I F$ with the structure of a tt-category. For this, choose a strictification of F, i.e., a strict 2-functor $G : I \to \mathcal{C}at$ together with a pseudonatural equivalence $\eta : F \to G$ (as pseudofunctors $I \to \mathcal{C}at$). Then use η pointwise to endow each category $G(i)$, where $i \in I$, with the structure of a tt-category, and each functor $G(\alpha)$, where $\alpha : i \to j$, with the structure of a tt-functor. In other words, make η into a pseudonatural equivalence of pseudofunctors $I \to 2\text{-}tt\mathcal{C}at$. Since $2\text{-}\varinjlim F \simeq 2\text{-}\varinjlim G$, we may assume without loss of generality that F is a strict 2-functor. But in this case the canonical functor $2\text{-}\varinjlim_I F \to \varinjlim_I F$ from the pseudocolimit to the (1-categorical) colimit is an equivalence (here we use the assumption that I is filtered see [SGA 4_2 1972, VI.6.8]). Then we can apply Lemma 8.1.[5]

Proposition 8.2 also holds in this nonstrict context. Notice first that nonstrict tt-functors induce maps on spectra exactly in the same way as strict ones. Moreover, isomorphic (nonstrict) tt-functors induce the same map. Therefore the statement of Proposition 8.2 makes sense also for pseudofunctors $I \to 2\text{-}tt\mathcal{C}at$. It is clear that $F \to 2\text{-}\varinjlim F$ satisfies the assumptions of Proposition 8.5 below, thus a homeomorphism

$$\operatorname{Spc}(2\text{-}\varinjlim F) \xrightarrow{\sim} \varprojlim_i \operatorname{Spc}(F(i)).$$

In order to generalize Proposition 8.2 we now abstract the pertinent properties of the relation between the system $(\mathcal{T}_i, f_{ij})$ and the "limit" $\mathcal{T}$.

Definition 8.4. Let $\mathcal{T}_\bullet : I \to 2\text{-}tt\mathcal{C}at$ be a pseudofunctor and $f : \mathcal{T}_\bullet \to \mathcal{T}$ a pseudonatural transformation, $\mathcal{T} \in 2\text{-}tt\mathcal{C}at$. We say that

- f is *surjective on morphisms* if for each morphism $\alpha : a \to b$ in $\mathcal{T}$ there exists $i \in I$, and a morphism $\alpha_i : a_i \to b_i$ in $\mathcal{T}_i$ such that $f_i(\alpha_i) \cong \alpha$.
- f *detects isomorphisms* if for each $a_i, b_i \in \mathcal{T}_i$ such that $f_i(a_i) \cong f_i(b_i)$ in $\mathcal{T}$ there exists $u : i \to j$ such that $\mathcal{T}_u(a_i) \cong \mathcal{T}_u(b_i)$.

The condition $f_i(\alpha_i) \cong \alpha$ here means that there are isomorphisms $a \cong f_i(a_i)$ and $b \cong f_i(b_i)$ such that

$$\begin{tikzcd}
a \arrow[r, "\alpha"] \arrow[d, dash, "\sim"'] & b \arrow[d, dash, "\sim"] \\
f_i(a_i) \arrow[r, "f_i(\alpha_i)"'] & f_i(b_i)
\end{tikzcd}$$

commutes. The transformation f being surjective on morphisms implies in particular that f is "surjective on objects" and even "surjective on triangles", in an obvious sense. Note also that detecting isomorphisms is equivalent to detecting zero objects.

[5] This is maybe not wholly satisfactory. In analogy to Lemma 8.1 one might expect the statement that $2\text{-}tt\mathcal{C}at \to 2\text{-}\mathcal{C}at$ creates filtered pseudocolimits. We won't need this at present, and leave it as a question for the interested reader.

In the following result a category I is said to be *conjoining* if

- I is nonempty, and
- for any $i, j \in I$ there exists $k \in I$ and $i \to k$, $j \to k$.

In contrast to a filtered category, it is not necessary that parallel morphisms can be equalized. (Of course, in applications I will often just be a directed poset.)

Proposition 8.5. *Let $\mathcal{T}_\bullet : I \to 2\text{-}tt\mathcal{C}at$ be a pseudofunctor with I conjoining and $f : \mathcal{T}_\bullet \to \mathcal{T}$ a pseudonatural transformation, $\mathcal{T} \in 2\text{-}tt\mathcal{C}at$. Assume that f is surjective on morphisms and detects isomorphisms. Then the induced map*

$$\varphi := \varprojlim_i f_i^{-1} : \operatorname{Spc}(\mathcal{T}) \to \varprojlim_i \operatorname{Spc}(\mathcal{T}_i)$$

is a homeomorphism.

Proof.

(1) We first prove injectivity. Let $\mathfrak{P} \neq \mathfrak{Q} \in \operatorname{Spc}(\mathcal{T})$, say $a \in \mathfrak{P} \backslash \mathfrak{Q}$. There exists $i \in I$ and $a_i \in \mathcal{T}_i$ such that $f_i(a_i) \cong a$ since f is surjective on objects. But then $a_i \in f_i^{-1}(\mathfrak{P}) \backslash f_i^{-1}(\mathfrak{Q})$ which implies $\varphi(\mathfrak{P}) \neq \varphi(\mathfrak{Q})$.

(2) For surjectivity, let $(\mathfrak{P}_i)_i \in \varprojlim \operatorname{Spc}(\mathcal{T}_i)$. Define

$$\mathfrak{P} = \{a \in \mathcal{T} \mid \exists i \in I, a_i \in \mathfrak{P}_i : a \cong f_i(a_i)\} \subset \mathcal{T}.$$

We claim that $\mathfrak{P}$ can also be described as

$$\mathfrak{P}' = \{a \in \mathcal{T} \mid \forall i \in I, a_i \in \mathcal{T}_i : a \cong f_i(a_i) \Rightarrow a_i \in \mathfrak{P}_i\}.$$

Indeed, if $a \in \mathfrak{P}'$ choose $i \in I$ and $a_i \in \mathcal{T}_i$ such that $a \cong f_i(a_i)$ which is possible since f is surjective on objects. By definition of $\mathfrak{P}'$ we must have $a_i \in \mathfrak{P}_i$, and therefore $a \in \mathfrak{P}$. Conversely, if $a \in \mathfrak{P}$, say $a \cong f_i(a_i)$ with $a_i \in \mathfrak{P}_i$, and we are given $a'_j \in \mathcal{T}_j$ such that $a \cong f_j(a'_j)$, let $k \in I$ and $u_i : i \to k$, $u_j : j \to k$. We have $f_k \mathcal{T}_{u_i}(a_i) \cong f_i(a_i) \cong a \cong f_j(a'_j) \cong f_k \mathcal{T}_{u_j}(a'_j)$ and so by assumption on f there exists $u : k \to l$ such that $\mathcal{T}_{uu_i}(a_i) \cong \mathcal{T}_u \mathcal{T}_{u_i}(a_i) \cong \mathcal{T}_u \mathcal{T}_{u_j}(a'_j) \cong \mathcal{T}_{uu_j}(a'_j)$. The former lies in $\mathfrak{P}_l$ hence so does the latter, and this implies $a'_j \in \mathfrak{P}_j$.

It is now straightforward to prove that $\mathfrak{P}$ is a prime ideal. For example, let $D : a \to b \to c \to^+$ be a triangle in $\mathcal{T}$ with $a, b \in \mathfrak{P}$. By assumption there exists $i \in I$ and a triangle $D_i : a_i \to b_i \to c_i \to^+$ in $\mathcal{T}_i$ such that $f_i(D_i) \cong D$. By what we just proved we must then have $a_i, b_i \in \mathfrak{P}_i$ and hence also $c_i \in \mathfrak{P}_i$. But then $c \cong f_i(c_i) \in \mathfrak{P}$. Since $\mathfrak{P}$ is clearly closed under translations, this shows that it is a triangulated subcategory.

For thickness we proceed similarly. Let $a, b \in \mathcal{T}$ such that $a \oplus b \in \mathfrak{P}$. We may find $i \in I$ and $a_i, b_i \in \mathcal{T}_i$ such that $a \cong f_i(a_i)$, $b \cong f_i(b_i)$. Then $f_i(a_i \oplus b_i) \cong a \oplus b \in \mathfrak{P}$ thus $a_i \oplus b_i \in \mathfrak{P}_i$ and this implies $a_i \in \mathfrak{P}_i$ or $b_i \in \mathfrak{P}_i$, i.e., $a \in \mathfrak{P}$ or $b \in \mathfrak{P}$. Primality is proven in exactly the same way as thickness.

Let $\pi_i : \varprojlim \operatorname{Spc}(\mathcal{T}_i) \to \operatorname{Spc}(\mathcal{T}_i)$ be the canonical projection so that $\pi_i \varphi = f_i^{-1}$. Then

$$\pi_i \varphi(\mathfrak{P}) = f_i^{-1}(\mathfrak{P}) = f_i^{-1}(\mathfrak{P}') = \mathfrak{P}_i$$

and this completes the proof of surjectivity.

(3) Since φ is continuous, it remains to show that it is open. A basis for the topology of $\mathrm{Spc}(\mathcal{T})$ is given by $U(a) = \mathrm{Spc}(\mathcal{T}) \setminus \mathrm{supp}(a)$, where a runs through the objects of $\mathcal{T}$. Fix $a \in \mathcal{T}$, say $a \cong f_i(a_i)$ with some $a_i \in \mathcal{T}_i$. We claim that $\varphi(U(a)) = \pi_i^{-1}(U(a_i))$ (which is open hence this would complete the proof).

Let $\mathfrak{P} \in U(a)$, which means $f_i(a_i) \cong a \in \mathfrak{P}$, or equivalently, $a_i \in f_i^{-1}(\mathfrak{P}) = \pi_i \varphi(\mathfrak{P})$, i.e., $\varphi(\mathfrak{P}) \in \pi_i^{-1}(U(a_i))$. Conversely, suppose $(\mathfrak{P}_i)_i \in \pi_i^{-1}(U(a_i))$, i.e., $a_i \in \mathfrak{P}_i$. By the proof of surjectivity in part (2), $(\mathfrak{P}_i)_i = \varphi(\mathfrak{P})$ with $a \in \mathfrak{P}$, i.e., $(\mathfrak{P}_i)_i \in \varphi(U(a))$. □

Remark 8.6. Certainly, these are not the only reasonable conditions on f which allow to deduce a homeomorphism on spectra. For example, it is likely that surjectivity on morphisms could be replaced by a nilfaithfulness assumption inspired by [Balmer 2017]. We mainly chose these conditions with easy applicability in mind.

We may apply this result to filtered modules, thereby concluding the second proof of Theorem 4.1.

Corollary 8.7. *If $\rho_R^\bullet : \mathrm{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)) \to \mathrm{Spc}^{\mathrm{h}}(R[\beta])$ is a homeomorphism for noetherian rings then it is a homeomorphism for all rings.*

Proof. Let R be an arbitrary ring and write it as the filtered colimit of its finitely generated subrings $R = \varinjlim_i R_i$. An inclusion $R_i \subset R_j$ induces a base change tt-functor $\otimes_{R_i} R_j : \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R_i) \to \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R_j)$ and we obtain a pseudofunctor $\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R_\bullet) : I \to 2\text{-}ttCat$ together with a pseudonatural transformation $f = \otimes R : \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R_\bullet) \to \mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)$. Let us check that f satisfies the assumptions of Proposition 8.5.

Note first that every free R-module comes from a free R_i-module by base change, for any i. Also, a morphism between finitely generated free R-modules is described by a matrix with entries in R. Adding these finitely many entries to R_i we see that morphisms also come from some R_i. In particular, this is true for idempotent endomorphisms of finitely generated free R-modules. We deduce that finitely generated projective R-modules also arise by base change from some R_i. The same is then true for objects and morphisms in $\mathrm{proj}_{\mathrm{fil}}(R)$ and therefore also in $\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R) = \mathcal{K}^b(\mathrm{proj}_{\mathrm{fil}}(R))$ (Corollary 3.5). In other words, f is surjective on morphisms. Moreover, a perfect filtered complex is 0 in $\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)$ if and only if it is nullhomotopic and such a homotopy again comes from some R_i. We conclude that f detects isomorphisms as well.

We may therefore apply Proposition 8.5 to deduce a commutative square

$$\begin{tikzcd}
\mathrm{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R)) \arrow[r] \arrow[d, "\rho_R^\bullet"'] & \varprojlim_i \mathrm{Spc}(\mathcal{D}_{\mathrm{fil}}^{\mathrm{perf}}(R_i)) \arrow[d, "(\rho_{R_i}^\bullet)_i"] \\
\mathrm{Spc}^{\mathrm{h}}(R[\beta]) \arrow[r] & \varprojlim_i \mathrm{Spc}^{\mathrm{h}}(R_i[\beta])
\end{tikzcd}$$

where the top horizontal map is a homeomorphism. Since the R_i are all noetherian rings, the right vertical map is a homeomorphism by assumption. And the bottom horizontal map is clearly a homeomorphism. We conclude that the left vertical map is too. □

Acknowledgment

I would like to thank Paul Balmer for his interest in this note, and his critical input on an earlier version. I am also grateful to the anonymous referee whose suggestions led to improved exposition.

References

[Balmer 2010a] P. Balmer, "Spectra, spectra, spectra: tensor triangular spectra versus Zariski spectra of endomorphism rings", *Algebr. Geom. Topol.* **10**:3 (2010), 1521–1563. MR Zbl

[Balmer 2010b] P. Balmer, "Tensor triangular geometry", pp. 85–112 in *Proceedings of the International Congress of Mathematicians, II* (Hyderabad, 2010), edited by R. Bhatia et al., Hindustan Book Agency, New Delhi, 2010. MR Zbl

[Balmer 2017] P. Balmer, "On the surjectivity of the map of spectra associated to a tensor-triangulated functor", *Bull. Lond. Math. Soc.* **50**:3 (2017), 487–495. Zbl

[Balmer and Schlichting 2001] P. Balmer and M. Schlichting, "Idempotent completion of triangulated categories", *J. Algebra* **236**:2 (2001), 819–834. MR Zbl

[Choudhury and Gallauer 2015] U. Choudhury and M. Gallauer Alves de Souza, "Homotopy theory of dg sheaves", 2015. to appear in *Comm. Algebra*. arXiv

[Cisinski and Déglise 2009] D.-C. Cisinski and F. Déglise, "Local and stable homological algebra in Grothendieck abelian categories", *Homology Homotopy Appl.* **11**:1 (2009), 219–260. MR Zbl

[Day 1972] B. Day, "A reflection theorem for closed categories", *J. Pure Appl. Algebra* **2**:1 (1972), 1–11. MR Zbl

[Day 1973] B. Day, "Note on monoidal localisation", *Bull. Austral. Math. Soc.* **8** (1973), 1–16. MR Zbl

[Dell'Ambrogio and Stevenson 2013] I. Dell'Ambrogio and G. Stevenson, "On the derived category of a graded commutative Noetherian ring", *J. Algebra* **373** (2013), 356–376. MR Zbl

[Dell'Ambrogio and Stevenson 2014] I. Dell'Ambrogio and G. Stevenson, "Even more spectra: tensor triangular comparison maps via graded commutative 2-rings", *Appl. Categ. Structures* **22**:1 (2014), 169–210. MR Zbl

[Gabriel and Zisman 1967] P. Gabriel and M. Zisman, *Calculus of fractions and homotopy theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete **35**, Springer, 1967. MR Zbl

[Gallauer 2017] M. Gallauer Alves de Souza, "tt-geometry of Tate motives over algebraically closed fields", submitted, 2017. arXiv

[Hasemeyer and Hornbostel 2005] C. Hasemeyer and J. Hornbostel, "Motives and etale motives with finite coefficients", *K-Theory* **34**:3 (2005), 195–207. MR Zbl

[Hochster 1967] M. Hochster, *Prime ideal structure in commutative rings*, Ph.D. thesis, Princeton University, 1967, Available at https://tinyurl.com/tinycutphd. Zbl

[Illusie 1971] L. Illusie, *Complexe cotangent et déformations, I*, Lecture Notes in Math. **239**, Springer, 1971. MR Zbl

[Johnstone 2002] P. T. Johnstone, *Sketches of an elephant: a topos theory compendium, II*, Oxford Logic Guides **44**, Oxford Univ. Press, 2002. MR Zbl

[Keller 1996] B. Keller, "Derived categories and their uses", pp. 671–701 in *Handbook of algebra, I*, edited by M. Hazewinkel, Handb. Algebr. **1**, Elsevier, Amsterdam, 1996. MR Zbl

[Neeman 1992] A. Neeman, "The connection between the K-theory localization theorem of Thomason, Trobaugh and Yao and the smashing subcategories of Bousfield and Ravenel", *Ann. Sci. École Norm. Sup.* (4) **25**:5 (1992), 547–566. MR Zbl

[Schapira and Schneiders 2016] P. Schapira and J.-P. Schneiders, "Derived categories of filtered objects", pp. 103–120 Astérisque **383**, Société Mathématique de France, Paris, 2016. MR Zbl

[Schneiders 1999] J.-P. Schneiders, *Quasi-abelian categories and sheaves*, Mém. Soc. Math. Fr. (N.S.) **76**, Société Mathématique de France, Paris, 1999. MR Zbl

[SGA 4_2 1972] M. Artin, A. Grothendieck, and J. L. Verdier, *Théorie des topos et cohomologie étale des schémas, Tome 2: Exposés V–VIII* (Séminaire de Géométrie Algébrique du Bois Marie 1963–1964), Lecture Notes in Math. **270**, Springer, 1972. MR Zbl

[Sjödin 1973] G. Sjödin, “On filtered modules and their associated graded modules”, *Math. Scand.* **33** (1973), 229–249. MR Zbl

Communicated by Yuri Manin
Received 2017-10-29 Revised 2018-03-13 Accepted 2018-06-12

gallauer@math.ucla.edu *Department of Mathematics, University of California, Los Angeles, Los Angeles, CA, United States*

ALGEBRA AND NUMBER THEORY 12:8 (2018)
dx.doi.org/10.2140/ant.2018.12.2005

The Euclidean distance degree of smooth complex projective varieties

Paolo Aluffi and Corey Harris

We obtain several formulas for the Euclidean distance degree (ED degree) of an arbitrary nonsingular variety in projective space: in terms of Chern and Segre classes, Milnor classes, Chern–Schwartz–MacPherson classes, and an extremely simple formula equating the Euclidean distance degree of X with the Euler characteristic of an open subset of X.

1. Introduction

The *Euclidean distance degree* (ED degree) of a variety X is the number of critical nonsingular points of the distance function from a general point u outside of X. This definition, tailored to real algebraic varieties, may be adapted to complex projective varieties X, by considering the critical points of the (complex-valued) function $\sum_i (x_i - u_i)^2$ on the smooth part of the affine cone over X. This is the context in which we will work in this paper. The ED degree is studied thoroughly in [Draisma et al. 2016], which provides a wealth of examples, results, and applications. In particular, [Draisma et al. 2016, Theorem 5.4] states that the ED degree of a complex projective variety $X \subseteq \mathbb{P}^{n-1}$ equals the sum of its "polar degrees", provided that the variety satisfies a technical condition related to its intersection with the *isotropic quadric,* i.e., the quadric Q with equation $x_1^2 + \cdots + x_n^2 = 0$. As a consequence, a formula is obtained [Draisma et al. 2016, Theorem 5.8] computing the Euclidean distance degree of a *nonsingular* variety X, assuming that X intersects Q transversally, i.e., under the assumption that $Q \cap X$ is a nonsingular hypersurface of X. This number is a certain combination of the degrees of the components of the *Chern class* of X (see (3-1)); we call this number the "generic Euclidean distance degree" of X, $\mathrm{gEDdeg}(X)$, since it equals the Euclidean distance degree of a general translate of X.

There are several directions in which this formula could be generalized. For example, the hypothesis of nonsingularity on X could be relaxed; it is then understood that the role of the Chern class of X is taken by the so-called *Chern–Mather* class of X, one of several generalizations of the notion of Chern class to possibly singular X. The resulting formula (see, e.g., [Aluffi 2018, Proposition 2.9]) gives the generic ED degree of an arbitrarily singular variety X. In a different direction, one could maintain the nonsingularity hypothesis, but attempt to dispose of any requirement regarding the relative position of Q and X, and aim at computing the "actual" ED degree of X.

MSC2010: primary 14C17; secondary 14N10, 57R20.
Keywords: algebraic optimization, intersection theory, characteristic classes, Chern–Schwartz–MacPherson classes.

The main result of this note is of this second type. It consists of formulas for the Euclidean distance degree of an *arbitrary nonsingular subvariety* of projective space; different versions are presented, in terms of different types of information that may be available on X. The simplest form of the result is the following:

Theorem 8.1. *Let X be a smooth subvariety of $\mathbb{P}^{n-1}$, and assume $X \not\subseteq Q$. Then*

$$\mathrm{EDdeg}(X) = (-1)^{\dim X} \chi(X \smallsetminus (Q \cup H)), \tag{1-1}$$

where H is a general hyperplane.

Here, χ is the ordinary topological Euler characteristic. This statement will be proven in Section 8; in each of Sections 5–7 we obtain an equivalent formulation of the same result. These serve as stepping stones in the proof of (1-1), and seem of independent interest. Theorems 5.1 and 6.1 will express $\mathrm{EDdeg}(X)$ as a "correction" $\gamma(X)$ of the generic Euclidean distance degree due to the singularities of $Q \cap X$. For example, it will be a consequence of Theorem 6.1 that when $Q \cap X$ has *isolated* singularities, then this correction equals the sum of the Milnor numbers of the singularities. (If X is a smooth hypersurface of degree $\neq 2$, the singularities of $Q \cap X$ are necessarily isolated; see Section 9.3.) Theorem 5.1 expresses $\gamma(X)$ in terms of the Segre class of the singularity subscheme of $Q \cap X$; this version of the result is especially amenable to effective implementation, using available algorithms for the computation of Segre classes [Harris 2017]. Theorem 7.1 relates the Euclidean distance degree to *Chern–Schwartz–MacPherson* classes, an important notion in the theory of characteristic classes for singular or noncompact varieties. In fact, $\mathrm{EDdeg}(X)$ admits a particularly simple expression, given in (7-3), in terms of the Chern–Schwartz–MacPherson class of the nonsingular, but noncompact, variety $X \smallsetminus Q$. Theorem 8.1, reproduced above, follows from this expression.

The progression of results in Sections 5–8 is preceded by a general formula, Theorem 4.3, giving the correction term $\gamma(X)$ for essentially arbitrary varieties X. Coupled with [Aluffi 2018, Proposition 2.9], this yields a general formula for $\mathrm{EDdeg}(X)$. This master formula is our main tool for the applications to nonsingular varieties obtained in the sections that follow; in principle it could be used in more general situations, but at this stage we do not know how to extract a simple statement such as formula (1-1) from Theorem 4.3 without posing some nonsingularity hypothesis on X.

Refining the techniques used in this paper may yield more general results, but this is likely to be challenging. Ultimately, the reason why we can obtain simple statements such as (1-1) is that Segre classes of singularity subschemes of hypersurfaces *of a nonsingular variety* X are well understood. In general, singularities of X will themselves contribute to the singularity subscheme of $Q \cap X$, even if the intersection of Q and X is (in some suitable sense) "transversal". In fact, for several of our formulas to hold it is only necessary that X be nonsingular along $Q \cap X$ (see Remarks 5.2 and 6.2).

The raw form of our result is a standard application of Fulton–MacPherson intersection theory, modulo one technical difficulty, which we will attempt to explain here. Techniques developed in [Draisma et al. 2016] express the ED degree as the degree of a projection map from a certain correspondence in

$\mathbb{P}^{n-1} \times \mathbb{P}^{n-1}$. Applying Fulton–MacPherson's intersection theory, one obtains a formula for the ED degree involving the Segre class of an associated subscheme Z_u^Δ in the conormal space of X (Theorem 4.3); this formula holds for arbitrary $X \not\subseteq Q$. In the nonsingular case, the formula may be recast in terms of the Segre class *in* X of a scheme supported on the singular locus of $Q \cap X$. A somewhat surprising complication arises here, since this scheme does *not* coincide with the singularity subscheme of $Q \cap X$. However, we can prove (Lemma 5.4) that the ideal sheaves of the two subschemes have the same *integral closure,* and deduce from this that their Segre classes coincide. This is key to our explicit formulas.

This technical difficulty is likely one of the main obstacles in extending the results of this paper to the case of more general subvarieties of projective space, by analogous techniques. We may venture the guess that a different approach, aiming at "understanding" (1-1) more directly, without reference to the theory of characteristic classes of singular varieties, may be more amenable to generalization. Finding such an approach would appear to be a natural project.

Preliminaries on the Euclidean distance degree are given in Section 2. In Section 3 we point out that (1-1), in its equivalent formulation (8-2), agrees with gEDdeg(X) when X is nonsingular and meets the isotropic quadric transversally. We find that this observation clarifies why a formula such as (1-1) may be expected to hold without transversality hypotheses. It is perhaps natural to conjecture that an analogue replacing ordinary Euler characteristics in (1-1) with the degree χ_{Ma} of the Chern–Mather class may hold for arbitrary varieties. Under the transversality hypothesis, an analogue of (8-2) does hold for possibly singular varieties, as we show in Proposition 3.1. The main body of the paper consists of Sections 4–8 Examples of applications of the results obtained here are given in Section 9.

2. Preliminaries on the Euclidean distance degree

As recalled in the introduction, the Euclidean distance degree of a variety in $\mathbb{R}^n$ is the number of critical nonsingular points of the (squared) distance function from a general point outside of the variety. We consider the complex projective version of this notion: for a subvariety $X \subseteq \mathbb{P}^{n-1} := \mathbb{P}(\mathbb{C}^n)$, we let EDdeg$(X)$ be the number of critical points of the (complex) function

$$(x_1 - u_1)^2 + \cdots + (x_n - u_n)^2 \tag{2-1}$$

which occur at nonsingular points of the cone over X, where $(u_1, \dots, u_n)$ is a general point.

Remark 2.1. If X is a subset of the isotropic quadric Q (with equation $x_1^2 + \cdots + x_n^2 = 0$), then the quadratic term in (2-1) vanishes, and (2-1) has *no* critical points. Therefore, EDdeg$(X) = 0$ in this case, and we can adopt the blanket convention that $X \not\subseteq Q$. With suitable positions, our results will hold without this assumption (see, e.g., Remark 5.8).

The definition of EDdeg(X) may be interpreted in terms of a *projective ED correspondence,* and this will be needed for our results. Our reference here is [Draisma et al. 2016, §5] (but we use slightly different notation). Consider the projective space $\mathbb{P}^{n-1}$ and its dual $\check{\mathbb{P}}^{n-1}$, parametrizing hyperplanes in $\mathbb{P}^{n-1}$. It is well-known that the projective cotangent space $\mathbb{T}^*\mathbb{P}^n := \mathbb{P}(T^*\mathbb{P}^{n-1})$ of $\mathbb{P}^{n-1}$ may be realized as the incidence correspondence $I \subseteq \mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1}$ consisting of pairs (p, H) with $p \in H$. Every subvariety $X \subsetneq \mathbb{P}^{n-1}$ has a (*projective*) *conormal space* $\mathbb{T}^*_X\mathbb{P}^{n-1}$, defined as the closure of the projective conormal variety to X; this may be realized as

$$\mathbb{T}^*_X\mathbb{P}^{n-1} := \overline{\{(p, H) \mid p \in X^{ns} \text{ and } T_pX \subseteq H\}} \subseteq I = \mathbb{T}^*\mathbb{P}^{n-1}.$$

Consider the subvariety $Z \subseteq \mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1} \times \mathbb{C}^n$ obtained as the image of

$$\{(x, y, u) \in (\mathbb{C}^n \smallsetminus \{0\})^2 \times \mathbb{C}^n \mid u = x + y\};$$

that is,

$$Z = \{([x], [y], u) \in \mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1} \times \mathbb{C}^n \mid \dim\langle x, y, u\rangle \leq 2\} \tag{2-2}$$

consists of points $([x], [y], u)$ such that $[x], [y], [u]$ are collinear. The (*projective joint*) *ED correspondence* $\mathcal{PE}$ (denoted $\mathcal{PE}_{X,Y}$ in [Draisma et al. 2016]) is the component of $(\mathbb{T}^*_X\mathbb{P}^{n-1} \times \mathbb{C}^n) \cap Z$ dominating $\mathbb{T}^*_X\mathbb{P}^{n-1}$. Thus, the fiber of $\mathcal{PE}$ over $([x], [y]) \in \mathbb{T}^*_X\mathbb{P}^{n-1}$ consists, for $[x] \neq [y]$, of the vectors $u \in \mathbb{C}^n$ in the span of x and y; this confirms that $\mathcal{PE}$ is irreducible and has dimension n. (Since $X \not\subseteq Q$ by our blanket assumption, there exist points $([x], [y]) \in \mathbb{T}^*_X\mathbb{P}^{n-1}$ with $[x] \neq [y]$.) The projection $\mathcal{PE} \to \mathbb{C}^n$ is in fact dominant, and we have the following result.

Lemma 2.2. *The Euclidean distance degree* EDdeg(X) *equals the degree of the projection* $\mathcal{PE} \to \mathbb{C}^n$.

Proof. This is implied by the argument in the proof of [Draisma et al. 2016, Theorem 5.4]. □

Lemma 2.2 suggests that one should be able to express EDdeg(X) in terms of an intersection with the fiber Z_u of Z over a general point $u \in \mathbb{C}^n$. We may view Z_u as a subvariety of $\mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1}$:

$$Z_u = \{([x], [y]) \in \mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1} \mid \dim\langle x, y, u\rangle \leq 2\},$$

where u is now fixed (and general). It is easy to verify that Z_u is an n-dimensional irreducible variety, and that

$$[Z_u] = h^{n-2} + h^{n-3}\check{h} + \cdots + \check{h}^{n-2} \tag{2-3}$$

in the Chow group $A_*(\mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1})$. Here, h, resp. $\check{h}$, denote the pull-back of the hyperplane class from $\mathbb{P}^{n-1}$, resp. $\check{\mathbb{P}}^{n-1}$. (For example, one may verify (2-3) by intersecting Z_u with suitably chosen $\mathbb{P}^i \times \check{\mathbb{P}}^j$ within $\mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1}$.) This implies the following statement; see [Draisma et al. 2016, Theorem 5.4].

Lemma 2.3. *For all* $u \in \mathbb{C}^{n-1}$ *and all subvarieties* $X \subseteq \mathbb{P}^{n-1}$,

$$Z_u \cdot \mathbb{T}^*_X\mathbb{P}^{n-1} = \sum_{i=0}^{n-2} \delta_i(X),$$

where the numbers $\delta_i(X)$ *are the* polar degrees *of* X.

Indeed, the polar degrees are (by definition) the coefficients of the monomials $h^{n-1-i}\check{h}^{i+1}$ in the class $[\mathbb{T}^*_X\mathbb{P}^{n-1}]$.

In view of Lemma 2.3, we define

$$\mathrm{gEDdeg}(X) := \sum_{i=0}^{n-2} \delta_i(X),$$

the "generic Euclidean distance degree" of X. By Lemma 2.2, $\mathrm{EDdeg}(X)$ is the contribution of the projective ED correspondence to the intersection number $\mathrm{gEDdeg}(X) = Z_u \cdot \mathbb{T}^*_X\mathbb{P}^{n-1}$ calculated in Lemma 2.3. It is a consequence of [Draisma et al. 2016, Theorem 5.4] that if X is in sufficiently general position, then this contribution in fact equals the whole intersection number, i.e., $\mathrm{EDdeg}(X) = \mathrm{gEDdeg}(X)$. Our goal is to determine a precise "correction term" evaluating the discrepancy between $\mathrm{EDdeg}(X)$ and $\mathrm{gEDdeg}(X)$ without any *a priori* hypothesis on X. In Section 4 we will formalize this goal and deduce a general formula for $\mathrm{EDdeg}(X)$ for an arbitrary variety X. In Sections 4–8 we will use this result to obtain more explicit formulas for $\mathrm{EDdeg}(X)$ under the assumption that X is nonsingular.

3. The generic Euclidean distance degree, revisited

This section is not used in the sections that follow, but should help motivating formula (1-1), which will be proven in Section 8. We also propose a possible conjectural generalization of this formula to arbitrary projective varieties.

We have defined the "generic" Euclidean distance degree of a subvariety $X \subseteq \mathbb{P}^{n-1}$ as the sum of its polar degrees. In [Draisma et al. 2016, Theorem 5.8] it is shown that if X is nonsingular, then

$$\mathrm{gEDdeg}(X) = \sum_{j=0}^{\dim X} (-1)^{\dim X + j} c(X)_j (2^{j+1} - 1); \tag{3-1}$$

this number may be interpreted as the Euclidean distance degree of a general translation of X, which will meet Q transversally by Bertini's theorem. Here $c(X)_j$ is the degree of the component of dimension j of the Chern class $c(TX) \cap [X]$ of X. Formula (3-1) holds for arbitrarily singular varieties X if one replaces $c(X)$ with the *Chern–Mather* class $c_{\mathrm{Ma}}(X)$ of X [Aluffi 2018, Proposition 2.9].

Assume first that X is nonsingular. As a preliminary observation, the reader is invited to perform the following calculus exercise:

For $0 \le j \le N$, the coefficient of t^N in the expansion of

$$\frac{t^{N-j}}{(1+t)(1+2t)}$$

is $(-1)^j(2^{j+1} - 1)$.

With this understood, we have the following computation:

$$\mathrm{gEDdeg}(X) = (-1)^{\dim X} \sum_{j=0}^{\dim X} c(X)_j (-1)^j (2^{j+1} - 1)$$

$$\begin{aligned}
&= (-1)^{\dim X} \sum_{j=0}^{\dim X} c(X)_j \int \frac{h^{\dim X - j}}{(1+h)(1+2h)} \cdot h^{\operatorname{codim} X} \cap [\mathbb{P}^{n-1}] \\
&= (-1)^{\dim X} \int \frac{1}{(1+h)(1+2h)} \cdot c(TX) \cap [X] \\
&= (-1)^{\dim X} \int \Bigl(1 - \frac{h}{1+h} - \frac{2h}{1+2h} + \frac{h \cdot 2h}{(1+h)(1+2h)}\Bigr) \cdot c(TX) \cap [X].
\end{aligned}$$

Assuming further that X is transversal to Q and that H is a general hyperplane, the last expression may be rewritten as

$$(-1)^{\dim X} \Bigl(\int c(TX) \cap [X] - \int c(T(X \cap H)) \cap [X \cap H] \\ - \int c(T(X \cap Q)) \cap [X \cap Q] + \int c(T(X \cap Q \cap H)) \cap [X \cap Q \cap H] \Bigr)$$

(by transversality, all of the loci appearing in this expression are nonsingular). The degree of the zero-dimensional component of the Chern class of a compact complex nonsingular variety is its topological Euler characteristic, so this computation shows

$$\operatorname{EDdeg}(X) = (-1)^{\dim X} (\chi(X) - \chi(X \cap H) - \chi(X \cap Q) + \chi(X \cap Q \cap H)), \tag{3-2}$$

if X is nonsingular and meets Q transversally (and where H is a general hyperplane).

Theorem 8.1 will amount to the assertion that (3-2) holds as soon as X is nonsingular, without any hypothesis on the intersection of Q and X. By the good inclusion-exclusion properties of the Euler characteristic, (3-2) is equivalent to (1-1).

While the computation deriving (3-2) from [Draisma et al. 2016, Theorem 5.8] is trivial under the transversality hypothesis, we do not know of any simple way to obtain this formula in the general case. The next several Sections (4–8) will lead to a proof of (3-2) for arbitrary nonsingular varieties.

The above computation can be extended to singular projective subvarieties. Just as the topological Euler characteristic of a nonsingular variety is the degree of its top Chern class, we can define an "Euler–Mather characteristic" of a possibly singular variety V by setting

$$\chi_{\mathrm{Ma}}(V) := \int c_{\mathrm{Ma}}(V),$$

the degree of the Chern–Mather class of V. This number is a linear combination of Euler characteristics of strata of V, with coefficients determined by the *local Euler obstruction* Eu, a well-studied numerical invariant of singularities.

Proposition 3.1. *For* any *subvariety $X \subseteq \mathbb{P}^{n-1}$ intersecting Q transversally*,

$$\operatorname{EDdeg}(X) = (-1)^{\dim X} \bigl(\chi_{Ma}(X) - \chi_{Ma}(X \cap Q) - \chi_{Ma}(X \cap H) + \chi_{Ma}(X \cap Q \cap H)\bigr), \tag{3-3}$$

where H is a general hyperplane.

Proof. Argue precisely as in the discussion leading to (3-2), using [Aluffi 2018, Proposition 2.9] in place of [Draisma et al. 2016, Theorem 5.8]. The only additional ingredient needed for the computation is the fact that if W is a nonsingular hypersurface intersecting a variety V transversally, then

$$c_{\mathrm{Ma}}(W \cap V) = \frac{W}{1+W} \cap c_{\mathrm{Ma}}(V).$$

For a much stronger result, and a discussion of the precise meaning of “transversality”, we address the reader to [Schürmann 2017], particularly Theorem 1.2. □

Of course (3-3) specializes to (3-2) when X is nonsingular, under the transversality hypothesis; but it does not do so in general, because $Q \cap X$ may be singular even if X is nonsingular, and $\chi_{\mathrm{Ma}}(Q \cap X)$ does not necessarily agree with $\chi(Q \cap X)$ in that case. Therefore, the transversality hypothesis in Proposition 3.1 is necessary. The signed Euler–Mather characteristic of the complement,

$$(-1)^{\dim X}\,\chi_{\mathrm{Ma}}(X \smallsetminus (Q \cup H)) = (-1)^{\dim X} \int c_*(\mathrm{Eu}_{X \smallsetminus (Q \cup H)})$$

(where c_* denotes MacPherson’s natural transformation) may be the most natural candidate as an expression for $\mathrm{EDdeg}(X)$ for arbitrary subvarieties $X \subseteq \mathbb{P}^{n-1}$, without smoothness or transversality hypotheses.

4. Intersection formula

In Section 2 we defined the projective ED correspondence $\mathcal{PE}$ to be one component of the intersection $(\mathbb{T}^*_X\mathbb{P}^{n-1} \times \mathbb{C}^n) \cap Z$, where Z is the variety of linearly dependent triples defined in (2-2). We next determine the union of the *other* irreducible components. We denote by Δ the diagonal in $\mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1}$. In this section, $X \subseteq \mathbb{P}^{n-1}$ is a subvariety (not necessarily smooth), and $X \not\subseteq Q$ (see Remark 2.1).

Lemma 4.1. *There exists a subscheme $Z^\Delta \subset \mathbb{T}^*_X\mathbb{P}^{n-1} \times \mathbb{C}^n$ such that*

$$(\mathbb{T}^*_X\mathbb{P}^{n-1} \times \mathbb{C}^n) \cap Z = \mathcal{PE} \cup Z^\Delta,$$

where the support of Z^Δ equals the support of $(\Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1}) \times \mathbb{C}^n$.

Proof. Consider the projection $(\mathbb{T}^*_X\mathbb{P}^{n-1} \times \mathbb{C}^n) \cap Z \to \mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1}$. We have already observed in Section 2 that the fiber over $([x],[y]) \in \mathbb{T}^*_X\mathbb{P}^{n-1}$, $([x],[y]) \notin \Delta$, consists of the span $\langle x, y\rangle$ in $\mathbb{C}^n$; it follows that $\mathcal{PE}$ is the only component of the intersection dominating $\mathbb{T}^*_X\mathbb{P}^{n-1}$. (Again note that since $X \not\subseteq Q$, there are points $([x],[y]) \in \mathbb{T}^*_X$, $([x],[y]) \notin \Delta$.)

We claim that if $([x],[x]) \in \mathbb{T}^*_X\mathbb{P}^{n-1}$, then the fiber of $(\mathbb{T}^*_X\mathbb{P}^{n-1} \times \mathbb{C}^n) \cap Z$ over $([x],[x])$ consists of the whole space $\mathbb{C}^{n-1}$; the statement follows immediately from this assertion.

Trivially, $([x],[x],u) \in \mathbb{T}^*_X\mathbb{P}^{n-1} \times \mathbb{C}^n$ for all u, so we simply need to verify that $([x],[x],u) \in Z$ for all $u \in \mathbb{C}^n$. But this is clear, since there are points $([x'],[y'],u)$ with $u \in \langle x', y'\rangle$ and $([x'],[y'])$ arbitrarily close to $([x],[x])$. □

The fact that Z contains $\Delta \times \mathbb{C}^n$ (used in the proof) may also be verified by observing that equations for Z in $\mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1} \times \mathbb{C}^n$ are given by the 3×3 minors of the matrix

$$\begin{pmatrix} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \\ u_1 & u_2 & \dots & u_n \end{pmatrix} \tag{4-1}$$

associated with a point $([x], [y], u) \in \mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1} \times \mathbb{C}^n$; the diagonal $\Delta \times \mathbb{C}^n$ obviously satisfies these equations.

Now we fix a general $u \in \mathbb{C}^n$. By Lemma 2.2, the fiber $\mathcal{PE}_u$ consists of $\mathrm{EDdeg}(X)$ simple points, which will be disjoint from the diagonal for general u. On the other hand, for all u, the fiber Z_u (when viewed as a subvariety of $\mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1}$) contains Δ. We deduce the following consequence of Lemma 4.1.

Corollary 4.2. *For a general $u \in \mathbb{C}^n$,*

$$Z_u \cap \mathbb{T}^*_X \mathbb{P}^{n-1} = \{\mathrm{EDdeg}(X) \textit{ simple points}\} \sqcup Z_u^\Delta$$

(*as subschemes of* $\mathbb{T}^*_X \mathbb{P}^{n-1}$), *where the support of Z_u^Δ agrees with the support of* $\Delta \cap \mathbb{T}^*_X \mathbb{P}^{n-1}$.

Taking into account Lemma 2.3 we obtain that

$$\mathrm{EDdeg}(X) = \mathrm{gEDdeg}(X) - \gamma(X), \tag{4-2}$$

where $\gamma(X)$ is the contribution of Z_u^Δ to the intersection product $Z_u \cdot \mathbb{T}^*_X \mathbb{P}^{n-1}$. This “correction term” $\gamma(X)$ does not depend on the chosen (general) u, and vanishes if $\Delta \cap \mathbb{T}^*_X \mathbb{P}^{n-1} = \varnothing$, since in this case $Z_u^\Delta = \varnothing$ by Corollary 4.2. This special case recovers the statement of [Draisma et al. 2016, Theorem 5.4], and indeed (4-2) is essentially implicit in [loc. cit.]. We are interested in computable expressions for the correction term $\gamma(X)$.

We will first obtain the following master formula, through a direct application of Fulton–MacPherson intersection theory. The diagonal Δ is isomorphic to $\mathbb{P}^{n-1}$, and we denote by H its hyperplane class, as well as its restrictions. (Thus, H agrees with the restriction of both h and $\check{h}$.)

Theorem 4.3. *With notation as above,*

$$\gamma(X) = \int (1+H)^{n-1} \cap s(Z_u^\Delta, \mathbb{T}^*_X \mathbb{P}^{n-1}) \tag{4-3}$$

for u general in $\mathbb{C}^n$.

Here, $\int$ denotes degree, and $s(-,-)$ is the *Segre class,* in the sense of [Fulton 1984, Chapter 4]. Segre classes are effectively computable by available implementations of algorithms (see, e.g., [Harris 2017]). However, the need to obtain explicit equations for the scheme Z_u^Δ, and conditions guaranteeing that a given u is general enough, limit the direct applicability of Theorem 4.3. Our task in the next several sections of this paper will be to obtain from (4-3) concrete computational tools, at the price of requiring X to be of a more specific type — we will assume in the following sections that $X \not\subseteq Q$ is *nonsingular,* but otherwise arbitrary.

The proof of Theorem 4.3 requires some additional information on Z_u, which we gather next. As noted in Section 2, Z_u is an irreducible n-dimensional subvariety of $\mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1}$. Equations for Z_u in $\mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1}$ are given by the 3×3 minors of the matrix (4-1), where now $u = (u_1, \dots, u_n)$ is fixed. The diagonal Δ is a divisor in Z_u.

Lemma 4.4. *The subvariety Z_u of $\mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1}$ is smooth at all points $([x], [y]) \neq ([u], [u])$.*

Proof. Given $(x, y, u) \in \mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1} \times \mathbb{C}^n$ we can change coordinates so that $u = (1, 0, \dots, 0)$. If either $[x]$ or $[y]$ is not $[u]$, we may without loss of generality assume that $x_n \neq 0$, and hence $x_n = 1$. The ideal of Z_u at this point $([x], [y])$ is generated by the 3×3 minors of

$$\begin{pmatrix} x_1 & x_2 & \dots & x_{n-1} & 1 \\ y_1 & y_2 & \dots & y_{n-1} & y_n \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix}$$

and among these we find the $n - 2$ minors

$$y_i - x_i y_n, \quad i = 2, \dots, n-1.$$

Near $([x], [y])$, these generate the ideal of an irreducible smooth complete intersection of dimension $n = \dim Z_u$, which must then coincide with Z_u in a neighborhood of $([x], [y])$, giving the statement. □

Denote complements of $\{([u], [u])\}$ by $^\circ$. Thus $Z_u^\circ = Z_u \smallsetminus \{([u], [u])\}$, $\Delta^\circ = \Delta \smallsetminus \{([u], [u])\}$, etc. By Lemma 4.4, Z_u° is a local complete intersection in $(\mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1})^\circ$, and we let N be its normal bundle.

Lemma 4.5. *With notation as above, $c(N)|_{\Delta^\circ} = (1 + H)^{n-1}$.*

Proof. Consider the rational map

$$\pi : \mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1} \to \mathbb{P}^{n-2} \times \mathbb{P}^{n-2}$$

defined by the linear projection from $[u]$ on each factor. Let $U \subseteq \mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1}$ be the complement of the union of $\{[u]\} \times \mathbb{P}^{n-1}$ and $\check{\mathbb{P}}^{n-1} \times \{[u]\}$; thus, $\pi|_U : U \to \mathbb{P}^{n-2} \times \mathbb{P}^{n-2}$ is a regular map, and U contains Δ°. A simple coordinate computation shows that $Z_u \cap U = \pi|_U^{-1}(\Delta')$, where Δ' is the diagonal in $\mathbb{P}^{n-2} \times \mathbb{P}^{n-2}$. It follows that

$$N|_{Z_u \cap U} = \pi|_U^*(N_{\Delta'}\mathbb{P}^{n-2} \times \mathbb{P}^{n-2}) \cong \pi|_U^*(T\Delta').$$

Since $\Delta' \cong \mathbb{P}^{n-2}$, $c(T\Delta') = (1 + H')^{n-1}$, where H' is the hyperplane class. The statement follows by observing that the pull-back of H' to Δ° agrees with the pull-back of H. This is the case, since the restriction $\pi|_{\Delta^\circ} : \Delta^\circ \cong \mathbb{P}^{n-1} \smallsetminus \{u\} \to \Delta' \cong \mathbb{P}^{n-2}$ is a linear projection. □

With these preliminaries out of the way, we can prove Theorem 4.3.

Proof of Theorem 4.3. Since $[u]$ is general, it may be assumed not to be a point of X. This ensures that $([u], [u]) \notin \mathbb{T}_X^*\mathbb{P}^{n-1}$; in particular

$$Z_u^\circ \cap \mathbb{T}_X^*\mathbb{P}^{n-1} = Z_u \cap \mathbb{T}_X^*\mathbb{P}^{n-1}.$$

It follows that, as a class in $A_*(Z_u \cap \mathbb{T}^*_X\mathbb{P}^{n-1})$, the (Fulton–MacPherson) intersection product of $\mathbb{T}^*_X\mathbb{P}^{n-1}$ by Z_u on $\mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1}$ equals the intersection product of $\mathbb{T}^*_X\mathbb{P}^{n-1}$ by Z_u° on $(\mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1})^\circ$.

Therefore, we can view $\gamma(X)$ as the contribution of Z_u^Δ to $Z_u^\circ \cdot \mathbb{T}^*_X\mathbb{P}^{n-1}$. Consider the fiber diagram

$$\begin{tikzcd}
Z_u \cap T^*_X\mathbb{P}^{n-1} \arrow[r] \arrow[d, "g"'] & T^*_X\mathbb{P}^{n-1} \arrow[d] \\
Z_u^\circ \arrow[r] & (\mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1})^\circ
\end{tikzcd}$$

By [Fulton 1984, §6.1] (especially Proposition 6.1(a), Example 6.1.1), this contribution equals

$$\int c(g|^*_{Z_u^\Delta} N) \cap s(Z_u^\Delta, \mathbb{T}^*_X\mathbb{P}^{n-1}),$$

where $N = N_{Z_u^\circ}(\mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1})^\circ$ as above. Since Z_u^Δ is supported on a subscheme of Δ°, $c(g|^*_{Z_u^\Delta} N)$ equals (the restriction of) $(1+H)^{n-1}$ by Lemma 4.5. The stated formula follows. □

Summarizing, we have proven that

$$\operatorname{EDdeg}(X) = \operatorname{gEDdeg}(X) - \int (1+H)^{n-1} \cap s(Z_u^\Delta, \mathbb{T}^*_X\mathbb{P}^{n-1}) \tag{4-4}$$

for all subvarieties $X \not\subseteq Q$ of $\mathbb{P}^{n-1}$. (If $X \subseteq Q$, then $\operatorname{EDdeg}(X) = 0$; see Remark 2.1.) The quantity $\operatorname{gEDdeg}(X)$ is invariant under projective translations, and may be computed in terms of the Chern–Mather class of X. The other term records subtle information concerning the intersection of X and Q, by means of the Segre class $s(Z_u^\Delta, \mathbb{T}^*_X\mathbb{P}^{n-1})$. We will focus on obtaining alternative expressions for this class.

5. Euclidean distance degree and Segre classes

Now we assume that $X \subseteq \mathbb{P}^{n-1}$ is a *smooth* closed subvariety. As recalled in Section 3, in this case $\operatorname{gEDdeg}(X)$ is given by a certain combination of the Chern classes of X:

$$\operatorname{gEDdeg}(X) = (-1)^{\dim X} \sum_{j=0}^{\dim X} (-1)^j c(X)_j (2^{j+1} - 1).$$

An application of Theorem 4.3, obtained in this section, will yield an explicit formula for the correction term $\gamma(X)$ (and hence for $\operatorname{EDdeg}(X)$). This result has two advantages over Theorem 4.3: first, the formula will not depend on the choice of a general u; second, its ingredients will allow us to draw a connection with established results in the theory of characteristic classes for singular varieties, leading to the results presented in Sections 6–8.

The main result of this section is the following. Recall that we are denoting by Q the isotropic quadric, i.e., the hypersurface of $\mathbb{P}^{n-1}$ with equation $\sum_{i=1}^n x_i^2 = 0$. By our blanket assumption that X should not be contained in Q, we have that $Q \cap X$ is a (possibly singular) hypersurface of X. We let $J(Q \cap X)$ denote its *singularity subscheme,* defined locally by the partial derivatives of its equation in X (or equivalently by the appropriate Fitting ideal of the sheaf of differentials of $Q \cap X$). Also, recall that h denotes the hyperplane class in $\mathbb{P}^{n-1}$.

Theorem 5.1. *Let X be a smooth subvariety of $\mathbb{P}^{n-1}$, and assume $X \not\subseteq Q$. Then*

$$\mathrm{EDdeg}(X) = \mathrm{gEDdeg}(X) - \int \frac{(1+2h)\cdot c(T^*X \otimes \mathscr{O}(2h))}{1+h} \cap s(J(Q \cap X), X). \tag{5-1}$$

The key ingredient in (5-1) is the Segre class $s(J(Q \cap X), X)$. This may be effectively computed by using the algorithm for Segre classes described in [Harris 2017].

Remark 5.2. It will be clear from the argument that it is only necessary to require X to be nonsingular in a neighborhood of $Q \cap X$. (Of course X must only have isolated singularities in this case.) Formula (5-1) holds as stated in this more general case; gEDdeg(X) may be computed using the same formula as in the smooth case (that is, (3-1)), but using the degrees of the component of the *Chern–Mather* class of X [Aluffi 2018, Proposition 2.9]. The hypothesis $X \not\subseteq Q$ is also not essential; see Remark 5.8.

The proof of Theorem 5.1 will rely on a more careful study of the schemes $\Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1}$ and Z^Δ_u encountered in Section 4. In Corollary 4.2 we have shown that these two schemes have the same *support;* here we will prove the much stronger statement that they have the same *Segre class* in $\mathbb{T}^*_X\mathbb{P}^{n-1}$. Since $\Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1}$ is closely related with $J(Q \cap X)$ (Lemma 5.3), this will allow us to recast Theorem 4.3 in terms of the Segre class appearing in (5-1), by means of a result of W. Fulton.

Recall that $\Delta \subseteq Z_u$ (in fact, Δ is a divisor in Z_u); it follows that

$$\Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1} \subseteq Z^\Delta_u.$$

These two schemes have the same support (Corollary 4.2); but they are in general different. It is straightforward to identify $\Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1}$ with a subscheme of X.

Lemma 5.3. *Let $\delta : \mathbb{P}^{n-1} \to \mathbb{P}^{n-1} \times \mathbb{P}^{n-1}$ be the diagonal embedding, and let X be a smooth subvariety of $\mathbb{P}^{n-1}$. Then $J(Q \cap X) = \delta^{-1}(\mathbb{T}^*_X\mathbb{P}^{n-1})$, i.e., δ maps $J(Q \cap X)$ isomorphically to $\Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1}$.*

Proof. Since $\mathbb{T}^*_X\mathbb{P}^{n-1} \subseteq \mathbb{P}(T^*\mathbb{P}^{n-1})$, we have

$$\Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1} = \Delta \cap \mathbb{P}(T^*\mathbb{P}^{n-1}) \cap \mathbb{T}^*_X\mathbb{P}^{n-1} = \mathbb{T}^*_Q\mathbb{P}^{n-1} \cap \mathbb{T}^*_X\mathbb{P}^{n-1}. \tag{5-2}$$

The diagonal δ restricts to an isomorphism $q : Q \xrightarrow{\sim} \mathbb{T}^*_Q\mathbb{P}^{n-1}$. By (5-2), we have that $\delta^{-1}(\mathbb{T}^*_X\mathbb{P}^{n-1})$ agrees with $q^{-1}(\mathbb{T}^*_X\mathbb{P}^{n-1})$, viewed as a subscheme of $\mathbb{P}^{n-1}$.

Now $q^{-1}(\mathbb{T}^*_X\mathbb{P}^{n-1})$ consists of points $[x]$ such that $[x] \in Q \cap X$ and $T_{[x]}Q \supseteq T_{[x]}X$, and these conditions define $J(Q \cap X)$ scheme-theoretically. The statement follows. □

Determining Z^Δ_u requires more work. We may assume without loss of generality that $u = (1, 0, \dots, 0)$, so that equations for Z^Δ_u are given by the 3×3 minors of

$$\begin{pmatrix} x_1 & x_2 & \dots & x_{n-1} & x_n \\ y_1 & y_2 & \dots & y_{n-1} & y_n \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix}$$

(defining Z_u) as well as the requirement that $([x],[y]) \in \mathbb{T}^*_X\mathbb{P}^{n-1}$. It is in fact useful to keep in mind that, for $([x],[y]) \in \mathbb{T}^*_X\mathbb{P}^{n-1}$, $\Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1}$ is defined by the 2×2 minors of

$$\begin{pmatrix} x_1 & x_2 & \dots & x_{n-1} & x_n \\ y_1 & y_2 & \dots & y_{n-1} & y_n \end{pmatrix}$$

while Z_u^Δ is defined (near the diagonal) by the 2×2 minors of

$$\begin{pmatrix} x_2 & \dots & x_{n-1} & x_n \\ y_2 & \dots & y_{n-1} & y_n \end{pmatrix}.$$

Let $\mathscr{I}_{\Delta\cap\mathbb{T}^*_X\mathbb{P}^{n-1}} \supseteq \mathscr{I}_{Z_u^\Delta}$ be the corresponding ideal sheaves on $\mathbb{T}^*_X\mathbb{P}^{n-1}$.

Lemma 5.4. *The ideal* $\mathscr{I}_{\Delta\cap\mathbb{T}^*_X\mathbb{P}^{n-1}}$ *is integral over* $\mathscr{I}_{Z_u^\Delta}$. *Therefore,*

$$s(Z_u^\Delta, \mathbb{T}^*_X\mathbb{P}^{n-1}) = s(\Delta\cap\mathbb{T}^*_X\mathbb{P}^{n-1}, \mathbb{T}^*_X\mathbb{P}^{n-1}). \tag{5-3}$$

Proof. The second assertion follows from the first; see the proof of [Aluffi 1995, Lemma 1.2]. The first assertion may be verified on local analytic charts, so we obtain an analytic description of $\mathbb{T}^*_X\mathbb{P}^{n-1}$ at a point $([x],[x])$ of the diagonal. Again without loss of generality we may let $[x] = (1:0:\cdots:0:i) \in Q\cap X$, and assume that the embedding $\iota : X \to \mathbb{P}^{n-1}$ has the following analytic description near this point:

$$\iota : (\underline{s}) = (s_2,\dots,s_d) \mapsto (1 : s_2 : \cdots : s_d : \varphi_{d+1}(\underline{s}) : \cdots : \varphi_n(\underline{s})).$$

Here $\underline{s}$ are analytic coordinates for X, centered at 0, and $\varphi_j(0) = 0$ for $j = d+1,\dots,n-1$, $\varphi_n(0) = i$. The tangent space to X at $(\underline{s})$ is cut out by the $n-d$ hyperplanes

$$\varphi_{j2}x_2 + \cdots + \varphi_{jd}x_d - x_j = \Phi_j x_1, \qquad j = d+1,\dots,n, \tag{5-4}$$

where $\varphi_{jk} = \partial\varphi_j/\partial s_k$ and

$$\Phi_j = \varphi_{j2}s_2 + \cdots + \varphi_{jd}s_d - \varphi_j.$$

The hyperplanes (5-4) span the fiber of $\mathbb{T}^*_X\mathbb{P}^{n-1}$ over the point $\iota(\underline{s})$. Therefore, $\Delta\cap\mathbb{T}^*_X\mathbb{P}^{n-1}$ is cut out by the 2×2 minors of the matrix

$$\begin{pmatrix} 1 & s_2 & \dots & s_d & \varphi_{d+1} & \dots & \varphi_n \\ \sum_j \lambda_j\Phi_j & -\sum_j \lambda_j\varphi_{j2} & \dots & -\sum_j \lambda_j\varphi_{jd} & \lambda_{d+1} & \dots & \lambda_n \end{pmatrix} \tag{5-5}$$

where $\lambda_{d+1},\dots,\lambda_n$ are homogeneous coordinates in the fibers of $\mathbb{T}^*_X\mathbb{P}^{n-1}$, while Z_u^Δ is cut out by the 2×2 minors of

$$\begin{pmatrix} s_2 & \dots & s_d & \varphi_{d+1} & \dots & \varphi_n \\ -\sum_j \lambda_j\varphi_{j2} & \dots & -\sum_j \lambda_j\varphi_{jd} & \lambda_{d+1} & \dots & \lambda_n \end{pmatrix}. \tag{5-6}$$

The last several minors in both matrices may be used to eliminate the homogeneous coordinates λ_j, giving $\lambda_j \propto \varphi_j$; in other words, we find that, near $([x],[x])$ both $\Delta\cap\mathbb{T}^*_X\mathbb{P}^{n-1}$ and Z_u^Δ lie in the local analytic section $\sigma : X \to \mathbb{T}^*_X\mathbb{P}^{n-1}$ defined by

$$\sigma(\underline{s}) : (\lambda_{d+1} : \cdots : \lambda_n) = (\varphi_{d+1}(\underline{s}) : \cdots : \varphi_n(\underline{s})).$$

Setting $\lambda_j = \varphi_j$ we obtain from (5-6) generators

$$s_k + \sum_j \varphi_j \varphi_{jk}, \quad k = 2, \dots, d \tag{5-7}$$

for the ideal of Z_u^Δ in $\sigma(X)$; the same generators, together with

$$1 - \sum_{j=d+1}^{n} \varphi_j \Phi_j \tag{5-8}$$

give the ideal of $\Delta \cap \mathbb{T}_X^* \mathbb{P}^{n-1}$ in $\sigma(X)$. It suffices then to verify that (5-8) is integral over the ideal generated by (5-7).

For this, note that the hypersurface $\iota^{-1}(Q \cap X)$ has the equation

$$G(\underline{s}) = 1 + s_2^2 + \cdots + s_d^2 + \varphi_{d+1}^2 + \cdots + \varphi_n^2.$$

Since $\partial G / \partial s_k = 2(s_k + \sum_j \varphi_j \varphi_{jk})$, the ideal generated by (5-7) is nothing but

$$\left(\frac{\partial G}{\partial s_2}, \dots, \frac{\partial G}{\partial s_d} \right). \tag{5-9}$$

On the other hand, (5-8) may be written as

$$\begin{aligned} 1 - \sum_{j=d+1}^{n} \varphi_j \Phi_j &= 1 - \sum_{j=d+1}^{n} \varphi_j (\varphi_{j2} s_2 + \cdots + \varphi_{jd} s_d - \varphi_j) \\ &= 1 - s_2 \Bigl(\sum_j \varphi_j \varphi_{j2} \Bigr) - \cdots - s_d \Bigl(\sum_j \varphi_j \varphi_{jd} \Bigr) + \varphi_{d+1}^2 + \cdots + \varphi_n^2 \\ &\sim 1 + s_2^2 + \cdots + s_d^2 + \varphi_{d+1}^2 + \cdots + \varphi_n^2 = G(\underline{s}) \end{aligned}$$

modulo (5-7). Since G is integral over (5-9) by [Huneke and Swanson 2006, Corollary 7.2.6], this shows that (5-8) is integral over (5-7), as needed. □

Remark 5.5. The argument also shows that the ideal of $\Delta \cap \mathbb{T}_X^* \mathbb{P}^{n-1}$ in $\sigma(X)$ equals $(G, \partial G/\partial s_2, \dots, \partial G/\partial s_d)$, that is, the (local analytic) ideal of $J(Q \cap X)$. This confirms the isomorphism $J(Q \cap X) \cong \Delta \cap \mathbb{T}_X^* \mathbb{P}^{n-1}$ obtained in Lemma 5.3.

Remark 5.6. The smoothness of X is needed in our argument, since it gives us direct access to the conormal space $\mathbb{T}_X^* \mathbb{P}^{n-1}$. However, it is reasonable to expect that (5-3) holds without this assumption, and it would be interesting to establish this equality for more general varieties.

By Theorem 4.3 and Lemma 5.4,

$$\gamma(X) = \int (1 + H)^{n-1} \cap s(\Delta \cap \mathbb{T}_X^* \mathbb{P}^{n-1}, \mathbb{T}_X^* \mathbb{P}^{n-1}) \tag{5-10}$$

if X is nonsingular and not contained in Q. We are now ready to prove Theorem 5.1.

Proof of Theorem 5.1. Our main tools are Lemma 5.3 and a result of W. Fulton. For a closed embedding $V \subseteq M$ of a scheme in a nonsingular variety M, Fulton [1984, Example 4.2.6] proves that the class

$$c_{\mathrm{F}}(V) := c(TM|_V) \cap s(V, M) \tag{5-11}$$

is *independent of M;*. We call $c_{\mathrm{F}}(V)$ the "Chern–Fulton class" of V.

By Lemma 5.3, the diagonal embedding $\delta : \mathbb{P}^{n-1} \to \mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1}$ restricts to an isomorphism $\delta|_{J(Q\cap X)} : J(Q \cap X) \xrightarrow{\sim} \Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1}$. Let $\pi' : \Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1} \to J(Q \cap X)$ be the natural projection, that is, the inverse of $\delta|_{J(Q\cap X)}$. Then

$$c_{\mathrm{F}}(\Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1}) = \pi'^* c_{\mathrm{F}}(J(Q \cap X)) \tag{5-12}$$

by Fulton's result. We proceed to determine this class. The Euler sequence for the projective bundle $\mathbb{T}^*_X\mathbb{P}^{n-1} = \mathbb{P}(T^*_X\mathbb{P}^{n-1}) \xrightarrow{\pi} X$,

$$0 \to \mathbb{O} \to \pi^* T^*_X\mathbb{P}^{n-1} \otimes \mathbb{O}(1) \to T(\mathbb{T}^*_X\mathbb{P}^{n-1}) \to \pi^* TX \to 0,$$

yields

$$c(T(\mathbb{T}^*_X\mathbb{P}^{n-1})) = c(\pi^* T^*_X\mathbb{P}^{n-1} \otimes \mathbb{O}(1)) \cdot \pi^* c(TX).$$

Pulling back and tensoring by $\mathbb{O}(1)$, the cotangent sequence defining the conormal bundle gives the exact sequence

$$0 \to \pi^* T^*_X\mathbb{P}^{n-1} \otimes \mathbb{O}(1) \to \pi^* T^*\mathbb{P}^{n-1} \otimes \mathbb{O}(1) \to \pi^* T^* X \otimes \mathbb{O}(1) \to 0,$$

implying

$$c(T(\mathbb{T}^*_X\mathbb{P}^{n-1})) = \frac{c(\pi^* T^*\mathbb{P}^{n-1} \otimes \mathbb{O}(1)) \cdot \pi^* c(TX)}{c(\pi^* T^* X \otimes \mathbb{O}(1))}.$$

The cotangent bundle $T^*\mathbb{P}^{n-1}$ may be identified with the incidence correspondence in the product $\mathbb{P}^{n-1} \times \check{\mathbb{P}}^{n-1}$, and $\mathbb{O}(1) = \mathbb{O}(h + \check{h})$ under this identification (see, e.g., [Aluffi 2018, §2.2]). Also, $c(T^*\mathbb{P}^{n-1}) = (1-h)^n$. It follows that

$$c(T(\mathbb{T}^*_X\mathbb{P}^{n-1})) = \frac{(1+\check{h})^n \cdot \pi^* c(TX)}{(1+h+\check{h}) \cdot c(\pi^* T^* X \otimes \mathbb{O}(h+\check{h}))}.$$

Now we restrict to the diagonal. As in Section 4, we denote by H the hyperplane class in $\Delta \cong \mathbb{P}^{n-1}$ (and its restrictions); note that $H = h \cdot \Delta = \check{h} \cdot \Delta$. Therefore

$$c(T(\mathbb{T}^*_X\mathbb{P}^{n-1})|_{\Delta\cap\mathbb{T}^*_X\mathbb{P}^{n-1}}) = \frac{(1+H)^n \cdot \pi'^* c(TX)}{(1+2H) \cdot c(\pi'^* T^* X \otimes \mathbb{O}(2H))},$$

where π' denotes the projection $\Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1} \to J(Q \cap X)$ as above (and we are omitting other evident restrictions). Since $H = \pi'^* h$, the Chern–Fulton class of $\Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1}$ must be

$$c_{\mathrm{F}}(\Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1}) = \pi'^* \left(\frac{(1+h)^n \cdot c(TX)}{(1+2h) \cdot c(T^* X \otimes \mathbb{O}(2h))} \right) \cap s(\Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1}, \mathbb{T}^*_X\mathbb{P}^{n-1}).$$

Using (5-12), this shows that

$$\begin{aligned} s(\Delta \cap \mathbb{T}^*_X\mathbb{P}^{n-1}, \mathbb{T}^*_X\mathbb{P}^{n-1}) &= \pi'^*\Big(\frac{(1+2h)\cdot c(T^*X \otimes \mathcal{O}(2h))}{(1+h)^n \cdot c(TX)} \cap c_{\mathrm{F}}(J(Q\cap X))\Big) \\ &= \pi'^*\Big(\frac{(1+2h)\cdot c(T^*X \otimes \mathcal{O}(2h))}{(1+h)^n} \cap s(J(Q\cap X), X)\Big). \end{aligned}$$

Since $\pi'^* = (\delta|_{Q\cap X})_*$ preserves degrees, and $H = \pi'^*(h)$, (5-10) gives

$$\gamma(X) = \int \frac{(1+2h)\cdot c(T^*X \otimes \mathcal{O}(2h))}{1+h} \cap s(J(Q\cap X), X) \tag{5-13}$$

and this concludes the proof. □

Remark 5.7. The very definition of the Euclidean distance degree relies on the square-distance function, $\sum_i (x_i - u_i)^2$, which is not a projective invariant. Therefore, $\mathrm{EDdeg}(X)$ *does* depend on the choice of coordinates in the ambient projective space $\mathbb{P}^{n-1}$. Formula (4-2),

$$\mathrm{EDdeg}(X) = \mathrm{gEDdeg}(X) - \gamma(X),$$

expresses the Euclidean distance degree of a variety in terms of a quantity that *is* projectively invariant, i.e., $\mathrm{gEDdeg}(X)$, and a correction term $\gamma(X)$ which is not. In fact, the coordinate choice determines the isotropic quadric Q; that is, $\sum_i x_i^2 = 0$ is a specific nonsingular quadric in $\mathbb{P}^{n-1}$. Theorem 5.1 prompts us to define a transparent "projective invariant version" $\mathrm{EDdeg}(Q, X)$ of the Euclidean distance degree, for smooth X: $\mathrm{EDdeg}(Q, X)$ could be defined by the right-hand side of (5-1), where now Q is an *arbitrary* nonsingular quadric in $\mathbb{P}^{n-1}$. (If X is not necessarily smooth, (4-3) could likewise be used to define such a notion.) The number $\mathrm{EDdeg}(Q, X)$ is determined by the pair $X \cap Q \subseteq X$ and does not depend on the choice of coordinates. What Theorem 5.1 shows is that if homogeneous coordinates $x_1, \dots, x_n$ are chosen so that the equation of Q is $\sum_{i=1}^n x_i^2$, then $\mathrm{EDdeg}(Q, X)$ equals the Euclidean distance degree of the variety X (in those coordinates). This fact is occasionally useful in computations; see Section 9.5.

Remark 5.8. As pointed out in Remark 2.1, $\mathrm{EDdeg}(X) = 0$ if $X \subseteq Q$. Theorem 5.1 is compatible with this fact, in the following sense. If $Q \cap X = X$, it is natural to set $J(Q\cap X) = X$, and hence $s(J(Q\cap X), X) = s(X, X) = [X]$. The reader can verify (using (3-1)) that

$$\int \frac{(1+2h)\cdot c(T^*X \otimes \mathcal{O}(2h))}{1+h} \cap [X] = \mathrm{gEDdeg}(X).$$

Therefore (5-1) reduces to $\mathrm{EDdeg}(X) = 0$ in this case, as expected.

6. Euclidean distance degree and Milnor classes

While the formula in Theorem 5.1 is essentially straightforward to implement, given the algorithm for the computation of Segre classes in [Harris 2017], it is fair to say that its "geometric meaning" is not too transparent. In this section and the following two we use results from the theory of characteristic classes of singular varieties to provide versions of the formula in terms of notions with a more direct geometric interpretation.

Our first aim is the following result. The *Milnor class* of a variety V is the signed difference

$$\mathcal{M}(V) := (-1)^{\dim V - 1}(c_{\mathrm{SM}}(V) - c_F(V))$$

between its Chern–Fulton class $c_F(V)$ (which we already encountered in Section 5) and its *CSM* (*"Chern–Schwartz–MacPherson"*) *class.*

We denote by $\mathcal{M}(V)_j$ the component of $\mathcal{M}(V)$ of dimension j. The Milnor class owes its name to the fact that if V is a hypersurface with at worst isolated singularities in a compact nonsingular variety, then the degree of its Milnor class is the sum of the Milnor numbers of its singularities [Parusiński and Pragacz 2001, Example 0.1].

The CSM class of a variety V is a "homology" class which agrees with the total Chern class of the tangent bundle of V when V is nonsingular, and satisfies a functorial requirement formalized by Deligne and Grothendieck. See [MacPherson 1974] for a definition inspired by this functorial requirement, and [Schwartz 1965a; 1965b] for an earlier equivalent definition motivated by the problem of extending theorems of Poincaré–Hopf type. An efficient summary of MacPherson's definition (upgraded to the Chow group) may be found in [Fulton 1984, Example 19.1.7]. With notation as in this reference (or as in [MacPherson 1974]), our $c_{\mathrm{SM}}(V)$ is $c_*(\mathbb{1}_V)$.

As an easy consequence of functoriality, the degree of $c_{\mathrm{SM}}(V)$ equals $\chi(V)$, the topological Euler characteristic of V. In fact the degrees of all the terms in $c_{\mathrm{SM}}(V)$ may be interpreted in terms of Euler characteristics [Aluffi 2013, Theorem 1.1], and this will be key for the version of the result we will present in Section 8.

If V is a hypersurface, then $c_{\mathrm{F}}(V)$ equals the class of the *virtual* tangent bundle of V; it may be interpreted as the limit of the Chern class of a smoothing of V in the same linear equivalence class. The terms in $c_{\mathrm{F}}(V)$ may therefore also be interpreted in terms of Euler characteristics (of smoothings of V). Roughly, the Milnor class measures the changes in the Euler characteristics of general hyperplane sections of V as we smooth it within its linear equivalence class.

Theorem 6.1. *Let X be a smooth subvariety of $\mathbb{P}^{n-1}$, and assume $X \not\subseteq Q$. Then*

$$\mathrm{EDdeg}(X) = \mathrm{gEDdeg}(X) - \sum_{j\geq 0}(-1)^j \deg \mathcal{M}(Q\cap X)_j. \tag{6-1}$$

Milnor classes are also accessible computationally; see [Aluffi 2003, Example 4.7].

Remark 6.2. It suffices to require X to be nonsingular in a neighborhood of $Q\cap X$; see Remark 5.2.

Proof. We begin by recalling an expression relating the Milnor class of a hypersurface V of a nonsingular variety M to the Segre class of its singularity subscheme $J(V)$. Letting $\mathscr{L} = \mathcal{O}(V)$,

$$\mathcal{M}(V) = (-1)^{\dim M}\frac{c(TM)}{c(\mathscr{L})} \cap (s(J(V), M)^{\vee} \otimes_M \mathscr{L}). \tag{6-2}$$

This is [Aluffi 1999a, Theorem I.4]. The notation used in this statement are as follows (see [Aluffi 1999a, §1.4] or [Aluffi 1994, §2]): if $A=\sum_{i\geq 0}a^i$ is a rational equivalence class in $V\subseteq M$, where a^i has codimension i in M, and $\mathscr{L}$ is a line bundle on V, then

$$A^\vee=\sum_{i\geq 0}(-1)^i a^i,\qquad A\otimes_M\mathscr{L}=\sum_{i\geq 0}\frac{a^i}{c(\mathscr{L})^i}$$

(note that the codimension is computed in the ambient variety M, even if the class may be defined in the Chow group of the subscheme V).

This notation satisfies several properties, for example a basic compatibility with respect to Chern classes of tensors of vector bundles. One convenient property is given in [Aluffi 2017, Lemma 3.1]: with notation as above, the term of codimension c in M in

$$c(\mathscr{L})^{c-1}\cap(A\otimes_M\mathscr{L})$$

is *independent of* $\mathscr{L}$. In particular,

$$\int c(\mathscr{L})^{\dim M-1}\cap(A\otimes_M\mathscr{L})$$

is independent of $\mathscr{L}$, and this implies (using [Aluffi 1994, Proposition 2])

$$\int c(\mathscr{L})^{\dim M-1}\cap A=\int A\otimes_M\mathscr{L}^\vee.$$

Apply this fact to $\gamma(X)$ (from (5-13)), viewed as

$$\gamma(X)=\int(1+2h)^{\dim X-1}\cap\left(\frac{c(T^*X\otimes\mathbb{O}(2h))}{(1+h)(1+2h)^{\dim X-2}}\cap s(J(Q\cap X,X))\right),$$

with $M=X$, $\mathscr{L}=\mathbb{O}(2h)$. We obtain

$$\begin{aligned}\gamma(X)&=\int\left(\frac{c(T^*X\otimes\mathbb{O}(2h))}{(1+h)(1+2h)^{\dim X-2}}\cap s(J(Q\cap X,X))\right)\otimes_M\mathbb{O}(-2h)\\&\overset{!}{=}\int\frac{c(T^*X)}{(1-h)(1-2h)}\cap\big(s(J(Q\cap X,X))\otimes_M\mathbb{O}(-2h)\big),\end{aligned}$$

where the equality $\overset{!}{=}$ follows by applying [Aluffi 1994, Proposition 1]. Since the degree of a class in X is the degree of its component of dimension 0, i.e., codimension $\dim X$, this gives

$$\begin{aligned}\gamma(X)&=(-1)^{\dim X}\int\left(\frac{1}{1-h}\cdot\frac{c(T^*X)}{1-2h}\cap\big(s(J(Q\cap X,X))\otimes_M\mathbb{O}(-2h)\big)\right)^\vee\\&=(-1)^{\dim X}\int\frac{1}{1+h}\cdot\frac{c(TX)}{1+2h}\cap\big(s(J(Q\cap X,X))^\vee\otimes_M\mathbb{O}(2h)\big).\end{aligned}$$

Finally, by (6-2) (with $M=X$, $V=Q\cap X$, $\mathscr{L}=\mathbb{O}(V)=\mathbb{O}(2h)$), we get

$$\gamma(X)=\int\frac{1}{1+h}\cap\mathcal{M}(Q\cap X),$$

and this implies (6-1). □

Corollary 6.3. *If $Q \cap X$ only has isolated singularities x_i, then*

$$\mathrm{EDdeg}(X) = \mathrm{gEDdeg}(X) - \sum_i \mu(x_i),$$

where $\mu(-)$ denotes the Milnor number.

We will see that this is in fact the case for most smooth hypersurfaces (Section 9.3).

7. Euclidean distance degree and Chern–Schwartz–MacPherson classes

Our next aim is to express the Euclidean distance degree of a nonsingular projective variety directly, rather than in terms of a correction from a "generic" situation. CSM classes provide a convenient means to do so. The formula presented in Section 8 may look more appealing, but the alternative (7-1) presented here, besides being a necessary intermediate result, is in a sense algorithmically more direct.

Theorem 7.1. *Let X be a smooth subvariety of $\mathbb{P}^{n-1}$. Then*

$$\mathrm{EDdeg}(X) = (-1)^{\dim X} \sum_{j \ge 0} (-1)^j (c(X)_j - c_{\mathrm{SM}}(Q \cap X)_j). \tag{7-1}$$

Here, $c_{\mathrm{SM}}(Q \cap X)_j$ denotes the degree of the j-dimensional component of $Q \cap X$. Again, (7-1) is straightforward to implement given available algorithms for characteristic classes (for example [Aluffi 2003; Jost 2015; Harris 2017]).

Remark 7.2. If $X \subseteq Q$, then $c_{\mathrm{SM}}(Q \cap X) = c_{\mathrm{SM}}(X) = c(X)$ by the basic normalization property of CSM classes, as X is nonsingular. In this case (7-1) gives $\mathrm{EDdeg}(X) = 0$, as it should (see Remark 2.1). Therefore, we will assume $X \not\subseteq Q$ in the proof.

Remark 7.3. In the proof we will use the fact that $c(X) = c_{\mathrm{F}}(X)$ if X is nonsingular. This prevents a straightforward generalization of the argument to the case in which X is only required to be nonsingular in a neighborhood of $Q \cap X$.

Proof. According to Theorem 6.1,

$$\mathrm{EDdeg}(X) = \mathrm{gEDdeg}(X) - (-1)^{\dim X} \sum_{j \ge 0} (-1)^j (c_{\mathrm{SM}}(Q \cap X)_j - c_{\mathrm{F}}(Q \cap X)_j).$$

By (3-1), therefore, $\mathrm{EDdeg}(X)$ equals

$$(-1)^{\dim X} \sum_{j \ge 0} (-1)^j \big((2^{j+1} - 1) c(X)_j + c_{\mathrm{F}}(Q \cap X)_j - c_{\mathrm{SM}}(Q \cap X)_j\big). \tag{7-2}$$

By definition,

$$c_{\mathrm{F}}(Q \cap X) = c(TX) \cap s(Q \cap X, X) = \frac{c(TX) \cdot 2h}{1+2h} \cap [X]$$

(after push-forward to X) since $Q\cap X$ is a hypersurface in X and $\mathbb{O}(Q)=\mathbb{O}(2h)$ as Q is a quadric. Therefore

$$\sum_{j\geq 0}(-1)^j c_{\mathrm{F}}(Q\cap X)_j=\int\frac{1}{1+h}\frac{2h}{1+2h}c(TX)\cap[X]=\sum_{j\geq 0}c(X)_j\int\frac{h^{\dim X-j}}{1+h}\frac{2h}{1+2h}\cap[\mathbb{P}^{\dim X}].$$

The coefficient of $c(X)_j$ in this expression equals the coefficient of h^j in the expansion of

$$\frac{2h}{(1+h)(1+2h)}=\sum_{j\geq 0}(-1)^{j+1}(2^{j+1}-2)h^j.$$

Therefore (7-2) gives

$$\mathrm{EDdeg}(X)=(-1)^{\dim X}\sum_{j\geq 0}(-1)^j\bigl(((2^{j+1}-1)-(2^{j+1}-2))c(X)_j-c_{\mathrm{SM}}(Q\cap X)_j\bigr)$$

and (7-1) follows. □

CSM classes may be associated with locally closed sets: if V is a locally closed set of a variety M, then $c_{\mathrm{SM}}(V)=c_*(\mathbb{1}_V)$ is a well-defined class in A_*M. (If $V=\overline{V}\smallsetminus W$, with W closed, then $c_{\mathrm{SM}}(V)=c_{\mathrm{SM}}(\overline{V})-c_{\mathrm{SM}}(W)$.) This notation allows us to express (7-1) in (even) more concise terms: if X is a smooth subvariety of $\mathbb{P}^{n-1}$, and $X\not\subseteq Q$, then

$$\mathrm{EDdeg}(X)=(-1)^{\dim X}\sum_{j\geq 0}(-1)^j c_{\mathrm{SM}}(X\smallsetminus Q)_j. \tag{7-3}$$

Indeed, $c(X)=c_{\mathrm{SM}}(X)$ since X is nonsingular.

If $Q\cap X$ is (supported on) a simple normal crossing divisor, (7-3) admits a particularly simple expression, given in the corollary that follows. An illustration of this case will be presented in Section 9.6.

Corollary 7.4. *Let $X\subseteq\mathbb{P}^{n-1}$ be a smooth subvariety, and assume the support of $Q\cap X$ is a divisor D with normal crossings and nonsingular components D_i, $i=1,\dots,r$. Then*

$$\mathrm{EDdeg}(X)=\int\frac{c(T^*X(\log D))}{1-H}\cap[X]=\int\frac{1}{1-H}\cdot\frac{c(T^*X)}{\prod_i(1-D_i)}\cap[X].$$

Proof. If D is a simple normal crossing divisor in X, then

$$c_{\mathrm{SM}}(X\smallsetminus D)=c(TX(-\log D))\cap[X] \tag{7-4}$$

(see [Aluffi 1999b], or [Goresky and Pardon 2002, Proposition 15.3]). Using this fact in (7-3), the stated formulas follow from simple manipulations and the well-known expression for $c(T^*X(\log D))$ when D is a simple normal crossing divisor. □

Liao has shown that (7-4) holds as soon as D is a free divisor that is locally quasihomogeneous [Liao 2012] or more generally with Jacobian of linear type [Liao 2018]. Therefore, $\mathrm{EDdeg}(X)=\int c(T^*X(\log D))/(1-H)\cap[X]$ as in Corollary 7.4 as soon as the support of $Q\cap X$ satisfies these less restrictive conditions.

8. Euclidean distance degree and Euler characteristics

Finally, we present a version of the main result which makes no use (in its formulation) of characteristic classes for singular varieties. This is the version given in the introduction.

Theorem 8.1. *Let X be a smooth subvariety of $\mathbb{P}^{n-1}$. Then*

$$\mathrm{EDdeg}(X) = (-1)^{\dim X} \chi(X \smallsetminus (Q \cup H)), \tag{8-1}$$

where H is a general hyperplane.

By the inclusion-exclusion property of the topological Euler characteristic, (8-1) is equivalent to

$$\mathrm{EDdeg}(X) = (-1)^{\dim X} (\chi(X) - \chi(X \cap Q) - \chi(X \cap H) + \chi(X \cap Q \cap H)) \tag{8-2}$$

which has the advantage of only involving closed subsets of $\mathbb{P}^{n-1}$. Any of the aforementioned implementations of algorithms for characteristic classes of singular varieties includes explicit functions to compute Euler characteristics of projective schemes from defining homogenous ideals, so (8-2) is also essentially trivial to implement. However, despite its conceptual simplicity, this expression is computationally expensive.

Remark 8.2. As in Section 7, we have to insist that X be smooth; only requiring it to be nonsingular in a neighborhood of $Q \cap X$ is not enough for the result to hold.

Proof. The statement is a consequence of Theorem 7.1 and of a result from [Aluffi 2013]. Collect the degrees of the components of the CSM class of a locally closed set $V \subseteq \mathbb{P}^N$ into a polynomial

$$\Gamma_V(t) = \sum_{j \geq 0} c_{\mathrm{SM}}(V)_j \, t^j;$$

and collect the signed Euler characteristics of generic linear sections of V into another polynomial

$$\chi_V(t) = \sum_{j \geq 0} (-1)^j \chi(V \cap H_1 \cap \cdots \cap H_j) \, t^j,$$

where the H_i's are general hyperplanes. Then according to [Aluffi 2013, Theorem 1.1] we have

$$\Gamma_V(t) = \mathscr{I}(\chi_V),$$

where $\mathscr{I}$ is an explicit involution. It follows from the specific expression of $\mathscr{I}$ that

$$\Gamma_V(-1) = \chi_V(0) + \chi_V'(0)$$

(see the paragraph preceding the statement of [Aluffi 2013, Theorem 1.1]). Therefore

$$\sum_{j \geq 0} (-1)^j c_{\mathrm{SM}}(V)_j = \chi(V) - \chi(V \cap H) \tag{8-3}$$

for every locally closed set V in projective space.

The statement of the theorem, in the form given in (8-2), follows by applying (8-3) to (7-1). □

9. Examples

9.1. ***Computations.*** The ingredients needed to implement the main theorems of this text on computer algebra systems such as [Sage] or [Macaulay2] are all available. One can compute Segre classes and Chern–Mather classes via [Harris 2017] and Chern–Schwartz–MacPherson classes via (for example) any of [Aluffi 2003; Marco-Buzunáriz 2012; Jost 2013; 2015; Helmer 2016; Harris 2017]. One issue in concrete examples is that computer algebra systems prefer to work with $\mathbb{Q}$-coefficients, and it is often difficult to write the defining equations of a variety which is tangent to the isotropic quadric without extending the field of coefficients. This difficulty can sometimes be circumvented by a suitable choice of coordinates; see Remark 5.7. Also see Section 9.5 below for a discussion of a template situation.

In many cases, the "standard" algorithm of [Draisma et al. 2016, Example 2.11] appears to be at least as fast as the alternatives obtained by implementing the results presented here. In some examples these alternatives are faster, particularly if they take advantage of the refinements which will be presented below. As an illustration, we can apply Proposition 9.2 (Section 9.4) to compute the ED degrees of plane curves in terms of a generator of their homogeneous ideal. A (non-optimal) implementation of this method in Macaulay2 can be coded as follows:

```
PP2 = QQ[x,y,z]; C = ideal( F )
S = QQ[s,t,i,Degrees=>{{1,0},{1,0},{0,1}}]/(i^2+1)
J = sub(C,{x=>s^2-t^2,y=>2*s*t,z=>i*(s^2+t^2)})
p = (first degrees radical J)#0 -- ignore degree of i
d = degree C
d*(d-2) + p
```

where $F = F(x, y, z)$ is the defining homogeneous polynomial for the curve.

For example, trial runs of computations of the ED degrees of Fermat curves $x^d + y^d + z^d = 0$ for all degrees $d = 3, \ldots, 40$ took an average of 4.5 seconds using this method (in a more efficient implementation), and 260 seconds by using the standard algorithm. However, direct implementations of the general formulas presented in this paper do not fare as well.

The interested reader can find the actual code used here, as well as implementations of the more general formulas at http://github.com/coreysharris/EDD-M2. At this stage, the value of the formulas obtained in Theorems 5.1–8.1 appears to rest more on their theoretical applications (in examples such as the ones discussed below in Section 9.6) than in the speed of their computer algebra implementations.

If the variety is known to be transversal to the isotropic quadric, then its Euclidean distance degree equals the *generic* Euclidean distance degree. This may be computed by using the algorithm for Chern(–Mather) classes in [Harris 2017], often faster than the standard algorithm. For example, let S be a general hyperplane section of the second Veronese embedding of $\mathbb{P}^3$ in $\mathbb{P}^9$. Then S is transversal to the isotropic quadric, and the implementation of the algorithm in [Harris 2017] computes its Euclidean distance degree (i.e., 36) in about 2 seconds. The standard algorithm appears to take impractically long on this example;

one can improve its performance by first projecting S to a general $\mathbb{P}^3$ (this does not affect the Euclidean distance degree, by [Draisma et al. 2016, Corollary 6.1]), and the computation then takes about a minute.

9.2. ***Quadrics and spheres.*** Let $X \subseteq \mathbb{P}^{n-1}$ be a nonsingular quadric hypersurface. We say that X is a *sphere* if it is given by the equation

$$x_1^2 + \cdots + x_{n-1}^2 = cx_n^2$$

with $c > 0$ a real number. It is clear from the definition in terms of critical points of the distance function that $\mathrm{EDdeg}(X) = 2$ if X is a sphere in $\mathbb{P}^{n-1}$, $n \geq 2$.

We use this example to illustrate some of the formulas obtained in this paper.

First, since X is a degree 2 hypersurface in $\mathbb{P}^{n-1}$,

$$c(TX) = \frac{c(T\mathbb{P}^{n-1}|_X)}{1+2h} = \frac{(1+h)^n}{1+2h}$$

(where h denotes the hyperplane class and its pull-backs, as in previous sections). Applying (3-1), one easily sees that

$$\mathrm{gEDdeg}(X) = 2n-2,$$

while

$$c(T^*X \otimes \mathbb{O}(2h)) = \frac{(1-h+2h)^n}{(1+2h)(1-2h+2h)} = \frac{(1+h)^n}{1+2h}.$$

For a sphere $X \subseteq \mathbb{P}^{n-1}$, the intersection $Q \cap X$ consists of a double quadric in $\mathbb{P}^{n-2}$, supported on the transversal intersection $X \cap H$ of X with a hyperplane. It follows that $J(Q \cap X) = X \cap H$, and therefore

$$s(J(Q \cap X), X) = \frac{h \cdot [X]}{1+h}.$$

According to Theorem 5.1, the correction term in this case is given by

$$\begin{aligned}\int \frac{(1+2h)\cdot c(T^*X \otimes \mathbb{O}(2h))}{1+h} \cdot s(J(Q\cap X), X) &= \int \frac{(1+2h)(1+h)^n}{(1+h)(1+2h)} \cdot \frac{h\cdot 2h}{1+h} \cap [\mathbb{P}^{n-1}] \\ &= \int 2(1+h)^{n-2} \cdot h^2 \cap [\mathbb{P}^{n-1}] \\ &= 2(n-2).\end{aligned}$$

By (5-1), $\mathrm{EDdeg}(X) = (2n-2) - 2(n-2) = 2$, as it should.

From the point of view of Theorem 8.1, we should deal with the topological Euler characteristics of X, $X \cap Q$, $X \cap H$, $X \cap Q \cap H$, where H is a general hyperplane (see (8-2)). If X is a sphere, then $X \cap Q$ is (supported on) a nonsingular quadric in $\mathbb{P}^{n-2}$; so is $X \cap H$, and $X \cap Q \cap H$ is a nonsingular quadric in $\mathbb{P}^{n-3}$. The Euler characteristic of a nonsingular quadric in $\mathbb{P}^N$ is $N+1$ if N is odd, N if N is even; therefore

$$\chi(X) - \chi(X \cap Q) - \chi(X \cap H) + \chi(X \cap Q \cap H) = \begin{cases} (n-1) - 2(n-1) + (n-3) = -2 & n \text{ odd}, \\ n - 2(n-2) + (n-2) = 2 & n \text{ even}. \end{cases}$$

and by Theorem 8.1,

$$\mathrm{EDdeg}(X) = (-1)^{\dim X}\chi(X \smallsetminus (Q \cup H)) = 2$$

for all n, as expected.

9.3. *Hypersurfaces.* The case of smooth hypersurfaces of degree ≥ 3 is more constrained than it may look at first.

Claim 9.1. If two smooth hypersurfaces of degree d_1, d_2 in projective space are tangent along a positive dimensional algebraic set, then $d_1 = d_2$.

(This is [Aluffi 2000, Claim 3.2].) It follows that if $X \subset \mathbb{P}^{n-1}$ is a smooth hypersurface of degree $d \neq 2$, then the intersection $Q \cap X$ necessarily has isolated singularities. We are then within the scope of Corollary 6.3, and we can conclude

$$\mathrm{EDdeg}(X) = \mathrm{gEDdeg}(X) - \sum_i \mu(x_i),$$

where the sum is over all singularities x_i of $Q \cap X$, and $\mu(-)$ denotes the Milnor number.

9.4. *Curves.* Let $C \subseteq \mathbb{P}^{n-1}$ be a nonsingular curve. Then

$$\mathrm{EDdeg}(C) = d + \#(Q \cap C) - \chi(C). \tag{9-1}$$

(This follows immediately from Theorem 8.1.)

For example, the twisted cubic parametrized by

$$(s:t) \mapsto (s^3 : \sqrt{3}s^2t : \sqrt{3}st^2 : t^3)$$

has EDdeg equal to 3: indeed, it meets the isotropic quadric at the images of the solutions of $s^6 + 3s^4t^2 + 3s^2t^4 + t^6 = (s^2+t^2)^3 = 0$, that is, at two points. More generally, the Euclidean distance degree of the rational normal curve of degree $n-1$ in $\mathbb{P}^{n-1}$ parametrized by

$$(s:t) \mapsto \left(\sqrt{\binom{n-1}{j}}s^{n-1-j}t^j\right)_{j=0,\dots,n-1}$$

is $(n-1)+2-2 = n-1$.

For *plane* curves, (9-1) admits a particularly explicit formulation.

Proposition 9.2. *Let C be a nonsingular plane curve, defined by an irreducible homogeneous polynomial $F(x,y,z)$. Then*

$$\mathrm{EDdeg}(C) = d(d-2) + R,$$

where R is the number of distinct factors of the polynomial

$$F(s^2-t^2, 2st, i(s^2+t^2)) \in \mathbb{C}[t]$$

and $d = \deg F$.

Proof. This follows immediately from (9-1), after observing that $d-\chi(C)=d-(2-(d-1)(d-2))=d(d-2)$ and that the isotropic conic $x^2+y^2+z^2=0$ is parametrized by $(s:t)\mapsto(s^2-t^2, 2st, i(s^2+t^2))$. □

For instance, consider the conic $x^2+2y^2+2iyz=0$. Since

$$(s^2-t^2)^2+2(2st)^2+2i(2st)(i(s^2+t^2))=(s-t)^4,$$

we have $R=1$, therefore its Euclidean distance degree is $2\cdot 0+1=1$.

For another example, the Fermat quintic C: $x^5+y^5+z^5=0$ has $R=8$ (as Macaulay2 can verify), therefore $\mathrm{EDdeg}(C)=5\cdot 3+8=23$ (see [Draisma et al. 2016, Example 2.5]). More generally, the Euclidean distance degree of the Fermat curve $x^d+y^d+z^d=0$ is $d(d-2)+R$, where R is the number of distinct factors of the polynomial

$$(s^2-t^2)^d+(2st)^d+(i(s^2+t^2))^d.$$

An explicit expression for the Euclidean distance degree of Fermat hypersurfaces in any dimension may be found in [Lee 2017, Theorem 4].

9.5. ***Surfaces.*** According to Theorem 8.1, if $S\subseteq\mathbb{P}^{n-1}$ is a smooth degree-d surface, and C is the support of the intersection $Q\cap S$ (which may very well be singular), then

$$\mathrm{EDdeg}(S)=\chi(S)-\chi(S\cap H)-\chi(C)+\deg(C),$$

where H is a general hyperplane. If $n-1=3$, then $\chi(S)=d(d^2-4d+6)$ and $\chi(S\cap H)=3d-d^2$; for $d\neq 2$, C is necessarily reduced (Claim 9.1), so $\deg(C)=2d$. In this case ($S\subseteq\mathbb{P}^3$ a smooth surface of degree $d\neq 2$, or more generally such that $S\cap Q$ is reduced),

$$\mathrm{EDdeg}(S)=d(d^2-4d+6)-(3d-d^2)-\chi(C)+2d=d(d^2-3d+5)-\chi(C). \tag{9-2}$$

If C is nonsingular, then $\chi(C)=-2d(d-2)$, and $\mathrm{EDdeg}(S)=\mathrm{gEDdeg}(S)=d(d^2-d+1)$.

If S is a plane in $\mathbb{P}^3$, tangent to the isotropic quadric Q, then $C=Q\cap S$ is a pair of lines, and (9-2) gives $\mathrm{EDdeg}(S)=0$. But note that the coefficients of the equation of this plane are necessarily not all real, so the enumerative interpretation of $\mathrm{EDdeg}(S)$ as the number of critical points of a "distance" function should be taken *cum grano salis*.

Next let S be a Veronese surface in $\mathbb{P}^5$, described parametrically by

$$(s:t:u)\mapsto(a_1s^2:a_2st:a_3su:a_4t^2:a_5tu:a_6u^2)$$

with $a_1\dots a_6\neq 0$. According to Theorem 8.1,

$$\mathrm{EDdeg}(S)=3-2-\chi(C)+2\deg C=2\deg C-\chi(C)+1,$$

where C is the support of the curve with equation

$$a_1^2x^4+a_2^2x^2y^2+a_3^2x^2z^2+a_4y^4+a_5y^2z^2+a_6z^4=0 \tag{9-3}$$

in the plane. (The degree of the image of C in $\mathbb{P}^5$ is $2 \deg C$.)

For example, if C is a smooth quartic (the "generic" case), then $\chi(C) = -4$ and $\mathrm{EDdeg}(S) = \mathrm{gEDdeg}(S) = 13$. If the rank of the matrix

$$\begin{pmatrix} 2a_1^2 & a_2^2 & a_3^2 \\ a_2^2 & 2a_4^2 & a_5^2 \\ a_3^2 & a_5^2 & 2a_6^2 \end{pmatrix} \tag{9-4}$$

is 1, then (9-3) is a double (smooth) conic, so that $\chi(C) = \deg(C) = 2$ and $\mathrm{EDdeg}(S) = 3$. For example, this is the case for

$$(s:t:u) \mapsto (s^2 : \sqrt{2}\,st : \sqrt{2}\,su : t^2 : \sqrt{2}\,tu : u^2). \tag{9-5}$$

If the rank of (9-4) is 2, then (9-3) factors as a product

$$(a'x^2 + b'y^2 + c'z^2)(a''x^2 + b''y^2 + c''z^2) = 0$$

and the factors are different and correspond to nonsingular conics. If these conics meet transversally, then $\mathrm{EDdeg}(S) = 9$; if they are "bitangent", then $\mathrm{EDdeg}(S) = 7$ (use Corollary 6.3, or again Theorem 8.1). Explicit examples of these two types are

$$(s:t:u) \mapsto (s^2 : \sqrt{3}\,st : 2\,su : \sqrt{2}\,t^2 : \sqrt{5}\,tu : \sqrt{3}\,u^2)$$

and

$$(s:t:u) \mapsto (s^2 : \sqrt{3}\,st : \sqrt{2}\,su : \sqrt{2}\,t^2 : \sqrt{3}\,tu : u^2).$$

More general Veronese embeddings are considered in Section 9.6.

Note that we could equivalently hold the surface $S = X$ fixed, choosing for example the standard Veronese embedding, parametrized by $(s:t:u) \mapsto (s^2 : st : su : t^2 : tu : u^2)$ with ideal

$$(x_1x_4 - x_2^2,\ x_1x_5 - x_2x_3,\ x_1x_6 - x_3^2,\ x_2x_5 - x_3x_4,\ x_2x_6 - x_3x_5,\ x_4x_6 - x_5^2)$$

in $\mathbb{P}^5_{(x_1:\cdots:x_6)}$, and consider a more general nonsingular quadric

$$Q : q_1x_1^2 + q_2x_2^2 + \cdots + q_6x_6^2 = 0,$$

$q_1 \dots q_6 \neq 0$, in place of the isotropic quadric. This corresponds to a change of coordinates $x_i \mapsto \sqrt{q_i}x_i$; i.e., $q_i = a_i^2$ with notation as above. The right-hand side $\mathrm{EDdeg}(Q, X)$ of (5-1) (or equivalently (6-1), (7-1), (8-1)) is independent of the coordinate choice; see Remark 5.7. For example, choosing

$$x_1^2 + 2x_2^2 + 2x_3^2 + x_4^2 + 2x_5^2 + x_6^2 = 0$$

for Q, along with the standard Veronese embedding, is equivalent to choosing the standard isotropic quadric along with the embedding (9-5) (hence $\mathrm{EDdeg}(Q, X) = 3$ in this case).

This observation may be useful in effective computations, since computer algebra systems prefer to work with $\mathbb{Q}$ coefficients.

9.6. ***Segre and Segre–Veronese varieties.*** Let X be the image of the usual Segre embedding

$$\mathbb{P}(\mathbb{C}^{m_1}) \times \cdots \times \mathbb{P}(\mathbb{C}^{m_p}) \to \mathbb{P}(\mathbb{C}^{m_1} \otimes \cdots \otimes \mathbb{C}^{m_p}),$$

that is, $\mathbb{P}^{m_1-1} \times \cdots \times \mathbb{P}^{m_p-1} \to \mathbb{P}^{m_1 \dots m_p-1}$. This embedding maps a point

$$((s_1^1 : \cdots : s_{m_1}^1), \dots, (s_1^p : \cdots : s_{m_p}^p))$$

to the point in $\mathbb{P}^{m_1 \dots m_p-1}$ whose homogeneous coordinates $(\underline{x})$ are all the monomials of multidegree $(1, \dots, 1)$ in the variables $\underline{s}^1, \dots, \underline{s}^p$. The equation $\sum_i x_i^2 = 0$ of the isotropic quadric pulls back to

$$\Bigl(\sum_i (s_i^1)^2\Bigr) \cdots \Bigl(\sum_i (s_i^p)^2\Bigr) = 0.$$

Let Q_i be the isotropic quadric in the i-th factor. Then this shows that

$$Q \cap X = (Q_1 \times \mathbb{P}^{m_2-1} \times \cdots \times \mathbb{P}^{m_p-1}) \cup \cdots \cup (\mathbb{P}^{m_1-1} \times \cdots \times \mathbb{P}^{m_{p-1}-1} \times Q_p).$$

It follows that $Q \cap X$ is a divisor with normal crossings and nonsingular components. Denoting by h_i the hyperplane class in the i-th factor, the class of the i-th component is $2h_i$. By Corollary 7.4,

$$\mathrm{EDdeg}(X) = \int \frac{1}{1 - h_1 - \cdots - h_p} \cdot \frac{(1-h_1)^{m_1} \dots (1-h_p)^{m_p}}{(1-2h_1) \dots (1-2h_p)} \cap [X]. \tag{9-6}$$

The conclusion is that $\mathrm{EDdeg}(\mathbb{P}^{m_1-1} \times \cdots \times \mathbb{P}^{m_p-1})$ equals the coefficient of $h_1^{m_1-1} \dots h_p^{m_p-1}$ in the expansion of

$$\frac{1}{1 - h_1 - \cdots - h_p} \cdot \prod_{i=1}^{p} \frac{(1-h_i)^{m_i}}{1-2h_i}.$$

Friedland and Ottaviani [2014, Theorem 4] (see [Draisma et al. 2016, Theorem 8.1]) obtain a different expression for the same quantity: they prove that it must equal the coefficient of $z_1^{m_1-1} \dots z_p^{m_p-1}$ in the expression

$$\prod_{i=1}^{p} \frac{\hat{z}_i^{m_i} - z_i^{m_i}}{\hat{z}_i - z_i}, \tag{9-7}$$

where $\hat{z}_i = (z_1 + \cdots + z_p) - z_i$. These coefficients must be equal, since they both compute the Euclidean distance degrees of Segre varieties. We note that, for example,

$$\mathrm{EDdeg}(\mathbb{P}^2 \times \mathbb{P}^8 \times \mathbb{P}^{11} \times \mathbb{P}^{13} \times \mathbb{P}^{24}) = 143046202777307645494624$$

according to *both* formulas.

The same technique may be used to deal with *Segre–Veronese varieties,* obtained by composing a Segre embedding with a product of Veronese embeddings:

$$\mathbb{P}(\mathbb{C}^{m_1}) \times \cdots \times \mathbb{P}(\mathbb{C}^{m_p}) \to \mathbb{P}(\mathrm{Sym}^{\omega_1} \mathbb{C}^{m_1}) \times \cdots \times \mathbb{P}(\mathrm{Sym}^{\omega_p} \mathbb{C}^{m_p}) \to \mathbb{P}(\mathrm{Sym}^{\omega_1} \mathbb{C}^{m_1} \otimes \cdots \otimes \mathrm{Sym}^{\omega_p} \mathbb{C}^{m_p}).$$

Using general coordinates for the Veronese embeddings, each Q_i (with notation as above) restricts to a smooth hypersurface of degree $2\omega_i$, and the resulting hypersurfaces of the product meet with normal crossings. The hyperplane class restricts to $\omega_1 h_1 + \cdots + \omega_p h_p$, therefore (again by Corollary 7.4) the EDdegree of this variety equals the coefficient of $h_1^{m_1-1} \dots h_p^{m_p-1}$ in the expansion of

$$\frac{1}{1-\omega_1 h_1 - \cdots - \omega_p h_p} \cdot \prod_{i=1}^{p} \frac{(1-h_i)^{m_i}}{1-2\omega_i h_i}. \tag{9-8}$$

Friedland and Ottaviani also consider Segre–Veronese varieties, but they choose suitably invariant coordinates in each factor; this is a different problem. (For $p = 1$, $m_1 = \omega_1 = 2$, this choice of coordinates is given by (9-5).) They prove [Friedland and Ottaviani 2014, Draisma et al. 2016, Theorem 8.6] that with these special coordinates the EDdegree is given again by the coefficient of $z_1^{m_1-1} \dots z_p^{m_p-1}$ in (9-7), but where now $\hat{z}_i = (\omega_1 z_1 + \cdots + \omega_p z_p) - z_i$. From our point of view, the choice of coordinates affects the restrictions of the isotropic quadrics Q_i to the factors. With the invariant coordinates used by Friedland and Ottaviani, each Q_i restricts to a multiple *quadric,* and this affects the denominator of (9-8): the resulting EDdegree equals the coefficient of $h_1^{m_1-1} \dots h_p^{m_p-1}$ in the expansion of

$$\frac{1}{1-\omega_1 h_1 - \cdots - \omega_p h_p} \cdot \prod_{i=1}^{p} \frac{(1-h_i)^{m_i}}{1-2h_i}. \tag{9-9}$$

Therefore, this coefficient must agree with the one obtained with the Friedland–Ottaviani formula. (It does not seem combinatorially trivial that this should be the case in general; it is easy to verify that both formulas yield $((\omega_1 - 1)^{m_1} - 1)/(\omega_1 - 2)$ for $p = 1$.)

Acknowledgments

This work was supported in part by NSA grant H98230-16-1-0016. The authors thank Giorgio Ottaviani for suggesting the example presented in Section 9.6.

References

[Aluffi 1994] P. Aluffi, "MacPherson's and Fulton's Chern classes of hypersurfaces", *Internat. Math. Res. Notices* 11 (1994), 455–465. MR Zbl

[Aluffi 1995] P. Aluffi, "Singular schemes of hypersurfaces", *Duke Math. J.* **80**:2 (1995), 325–351. MR Zbl

[Aluffi 1999a] P. Aluffi, "Chern classes for singular hypersurfaces", *Trans. Amer. Math. Soc.* **351**:10 (1999), 3989–4026. MR Zbl

[Aluffi 1999b] P. Aluffi, "Differential forms with logarithmic poles and Chern–Schwartz–MacPherson classes of singular varieties", *C. R. Acad. Sci. Paris Sér. I Math.* **329**:7 (1999), 619–624. MR Zbl

[Aluffi 2000] P. Aluffi, "Weighted Chern–Mather classes and Milnor classes of hypersurfaces", pp. 1–20 in *Singularities — Sapporo* 1998, edited by J.-P. Brasselet and T. Suwa, Adv. Stud. Pure Math. **29**, Kinokuniya, Tokyo, 2000. MR Zbl

[Aluffi 2003] P. Aluffi, "Computing characteristic classes of projective schemes", *J. Symbolic Comput.* **35**:1 (2003), 3–19. MR Zbl

[Aluffi 2013] P. Aluffi, "Euler characteristics of general linear sections and polynomial Chern classes", *Rend. Circ. Mat. Palermo* (2) **62**:1 (2013), 3–26. MR Zbl

[Aluffi 2017] P. Aluffi, "Tensored Segre classes", *J. Pure Appl. Algebra* **221**:6 (2017), 1366–1382. MR Zbl

[Aluffi 2018] P. Aluffi, "Projective duality and a Chern–Mather involution", *Trans. Amer. Math. Soc.* **370**:3 (2018), 1803–1822. MR Zbl

[Draisma et al. 2016] J. Draisma, E. Horobeţ, G. Ottaviani, B. Sturmfels, and R. R. Thomas, "The Euclidean distance degree of an algebraic variety", *Found. Comput. Math.* **16**:1 (2016), 99–149. MR Zbl

[Friedland and Ottaviani 2014] S. Friedland and G. Ottaviani, "The number of singular vector tuples and uniqueness of best rank-one approximation of tensors", *Found. Comput. Math.* **14**:6 (2014), 1209–1242. MR

[Fulton 1984] W. Fulton, *Intersection theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **2**, Springer, 1984. MR Zbl

[Goresky and Pardon 2002] M. Goresky and W. Pardon, "Chern classes of automorphic vector bundles", *Invent. Math.* **147**:3 (2002), 561–612. MR Zbl

[Harris 2017] C. Harris, "Computing Segre classes in arbitrary projective varieties", *J. Symbolic Comput.* **82** (2017), 26–37. MR Zbl

[Helmer 2016] M. Helmer, "Algorithms to compute the topological Euler characteristic, Chern–Schwartz–MacPherson class and Segre class of projective varieties", *J. Symbolic Comput.* **73** (2016), 120–138. MR Zbl

[Huneke and Swanson 2006] C. Huneke and I. Swanson, *Integral closure of ideals, rings, and modules*, London Mathematical Society Lecture Note Series **336**, Cambridge University Press, 2006. MR Zbl

[Jost 2013] C. Jost, "An algorithm for computing the topological Euler characteristic of complex projective varieties", 2013. arXiv

[Jost 2015] C. Jost, "Computing characteristic classes and the topological Euler characteristic of complex projective schemes", *J. Softw. Algebra Geom.* **7** (2015), 31–39. MR

[Lee 2017] H. Lee, "The Euclidean distance degree of Fermat hypersurfaces", *J. Symbolic Comput.* **80**:2 (2017), 502–510. MR Zbl

[Liao 2012] X. Liao, "Chern classes of logarithmic vector fields for locally-homogenous free divisors", 2012. arXiv

[Liao 2018] X. Liao, "Chern classes of logarithmic derivations for free divisors with Jacobian ideal of linear type", *J. Math. Soc. Japan* **70**:3 (2018), 975–988. MR

[Macaulay2] D. R. Grayson and M. E. Stillman, "Macaulay2, a software system for research in algebraic geometry", available at http://www.math.uiuc.edu/Macaulay2/.

[MacPherson 1974] R. D. MacPherson, "Chern classes for singular algebraic varieties", *Ann. of Math.* (2) **100** (1974), 423–432. MR Zbl

[Marco-Buzunáriz 2012] M. A. Marco-Buzunáriz, "A polynomial generalization of the Euler characteristic for algebraic sets", *J. Singul.* **4** (2012), 114–130. MR

[Parusiński and Pragacz 2001] A. Parusiński and P. Pragacz, "Characteristic classes of hypersurfaces and characteristic cycles", *J. Algebraic Geom.* **10**:1 (2001), 63–79. MR Zbl

[Sage] The Sage developers, "Sagemath, the Sage mathematics software system", available at http://www.sagemath.org.

[Schürmann 2017] J. Schürmann, "Chern classes and transversality for singular spaces", pp. 207–231 in *Singularities in geometry, topology, foliations and dynamics*, edited by J. L. Cisneros-Molina et al., Springer, Cham, Switzerland, 2017. MR

[Schwartz 1965a] M.-H. Schwartz, "Classes caractéristiques définies par une stratification d'une variété analytique complexe I", *C. R. Acad. Sci. Paris* **260** (1965), 3262–3264. MR Zbl

[Schwartz 1965b] M.-H. Schwartz, "Classes caractéristiques définies par une stratification d'une variété analytique complexe II", *C. R. Acad. Sci. Paris* **260** (1965), 3535–3537. MR Zbl

Communicated by Ravi Vakil
Received 2017-11-03 Revised 2018-05-03 Accepted 2018-06-19

aluffi@math.fsu.edu *Department of Mathematics, Florida State University, Tallahassee, FL, United States*

coreyharrismath@gmail.com *Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany*

mathematical sciences publishers msp

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.

Length There is no a priori limit on the length of an *ANT* article, but *ANT* considers long articles only if the significance-to-length ratio is appropriate. Very long manuscripts might be more suitable elsewhere as a memoir instead of a journal article.

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@msp.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 12 No. 8 2018

1937-0652(2018)12:8;1-9